

# **EMIPRA**

12. Jenaer Empiriepraktikumskongress

7. Juli 2017  
Friedrich-Schiller Universität Jena



# Danksagung

Mit dem heutigen Tag findet der Jenaer Emprakongress bereits zum 12. Mal statt. Und auch dieses Jahr lässt sich rückblickend festhalten, dass dieser Kongress nicht zustande gekommen wäre ohne die Unterstützung und das Engagement vieler verschiedener Personen. Wieder haben sich engagierte Juroren aus der Professorenschaft, dem Mittelbau und aus dem Kreis der Studierenden gefunden, die die einzelnen Arbeiten mit viel Engagement und Sorgfalt begutachten. Herzlichen Dank dafür.

Darüber hinaus danken wir allen teilnehmenden Gruppen und den Dozenten, die durch die bunte Themenvielfalt, die kreativen Forschungsideen und nicht zuletzt durch die Freude am Austausch und dem Generieren neuer Ideen den Emprakongress prägen und unvergesslich machen.

Für den Gastvortrag zum Thema *Is it All in Your Head? Der Einfluss subjektiver Theorien auf die Selbstregulation* haben wir Dr. Veronika Job (Universität Zürich) gewinnen können, die außerdem als Jurorin die Begutachtung der Arbeiten unterstützt. Für diese Bereicherung des Kongressprogramms danken wir ihr herzlich.

Ein besonderer Teil des Emprakongresses ist die Auszeichnung und Prämierung der besten Arbeiten. Für die finanzielle Unterstützung bedanken wir uns beim Institut für Psychologie. Ganz besonders freuen wir uns, auch in diesem Jahr wieder Gutscheine für die Jenaer Universitätsbuchhandlung *Thalia* vergeben zu können und danken Herrn Dieter Hartendorf für die langjährige Unterstützung.

Des Weiteren danken wir den Studierenden aus der Fachschaft Psychologie für die Organisation und Betreuung des Getränkeverkaufs während des Kongresses.

Ein letztes großes Dankeschön geht an die Hilfskräfte des Lehrstuhls für Allgemeine Psychologie II, die sich stets in vielfältiger Weise an der Organisation beteiligen und damit einen reibungslosen Ablauf des Kongresses erst ermöglichen.

Für die Abteilung Allgemeine Psychologie II,  
Nicolas Koranyi, Florian Müller, Franziska Meißner

# Inhaltsverzeichnis

Gruppe 1	6
<i>Face a thousand Points of Light – Der Zusammenhang zwischen der Wahrnehmung Biologischer Bewegung und der Erkennung von Gesichtern</i>	
Lena Krolopp, Cosima Lukas, Marta Tomé	
Gruppe 2	8
<i>Dyskalkulie und Zeitwahrnehmung</i>	
<i>Ein Kurzbericht im Rahmen des empirischen Forschungsseminars „Die Entwicklung kognitiver Fertigkeiten aus neurowissenschaftlicher und klinischer Sicht“</i>	
Gyde Lena Berger, Paula Fehrmann, Julia Früh, Emily Schindler, Franziska Siblewski, Miriam Steinmann, Irina Barnaveli	
Gruppe 3	10
<i>Verarbeitung neutraler Adjektive im Verlauf einer Schmerzconditionierung und anschließender Extinktion</i>	
Luisa Bärwolf, Beatrix Eichert, Simone Fiedler, Kira Franke, Jolanda Krok, Daniel Rezmer	
Gruppe 4	12
<i>Konservative Psychologie-Studenten finden Trump besser als Statistik</i>	
Sandrine Hinrichs, Sina Hinzmann, Friederike Jakob, Clarissa Wendel, Robin Wester	
Gruppe 5	14
<i>„p-hacking: Zeig mir dein Handy und ich sag dir, wer du bist“</i>	
Anna El-Serazh, Alicia Hentschel, Lara Lietge, Elina Mizgir, Matthias Gehlert	
Gruppe 6	16
<i>“It don’t matter if you’re black or white?”</i>	
<i>Der Einfluss von Bekanntheit und Ethnie auf die kategoriale Wahrnehmung von Gesichteridentität</i>	
Isabelle Ewe, Philipp Laue, Vanessa Mathieu, Charlotte Schüßler, Alisa Shugaley	
Gruppe 7	18
<i>Der Einfluss von semantischer Information und Emotion auf die kategoriale Gesichtererkennung</i>	
Lukas Erchinger, Almut Hellwig, Fabian Holdmann, Max Meinhardt, Anna Pscheidt	
Gruppe 8	20
<i>Fake it `til you make it: Increasing the AMP Validity by Faking Instructions</i>	
Alexia Dalski, Lukas Knitter, Ole Kölbl, Toya Rosenkranz, Annika Vogel	
Gruppe 9	22
<i>Kriminalität in Deutschland - Wann ist eine harte Strafe gerechtfertigt?</i>	
<i>Soziale Funktionen von Bestrafung</i>	
Solveig Becker, Martin Jäger, Jannis Lobs, Aline Schönenberg, Dario Urban	
Gruppe 10	24
<i>Wenn ich ehrlich bin, dann lüg‘ ich nicht: Der Einfluss von Unglaube und Unruhe auf das Lügengedächtnis</i>	
Sara Bechtold, Benita Ebersbach, Jerome Hoffmann, Kristin Meißner, Lisa Stärker, Annika Stahn	

Gruppe 11	26
<i>Hör mal, wer da spricht!</i>	
<i>Der Einfluss von Distinktheit eines Sprechers auf die Erinnerungsleistung von Sprecher und Inhalt</i>	
Sarenka Allmacher, Charlotte Deckert, Madlen Schiewek, Theresa Strübe, Nadine Zscherper	
Gruppe 12	28
<i>Der erste Eindruck zählt?</i>	
<i>Wahrnehmung von Attraktivität bei Gesichtern und Stimmen</i>	
Christina Biesenbaum, Anne Jüngling, Johanna Menke, Anne-Maria Nagel, Ketrin Malvina Nikitina, Juliane Kröhn, Maria Helena Szabo, Ramona Schunke, Amelie Wild, Jasmin Wohlrab	
Gruppe 13	30
<i>Auswirkungen der Täter- und Opferzugehörigkeit auf Bestrafungstendenzen bei gruppenspezifischen Normverletzungen</i>	
Isabel Jaki, Theresa Leistner, Anna Lumaschi, Julia Müller, Christina Schulte	
Gruppe 14	32
<i>„Don't bias – be happy“ – Auswirkungen positiver Emotionen auf den Own-Race Bias</i>	
Klara Steinberg, Anna Kalinichenko, Camille Maennig, Pauline Gürtler	
Gruppe 15	34
<i>Ein Zeuge, der lügt, kennt die Wahrheit nicht – oder doch?</i>	
<i>Einfluss der Gedächtnisstärke auf den Lügen-Effekt</i>	
Theresa Epperlein, Nils Frisius, Stefanie Hunger, Sarah Jakob, Laura Sterba	
Gruppe 16	36
<i>Need for Speed – wie Risikokognitionen unsere Bestrafungstendenz beeinflussen</i>	
Paul Otto Löbl, Katharina Rinke, Rica Volmerich, Clara Ritter, Sophie Gräßer-Dietsch	
Gruppe 17	38
<i>“Ebony and Ivory live together in holistic harmony”:</i>	
<i>Der Einfluss von Ethnie und Bekanntheit auf die Composite Face Illusion</i>	
Gesine Franke, Bianka Knorr, Alexandra Kruppa, Katharina Kühne, Nandiyin Nandiyin	
Gruppe 18	40
<i>“I've been looking for freedom” – Effekte des IAT auf den freien Willen, Kontrollüberzeugungen und aggressives Verhalten</i>	
Paula Kliefert, Wilma Middendorf, Stella Pfund, Deborah Löschner, Alexander Fuchs	
Gruppe 19	42
<i>„Und führe uns nicht in Versuchung...“ -</i>	
<i>Implizites ‚wanting‘ und ‚liking‘ für Ziele und Versuchungen m Studienkontext</i>	
Luisa Borsche, Jennifer Graf, Fenna Plume, Maja Sindermann, Laura Wodara	
Gruppe 20	44
<i>Du bist nicht du, wenn du gestresst bist! Wie Stress das Entscheidungsverhalten bei Männern und Frauen beeinflusst</i>	
Franziska Sophia Roller, Rico Stecher, Heide Wenzel	
Gruppe 21	46
<i>Das aufgePEPte Reaktionszeitexperiment, does sex really sell?</i>	
<i>Einfluss von Stereotypen auf die Bewerberauswahl</i>	
Julian Kern, Doreen Krall, Tobias Kufner, Laura Luxen, Christiane Petzold, Delina Ubben	

# **Face a thousand Points of Light – Der Zusammenhang zwischen der Wahrnehmung Biologischer Bewegung und der Erkennung von Gesichtern**

**Lena Krolopp, Cosima Lukas, Marta Tomé**

Leitung: Prof. Dr. Christian Dobel

## 1. Einleitung

Die Fähigkeit der Gesichtserkennung stellt einen zentralen Bestandteil der menschlichen Kommunikation dar und gelingt meist mühelos. So können eine Vielzahl unterschiedlicher Gesichter unterschieden, erinnert oder wiedererkannt werden. Liegt eine isolierte Störung der Fähigkeit zur Gesichtserkennung vor, bei ansonsten intakten sensorischen Fähigkeiten, spricht man von Prosopagnosie (Behrmann & Avidan, 2005). Im Fokus dieser Untersuchung steht ausschließlich die kongenitale Prosopagnosie, bei der keine organischen Pathologien vorliegen. Diese kann einerseits als ein stimulus-spezifisches Defizit hinsichtlich der Verarbeitung von Gesichtern betrachtet werden. Andererseits könnte ihr jedoch auch ein eher generelles Wahrnehmungsdefizit zugrunde liegen, das sich auf ein breiteres Spektrum von Reizen auswirkt. Dafür sprechen beispielsweise Befunde, in denen Patienten, die unter Prosopagnosie leiden, ebenfalls Beeinträchtigungen in der Wahrnehmung von Reizen zeigten, deren globale Form aus mehreren kleinen, lokalen Elementen zusammengesetzt ist (Behrmann & Avidan, 2005), oder in denen Unregelmäßigkeiten bei der Wahrnehmung statischer menschlicher Körper aufgetreten sind (B. C. Duchaine & K. Nakayama, 2006). Auch der Befund, dass gesichtssensitive Hirnareale bei Personen mit kongenitaler Prosopagnosie eine normale Aktivierung für Gesichter zeigen können (Steeves et al., 2006), spricht für eine Dysfunktion innerhalb eines weitaus komplexeren Netzwerkes der Gesichtserkennung.

Unter der Annahme eines weitergefassten Wahrnehmungsdefizits und in Anlehnung an frühere Experimente (Lange et al., 2009), beschäftigt sich diese Studie mit einem möglichen Zusammenhang zwischen kongenitaler Prosopagnosie und der Wahrnehmung Biologischer Bewegung. Im Vergleich zu Gesichtern stellt ein Point Light Walker, mit dessen Hilfe in unserem Experiment die Erkennungsleistung biologischer Bewegung erfasst wurde, eine völlig gegenteilige Art der Übermittlung visueller Stimuli dar, da die optischen Reize auf ein sehr spärliches Maß reduziert werden. Zusätzlich handelt es sich hierbei um eine dynamische Darstellung. Vor dem Hintergrund eines gemeinsamen Inversionseffektes (Reed, Stone, Bozova, & Tanaka, 2003), der das Phänomen einer erschwerten konfiguralen Verarbeitung bei auf dem Kopf stehenden Stimuli darstellt, macht gerade diese Diskrepanz die Biologische Bewegung als zu vergleichenden Reiz relevant. Das Ziel unserer Arbeit ist es zu untersuchen, ob die Defizite der Prosopagnosie lediglich auf die Erkennung von Gesichtern einzugrenzen sind oder vielmehr ebenfalls die Wahrnehmung biologischer Bewegung beeinträchtigen. Besteht ein Zusammenhang zwischen den Leistungen in beiden Wahrnehmungsaufgaben, so könnte dies auf einen gemeinsamen perzeptuellen Mechanismus hinweisen und damit auch für ein wesentlich umfassenderes Wahrnehmungsdefizit bei Prosopagnosie sprechen.

## 2. Methode

Da früheren Untersuchungen nur eine sehr kleine und selektive Stichprobe zugrunde lag, soll die Wahrnehmung von Gesichtern in Zusammenhang mit biologischer Bewegung im Rahmen einer korrelativen Studie, anhand einer größeren Stichprobe, erhoben werden. Im Rahmen unserer Studie nahmen 128 ProbandenInnen (69% ♀, 28% ♂, 3% n.a., 89,2% Rechtshänder, Bildungsabschluss: 91% mit Abitur oder höher) an einer Online-Befragung teil. Das durchschnittliche Alter der Teilnehmer betrug 23,4 Jahre (SD = 4,8). 56 Personen brachen das Experiment vorzeitig ab.

Anschließend wurden den Probanden 2 Aufgabentypen präsentiert: Die Fähigkeit zur Erkennung bekannter Gesichter wurde mithilfe des Famous Faces Tests (B. Duchaine & K. Nakayama, 2006) erfasst. Die Erkennungsleistung biologischer Bewegung wurde im Rahmen einer Coherence/Incoherence Discrimination Task erfasst, bei der den Probanden zuvor mitgeteilt wurde, dass sie nachfolgend einen sich vorwärts bewegendem, Point Light Walker sehen werden und anschließend entscheiden sollen, ob die obere und untere Körperhälfte der dargestellten Figur in dieselbe oder entgegengesetzte Richtung orientiert sind. Eine Hälfte von insgesamt 44 Trials zeigte kongruente Bedingungen, die andere Hälfte inkongruente Bedingungen. Um die Vertrautheit der Stimuli zu gewährleisten, wurde zunächst jeweils ein Beispielitem für jede Bedingung gezeigt.

### 3. Ergebnisse

Famous Faces Test:  $\bar{x} = 19,2$ ;  $SD = 10,4$ ; range 0- 43 richtig gelöste Items. Die Pearson's Product-Moment Korrelation der Testleistung mit dem Alter der Probanden war signifikant positiv und betrug  $r = ,23$  mit  $p = ,011$ . Erkennen biologischer Bewegung durch Diskriminierung kohärenter und inkohärenter Bewegung:  $\bar{x} = 32,33$ ;  $SD = 6,27$ ; range 12- 40 korrekt erkannte Items. Die Testleistung korreliert nicht signifikant mit dem Alter ( $r = -,055$ ;  $p = ,538$ ).

Die Korrelation der Testleistung zwischen beiden Aufgabentypen betrug  $r = ,063$  und ist mit einem p-Wert von  $p = 0,481$  nicht signifikant.

### 4. Diskussion

Auf der Grundlage der Daten kann keine Korrelation zwischen der Erkennung von Gesichtern und dem Erkennen von biologischer Bewegung abgeleitet werden. Die Hypothese, dass beiden perzeptuellen Aufgaben eine ähnliche Art der Wahrnehmung und Verarbeitung zugrunde liegt und die Prosopagnosie über stimulus-spezifische Aspekte hinaus, zumindest teilweise größere Wahrnehmungsbereiche beeinträchtigt, kann - entgegen der Ergebnisse früherer Studien (Lange et al., 2009) - nicht bestätigt werden. Vor dem Hintergrund des überwiegend weiblichen Teilnehmeranteils, sowie des eher homogenen Durchschnittsalters, kann lediglich von einer bedingten Repräsentativität der Stichprobe ausgegangen werden. In zukünftigen Untersuchungen sollte zudem darauf geachtet werden, dass eine aktualisierte Form des Famous Faces Tests verwendet wird, um sicherzustellen, dass die Fähigkeiten zur Gesichtererkennung in der untersuchten Probandengruppe optimal abgebildet werden. Besonders bei einer Online-Studie tritt das Problem einer hohen Abbruchrate auf. Dies könnte die vorliegenden Daten dahingehend verzerrt haben, dass gerade Probanden, die keines der Gesichter erkannten oder nur wenige Aufgaben zur biologischen Bewegung richtig lösen konnten, zum vorzeitigen Abbruch der Studie neigten.

### 5. Literatur

- Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends in Cognitive Sciences*, 9(4), 180-187. doi:10.1016/j.tics.2005.02.011
- Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4), 576-585. doi:10.1016/j.neuropsychologia.2005.07.001
- Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: a window to content-specific face processing. *Curr Opin Neurobiol*, 16(2), 166-173. doi:10.1016/j.conb.2006.03.003
- Lange, J., de Lussanet, M., Kuhlmann, S., Zimmermann, A., Lappe, M., Zwitserlood, P., & Dobel, C. (2009). Impairments of biological motion perception in congenital prosopagnosia. *PLoS One*, 4(10), e7414. doi:10.1371/journal.pone.0007414
- Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychol Sci*, 14(4), 302-308. doi:10.1111/1467-9280.14431
- Steeves, J. K., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., . . . Goodale, M. A. (2006). The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia*, 44(4), 594-609. doi:10.1016/j.neuropsychologia.2005.06.013

## **Dyskalkulie und Zeitwahrnehmung**

### **Ein Kurzbericht im Rahmen des empirischen Forschungsseminars „Die Entwicklung kognitiver Fertigkeiten aus neurowissenschaftlicher und klinischer Sicht“**

**Berger G. L., Fehrmann P., Früh J. N., Schindler E. I. E., Siblewski F., Steinmann M., Barnaveli I.**

Leitung: Univ. Prof. Dr. Christian Dobel

#### 1. Einleitung

Wie lang ist der Weg zur nächsten Bushaltestelle? Stelle ich meinen Wecker auf 6.15 Uhr oder doch 15.06 Uhr? All dies sind Fragen, welche für die Bewältigung unseres Alltags relevant und normalerweise einfach zu beantworten sind. Zirka 4,4% - 6,7% der deutschen Bevölkerung leiden jedoch unter Dyskalkulie (Jacobs & Petermann, 2003). Diese zählt zu den Entwicklungsstörungen und besteht in einer umschriebenen Beeinträchtigung von Rechenfertigkeiten, die nicht allein durch eine allgemeine Intelligenzminderung oder eine unangemessene Beschulung erklärbar ist. Das Defizit betrifft vor allem die Beherrschung grundlegender Rechenfertigkeiten, weniger die Höheren ("ICD-10-GM-2017 F81.2 Rechenstörung - ICD10," 2017). Durch diese Störung könnte auch das Schätzen von Zeit ein Problem darstellen. Im Fokus der Forschung liegt vor allem die Thematik Dyskalkulie im Schulkontext, im Erwachsenenalter ist sie jedoch kaum erforscht. Aus diesem Grund entschlossen wir uns die Studie „time perception in children with developmental dyscalculia“ zur Zeitwahrnehmung bei Kindern mit Dyskalkulie mit Erwachsenen zu replizieren (Pellerone, 2013). Die Ergebnisse der Studie implizierten, dass Kinder mit Dyskalkulie, im Vergleich zu anderen Kindern, Defizite in der Zeitschätzung aufwiesen.. Deshalb interessieren wir uns speziell dafür, ob Erwachsene, die an Dyskalkulie leiden, Strategien entwickelt haben um ihre Defizite in diesem Bereich auszugleichen.

#### 2. Methode

Die Experimentalgruppe setzte sich aus zwei Personen, welche an Dyskalkulie leiden (Alter in Jahren  $M = 26 \pm 1.3$ ; Schulabschluss in Jahren  $M = 11 \pm 1.4$ ; letzte Mathenote in Punkten  $M = 2.5, \pm 3.5$ ), zusammen. Die Kontrollgruppe bestand aus 24 Personen (11 Männer, 12 Frauen, 1 Person ohne Angabe; Alter in Jahren  $M = 25 \pm 3.6$ ; Schulabschluss in Jahren  $M = 13.9 \pm 1.9$ , letzte Mathenote in Punkten  $M = 9.7 \pm 3.6$ ). Es wurde mit allen Probanden der MWT-B durchgeführt (Lehrl S., 2005). Alle Partizipierenden wiesen einen durchschnittlichen Intelligenzquotient auf. Das anschließend durchgeführte Experiment bestand aus einer Zeitvergleichsaufgabe. Den Teilnehmenden wurde auf einem Computerbildschirm ein Standardreiz (roter Kreis) präsentiert. Danach erschien ein Teststimulus (schwarzer Kreis) und die Aufgabe bestand darin einzuschätzen, ob das Testintervall (Darbietungszeit des schwarzen Kreises) länger oder kürzer war als das Referenzintervall (Präsentationsdauer des roten Kreises) (Pellerone, 2013). Die Versuchsperson sollte nun durch das Drücken der Tasten D und L angeben, ob der Testreiz länger oder kürzer als der Standardreiz präsentiert wurde. Die Belegung der Tasten wurde zwischen den Personen randomisiert. Die Aufgabe setzte sich aus zwei separaten und konsekutiven Blocks zusammen. Zusätzlich unterschied man zwischen zwei Bedingungen. Die erste bestand aus „sub-second“ Intervallen, das heißt Darbietungszeiten von unter einer Sekunde (Testreiz: 300, 500 ms und Standardreiz: 400 ms). Bedingung zwei („supra-second“) hingegen beinhaltete Zeitintervalle von über einer Sekunde (Testreiz: 1200, 1600 ms, Standardreiz: 1400 ms). Beide Blöcke bestanden aus jeweils 60 Durchgängen, in denen beide Bedingungen in randomisierter Reihenfolge präsentiert wurden. Der Standardstimulus wird in allen Durchgängen unmittelbar vor dem Teststimulus

präsentiert. Dieser ist in 50% der Fälle länger beziehungsweise kürzer als der Referenzreiz.. Erhoben wurden die Antwortgenauigkeit sowie die Reaktionszeit.

### 3. Ergebnisse

Ausgewertet wurde mittels einer ANOVA mit den Faktoren „sub-“ vs. „suprasecond“, kürzere vs. längere Präsentation des Testreizen im Vergleich zum Standardreiz, sowie die Reaktionszeit. Es zeigten sich zwei signifikante Haupteffekte in der Kontrollgruppe. Wurde der Standardreiz länger als eine Sekunde präsentiert, benötigten die Versuchspersonen signifikant länger als in der „sub-second“ Bedingung ( $F(1, 23) = 31.9, p < .001$ ). Zudem war die Reaktionszeit signifikant länger, wenn der Testreiz kürzer als der Standardreiz präsentiert wurde ( $F(1, 23) = 28.1, p < .001$ ). Eine signifikante Interaktion konnte nicht gefunden werden. Eine ANOVA der Faktoren „sub-“ vs. „suprasecond“, kürzere vs. längere Präsentation des Testreizen im Vergleich zum Standardreiz und der Antwortgenauigkeit wies auf einen Haupteffekt in der Gruppe Testreiz kürzer vs. länger im Vergleich zum Standardreiz hin. War die Präsentationszeit des Testreizes geringer als die des Standardreizes fiel die Antwortgenauigkeit geringer aus ( $F(1, 23) = 6.50, p < .002$ ). Aufgrund der mangelnden Anzahl an Probanden in der Experimentalgruppe, wurden diese Fälle mittels einer Single-Case-Statistik ausgewertet. Es zeigte sich kein signifikanter Unterschied bezüglich der Antwortgenauigkeit zwischen Kontrollgruppe und Experimentalgruppe. Bei Betrachtung der Reaktionszeit unterschieden sich beide Gruppen signifikant. Person 1 der Experimentalgruppe wies in der Bedingung „suprasecond“ eine überzufällig längere Reaktionszeit im Vergleich zur Kontrollgruppe auf („suprasecond“, Testreiz kürzer als Standardreiz: einseitige Testung  $p < .03$ , zweiseitige Testung  $p = .05$ ; „suprasecond“, Testreiz länger als Standardreiz: einseitige Testung  $p < .01$ , zweiseitige Testung  $p < .02$ ). Person 2 der Experimentalgruppe wies ebenfalls in der Bedingung „suprasecond“ eine signifikant längere Reaktionszeit als die Kontrollgruppe auf („suprasecond“, Testreiz kürzer als Standardreiz: einseitige Testung  $p < .02$ , zweiseitige Testung  $p < .04$ ; „suprasecond“, Testreiz länger als Standardreiz: einseitige Testung  $p < .03$ , zweiseitige Testung  $p < .05$ ). Zudem zeigte Person 2 in der „subsecond“ Bedingung einen signifikanten Effekt, wenn der Testreiz kürzer als der Standardreiz präsentiert wurde (einseitige Testung  $p < .001$ , zweiseitige Testung  $p < .02$ ).

### 4. Diskussion

Die Ergebnisse unterscheiden sich somit von der replizierten Studie. Menschen die an Dyskalkulie leiden, machen bei der Zeitschätzung nicht mehr Fehler, sondern haben längere Reaktionszeiten. Für zukünftige Studien wäre eine größere Experimentalgruppe erstrebenswert, um repräsentativere Ergebnisse zu erzielen. Das Thema der Dyskalkulie im Erwachsenenalter ist trotz hoher Alltagsrelevanz wenig erforscht, weshalb weitere Forschung in diesem Bereich anzustreben ist.

### 5. Literatur

- Frassinetti, F., Magnani, B., & Oliveri, M. (2009). Prismatic lenses shift time perception. *Psychological Science*, 20, 949–954.
- ICD-10-GM-2017 F81.2 Rechenstörung - ICD10. (2017). Retrieved from <http://www.icd-code.de/icd/code/F81.2.html>
- Jacobs, C., & Petermann, F. (2003). Dyskalkulie - Forschungsstand und Perspektiven. *Kindheit und Entwicklung*, 12(4), 197–211. <https://doi.org/10.1026//0942-5403.12.4.197>
- Koch, G., Oliveri, M., Carlesimo, G. A., & Caltagirone, C. (2002). Selective deficit of time perception in a patient with rightprefrontal cortex lesion. *Neurology*, 59, 1658–1659.
- Lehrl S. (2005). Mehrfachwahl-Wortschatz-Intelligenztest MWT-B.
- Pellerone, M. (2013). Time Perception in Children with Developmental Dyscalculia. *Procedia - Social and Behavioral Sciences*, 103, 1220–1227. <https://doi.org/10.1016/j.sbspro.2013.10.450>
- Tregellas, J. R., Davalos, D. B., & Rojas, D. C. (2006). Effect of task difficulty on the functional anatomy of temporal processing. *NeuroImage*, 32, 307–315.

## Verarbeitung neutraler Adjektive im Verlauf einer Schmerzconditionierung und anschließender Extinktion

Luisa Bärwolf, Beatrix Eichert, Simone Fiedler, Kira Franke, Jolanda Krok, Daniel Rezmer

Leitung: Prof. Dr. Thomas Weiß

### 1. Einleitung

Wie hängt Schmerzwahrnehmung mit semantischem Material zusammen? Befunde zeigen, dass schmerzbezogene Wörter eine erhöhte Aktivierung gegenüber positiven und negativen Wörtern in bekannten Clustern der Schmerzmatrix (z.B. DLPFC, linker IPC) erzeugen können (Richter, Eck, Straube, Miltner, & Weiss, 2010). Schmerzbezogene Wörter sind also in der Lage, bestimmte Areale der kortikalen Schmerzmatrix zu bahnen. Außerdem konnten Montoya, Larbig, Pulvermüller, Flor & Birbaumer (1996) erhöhte kortikale ereigniskorrelierte Potentiale bei der visuellen Präsentation von Pseudowörtern, die zuvor mit einem schmerzhaften elektrischen Schock gepaart wurden, beobachten. Beide Studien weisen damit darauf hin, dass neutrales verbales Material durch Kopplung mit Schmerz zu „schmerzvollem“ werden kann und sich dies in einer Aktivierung bestimmter Hirnareale zeigt.

Unser Ziel ist zu untersuchen, inwieweit bei ursprünglich neutralen Wörtern, die im Laufe des Experiments durch Assoziation mit Schmerzreizen voraussichtlich eine Valenzänderung erfahren, Wortwahrnehmung und Schmerzverarbeitung miteinander in Beziehung stehen. Dazu wurden folgende Hypothesen aufgestellt:

- I) Durch die Paarung eines Sets von Adjektiven mit Schmerzstimuli verändert sich die Valenz der Adjektive in Abhängigkeit der Experimentalphase (Konditionierung, Extinktion).
- II) Im Vergleich zu den nicht konditionierten Adjektiven werden bei der Präsentation der konditionierten Wörter in einer ersten Extinktionsphase differente Hirnregionen aktiviert. Potentiell gehören dazu Wortverarbeitungs- und schmerzassoziierte Gehirnregionen. Bei fortschreitender Extinktion nimmt die Hirnaktivierung ab.

### 2. Methode

Es wurden 33 rechtshändige Versuchspersonen (VP) ohne neurologische oder psychiatrische Vorerkrankungen untersucht, von denen 19 ausgewertet werden konnten. Bei dem Experiment handelte es sich um eine fMRT Studie. Diese teilte sich in die *Phasen* Lernen, Konditionierung und Extinktion 1-4 auf. Zu Beginn wurde bei den VP eine Stimulationselektrode intrakutan am linken Mittelfinger angebracht. Die individuelle Wahrnehmungs- und Schmerzschwelle wurden bestimmt. In der Lernphase wurden den Probanden 40 Adjektive, deren Neutralität zuvor geprüft worden war, präsentiert. Diese sollten möglichst gut gelernt werden. Die Valenz der Wörter wurde von den VP auf einer 9-stufigen Skala eingeschätzt (1 = sehr angenehm, 5 = neutral, 9 = sehr unangenehm). Anschließend wurden zwei Probereize oberhalb der Schmerzschwelle verabreicht, deren Stärke die VP einschätzen sollten. In der Phase der Konditionierung wurden die 40 Adjektive der Lernphase ein weiteres Mal gezeigt, wobei 20 mit Schmerzreizen gepaart wurden, die andere Hälfte wurde ohne Schmerzreize präsentiert (*Wortgruppe*: konditioniert vs. unkonditioniert). Es gab 2 Konditionierungssets, die den VP abwechselnd zugeordnet wurden. Die Valenz der Wörter wurde erneut abgefragt. In den folgenden vier Extinktionsphasen wurden den bekannten 40 Adjektiven jeweils 40 weitere hinzugefügt. In allen vier Phasen wurde der Lernerfolg geprüft, indem die VP das jeweils gezeigte Wort als gelernt vs. nicht gelernt einordneten. Des Weiteren erfolgte in Extinktion 1 und 4 eine Valenzabfrage. In Extinktion 2 und 3 wurde hierauf und auf MRT-

Aufnahmen aus Zeitgründen verzichtet. Schließlich hatten die VP 90 Sekunden Zeit, sich an so viele Adjektive wie möglich zu erinnern und diese aufzuschreiben.

In der folgenden Verhaltensanalyse wurde für jede VP über die vier relevanten Phasen hinweg jeweils die mittlere Valenz der konditionierten und der unkonditionierten Adjektive mit Hilfe einer zweifaktoriellen Varianzanalyse mit Messwiederholung und den Faktoren *Phase* und *Wortgruppe* ausgewertet. Außerdem wurden auf Ebene der Verhaltensdaten, Phasen und wortgruppenspezifische Kontraste mittels (paarweisen) t-tests gerechnet und auf Ebene der fMRT-Daten Kontraste zwischen den beiden Wortgruppen in der Extinktionsphase betrachtet.

### 3. Ergebnisse

In der Varianzanalyse wurden keine signifikanten Haupteffekte für *Phase* ( $F = 2.671$ ,  $p = 0.058$ ) oder *Wortgruppe* gefunden ( $F = 0.183$ ,  $p = 0.675$ ), die Interaktion der beiden Faktoren hatte einen signifikanten Einfluss auf die durchschnittliche Valenzbewertung ( $F = 4.808$ ,  $p = 0.005$ ).

In der folgenden post-hoc Analyse für spezifische Kontraste wurde ein signifikanter Unterschied im Kontrast der Veränderung der beiden Wortgruppen von Lernen zu Konditionierungsphase gefunden ( $t = -2.982$ ,  $p = 0.009$ ). Der Vergleich der durchschnittlichen Valenzratings der beiden Wortgruppen in der Extinktionsphase erbrachte aber keine signifikanten Ergebnisse ( $t = -0.037$ ,  $p = 0.971$ ). Im fMRT fand sich in der Konditionierungsphase bei den mit Schmerzreiz gepaarten Wörtern im Kontrast zu den unkonditionierten Wörtern eine signifikant ( $q(\text{FDR}) < 0.010$ ,  $p = 0.0007$ ) höhere Aktivierung der neuronalen Schmerzmatrix (ACC, Insel, S2), Teilen des medialen PFC und des höheren visuellen Areal.

Der Vergleich zwischen Lernen und Extinktionsphase für beide Wortgruppen erbrachte auch in den fMRT-Daten keine signifikanten Ergebnisse.

### 4. Diskussion

In der Studie wurde eine signifikante Interaktion *Phase* x *Wortgruppe* der Valenzbewertung nachgewiesen, wobei diese Interaktion durch Veränderungen der Valenz während der Konditionierung, nicht aber in den Extinktionsphasen zustande kommt. Übereinstimmend dazu zeigte sich auch in den Ergebnissen der fMRT-Daten im Unterschied zu den Befunden von Richter et al. (2010) Aktivierungen nur in der Konditionierungsphase, aber keine differente Aktivierung schmerzbezogener Areale zwischen erster Extinktionsphase und Lernphase für die konditionierten im Vergleich zu den unkonditionierten Adjektiven. Dies steht auch in Widerspruch zu einer Untersuchung mit einmaliger Kopplung zwischen Hitzeschmerz und Bildern (Forkmann, Wiech, Sommer & Bingel, 2015).

Neben der relativ klein ausfallenden Stichprobe sind die Ergebnisse vermutlich auf eine schwache Effektstärke bei einmaliger Kopplung zurückzuführen. Zudem könnte das Paradigma nicht optimal sein, da die einmalige Präsentation der Adjektive keinen prädiktiven Wert hat beziehungsweise den gleichen prädiktiven Wert wie die Phase an sich (Kontextkonditionierung). Ein weiteres Problem stellt die Neutralität der Adjektive dar, welche wegen interindividuell variierender Bewertungen durch Probanden nicht zwangsweise gegeben war.

Die einmalige Assoziation von neutralen Wörtern und Schmerzstimulation verändert die Wahrnehmung der Wörter (Valenz, kortikale Verarbeitung) nur während der Assoziation, nicht aber in der folgenden Extinktion. Darin unterscheiden sich möglicherweise Wort- und Bildverarbeitung.

### 5. Literatur

- Forkmann, K., Wiech, K., Sommer, T., & Bingel, U. (2015). Reinstatement of pain-related brain activation during the recognition of neutral images previously paired with nociceptive stimuli. *Pain*, 156(8):1501-10.
- Montoya, P., Larbig, W., Pulvermüller, F., Flor, H., & Birbaumer, N. (1996). Cortical correlates of semantic classical conditioning. *Psychophysiology*, 33(6), 644-649.
- Richter, M., Eck, J., Straube, T., Miltner, W. H., & Weiss, T. (2010). Do words hurt? Brain activation during the processing of pain-related words. *Pain*, 148(2), 198-205.

# **Konservative Psychologie-Studenten finden Trump besser als Statistik**

**Sandrine Hinrichs, Sina Hinzmann, Friederike Jakob, Clarissa Wendel, Robin Wester**

Leitung: Andreas Neudecker

## 1. Einleitung

Konservativ eingestellte Menschen sind im Allgemeinen weniger interessiert an der Wissenschaft und vertrauen ihr weniger als Liberale. Diese wissenschaftskritische Haltung von konservativen Menschen konnte durch Studien belegt werden (Carl & Cofnas, 2016).

Donald Trump steht für Protektionismus (Medick & Pitzke, 2015), was ein typisches Merkmal von konservativen Politikern darstellt (engl. protection=beschützen; lat. conservare=bewahren). Ebenso ist durch bspw. die Ankündigung seines Vorhabens, „Budgets für Forschungsprogramme zusammenzustreichen“ (Timo Peitsch, 2017) deutlich geworden, dass Donald Trump wissenschaftskritisch, wenn nicht sogar wissenschaftsverleugnend eingestellt ist, was die eben genannten Befunde bestätigen und verdeutlichen.

Auf der anderen Seite zeigen Studien, dass unter Psychologiestudenten das Fach Statistik unbeliebt ist (Williams, 2015). Da eine ablehnende Haltung gegenüber Wissenschaft und Statistik Hand in Hand gehen, stellte sich für uns die spannende Frage: Sind Psychologiestudenten gegenüber Trump positiver eingestellt als gegenüber Statistik? Die Hypothese, die sich aus dieser Fragestellung ergab, konnte durch unser Experiment bestätigt werden.

## 2. Methode

In unserem Experiment nahmen 40 Psychologie-Studenten (26 weiblich, 14 männlich) zunächst an einem zehnminütigen Computer-Experiment teil. Valente Adjektive sollten per Tastendruck als positiv oder negativ kategorisiert werden, wobei vorher im Zuge des hier angewendeten affektiven Primings (Augustine, Larsen & Elliot, 2013) für 12 Millisekunden entweder ein Bild von Donald Trump (Trump-Bedingung) oder das einer statistischen Formel bzw. eines Statistik-Dozenten der FSU Jena (Statistik-Bedingung) erschien. Anschließend füllten sie einen Fragebogen zum Thema Konservatismus aus (Schiebel, Riemann & Mummendey, 1983).

Unsere Hypothese des Experiments war nun, dass konservative Psychologie-Studenten schneller positive Wörter korrekt zuordnen, die mit Trump-Bildern präsentiert werden, als solche, die einem mit Statistik assoziiertem Bild folgen.

Zur Auswertung der Daten ermittelten wir für jeden einzelnen Probanden die Mediane seiner Reaktionszeiten auf positive Wörter für jeweils beide Bedingungen. Der Median wurde statt des Mittelwerts genutzt, um den Einfluss von Ausreißerwerten zu verringern.

## 3. Ergebnisse

Die Ergebnisse unseres Experiments konnten unsere Hypothese bestätigen.

Durch einen Zweigruppen-t-Test stellten wir fest, dass der Mittelwert der Reaktionszeit zur Trump-Bedingung beim konservativsten Drittel unserer Probanden signifikant ( $p$ -Wert=0,038) kleiner war als der Mittelwert der Statistik-Bedingung. Daraus schlussfolgern wir, dass bei konservativen Psychologie-Studenten Donald Trump besser anzukommen scheint als Statistik.

## 4. Diskussion

Bei den von uns berichteten Ergebnissen handelt es sich in Wahrheit um ein statistisches Artefakt, das wir mittels  $p$ -hacking ‚gefunden‘ haben. Das eigentliche Ziel unserer Studie war, die Auswirkungen von  $p$ -hacking zu veranschaulichen, wobei wir uns von dem in der Wissenschaft

teilweise verbreiteten Leitsatz „The more startling the claim, the better“ (Young & Karr, 2011, S. 116) leiten ließen und mit Donald Trump ein kontroverses Thema wählten.

Es ist ein mittlerweile bekanntes Problem, dass viele Studien, v.a. solche mit umwälzenden Ergebnissen, nicht repliziert werden können. Dies führt manche Wissenschaftler zur Einschätzung, dass die Mehrzahl der veröffentlichten Studien falsche Ergebnisse berichtet (Ioannidis, 2005). *p*-hacking ist in großen Teilen verantwortlich für dieses Problem, wie Untersuchungen der Verteilung von *p*-Werte zeigen können.

*p*-hacking bezeichnet im weitesten Sinne die Erzeugung von signifikanten Ergebnissen, ohne dass Daten direkt gefälscht werden. Es geschieht, sobald „researchers collect or select data or statistical analyses until nonsignificant results become significant“ (Head et al., 2015, S. 1) und kann demnach durch eine Vielzahl von Methoden erreicht werden.

In unserer Untersuchung setzten wir das multiple Testen ein: Wir legten nicht vor der Untersuchung fest, wann Probanden als konservativ gelten, sondern rechneten *t*-Tests für alle Konservatismus-Werte (Abb. 1) und wählten diejenige Gruppe aus, deren Ergebnis signifikant wurde. Die Wahrscheinlichkeit, dass für eine dieser 39 Gruppen ein zu einem Alpha-Niveau von 0.05 signifikantes Ergebnis erfolgt, steigt damit eklatant und verletzt das eigentliche Prinzip des „null hypothesis significance testing“ nach Fischer (Head et al., S. 2).

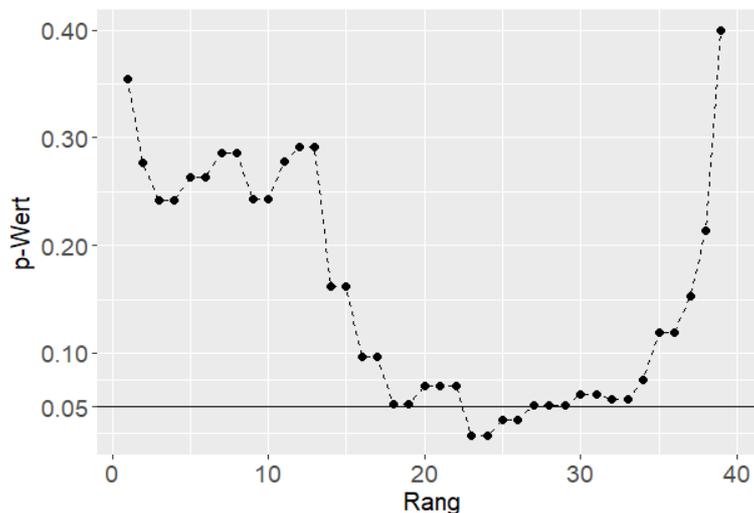


Abbildung 1: Verteilung der *p*-Werte bei unterschiedlicher Größe der als konservativ bezeichneten Gruppe. Rang 10 bedeutet, dass nur Personen ab einem Konservatismus-Rang von 10 in die Gruppe aufgenommen wurden.

## 5. Literatur

- America first, America first: Donald Trump ist neuer US-Präsident. (2017, Juni 19). Abgerufen am 19. Juni 2017. Von: <http://www.n-tv.de/politik/Donald-Trump-ist-neuer-US-Praesident-article19607557.html>
- Augustine, A., Larsen, R., & Elliot, A. (2013). Affect is greater than, not equal to, condition: Condition and person effects in affective priming paradigms. *Journal of Personality*, 81(4), 355–364.
- Carl, N., & Cofnas, N. (2016). Scientific literacy, optimism about science and conservatism. *Personality and Individual Differences*, 94, 299–302.
- Head, Megan L., Holman, Luke, Lanfear Rob, Kahn, Andrew T. & Jennions, Michael D. (2015). The Extent and Consequences of P-Hacking in Science. *PLOS Biology*, 13(3), e1002106.
- Ioannidis, John P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8), e124.
- Medick, V. & Pitzke, M. (2015, 8. August). Das Programm des Republikaners: Trumps Welt. *Spiegel Online*, abgerufen am 19. Juni 2017, von <http://www.spiegel.de/politik/ausland/donald-trump-so-saehen-die-usa-unter-praesident-trump-aus-a-1048913.html>
- Peitsch, T. (2017, 22. April). Demos in Städten weltweit: Marsch für die Wissenschaft - und gegen Trump. *heute-Nachrichten*, abgerufen am 19. Juni 2017, von <http://www.heute.de/march-for-science-marsch-fuer-die-wissenschaft-und-gegen-trump-47015110.html>
- Schiebel, B., Riemann, R., & Mummendey, H. (1984). Eine aktualisierte deutschsprachige Form der Konservatismus-Skala von Wilson & Patterson. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 5(4), 311-321.
- Williams, A. S. (2015). Statistics Anxiety and Worry: The Roles of Worry Beliefs, Negative Problem Orientation, and Cognitive Avoidance. *Statistics Education Research Journal*, 14(2), 53–75.
- Young, S. Stanley, Karr, Alan (2011). Deming, data and observational studies. *Significance*, 8(3), 116-120.

## „p-hacking: Zeig mir dein Handy und ich sag dir, wer du bist“

Anna El-Serazh, Alicia Hentschel, Lara Lietge, Elina Mizgir, Matthias Gehlert

Leitung: Andreas Neudecker

### 1. Einleitung

In der heutigen, digitalisierten Gesellschaft ist ein Leben ohne Handy unvorstellbar. Es ist unser alltäglicher Begleiter und kennt uns teilweise sogar besser als unsere Familie. Doch wie viel von unserer Persönlichkeit steckt in diesem einfachen Gegenstand? Was verrät unser Handy über uns?

Die Persönlichkeit steht seit vielen Jahren schon im Fokus der Forschung. So konnten bereits einige Variablen identifiziert werden, die Persönlichkeitsmerkmale vorhersagen. Trägt nun vielleicht auch das Handy zur Varianzaufklärung bei Persönlichkeitsunterschieden bei? Oder reicht, noch provokanter formuliert, bereits ein Blick auf das gewählte Handyhintergrundbild aus, um die Persönlichkeit eines Menschen in all ihren Facetten zu erfassen?

*Hypothese 1.* Eine glückliche Beziehung gehört wohl zu den wichtigsten Zielen des menschlichen Zusammenlebens. Soziale Medien bieten eine Plattform um diese Gefühle mit anderen zu teilen. So konnte in empirischen Untersuchungen bereits ein positiver Zusammenhang zwischen der Darstellung des Partners auf Facebook und der eigenen Beziehungszufriedenheit gezeigt werden (Papp, Danielewicz & Cayemberg, 2012). Wir gehen jetzt noch einen Schritt weiter und vermuteten einen positiven Zusammenhang zwischen Beziehungszufriedenheit und der Wahl des Partners als Handyhintergrundbild im Vergleich zu anderen Hintergrundbildern.

*Hypothese 2.* Kent et al. (2001) haben in ihrer Studie unter anderem untersucht, welche Folgen die Angst vor negativer Bewertung für Patienten mit Psoriasis (dermatologische Erkrankung) haben kann. Dabei wurde die Angst vor negativer Bewertung bei Unterstützung von Seiten der Familie (z.B. durch den Ehemann) signifikant verringert. Basierend darauf haben wir untersucht, ob die Angst vor negativer Bewertung für die Personen mit Hintergrundbildern von Familien und Freunden größer ist, als für andere Hintergrundbilder.

*Hypothese 3.* Schließlich sollten sich auch dunkle Persönlichkeitseigenschaften, wie die aus dem Dirty Dozen in der Wahl des Handyhintergrundbildes zeigen. So untersuchten bereits Kowalski et al. (2017) ob die Ausprägungen dieser Persönlichkeitsmerkmale einen Einfluss auf berufliche Interessen haben können. Sie fanden dabei einen Zusammenhang von künstlerischem Interesse und Narzissmus. Zuletzt überprüften wir daher, ob Psychopathie nach der Theorie der Dark Triad stärker ausgeprägt ist bei Personen, die Kunst oder Comics als Hintergrundbild aufwiesen, als bei Personen, die ein anderes Motiv als Hintergrund verwendeten.

### 2. Methode

*Stichprobe.* Insgesamt wurden Daten von 185 Teilnehmern erhoben und ausgewertet. Davon sind 69.73% weibliche Probanden. Das Durchschnittsalter der Stichprobe beträgt  $M=22.03$  ( $SD=3.368$ ). Der Anteil der Psychologiestudenten liegt bei 47.06%.

*Fragebogen.* Der Fragebogen konnte über die Onlineplattform Sosci-Survey ([soscisurvey.de/phacking](https://soscisurvey.de/phacking)) ausgefüllt werden. Dies war den Probanden von zu Hause oder im Foyer am Campus der Friedrich-Schiller-Universität möglich.

*Tests.* Insgesamt bestand der Fragebogen aus fünf Teilen. Am Anfang wurden einige allgemeine demografische Daten gesammelt. Durch das Freiburger Persönlichkeitsinventar (FPI-R, 12 Skalen) wurden die Persönlichkeitsfacetten erfasst. Der SANB-5 (5 Items) maß die Angst vor negativer Bewertung. Mit den Dirty Dozen (12 Items, 3 Skalen) erhoben wir die interpersonell maladaptiven Konstrukte Machiavellismus, Narzissmus und Psychopathie. Zusätzlich wurden Beziehungsstatus, Beziehungszufriedenheit und Informationen zum Handy berücksichtigt. Als letztes wählte der

Proband eine passende Kategorie für sein Handyhintergrundbild aus insgesamt fünfzehn Kategorien.

### 3. Ergebnisse

*Auswertung.* Der FPI wurde der Auswertungsschablone der Testautoren entsprechend für alle 12 Persönlichkeitsmerkmale ausgewertet. Für den SANB-5 und die Dirty Dozen summierten wir die Skalenwerte der verschiedenen Antworten. Mit R führten wir unserer Hypothese entsprechend in jedem der Teilfragebögen einen einseitigen *t*-Test ( $\alpha$ -Niveau = 0.05) durch.

*Hypothese 1.* Es besteht eine signifikant höhere Beziehungszufriedenheit bei Personen, die den Partner als Hintergrundbild wählten ( $p < 0.001$ ).

*Hypothese 2.* Die Angst vor negativer Bewertung ist in der Gruppe, die Familie oder Freunde als Hintergrundbild haben ( $p = 0.028$ ) signifikant größer als bei anderen Hintergrundbildern

*Hypothese 3.* In der Ausprägung Psychopathie gibt es signifikante Unterschiede ( $p = 0.0056$ ) zwischen Personen mit Kunst und Comic-Hintergrundbildern und Personen ohne besagten Motiven.

### 4. Diskussion

In dieser Studie konnten wir zeigen, dass die Wahl des Handyhintergrundbildes Rückschlüsse auf die Persönlichkeit zulässt. Hieraus ergeben sich Ansatzpunkte für weitere Studien, die die Persönlichkeit mit der medialen Nutzung und Selbstpräsentation in Verbindung bringen. Einen Beitrag dazu konnten wir mit unseren Ergebnissen schon leisten.

Allerdings beschäftigten wir uns im Rahmen dieses Forschungsseminars mit dem Thema ‚p-hacking‘. Dies beschreibt eine Vorgehensweise, bei der man die Ergebnisse seiner Studie bewusst oder unbewusst beeinflusst und damit verfälscht. Damit ist es prinzipiell möglich Ergebnisse den eigenen Vorstellungen entsprechend anzupassen. So hatten wir in unserer Studie anfänglich eine Vielzahl von möglichen Hypothesen ( $n \approx 60$ ). Bei einem gewählten  $\alpha$ -Niveau von 0.05 müssen nun der Definition entsprechend einige davon fälschlicherweise signifikant werden. Unsere bewusst herausgesuchten Ergebnisse für diesen Bericht sind daher vermutlich reine Zufallsprodukte und müssen erneut geprüft werden.

In der Forschungspraxis finden auch noch andere Möglichkeiten des ‚p-hackings‘ Anwendung. Neben der nachträglichen Generierung von Hypothesen sind auch die Erhebung von zusätzlichen Probanden (Nacherhebung), das Entfernen von Teilnehmern (Ausreißer-Analyse) und eine gezielte Teilstestung (Subgruppentests) möglich. Die letzten beiden Möglichkeiten haben wir auch angewendet um eine theorieferne Hypothese testen und signifikant werden zu lassen. Bei der Begrenzung auf fünfzig Versuchsteilnehmer und der Betrachtung ausschließlich männlicher Teilnehmer, fanden wir einen Zusammenhang zwischen dem Besitz eines Samsunghandys und körperlichen Beschwerden nach dem FPI ( $p < 0.01$ ).

Aus diesem Grund, möchten wir mit unserer Arbeit darauf aufmerksam machen, dass ein empirisches Ergebnis nicht immer eindeutig ist und das Phänomen des ‚p-hackings‘ ein stärkeres Bewusstsein verdient hat.

### 5. Literatur

- Kent, G & Keohane, S (2001). Social anxiety and disfigurement: The moderating effects of fear of negative evaluation and past experience. *British Journal of Clinical Psychology* (2001), 40, 25-54.
- Papp, L. M., Danielewicz, J., & Cayemberg, C. (2012). “Are we Facebook official?” Implications of dating partners’ Facebook use and profiles for intimate relationship satisfaction. *Cyberpsychology, Behavior, And Social Networking*, 15(2), 85-90. doi:10.1089/cyber.2011.0291
- Simonsohn, U., Nelson, L. D. & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143 (2), 534–547. doi: 10.1037/a0033242
- Kowalski, C. M., Vernon, P. A., & Schermer, J. A. (2017). Vocational interests and dark personality: Are there dark career choices?. *Personality and Individual Differences*, 104, 43-47

# **“It don’t matter if you’re black or white?”**

## **Der Einfluss von Bekanntheit und Ethnie auf die kategoriale Wahrnehmung von Gesichteridentität**

**Isabelle Ewe, Philipp Laue, Vanessa Mathieu, Charlotte Schüßler, Alisa Shugaley**

Leitung: Dr. Jürgen M. Kaufmann

### 1. Einleitung

Kategoriale Wahrnehmung bezeichnet das Phänomen, dass Stimuli trotz kontinuierlicher Veränderung auf der Reizseite in eindeutigen Kategorien wahrgenommen werden können. Dieser Mechanismus vereinfacht u.a. die Wahrnehmung gesprochener Sprache (Kuhl, 2004), als auch die Erkennung der Identität von Gesichtern (Beale & Keil, 1995). In der vorliegenden Studie sollte untersucht werden, ob der „other-race effect“ (Gesichter anderer Ethnien als der eigenen werden meist schlechter gelernt und wiedererkannt; siehe z.B. Meissner & Brigham, 2001), mit einer weniger stark ausgeprägten kategorialen Wahrnehmung von Gesichteridentität in Zusammenhang steht. Von zusätzlichem Interesse war, ob mögliche Unterschiede zwischen „same“ und „other-race“ Gesichtern durch Bekanntheit aufgehoben werden, und ob individuelle Unterschiede in der Fähigkeit, Gesichter zu lernen mit der kategorialen Wahrnehmung von Identität in Zusammenhang stehen. Es wurden folgende Hypothesen untersucht: a) Bekannte Gesichter werden stärker kategorial verarbeitet als unbekanntes, b) Unbekannte Gesichter der eigenen Ethnie werden stärker kategorial verarbeitet, als unbekanntes Gesichter einer anderen Ethnie, c) Für bekannte Gesichter verschwindet der Ethnieneffekt, und d) Die kategoriale Verarbeitung bekannter Gesichter ist bei guten Gesichterlernern stärker ausgeprägt.

### 2. Methode

*Stichprobe:* Daten von 29 Versuchspersonen (22 weiblich) gingen in die Analyse ein (Alter:  $M = 21.5$  Jahre;  $SD = 2.8$ ).

*Stimulusmaterial:* Grundlage bildeten jeweils acht frontale Bilder unbekannter und bekannter (prominenter) Personen (davon jeweils 4 mitteleuropäisch und 4 afrikanisch aussehende, jeweils 2 männliche und 2 weibliche), die zunächst mit Gimp 2.0 und Adobe Photoshop CS6 bearbeitet und formatiert wurden. Mittels Psychomorph (Tiddeman, 2005) wurden Kontinua über die Identität jeweils zweier Gesichter erstellt, indem in 10% Schritten gemorphet wurde. Innerhalb eines Morphkontinuums waren Bekanntheit, Ethnie und Geschlecht konstant gehalten. Für die Übungsphase wurden zusätzliche Stimuli hergestellt.

*Prozedur:* Zunächst wurde der *Cambridge Face Memory Test* (CFMT; Duchaine & Nakayama, 2005) zur Erhebung der individuellen Gesichterlernfähigkeit durchgeführt. Im anschließenden Experiment wurden Antwortgenauigkeiten und Reaktionszeiten der Versuchspersonen in einer Diskriminationsaufgabe gemessen: In 50% der Trials sahen die VP jeweils zwei Stimuli eines Morphkontinuums, die um 20% auseinanderlagen, also z.B. 10%: oder 30% Anteile der Person A enthielten. Obwohl der physikalische Unterschied also immer gleich war, wurde für einige Morphpaare die Identitäts-Kategoriengrenze überschritten (z.B. 40% vs. 60% Anteile von Person B), und für andere nicht (z.B. 10% vs. 30% Anteile von Person B). Für die andere Hälfte der Trials wurden zwei identische Bilder einer Morphstufe gezeigt. Per Tastendruck musste entschieden werden, ob die beiden Bilder gleich oder unterschiedlich waren. Von Interesse für die Auswertung waren v.a. die Trials, in denen unterschiedliche Bilder gezeigt wurden. Abschließend sollten die Versuchspersonen zur Kontrolle die bekannten Personen mithilfe eines Fragebogens identifizieren.

### 3. Ergebnisse

Für die Trials mit unterschiedlichen Bildern wurde jeweils für Antwortgenauigkeiten und Reaktionszeiten eine 9x2x2-ANOVA mit Messwiederholung auf den Faktoren Morphpaar, Bekanntheit (unbekannt vs. bekannt) und Ethnie (schwarz vs. weiß), und dem Zwischengruppenfaktor CFMT\_Gruppe (gute vs. schlechte Gesichterlerner) durchgeführt. Bei Signifikanz des Mauchly-Tests wurde jeweils der Huynh-Feldt-korrigierte  $p$ -Wert benutzt. Für die Antwortgenauigkeit zeigten sich Haupteffekte für die Faktoren *Ethnie*,  $F(1,27) = 10.585$ ,  $p = .003$ ,  $\eta_p^2 = .282$ , und *Morphpaar*,  $F(8,216) = 10.937$ ,  $p < .001$ ,  $\eta_p^2 = .288$ . Diese Effekte wurden durch eine zweifach-Interaktion zwischen *Bekanntheit* und *Ethnie*,  $F(1,27) = 17.833$ ,  $p < .001$ ,  $\eta_p^2 = .398$ , eine zweifach-Interaktion zwischen *Bekanntheit* und *Morphpaar*,  $F(8,216) = 5.927$ ,  $p < .001$ ,  $\eta_p^2 = .18$ , eine zweifach-Interaktion zwischen *Ethnie* und *Morphpaar*,  $F(8,216) = 2.905$ ,  $p = .004$ ,  $\eta_p^2 = .097$ , sowie eine Dreifachinteraktion zwischen *Morphpaar*, *Bekanntheit* und *Ethnie*,  $F(8,216) = 6.452$ ,  $p < .001$ ,  $\eta_p^2 = .193$  moduliert. Eine Interaktion zwischen CFMT\_Gruppe, Ethnie und Morphpaar erreichte nur Trendniveau,  $F(8,216) = 1.873$ ,  $p = .066$ ,  $\eta_p^2 = .065$ .

Für die *Reaktionszeiten* zeigte sich ein Haupteffekt für *Morphpaar*,  $F(8,80) = 2.75$ ,  $p = .010$ ,  $\eta_p^2 = .216$ , eine Interaktion zwischen *Morphpaar* und *Bekanntheit*,  $F(8,80) = 3.073$ ,  $p = .006$ ,  $\eta_p^2 = .235$ , sowie eine Dreifachinteraktion zwischen *Morphpaar*, *Bekanntheit* und *Ethnie*,  $F(8,80) = 2.271$ ,  $p = .030$ ,  $\eta_p^2 = .185$ . Bezüglich des CFMTs ergaben sich keine signifikanten Haupteffekte, und, abgesehen von einer Interaktion zwischen *CFMT\_Gruppe* und *Bekanntheit*,  $F(1,10) = 5.07$ ,  $p = .048$ ,  $\eta_p^2 = .336$  auch keine weiteren Interaktionen.

Eine Weiterverfolgung der Interaktionen durch separate Nachtests für bekannte und unbekannte Gesichter legte stärkere kategoriale Verarbeitungseffekte für bekannte im Vergleich zu unbekanntem Gesichtern nahe. Diese Effekte waren für bekannte Gesichter der eigenen Ethnie etwas stärker ausgeprägt.

#### 4. Diskussion

Zunächst einmal konnte das Experiment bestehende Erkenntnisse zur kategorialen Wahrnehmung replizieren. Wie erwartet zeigte sich sowohl in der Antwortgenauigkeit als auch in der Reaktionszeit Evidenz für eine kategoriale Verarbeitung von Gesichtern. Dieser Effekt des Morphpaars wurde durch die Faktoren Ethnie und Bekanntheit moduliert. Die Nachtests legten nahe, dass es sowohl für die bekannten „same“, als auch für die bekannten „other-race“ Gesichter kategoriale Wahrnehmungseffekte gibt, diese für erstere aber stärker ausgeprägt waren. Dieses Ergebnis widerspricht unserer Hypothese, dass Ethnienunterschiede bei Bekanntheit komplett verschwinden sollten. Für die unbekanntem Gesichter zeigte sich, unabhängig von der Ethnie, keine klare Evidenz für eine kategoriale Wahrnehmung der Identität. Dies stützt die aufgestellte Hypothese, dass bekannte Gesichter generell kategorialer verarbeitet werden als unbekanntem. Im Gegensatz zu unseren Erwartungen wurden unbekanntem Gesichter der eigenen Ethnie aber nicht stärker kategorial verarbeitet, als unbekanntem Gesichter einer anderen Ethnie. Interessanterweise legen die Ergebnisse also nahe, dass der Hauptfaktor für kategoriale Wahrnehmung eher Bekanntheit als Ethnie ist. Da dieses Ergebnismuster nicht signifikant von den Leistungen im CFMT abhing schließen wir daraus, dass gute und schlechte Gesichterlerner sich bezüglich der kategorialen Verarbeitung von „same“ und „other-race faces“ nicht qualitativ unterscheiden.

#### 5. Literatur

- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, 57(3), 217–239.
- Duchaine, B., & Nakayama, K. (2006). The Cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44, 576–585.
- Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, 5(11), 831–843.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology Public Policy and Law*, 7(1), 3–35.
- Tiddeman, B., Stirrat M., & Perrett, D. I., (2005). Towards realism in facial transformation: results of a wavelet MRF method. *Computer Graphics Forum, Eurographics conference issue*, 24, 1–5.

# Der Einfluss von semantischer Information und Emotion auf die kategoriale Gesichtererkennung

Lukas Erchinger, Almut Hellwig, Fabian Holdmann, Max Meinhardt, Anna Pscheidt

Leitung: Dr. Jürgen M. Kaufmann

## 1. Einleitung

Unter kategorialer Wahrnehmung wird ein entscheidender Mechanismus in der menschlichen Perzeption verstanden. Menschen nehmen z.B. die meisten visuellen oder auditiven Sinneseindrücke nicht als ineinanderfließende Übergänge wahr, sondern grenzen verschiedene Eindrücke durch das Bilden von Kategorien voneinander ab (Goldstone & Hendrickson, 2010). Diese Kategorien bilden sich erst durch Erfahrung, wobei Eindrücke aus derselben Kategorie weniger deutlich voneinander unterscheidbar sind als Eindrücke unterschiedlicher Kategorien. Dadurch wird eine effizientere Verarbeitung von Sinneseindrücken möglich. Das Resultat von Kategorisierung ist unter anderem für die Erkennung von Gesichtern bekannt. Sie unterstützt nicht nur die Differenzierung von Personen, sondern auch die Erkennung derselben Person in verschiedensten Situationen (Burton, Jenkins, & Schweinberger, 2011).

Im letzten Semester wurde von unserer Studiengruppe bereits der Einfluss von semantischer Information auf die kategoriale Gesichtererkennung untersucht. Dabei zeigte sich, dass zusätzliche semantische Information während des Lernens prä-experimentell unbekannter Gesichter die kategoriale Wahrnehmung verstärken kann. So fanden auch Kaufmann, Schweinberger und Burton (2009), dass die Wiedererkennung von Gesichtern durch zusätzliche semantische Informationen verbessert wird. Semantisches Wissen umfasst explizit gespeichertes Wissen im deklarativen Gedächtnis, allerdings wurden semantische Informationen in der Forschung unterschiedlich operationalisiert.

In der Vorstudie wurden in der semantischen Bedingung humorvolle Texte mit den Gesichtern in der Lernphase präsentiert. Fraglich blieb dabei, inwieweit der Effekt auf die semantische Information an sich oder die emotional anregende Wirkweise zurückging. In dieser Studie sollte dies näher betrachtet werden, indem der Einfluss der Präsentation neutral-semantischer, emotional-semantischer und rein emotionaler Information auf die kategoriale Gesichtererkennung untersucht wurde. Die Bedingung „semantisch“ der Vorstudie entsprach dabei der emotional-semantischen Bedingung in dieser Studie.

*Die Studie testete folgende Hypothesen:*

Semantische Information verstärkt die kategoriale Gesichtererkennung.

Emotion verstärkt die kategoriale Gesichtererkennung.

Der Einfluss von Semantik und Emotion auf die kategoriale Gesichtererkennung wirkt kumulativ: einzeln haben beide einen positiven Einfluss, gemeinsam ist der Effekt noch stärker.

## 2. Methode

An der Studie nahmen insgesamt 36 Personen (davon 13 Männer) im Alter von 18 bis 33 Jahren teil ( $M = 22.61$ ). Für den Manipulationscheck wurden 16 Bilder von Frauen und Männern verwendet. Für die Testphase wurden Bilder von Frauen und Männern (aus denen 16 Morphpaare des gleichen Geschlechts gebildet wurden, die in 10% Schritten ineinander übergingen), für die Lernphase wurden entweder die Bilder allein oder zusätzlich ein „nerviger“ Ton oder humoristisch oder neutral formulierte semantische Informationstexte verwendet. Das Experiment wurde mit der Software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) erstellt und dauerte 45 Minuten. Nach Angabe von Alter, Geschlecht und Händigkeit nach dem „Edinburgh Handedness Inventory“ folgte die Lernphase: 16 Gesichter wurden je 20 Sekunden präsentiert; 4 mit „nervigem“ Ton (Bedingung „Emotional“), 4 mit neutraler semantischer Information (Bedingung „Semantisch-

Neutral“), 4 mit humoristischer semantischer Information (Bedingung „Semantisch-Emotional“) und 4 ohne weiteres Stimulusmaterial (Kontrollbedingung). Im Manipulationscheck wurden gelernte und ungelernte Gesichter je 750 ms präsentiert. Danach sollte entschieden werden, ob diese gelernt oder nicht gelernt waren. In der Testphase wurden Morphpaare unterschiedlicher Morphlevel je 2000 ms lang präsentiert, worauf entschieden werden sollte, ob diese gelernt oder nicht gelernt waren. Es gab 8 verschiedene Versionen des Experiments, in denen die Tastenbelegung und die Bildstimuli über die Versuchspersonen ausbalanciert wurden.

### 3. Ergebnisse

Es wurde eine ANOVA mit den Faktoren Lernbedingung und Morphlevel durchgeführt. Der Mauchly-Test auf Sphärizität wurde für die Bedingung Morphlevel nicht signifikant, wohl aber für die Lernbedingung und die Interaktion zwischen Morphlevel und Lernbedingung. Deshalb wird beim Test der Innersubjekteffekte für das Ergebnis der Lernbedingung und der Interaktion der Hyhn-Feldt-Wert angegeben. Für die Lernbedingung ergab sich kein signifikanter Effekt ( $F [3,105] = 1.415, p = 0.243, \eta_p^2 = 0.039$ ). Für Morphlevel ergab sich ein Haupteffekt ( $F [10,350] = 562.022, p \leq 0.001, \eta_p^2 = 0.941$ ). Die Interaktion zwischen Lernbedingung und Morphlevel wurde nicht signifikant ( $F [30,1050] = 1.275, p = 0.186, \eta_p^2 = 0.035$ ).

### 4. Diskussion

Es ergab sich ein Haupteffekt für Morphlevel. Das Muster der lange weitgehend konstanten und sich dann rasch ändernden Bewertungen spricht für die kategoriale Wahrnehmung von Gesichtern in dieser Studie. Die Interpretationsmöglichkeiten der weiteren Ergebnisse sind jedoch durch fehlende Signifikanz eingeschränkt. Die Hypothesen konnten somit nicht bestätigt werden: Weder die semantische noch die emotionale Information hatten einen signifikanten Effekt auf die kategoriale Gesichtererkennung. Die Daten haben allerdings interessante Tendenzen erkennen lassen: die deskriptiven Ergebnisse der Akkuratheit weisen darauf hin, dass in der semantisch-neutralen Bedingung ein etwas größerer Einfluss auf die kategoriale Gesichtererkennung besteht als in der semantisch-emotionalen und der rein emotionalen Bedingung.

Entgegen unserer ursprünglichen Hypothese könnten die Ergebnisse darauf hinweisen, dass ein emotionaler Lernfaktor die kategoriale Gesichtererkennung eher behindert als fördert. Insgesamt schien der Effekt der Vorstudie eher aus den semantischen Informationen hervorzugehen.

Die fehlende Signifikanz in dieser Studie ist womöglich durch eine geringere Teststärke als in der Vorstudie erklärbar. Dafür spricht, dass kein Effekt für die semantischen Stimuli im Vergleich zur Kontrollgruppe gefunden wurde, während die Bedingung „Semantisch“ in der Vorstudie signifikant wurde. Daneben mussten die Probanden in dieser Studie vier Bedingungen bearbeiten, wodurch die Anzahl der zu lernenden Gesichter verdoppelt wurde. Möglicherweise war deshalb die Qualität der Lernleistung beeinträchtigt.

Aus den Trends dieser Studie ergaben sich für weitere Untersuchungen andere Erwartungen für die Effekte der Lernbedingungen: es scheint die semantische Information an sich zu sein, die die kategoriale Gesichterverarbeitung fördert. Es bleibt offen, wie genau sich das Lernen von Gesichtern in Verbindung mit emotionalen Inhalten auf die kategoriale Verarbeitung auswirkt.

### 5. Literatur

- 1234bjoern. (2010,16.08.). Nerviger Ton, annoying tone [Onlinevideo]. Abgerufen von [<https://www.youtube.com/watch?v=L6mLEVNvDMU>] am 15.05.2017.
- Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces: Mental representations of familiar faces. *British Journal of Psychology*, 102(4), 943–958.
- Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(1), 69–78.
- Kaufmann, J. M., Schweinberger, S. R., & Burton, A. M. (2009). N250 ERP Correlates of the Acquisition of Face Representations across Different Images. *Journal of Cognitive Neuroscience*, 21(4), 625–641.

# **Fake it `til you make it: Increasing the AMP Validity by Faking Instructions**

**Alexia Dalski, Lukas Knitter, Ole Kölbel, Toya Rosenkranz, Annika Vogel**

Leitung: Dr. Franziska Meißner

## 1. Einleitung

Die Affekt-Misattributions-Prozedur (AMP, Payne Cheng, Govorun, & Stewart, 2005) ist eine in der psychologischen Forschung häufig angewandte Methode zur impliziten Einstellungsmessung. Doch der AMP sieht sich dem Vorwurf ausgesetzt, fälschbar zu sein. Dass Probanden dazu in der Lage sind, konnte von Teige-Mocigemba, Penzl, Becker, Henn und Klauer (2016) gezeigt werden. Dazu erhielt die Experimentalgruppe (EG) vor dem eigentlichen Test eine Faking-Instruktion: Probanden sollten sich in bestimmten Trials entgegen der eigenen Einstellung darstellen. Das wirkte sich nicht nur auf das beobachtete Antwortverhalten im AMP aus, auch Korrelationen mit expliziten Einstellungsmaßen brachen mit der Faking-Instruktion ein. Auch wenn Teige-Mocigemba et al. damit zeigten, dass die Validität des AMP kritisch hinterfragt werden sollte, deutete sich gleichzeitig eine Möglichkeit an, wie man das Verfahren verbessern könnte. In der EG zeigten sich nämlich in den Kontrolltrials, in welchen nicht gefälscht werden sollte, tendenziell bessere Ergebnisse als in der Kontrollgruppe (KG). Diesen Nebenbefund nahmen wir als Ausgangspunkt unserer Studie, anhand dessen wir folgende Hypothese aufstellten: Wenn kognitive Ressourcen in Form einer Faking-Instruktion für Ablenkungsstimuli im AMP gebunden werden, *erhöht* sich die Validität der impliziten Einstellungsmessung bezüglich der forschungsrelevanten Stimuli.

## 2. Methode

Wir konstruierten ein vierstufiges Experiment, in welchem wir die Parteien "CDU" und "Die Linke" als Ablenkungsstimuli und die Marken "Apple" und "Samsung" als forschungsrelevante Stimuli wählten. Auf einen Selbstbericht (9-stufiges Gefühlsthermometer, Entscheidungsfrage) bzgl. der Einstellung zu den Parteien folgte für die EG eine Faking-Instruktion („Du hast eben angegeben, dass du die CDU besser findest als die LINKE. Bitte stell dich nun im folgenden Verfahren so dar, als hättest du eine gegenteilige politische Einstellung. [...]“). Im anschließenden AMP sollten chinesische Schriftzeichen per Tastendruck als positiv oder negativ bewertet werden. Vor jedem chinesischen Zeichen wurde jeweils ein Parteien- oder Markenlogo als Prime gezeigt. Der AMP folgte einem 2 (Trialtyp: Parteien vs. Marken) x 2 (Pol: im Selbstbericht präferierte Partei/Marke vs. im Selbstbericht abgelehnte Partei/Marke) x 2 (Gruppe: EG vs. KG) Design. Trialtyp und Pol waren within-subjects Faktoren, Gruppe ein between-subjects Faktor. Die Tastenzuweisung wurde ausbalanciert. Abhängige Variable war die relative Häufigkeit von „positiv“-Antworten auf die chinesischen Schriftzeichen. Auf den AMP folgte letztlich ein dreiteiliger Selbstbericht (9-stufiges Gefühlsthermometer, Entscheidungsfrage, Semantisches Differential) bezüglich der Einstellung zu den Marken „Apple“ und „Samsung“.

63 Studenten der FSU Jena nahmen am Experiment teil, von welchen wir vier wegen Chinesisch-Kenntnissen bzw. nicht instruktionskonformen Verhaltens aus den Analysen ausschlossen.

## 3. Ergebnisse

Das Antwortverhalten im AMP, d.h. der Anteil positiver Antworten, wurde in einer 2x2x2-ANOVA betrachtet. Die erwartete Dreifach-Interaktion Trialtyp x Pol x Gruppe wurde signifikant,  $F(1,57) = 6.56, p = .013, \eta^2_p = .10$ . Um den Effekt zu verstehen, rechneten wir getrennte Folgeanalysen für die Parteien- und Markentrials. In den Parteientrials zeigte sich die erwartete Interaktion von Gruppe x

Pol,  $F(1,57) = 5.52, p = .022, \eta^2_p = .09$ . In den Markentrials dagegen zeigte sich keine Interaktion,  $F(1,57) = 0.27, p = .602, \eta^2_p = .01$ , sondern lediglich ein Haupteffekt von Pol,  $F(1,57) = 9.58, p = .003, \eta^2_p = .14$ . Das Antwortverhalten im AMP wurde demnach auch bei uns spezifisch gefälscht.

Für die entscheidenden, korrelativen Analysen berechneten wir AMP-Effekte, die die Parteienpräferenz (Differenz der „positiv“-Antworten nach CDU- vs. Linke-Primes) und die Markenpräferenz (Differenz der „positiv“-Antworten nach Apple- vs. Samsung-Primes) widerspiegeln (für KG & EG,  $0.83 \leq \text{Cronbach's } \alpha \leq 0.98$ ). Um die Validität zu testen, überprüften wir den Zusammenhang zwischen diesen AMP-Effekten und der in den Gefühlsthermometern gemessenen Einstellung. In den Partientrials fanden wir in der KG eine mittlere ( $r = .42, p = .022$ ), in der EG eine Null-Korrelation ( $r = .01, p = .942$ ). Eine Moderationsanalyse ergab jedoch keine Signifikanz für den korrelativ gefundenen Unterschied,  $\beta = -.38, t(55) = -1.45, p = .154$ . In den Markentrials ergab sich deskriptiv die erwartete geringere Korrelation für die KG ( $r = 0.45, p = .014$ ) als für die EG ( $r = 0.61, p < .001$ ), in der Moderationsanalyse wurde dieser Unterschied jedoch erneut nicht signifikant,  $\beta = .06, t(55) = 0.29, p = .772$ . Darum untersuchten wir post-hoc noch die inkrementelle Validität. Wir prüften, ob der AMP-Effekt bzgl. der Markentrials *zusätzlich* zur selbstberichteten Einstellung noch Varianz im Gefühlsthermometer erklären kann. Dazu wurde die eben berichtete Moderationsanalyse durch einen weiteren Prädiktor, das Semantische Differential, erweitert. Neben diesem,  $\beta = .79, t(54) = 11.09, p < .001$ , wurde nun auch die erwartete Moderation signifikant,  $\beta = .25, t(54) = 2.28, p = .027$ . Folgeanalysen ergaben, dass in der KG der AMP-Effekt über das Semantische Differential hinaus *keinen* Beitrag zur Vorhersage des Gefühlsthermometers leistete,  $\beta = .01, t(26) = 0.08, p = .940$ . In der EG dagegen konnte der AMP-Effekt über das Semantische Differential hinaus die im Gefühlsthermometer angegebene Einstellung vorhersagen,  $\beta = .32, t(27) = 3.33, p = .003$ .

#### 4. Diskussion

In unserer Untersuchung konnten wir die Ergebnisse der Vorgängerstudie weitgehend replizieren: Die Antworten im AMP sind spezifisch fälschbar und der beobachtete AMP-Effekt bzgl. der Parteien (=Ablenkungsstimuli) ist nach Faking-Instruktion zumindest deskriptiv weniger valide. In Bezug auf unsere Forschungsfrage fanden wir, dass Faking-Instruktionen die konvergente Validität des AMP-Effekts bzgl. der Marken (=forschungsrelevante Stimuli) deskriptiv tatsächlich *erhöhen*. Dass die beobachteten Gruppenunterschiede nicht signifikant wurden, könnte unserer relativ kleinen Stichprobe geschuldet sein. Bezüglich der inkrementellen Validität wurde der Nutzen ablenkender Faking-Instruktionen aber statistisch signifikant. Insgesamt stützen die Ergebnisse die Annahme, dass die Bindung von Ressourcen durch Faking-Instruktionen für Ablenkungsstimuli dazu führt, dass die Validität des AMP-Effekts bezogen auf forschungsrelevante Stimuli steigt.

Es ist möglich, dass sich durch die Faking-Instruktion das Bewusstsein, den AMP kontrolliert beeinflussen zu können, auf die anderen Trials übertragen hat. Dass nach Faking-Instruktion einfach alle gezeigten Logos kontrolliert bewertet werden, wird scheinbar durch die höhere Reliabilität in der EG gestützt. Allerdings sprechen die Befunde bzgl. der inkrementellen Validität eher gegen diese Annahme. Dennoch sollten die Ergebnisse mit weiteren Validitätskriterien (andere implizite Maße, spontane Verhaltensmessung) abgesichert werden. In dem Zuge wäre es auch interessant, zu erfragen, was die Instruktion in den Probanden auslöst und ob es Transfereffekte gibt. Zukünftige Studien sollten außerdem auch sensible Themen als Untersuchungsgegenstand wählen, um herauszufinden, inwieweit soziale Erwünschtheit einen Einfluss auf die gefundenen Effekte hat.

#### 5. Literatur

- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*(3), 277-293.
- Teige-Mocigemba, S., Penzl, B., Becker, M., Henn, L., & Klauer, K. C. (2016). Controlling the “uncontrollable”: Faking effects on the affect misattribution procedure. *Cognition and Emotion, 30*(8), 1470-1484.

# Kriminalität in Deutschland - Wann ist eine harte Strafe gerechtfertigt? Soziale Funktionen von Bestrafung

Solveig Becker, Martin Jäger, Jannis Lobs, Aline Schönenberg, Dario Urban

Leitung: Stefanie Hechler

## 1. Einleitung

„Im Namen des Volkes“ lautet der Beginn einer Urteilsverkündung an deutschen Gerichtshöfen. Oftmals fragen wir uns dann, ob die verkündete Strafe gerechtfertigt ist und wie nachhaltig sie wirken kann, um Wiederholungstaten vermeiden zu können. Welche Funktionen gehen damit einher?

Gollwitzer et al. (2010) postulierten, dass schwere Vergehen härter geahndet werden als vergleichsweise harmlose Taten. Dies lässt vermuten, dass es der Gesellschaft wichtig erscheint, verletzte Normen wiederherzustellen, in dem klargemacht wird, dass diese Tat nicht geduldet wird. Der Wertekonsens einer Gruppe wird somit durch die Strafe verdeutlicht. Damit einher geht auch die wahrgenommene Sicherheit der Gesellschaft, wie Tyler und Boeckmann (1997) zeigten. Eine harte Strafe für Wiederholungstäter wird gefordert, um den Werteverfall der Gesellschaft zu unterbinden.

Dass Wiederholungstäter gefährlicher und unglaubwürdiger hinsichtlich ihrer Absichten, ein Verbrechen in Zukunft zu vermeiden gegenüber Ersttätern wirken, hielten Gollwitzer et al. (2010) fest. Auch eine härtere Bestrafung sollte ihnen zukommen. Soll eine Strafe also den Täter bekehren, ihn vielleicht sogar Buße tun lassen? Funk et al. (2014) bestätigten, dass Bestrafung eine kommunikative Funktion besitzt („Was du getan hast, finden wir schlimm.“). Dies stellt die Abgrenzung zur Tat sicher und verfolgt das Ziel, eine Einstellungsveränderung beim Täter hervorzurufen. Hat eine Strafe demnach ihre sozialen Funktionen verfehlt, wenn die Tat wiederholt wird? Um diesen Fragen auf den Grund zu gehen, formulierten wir folgende Hypothesen:

H1: Die Tat eines Wiederholungstäters löst mehr Unzufriedenheit bei Beobachtern aus als erstmalige Taten. Egal ob dieser bestraft wurde oder nicht.

H2: Personen sind zufriedener, wenn ein Wiederholungstäter für die Ersttat bestraft wurde vs. wenn er nicht bestraft wurde.

## 2. Methode

An der Studie haben insgesamt 162 Personen mit einem Durchschnittsalter von 24.2 Jahren (SD=9.52) teilgenommen (33 männlich, 106 weiblich, 3 sonstiges). 19 Personen wurden aufgrund von Falschantworten im Lesecheck ausgeschlossen.

In der Fragebogenstudie gab es 8 verschiedene Bedingungen, denen die Probanden randomisiert zugeteilt wurden. In jeder Bedingung wurde den Probanden eine andere Situation geschildert. Anschließend sollten sie zu dieser Situation einige Fragen beantworten. Die geschilderten Situationen handelten entweder von einem Diebstahl oder Raubüberfall. Es wurde variiert, ob es sich um einen Wiederholungstäter oder einen Ersttäter handelte. Dieser Täter wurde in manchen Bedingungen bestraft, in anderen wurde die Anzeige fallengelassen.

Es hat sich somit ein 2x2x2 Between-Subjects-Design mit den unabhängigen Variablen „Wiederholungstäter“ (ja/nein), „Bestrafung“ (ja/nein) und „Schwere des Vergehens“ (leicht/schwer) ergeben. Die abhängige Variable war die empfundene gerechtigkeitsbezogene Zufriedenheit. Nachdem die Probanden die Texte gelesen hatten, folgten einige Fragen zu ihrer gerechtigkeitsbezogenen Zufriedenheit. Anschließend gab es eine Skala zur wahrgenommenen Schwere der Tat und einen Lesetest als Manipulationscheck. Eine weitere Skala zum Ziel der

Bestrafung (Retribution, Restauration des Wertekonsenses, Schutz der Gesellschaft, Täterreformation) wurde aus explorativen Gründen ebenfalls abgefragt.

### 3. Ergebnisse

Die interne Konsistenz wurde über Cronbachs Alpha berechnet. Hierbei erhielten wir  $\alpha=.78$  für die Skala „Schwere der Tat“,  $\alpha=.81$  für die Zufriedenheitsskala,  $\alpha=.81$  für die Retributionsskala,  $\alpha=.79$  für die Restaurationsskala,  $\alpha=.83$  für die Reformationsskala und  $\alpha=.89$  für die Schutzskala.

Varianzanalysen ergaben eine signifikante Veränderung der Zufriedenheit bei bestraften Tätern ( $F=15.1856$ ,  $p=0.0001436$  \*\*\*) und Wiederholungstätern ( $F=25.8302$ ,  $p=1.042e-06$  \*\*\*). Dies bestätigt unsere in der H1 getätigte Annahme.

Hypothese zwei musste aufgrund fehlender Signifikanz abgelehnt werden ( $F=0.1675$ ,  $p=0.6836$ ).

Des Weiteren haben wir einen signifikanten Effekt von der Schlimme der Tat ( $F=24.8559$ ,  $p=1.61e-06$  \*\*\*) und der Strafe ( $F=11.4501$ ,  $p=0.0009016$  \*\*\*) auf die Zufriedenheit gefunden.

Bei einer weniger schlimmen Tat gab es bei der Zufriedenheit nur einen signifikanten Unterschied von zwischen bestraften und unbestraften Tätern ( $F=8.1636$ ,  $p=0.00551$  \*\*), jedoch haben wir hier auch einen Interaktionseffekt gefunden ( $F=4.8170$ ,  $p=0.03124$  \*).

Bei einer schlimmeren Tat gab es weder einen signifikanten Effekt durch die Strafe, noch durch eine Wiederholungstat auf die Zufriedenheit.

### 4. Diskussion

Aufgrund dieser Ergebnisse finden wir unsere erste Hypothese bestätigt, während wir die zweite ablehnen müssen. Als Beobachter ist man also unzufriedener mit einem Wiederholungs- als mit einem Ersttäter. Auch eine Strafe mildert diese Unzufriedenheit bei Wiederholungstätern nicht ab, entgegen unserer Annahme. Hieran lässt sich erkennen, dass die Täterreformation für die Zufriedenheit des Beobachters eine wesentliche Rolle spielt. Denn eine Wiederholungstat demonstriert die mangelnde Einsicht des Täters. Selbst das Vorhandensein einer Strafe kann die Unzufriedenheit mit dem unverbesserlichen Täter nicht abmildern. Dies deckt sich auch mit den Befunden unserer ersten Studie, in der besonders der Sinneswandel („moral change“) des Täters die Beobachter zufriedener stimmte.

Des Weiteren stieg die Unzufriedenheit mit der Schlimme der Tat. Betrachtet man aber die weniger schlimme Tat für sich genommen, so steigert nur eine Strafe die Zufriedenheit. Ob der Täter erstmalig oder wiederholt eine Tat begeht, hat hier auf die Zufriedenheit keinen Einfluss. In der Interaktion zwischen Strafe und Wiederholungstat auf die Zufriedenheit gewinnt die Wiederholungstat wieder an Wichtigkeit: Am zufriedensten waren die Menschen mit einem bestraften Ersttäter und am unzufriedensten mit einem unbestraften Wiederholungstäter.

Im explorativen Teil unserer Studie stellen wir die jeweilige Korrelation der verschiedenen sozialen Funktionen von Strafe (Retribution, Restauration, Reform, Schutz) mit der Zufriedenheit der Personen dar. Auch hier zeigte sich deutlich, dass die Reformation des Täters den größten Einfluss auf die Zufriedenheit hat.

### 5. Literatur

- Gollwitzer, M., & Keller, L. (2010). What you did only matters if you are one of us: Offenders' group membership moderates the effect of criminal history on punishment severity. *Social Psychology*, 41(1), 20-26.
- Funk, F., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin*, 40(8), 986-997.
- Tyler, T., Boeckmann, R. (1997). Three strikes and you are out, but why? The psychology of public support for punishing rule breakers. *Law & Society Review*, 31(2), 237-266.

# Wenn ich ehrlich bin, dann lüg' ich nicht: Der Einfluss von Unglaube und Unruhe auf das Lügengedächtnis

Sara Bechtold, Benita Ebersbach, Jerome Hoffmann, Kristin Meißner, Lisa Stärker, Annika Stahn

Leitung: M.Sc. Franziska Schreckenbach

## 1. Einleitung

Wie stellen wir uns typische Reaktionen eines Lügners vor? Sofort kommen einem dabei nervöse Blicke, feuchte Hände und allgemeine Unruhe in den Sinn. Natürlich möchte auch jeder Lügner, dass seine Falschaussage so lange wie möglich unentdeckt bleibt und muss sie deshalb im Gedächtnis behalten. Doch was passiert, wenn wir befürchten müssen, erwischt zu werden?

Die Grundlage unserer Studie bildet die Hypothese von Koranyi, Schreckenbach & Rothermund (2015), dass ein implizit arbeitendes Lügengedächtnis existiert. Nach der *Instance Theory of Automatization* sind die Speicherung im Gedächtnis sowie der Gedächtnisabruf zwingende und unvermeidliche Konsequenzen von Aufmerksamkeit (Logan, 1988). Für Lügen sind einzelne Episoden, sogenannte Instances, relevant. Diese umfassen Umgebungsreize, Handlungsziele sowie Handlungen und speichern diese langfristig im Gedächtnis. Es erfolgt eine Bindung zwischen dem Stimulus und dem Wissen über die Lüge, sodass diese unter geringerem kognitiven Aufwand reproduziert werden kann (Koranyi et al., 2015).

In unserer Studie entwickelten wir ein Reaktionszeitexperiment, um herauszufinden, wie das Lügengedächtnis davon beeinflusst wird, ob man von einem Menschen oder einer Maschine überwacht wird. Diese beiden Manipulationen stellten wir durch einen vermeintlichen Experten für Mikroexpressionen und einen vermeintlichen Lügendetektor dar. Wir stellten die Vermutung an, dass diese Beobachtung während des Lügens einen Einfluss auf die Nervosität der Versuchsperson und dadurch auch auf die Ausbildung eines Lügengedächtnisses hat. Dazu haben wir ein Selbstbeurteilungsmaß für Nervosität (den Mehrdimensionalen Befindlichkeitsfragebogen; MDBF) mit einem Reaktionszeitexperiment am PC kombiniert. Dort erwarteten wir einen Kongruenzeffekt: Die Reaktion auf kongruente Prime-Target-Kombinationen sollte schneller erfolgen als auf inkongruente Prime-Target-Kombinationen.

## 2. Methode

*Teilnehmer und Design:* Nach Ausschluss von drei Versuchspersonen (VPN) wegen zu hohen Fehlerraten oder zu langsamen Reaktionszeiten setzte sich die Stichprobe aus 59 Probanden zusammen. Das Durchschnittsalter betrug 22.45 Jahre ( $SD = 4.0$ ) mit 43 weiblichen Teilnehmern.

*Durchführung und Materialien:* Zunächst mussten die Teilnehmenden die Kurzform A des Mehrdimensionalen Befindlichkeitsfragebogen (MDBF) zur Erfassung ihrer aktuellen Anspannung anhand der drei Dimensionen gute vs. schlechte Stimmung, Wachheit vs. Müdigkeit und Ruhe vs. Unruhe ausfüllen und einen von uns verfassten Informationstext zum Lügendetektor (Bedingung 1) bzw. Mikroexpressionen (Bedingung 2) lesen. Anschließend erhielten sie am PC die Instruktion zu welchem von unseren drei Themen sie im folgenden Interview lügen und zu welchem sie die Wahrheit sagen sollten. In beiden Bedingungen kamen die VPN in einen Raum mit zwei Testleitern, wo sie in Bedingung 1 an den vermeintlichen Lügendetektor angeschlossen und in Bedingung 2 von einem angeblichen Experten für Mikroexpressionen überwacht wurden. Nachdem wir ihnen vier Baselinefragen gestellt haben, mussten die Teilnehmer die Kurzform B des MDBF ausfüllen. Anschließend wurde das Interview zu den vorher festgelegten Themen durchgeführt. Danach folgte das Reaktionszeitexperiment am PC. In jedem Durchgang musste die VPN auf ein Target (die Wörter „ehrlich“ bzw. „gelogen“, degradiert durch Sonderzeichen) per Tastendruck reagieren, das immer auf einen Prime (eine Frage aus dem Interview) folgte. Zum Abschluss

beantworteten die Versuchspersonen noch Fragen zur wahrgenommenen Professionalität des Experimentalaufbaus, zum Glauben an diese Technik zur Aufdeckung von Lügen und zur subjektiven Wahrscheinlichkeit, dass die eigenen Lügen aufgedeckt wurden.

### 3. Ergebnisse

Um zu überprüfen, wie sich die Nervosität der VPN über die Zeit und in den verschiedenen Bedingungen entwickelte, rechneten wir eine 2 (Bedingung: Lügendetektor vs. Mikroexpression) x 2 (MDBF: ruhig vs. unruhig) ANOVA. Es ergab sich lediglich ein signifikanter Haupteffekt für den Zeitpunkt  $F(1,57)=18.69$ ,  $p<.0001$ , alle anderen  $F<3.03$ : Die VPN in beiden Bedingungen wurden also vom ersten zum zweiten Messzeitpunkt signifikant unruhiger. Anschließend rechneten wir eine 2 (Prime: ehrlich vs. gelogen) x 2 (Target: ehrlich vs. gelogen) x 2 (Bedingung: Lügendetektor vs. Mikroexpression) ANOVA. Die Interaktion zwischen Prime und Target wurde marginal signifikant,  $F(1,57)=2.815$ ,  $p=0.099$ . Alle anderen Haupt- und Interaktionseffekte wurden nicht signifikant, alle  $F<2.4$ . Dies deutet darauf hin, dass die VPN sich auch im Kongruenzeffekt nicht zwischen den Bedingungen unterscheiden. Deshalb überprüften wir, ob der Grad der Nervosität einen Einfluss darauf hatte. Dafür rechneten wir eine Regression auf die Effektvariable (RT-inkongruente Durchgänge minus RT-kongruente Durchgänge) als abhängige Variable mit den Prädiktoren „Ruhe“ und „Wahrscheinlichkeit des Aufdeckens“: Das Modell wurde signifikant mit  $F(3,55)=3.294$ ,  $p=0.027$ ,  $R^2=0.1523$ . Zudem wurde der Haupteffekt auf „Ruhe“ signifikant,  $t(55)=-2.919$ ,  $p=0.005$  und die Interaktion zwischen beiden Prädiktoren,  $t(55)=3.094$ ,  $p=0.003$ . Unruhe in Kombination mit der Vermutung, erwischt zu werden, scheint also den Kongruenzeffekt zu beeinflussen.

### 4. Diskussion

Ziel unserer Studie war es, herauszufinden, ob Unruhe der VPN einen Einfluss auf das Lügengedächtnis hat und ob es einen Unterschied macht, ob ein Mensch oder eine Maschine die Lügen aufdecken. Es ergab sich ein signifikanter Haupteffekt für Unruhe: Die VPN wurden während des Experiments zunehmend nervöser, allerdings fanden wir dabei keinen Unterschied zwischen den Bedingungen. Dies könnte daran liegen, dass das Lügen im Allgemeinen eine ungewohnte Situation ist, und somit alle Personen nervöser und unruhig werden, unabhängig von der Bedingung. Auch könnte es einen Einfluss gehabt haben, dass die VPN von uns die Information bekommen haben, dass sie durch einen Lügendetektor oder Experten für Mikroexpressionen beobachtet werden. Um dies genauer zu überprüfen, könnte man in zukünftigen Studien eine Kontrollgruppe ohne Manipulationen der Beobachtung einführen, die eventuell weniger unruhig ist. Insgesamt zeigt unsere Regressionsanalyse ein überraschendes Ergebnis: Personen, die unruhiger waren, zeigten den Kongruenzeffekt nur, wenn sie glaubten, dass ihre Lüge nicht aufgedeckt wurde. Dies ist naheliegend, wenn man bedenkt, dass nur dann auch in späteren Situationen wieder gelogen werden muss, wenn man nicht auffliegt. Im Gegensatz dazu zeigen ruhiger gewordene VPN den Effekt nur, wenn sie glauben, erwischt worden zu sein. Hierbei könnte Selbstsicherheit eine Rolle spielen: Solchen Personen scheint es egal zu sein, wenn man sie erwischt, aber sie behalten die Lüge dennoch für eine weitere Situation, in der diese von Nutzen sein könnte, im Gedächtnis. Andererseits könnten dies auch Personen sein, die allgemein ruhiger an Aufgaben herangehen, und denen die Lüge somit nur dann als salient im Gedächtnis bleibt, wenn sie das Gefühl haben aufgedeckt zu werden. Ein wichtiger Ansatzpunkt für weitere Studien wäre deshalb, dass man untersucht, wie sich die Unterschiede zwischen ruhigen und unruhigen VPN erklären lassen.

### 5. Literatur

- Koranyi, N., Schreckenbach, F., Rothermund K., (2015): The Implicit Cognition of Lying: Knowledge about Having Lied to a Question is Retrieved Automatically. *Social Cognition*
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527. doi: 10.1037/0033-295X.95.4.492

# Hör mal, wer da spricht!

## Der Einfluss von Distinktheit eines Sprechers auf die Erinnerungsleistung von Sprecher und Inhalt

Sarenka Allmacher, Charlotte Deckert, Madlen Schiewek, Theresa Strübe,  
Nadine Zscherper

Leitung: Denise Humble

### 1. Einleitung

Jeder kennt das: Man hört Radio und erkennt eine Stimme, von der man weiß, dass man sie kennt, aber es fällt einem nicht ein, woher. Woran liegt es, dass wir uns einige Stimmen besonders gut merken können und andere nicht?

Nygaard, Sommers & Pisoni (1994) fanden bereits heraus, dass das Wiedererkennen von Wörtern im Zusammenhang mit der Bekanntheit des Sprechers steht. In der Forschung zur Wahrnehmung von Gesichtern zeigte sich, dass Gesichter akkurater zugeordnet wurden, wenn simultan eine distinkte Stimme präsentiert wird (Bülhoff & Newell, 2015). Andersherum wurde für das Stimmenlernen berichtet, dass eine Kombination des auditiven Stimulus mit einem visuellen Stimulus (Gesicht) in der Lernphase den Prozess des Einprägens der Stimme eher behindert (z.B. Zäske et.al., 2015). Dieses Phänomen wird auch Overshadowing Effekt genannt. Die Stimmenerkennung wird durch die Typikalität der Stimme beeinflusst (Mullinex et.al, 2011), was wiederum die Inhaltswahrnehmung beeinflussen kann.

Das Ziel bestand darin, herauszufinden, ob und welchen Einfluss die Distinktheit der Stimme auf die Wiedererkennungsleistung dieser, deren Inhalt und der Assoziation zwischen Sprecher und Inhalt hat. Dabei wird postuliert, dass eine distinkte Stimme akkurater wiedererkannt wird. Für die Wiedererkennungsleistung des sprachlichen Inhalts sowie die Assoziation zwischen Sprecher und Inhalt wird hingegen eine höhere Akkuratheit für nicht-distinkte Stimmen aufgrund des Overshadowing Effektes erwartet.

### 2. Methoden

Am Computereperiment nahmen  $N = 24$  Probanden teil (16 weiblich,  $M_{\text{Alter}} = 22,14$ ,  $SD_{\text{Alter}} = 2,44$ ). Zwei der Probanden mussten aufgrund einer fehlerhaften Experimentprogrammierung ausgeschlossen werden. Die Stimuli wurden aus dem Jena Speaker Set entnommen. Im Durchschnitt waren die Sprecher  $M_{\text{Alter}} = 21,42$  Jahre alt ( $SD_{\text{Alter}} = 2,25$ ). 12 von insgesamt 24 Stimuli waren mittels eines Distinktheitsratings als besonders markant charakterisiert worden. Das verwendete Rating erfasste die deviation-based distinctiveness. Das heißt es wurde erfragt, wie stark abweichend eine Stimme von anderen Stimmen wahrgenommen wurde. Die Probanden hörten die Stimuli via Kopfhörer. Das Experiment bestand aus einer Lern- und einer Testphase, die nach einmaligem Absolvieren wiederholt wurden. In der Lernphase wurden acht Stimmen zweimal mit jeweils einem Satz präsentiert, die sich die Probanden merken sollten. Von diesen Stimmen waren vier distinkt und vier weiblich. Die gesprochenen Sätze waren außerdem nach dem Artikel ausbalanciert (Bsp.: „Der Fahrer lenkt den Wagen“). In der Stimmentestphase wurden die gelernten Stimmen sowie acht unbekannte Stimmen, die in ihrer Ausbalancierung von Geschlecht und Distinktheit dem gelernten Sprecherset entsprachen, vorgespielt, wobei die Sprecher nun alle den gleichen, neutralen Satz sagten (Bsp.: „Keine Antwort ist auch eine Antwort“). Die Probanden sollten entscheiden, ob der Sprecher in der Lernphase vorkam oder nicht (bekannt oder unbekannt). In der Inhaltstestphase wurden die in der Lernphase gesprochenen Satzsubjekte sowie acht unbekannte Wörter lediglich visuell präsentiert. Die Probanden hatten die Aufgabe, die Subjekte aus der Lernphase wiederzuerkennen. In der Assoziationstestphase sprachen die bekannten Sprecher nochmals einen anderen neutralen Satz (z.B. „Dichter und Denker dachten

dasselbe“), worauf anschließend ein aus der Lernphase bekanntes Subjekt oder ein unbekanntes visuell erschien. Hier sollte von den Probanden eine Rekognition dieses bestimmten Sprechers zu jenem Inhalt erfolgen.

### 3. Ergebnisse

Zur statistischen Auswertung wurde das Programm IBM® SPSS® Statistics 22 genutzt und für alle durchgeführten Tests ein Signifikanzniveau von  $\alpha = .05$  zugrunde gelegt. Für den Vergleich distinkter und nicht-distinkter Stimmen wurden zunächst d-Prime-Werte berechnet. Der dafür durchgeführte  $t$ -Test ergab keinen signifikanten Unterschied ( $t [21] = 1.77, p = .091$ ). Jedoch unterschieden sich beide signifikant vom Rateniveau (mit  $t [21] = 7.819, p < .001$  für distinkte Stimmen und mit  $t [21] = 3.611, p = .002$  für nicht-distinkte Stimmen). Für die Akkuratheit des Sprechertests wurde eine zweifaktorielle ANOVA mit den Messwiederholungsfaktoren Distinktheit (distinkt vs. nicht distinkt) und Bekanntheit (bekannt vs. unbekannt) gerechnet. Diese ergab eine Zweifach-Interaktion ( $F[1,21] = 14.121, p = .001, \eta_p^2 = .402$ ). Post-Tests zeigten dabei einen Haupteffekt für Distinktheit innerhalb der unbekannt Stimmen ( $F[1,21] = 15.174, p = .001, \eta_p^2 = .419$ ). Distinkte Stimmen wurden dabei signifikant besser als unbekannt zurückgewiesen ( $M_{dis} = .744, SD = .147; M_{ndis} = .544, SD = .169$ ). Für die Akkuratheit der Inhalts- und Assoziationstestphase zeigten sich keine signifikanten Effekte.

### 4. Diskussion

Das Ziel der Studie war es, den Einfluss von Distinktheit auf Wahrnehmung und assoziatives Lernen von Sprachinhalt und Sprecheridentität zu untersuchen. Entgegen unserer Annahme ergab sich kein signifikanter Unterschied zwischen distinkten und nicht-distinkten Stimmen, außer die signifikant bessere Zurückweisung von distinkten unbekannt Stimmen. Für den Inhalts- und Assoziationstest ergaben sich keine signifikanten Effekte. Eine mögliche Ursache für die nicht signifikanten Effekte liegt in der Aufgabenschwierigkeit. Die durchgehend sehr guten Leistungen der Probanden könnten auf Deckeneffekte und somit auf eine zu leichte Aufgabenstellung hinweisen. Die Sprecher wurden unabhängig von der Distinktheit gut gelernt. Die signifikant bessere Zurückweisung von distinkten unbekannt Stimmen zeigt aber, dass die Distinktheit einen Einfluss auf das Lernen und Wiedererkennen von Stimmen hat. Daraus ergibt sich die Frage, wie Distinktheit überhaupt definiert wird. Das in dieser Untersuchung verwendete Rating erklärt nicht, warum distinkte Stimmen als solche bewertet wurden. Eine Lösung dafür könnte die Definition von Distinktheit als eine große Abweichung der Grundfrequenz eines Sprechers von der durchschnittlichen Grundfrequenz des Sprechersets sein. Bei der Wahrnehmung einer Stimme erfolgt zuerst eine Analyse des Sprachsignals und ein Abgleich mit bereits bekannten Stimmen. Sobald eine Stimme als bekannt klassifiziert wird, werden die Person Identity Units aktiviert, die spezifische Informationen über die Identität einer Person enthalten. (Belin, Fecteau & Bédard, 2004). Die im Experiment verwendeten neutralen Stimuli enthalten keine zusätzlichen Informationen über den Sprecher. Wir vermuten, dass für zukünftige Designs die Einbeziehung von personenbezogenen Inhalten die Assoziation zwischen Stimme und Inhalt begünstigen könnte.

### 5. Literatur

- Belin, P., Fecteau S. & Bédard C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Science*, Vol.8 No.3, 129-135.
- Mullinex, J.W., Ross, A., Smith, C., Kuykendall, K., Conard, J. & Barb, S. (2011). Typicality effects on memory for voice: Implications for eyewitness testimony. *Applied Cognitive Psychology*. 25: 29–34.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42–46.
- Zäske, R., Mühl, C. & Schweinberger, S. R. (2015). Benefits for voice learning caused by concurrent faces develop over time. *Plos One*, November 20, 2015.

## **Der erste Eindruck zählt? Wahrnehmung von Attraktivität bei Gesichtern und Stimmen**

**Christina Biesenbaum, Anne Jügling, Johanna Menke, Anne-Maria Nagel, Ketrin Malvina Nikitina, Juliane Kröhn, Maria Helena Szabo, Ramona Schunke, Amelie Wild, Jasmin Wohlrab**

Leitung: Dr. Romi Zäske & Denise Humble

### 1. Einleitung

In sozialen Kontexten spielen Stimmen und Gesichter eine zentrale Rolle. Sie tragen Informationen über verschiedene Eigenschaften einer Person, wie z.B. deren Attraktivität, und beeinflussen den ersten Eindruck. Hört man eine Stimme, macht man sich sofort ein Bild über den Sprecher - das funktioniert sogar schon bei kurzen Wörtern (McAleer et al., 2014). Doch wann bildet sich der erste Eindruck und ab wann wird er stabil? Während dies für Stimmen völlig unbekannt ist, zeigen Untersuchungen mit Gesichtern, dass sich der erste Eindruck möglicherweise schon bei 33ms bildet und innerhalb von 100ms stabilisiert (Willis & Todorov, 2006; Todorov et al., 2009). Am Beispiel von Attraktivität wird hier untersucht, wann sich der erste Eindruck anhand der Stimme im Vergleich zum Gesicht stabilisiert. Hierzu beurteilten Hörer die Attraktivität von Stimmen und Gesichtern unbekannter Personen, wobei die Darbietungszeit (DZ) unbegrenzt (Vorstudie) oder begrenzt sein konnte (Hauptstudie). Da sich im Gegensatz zu einem statischen Bild eines Gesichts, stimmliche Information über eine Äußerung hinweg entfaltet, wurde angenommen, dass sich das Attraktivitätsurteil für Stimmen später ( $> 100$  ms) als für Gesichter stabilisiert. Zudem wurde erwartet, dass Gesichter und Stimmen mit steigender DZ weniger attraktiv erscheinen, die Urteilsicherheit sich jedoch mit steigender DZ erhöht (Todorov et al., 2006).

### 2. Methode

Stichprobe: Je 24 Probanden nahmen an der Vorstudie (WiSe: 18 weiblich,  $M_{Alter} = 21,1$ ;  $range = 18-34$  Jahre) und Hauptstudie (SoSe: 12 weiblich,  $M_{Alter} = 21,6$ ;  $range = 19-30$  Jahre) teil.

Stimuli: Als Stimuli dienten Stimmen (je 1 Satz) und Gesichter (Fotos) von 123 unbekanntem Personen (61 weiblich; davon etwa die Hälfte jung [18-25 Jahre] und alt [60-81 Jahre]) aus dem Stimuluspool des Lehrstuhls Allgemeine I der FSU Jena.

Prozedur: Beide Computerexperimente bestanden aus 2 Blöcken (getrennt nach Modalitäten: Gesicht/ Stimme), denen jeweils ein Übungsdurchgang vorgeschaltet war. Diese 2 Blöcke waren in je 4 kleinere Blöcke unterteilt (getrennt nach Alter und Geschlecht der Stimuli), innerhalb derer alle Stimuli genau einmal randomisiert präsentiert wurden. Während die DZ der Stimuli in der Vorstudie unbegrenzt war, variierte sie in der Hauptstudie: Gesicht = 50, 100, 500, 2000ms; Stimme = 100, 500, 1000, 2000ms. Nach einer visuellen oder auditiven Rückwärtsmaske (50ms) sollten die Probanden alle Stimuli nach Attraktivität bewerten und angeben, wie sicher sie sich in ihrer Entscheidung waren. Dazu wurde jeweils eine Skala von 1-6 (1= sehr unattraktiv/unsicher; 6=sehr attraktiv/sicher) verwendet. Es gab keine Zeitbegrenzung für die Bewertung der Stimuli.

### 3. Ergebnisse

Zunächst wurden in SPSS bivariate, zweiseitige Korrelationen nach Spearman zwischen Attraktivitätsbeurteilungen ohne (Vorstudie) und mit Zeitbegrenzung (Hauptstudie) für beide Modalitäten berechnet. Bei Gesichtern ergab sich für alle vier DZ ein signifikanter positiver Zusammenhang (alle  $\rho \geq .522$ ;  $p < .001$ ), wobei die Korrelationskoeffizienten mit steigender DZ anstiegen. Für Stimmen ergaben sich ebenfalls signifikante Korrelationen bei allen DZ ( $\rho \geq .448$ ;  $p < .001$ ), mit Ausnahme der kürzesten DZ (100ms:  $\rho = .056$ ;  $p = .539$ ). Bezüglich der Urteilsstabilisierung wurden die Korrelationen zwischen allen DZ in der jeweiligen Modalität

verglichen (Diedenhofen und Musch, 2015): Für Gesichter war kein Vergleich signifikant (Ausnahme: 2000ms>50ms,  $z=-2.896$ ;  $p=.004$ ). Bei Stimmen erzielten alle Korrelationsvergleiche signifikante Unterschiede (alle  $z\leq-2.207$ ;  $p<.001$ ), außer zwischen 1000ms und 2000ms ( $z=0.047$ ;  $p=.962$ ). Der Korrelationsvergleich zwischen den Modalitäten für gemeinsame DZ (100, 500 und 2000ms) zeigte signifikant größere Korrelationen bei Gesichtern im Vergleich zu Stimmen ( $z\leq-2.297$ ;  $p\leq .022$ ), außer bei 2000ms ( $z=-1.068$ ;  $p=.289$ ).

Ein Vergleich der mittleren Attraktivitätseinschätzung mittels ANOVA mit Messwiederholung (Modalität [2] x DZ [3]) ergab signifikante Haupteffekte von Modalität ( $F[1,122]=82,389$ ;  $p<.001$ ; höhere Attraktivitätsurteile für Stimmen als für Gesichter) und DZ ( $F[2,244]=3,166$ ;  $p=.046$ ), sowie eine Interaktion von Modalität x DZ ( $F[2,244]=30,556$ ;  $p<.001$ ; mit steigender DZ wurden Stimmen attraktiver eingeschätzt, Gesichter unattraktiver). Eine äquivalente ANOVA für die Konfidenzbewertung ergab signifikante Haupteffekte von Modalität ( $F[1,122]=697,642$ ;  $p<.001$ ; sichere Urteile für Gesichter als für Stimmen) und DZ ( $F[2,244]=631,398$ ;  $p<.001$ ; Zunahme der eigenen Urteilssicherheit mit steigender DZ), sowie eine Interaktion von Modalität x DZ ( $F[2,244]=179,513$ ;  $p<.001$ ): der Modalitätsunterschied in der Konfidenzbewertung nimmt mit steigender DZ ab und verschwindet bei 2000ms ( $t[122] = 2.170$ ,  $p=.096$ , Bonferroni-korrigiert).

#### 4. Diskussion

Die Korrelationsergebnisse bestätigen eine Stabilisierung des ersten Eindrucks der Attraktivität für Gesichter innerhalb der ersten 100ms (Willis & Todorov, 2006; Todorov et al., 2009). Für Stimmen stabilisierte sich der erste Eindruck erst später (ab ~ 1000ms). Dies passt zu Befunden, wonach auch die Sprechererkennung anhand von kurzen Sätzen am stärksten von einer Erhöhung der DZ innerhalb der ersten 1000ms profitiert (Schweinberger et al., 1997). Mit steigender DZ wurden Gesichter im Schnitt zunehmend unattraktiver bewertet (vgl. Willis & Todorov, 2006), während Stimmen zunehmend attraktiver bewertet wurden. Letzteres könnte damit erklärt werden, dass auditorische Stimuli mit zunehmender DZ mehr prosodische Informationen enthalten, natürlicher und damit attraktiver wirken (Ferdenzi et al., 2013). Je länger Stimmen sowie Gesichter präsentiert wurden, desto sicherer fiel das Urteil aus (vgl. Willis & Todorov, 2006). Dabei waren die Attraktivitätsurteile bei kurzen DZ (100ms und 500ms) signifikant sicherer für Gesichter als für Stimmen, unterschieden sich jedoch nicht mehr bei 2000ms. Dies scheint plausibel aufgrund der relativ späten Stabilisierung von Stimmenurteilen (ab ~1000ms). Da hier statische Gesichter mit dynamischen Stimmsignalen verglichen wurden, könnten die gefundenen Unterschiede in der Eindrucksbildung und -stabilisierung auf die unterschiedliche zeitliche Entwicklung des Informationsgehalts in beiden Reizklassen zurückgehen. In zukünftigen Untersuchungen könnte dies kontrolliert werden etwa durch die Einbeziehung von dynamischen Gesichtsstimuli (Videos) oder „statischen“ Stimmsignalen (z.B. Vokale), da statische und dynamische Stimuli verschiedenen Bewertungsstandards unterliegen (Ferdenzi et al., 2013; Rubenstein, 2005).

#### 5. Literatur

- Diedenhofen, B., & Musch, J. (2015). A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945.
- Ferdenzi, C., Patel, S., Mehu-Blantar, I., Khidasheli, M., Sander, D., Delplanque, S. (2013). Voice Attractiveness: Influence of Stimulus Duration and Type. *Behavior Research Methods*, *45* (2), 405-413.
- McAleer, P., Todorov, A., Belin, P. (2014). How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *PLoS ONE*, *9*(3), e90779.
- Rubenstein, A.J. (2005). Variation in Perceived Attractiveness: Differences between Dynamic and Static Faces. *Psychological Science*, *16*(12), 759-762.
- Schweinberger, S., & Sommer, W. (1997). Recognizing Famous Voices: Influence of Stimulus Duration and Different Types of Retrieval Cues. *JSLHR*, *40*, 453-463
- Todorov, A., Pakrashi, M., Oosterhof, N.N. (2009). Evaluating Faces on Trustworthiness after Minimal Time Exposure. *Social Cognition*, *27*(6), 813-833.
- Willis, J. & Todorov, A. (2006). First Impressions: Making up your Mind after a 100-ms Exposure to a Face. *Psychological Science*, *17*(7), 592-598.

# **Auswirkungen der Täter- und Opferzugehörigkeit auf Bestrafungstendenzen bei gruppenspezifischen Normverletzungen**

**Isabel Jaki, Theresa Leistner, Anna Lumaschi, Julia Müller, Christina Schulte**

Leitung: Stefanie Hechler

## 1. Einleitung

Wir bestrafen Normverletzungen, um Werte und Normen wiederherzustellen und so die Identität unserer Gruppe zu wahren. Dafür wenden Bestrafende auch eigene Ressourcen auf. Diese kostspielige Bestrafung fällt bei Verstößen von Mitgliedern aus der Eigengruppe höher aus, als bei Verstößen von Fremdgruppenmitgliedern (Mendoza, Lane & Amodio, 2014). Gruppenmitglieder gehen davon aus, dass andere Mitglieder die Gruppennormen achten. Normverstöße innerhalb der Gruppe gefährden die Identität der Gruppe. Restaurative und Retributive Bestrafung spielen in Gruppenkontexten ebenfalls eine große Rolle (Wenzel, Okimoto und Cameron, 2012). Erstere zielt auf die Wiederherstellung von Gerechtigkeit durch ein gemeinsames Verständnis der Normen und die Bestätigung relevanter Werte ab. Bei retributiven Bestrafungsmaßnahmen wird durch eine einseitige Zufügung der Strafe die Unterwerfung des Täters erzwungen, um den Status/die Macht des Opfers wiederherzustellen. Wie bestraft wird, hängt von der wahrgenommenen symbolischen Bedeutung der Tat ab. Ein Verstoß innerhalb der Eigengruppe wird meist als Werte-/Normbruch eingestuft. Vorfälle zwischen Gruppen gelten eher als Status-/Machtverletzung (Wenzel & Okimoto, 2010). Emotionen moderieren Handlungsintentionen (Yzerbyt, Dumont, Wigboldus & Gordijn, 2003). Wut steigert Bestrafungstendenzen und wird stärker empfunden, wenn das Opfer unserer Gruppe angehört und wir uns stark mit dieser identifizieren. Schadenfreude wird empfunden, wenn das Verhalten eines Fremdgruppenmitglieds als ungerechtfertigt beurteilt wird und dies den Status der gegnerischen Gruppe mindert (Leach, Spears, Branscombe & Doosje, 2003).

Unsere Fragestellung war, wie Personen auf einen Verstoß gegen gruppenspezifische Normen aus der Eigen- oder Fremdgruppe reagieren, der sich entweder gegen ein Mitglied der eigenen Gruppe, oder der anderen Gruppe richtet. In der gegenwärtigen Studie stellten wir folgende Hypothesen auf: In der Eigengruppe wird mehr kostspielige Bestrafung aufgewendet. Eigengruppentäter werden eher restaurativ bestraft, während Fremdgruppentäter eher retributiv bestraft werden. Außerdem lösen Fremdgruppentäter mehr Schadenfreude aus als Eigengruppentäter.

## 2. Methode

Um unsere Hypothesen zu prüfen, gingen die Daten von 117 Versuchspersonen, von denen 64 Studierende waren (Alter  $M=26,04$ ,  $SD=10,51$ , Geschlecht: 87 weiblich, 27 männlich, 3 keine Angabe) in unsere Analyse ein. Die unabhängigen Variablen Gruppenzugehörigkeit des Täters und Gruppenzugehörigkeit des Opfers wurden in einem 2x2 between-subject Design untersucht. Unser Experiment lief als Onlinefragebogen bestehend aus einer Kolumne. Diese thematisierte die Meinungsfreiheit in Deutschland (Eigengruppe). Zunächst schilderte die Kolumne eine Situation, in welcher die Meinungsfreiheit verletzt wurde, durch einen Fremd- oder Eigengruppentäter. Folgend wurden als erste abhängige Variable die Emotionen erfragt, insbesondere Wut und Schadenfreude (5 Items,  $\alpha=0.9$ , Beispielitem: *“verärgert”* und 4 Items,  $\alpha=0.89$ , Beispielitem: *“zufrieden”*). Infolgedessen erfragten wir als weitere abhängige Variablen den Wunsch nach kostspieliger (6 Items,  $\alpha=0.77$ , Beispielitem: *“Ich würde noch länger bei der Konferenz bleiben als von mir erwartet wird, um Oliver zur Rede zu stellen.”*), sowie retributiver und restaurativer Bestrafungorientierung (jeweils 6 Items,  $\alpha=0.93$  und  $\alpha=0.83$ , Beispielitems: *“Oliver verdient es, bestraft zu werden.”* und *“Ohne die aufrichtige Beteuerung Olivers, dass er unangebracht gehandelt hat, ist die Gerechtigkeit nicht komplett wiederhergestellt.”*). Alle Items wurden auf einer

7-stufigen Likert-Skala erhoben ( $I$ =gar nicht bis  $7$ =sehr). Um zu überprüfen, ob alle Probanden die Gruppenzugehörigkeit des Täters bzw. Opfers richtig erinnerten, führten wir noch einen Manipulationscheck durch und überprüften, ob die Tat tatsächlich als Werte-/Normverstoß oder Angriff auf den Status/die Macht angesehen wurde.

### 3. Ergebnisse

Durch den Manipulationscheck wurden neun Personen aus den Analysen ausgeschlossen, da sie sich nicht korrekt an die Gruppenzugehörigkeit des Täters oder Opfers erinnerten. Des Weiteren überprüften wir, ob die Normen und Werte oder der Status der Gruppe als Angriff wahrgenommen wurden. Die Probanden empfanden den Angriff auf das Opfer generell als moralisch verwerflich ( $M = 4.96$ ,  $SD = 1.58$ ), es ließ sich jedoch kein Unterschied zwischen den Gruppen erkennen. Die Werte und Normen des eigenen Landes wurden besonders in den Gruppen als verletzt wahrgenommen, in denen der Täter der Eigengruppe angehörte. Um dies zu überprüfen berechneten wir eine zweifaktorielle ANOVA, dessen Haupteffekt bezüglich der Tätermitgliedschaft signifikant wurde ( $F(112) = 20.533$ ,  $p < 0.05$ ). Das Item zur Erfassung der Statusbeeinträchtigung des Herkunftslandes des Täters wurde aus unseren Analysen ausgeschlossen. Wenn das Opfer der Fremdgruppe zugehörig war, wurde der Verstoß eher als Statusbeeinträchtigung des Täterlandes wahrgenommen als wenn das Opfer in der Eigengruppe war. Es ließen sich keine Haupteffekte der Täter- und Opfermitgliedschaft bezüglich kostspieliger Bestrafung (Täter:  $F(111) = 1.12$ ,  $p = 0.29$ ; Opfer:  $F(111) = 0.07$ ,  $p = 0.79$ ) erkennen, allerdings war der Interaktionseffekt signifikant ( $F(111) = 7.77$ ,  $p < .01$ ). Die Probanden nahmen am meisten kostspielige Bestrafung in Kauf, wenn der Täter und das Opfer jeweils unterschiedlichen Gruppen zugehörig waren ( $TEOE: M = 2.6$ ,  $SD = 0.99$ ;  $TFOE: M = 3.42$ ,  $SD = 1.25$ ;  $TEOF: M = 3.22$ ,  $SD = 1.21$ ;  $TFOF: M = 2.89$ ,  $SD = 0.93$ ). Für die abhängigen Variablen Schadenfreude und Wut ließen sich jeweils keine signifikanten Effekte berichten. Allerdings korrelierte das Ausmaß an kostspieliger Bestrafung positiv mit der empfundenen Wut ( $r = .27$ ,  $p < .001$ ). Bei Betrachtung des Ausmaßes an geforderter retributiver oder restaurativer Bestrafung ließen sich keine signifikanten Unterschiede in Abhängigkeit der Gruppenzugehörigkeiten ernennen. Retribution und kostspielige Bestrafung korrelierten positiv ( $r = .24$ ,  $p.01$ ). Und auch der Zusammenhang zwischen Restauration und kostspieliger Bestrafung war positiv korreliert ( $r = .31$ ,  $p < .001$ ).

### 4. Diskussion

Die Ergebnisse konnten unsere Hypothesen nur teilweise bestätigen. Die kostspielige Bestrafung war in den Intergruppensituationen höher. Zu erwarten war laut Theorie, dass zumindest bei Vorfällen, bei denen ein Eigengruppenmitglied zu Schaden kommt, Täter der eigenen Gruppe signifikant mehr bestraft werden als Fremdgruppentäter. Die Korrelation von Wut und kostspieliger Bestrafung zeigt, dass diese tatsächlich durch Emotionen verstärkt wird. Für Schadenfreude fanden wir keine Effekte.

Die erwartete Wahl der Bestrafungsart nach Tätermitgliedschaft hat sich ebenso nicht bestätigt.

### 5. Literatur

- Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: schadenfreude at the suffering of another group. *Journal of personality and social psychology*, 84(5), 932.
- Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For members only: ingroup punishment of fairness norm violations in the ultimatum game. *Social Psychological and Personality Science*, 5(6), 662-670.
- Okimoto, T. G., & Wenzel, M. (2010). The symbolic identity implications of inter and intra-group transgressions. *European Journal of Social Psychology*, 40(3), 552-562.
- Wenzel, M., Okimoto, T. G., & Cameron, K. (2012). Do retributive and restorative justice processes address different symbolic concerns? *Critical Criminology: An International Journal*, 20, 25-44.
- Yzerbyt, V., Dumont, M., Wigboldus, D., & Gordijn, E. (2003). I feel for us: The impact of categorization and identification on emotions and action tendencies. *British Journal of Social Psychology*, 42(4), 533-549.

# **„Don't bias – be happy“ – Auswirkungen positiver Emotionen auf den Own-Race Bias**

**K. Steinberg, A. Kalinichenko, C. Maennig, P. Gürtler**

Leitung: Dr. Jürgen M. Kaufmann

## 1. Einleitung

„Die sehen ja alle gleich aus!“ – was kann der Grund dafür sein, dass Gesichter anderer ethnischer Gruppen weniger gut unterschieden und erkannt werden können als die der eigenen Ethnie? In der Wissenschaft ist dieser Effekt als Own-Race Bias bekannt. Ist dieser unveränderlich? In vielen Studien wurden verschiedene Bedingungen wie persönliche Eigenschaften, Stimmung und physische Merkmale hinsichtlich ihres Einflusses auf den Own-Race Bias untersucht. Zum Beispiel besagt die sogenannte „Broaden-Hypothese“, dass positive Emotionen einen steigenden Einfluss auf die Aufmerksamkeit, die Kognition und das Handeln eines Probanden haben (Fredrickson & Branigan, 2005).

In einer anderen Studie untersuchte Chen (2014), inwiefern Emotionen und Ethnie die Wiedererkennung von Gesichtern beeinflussen. So fand man heraus, dass fröhliche Gesichter besser als neutrale und negative Gesichter erkannt wurden.

Außerdem fanden Johnson und Fredrickson (2005) heraus, dass positive Emotionen den Own-Race Bias reduzieren. Inwieweit evozierte Emotionen (fröhlich/neutral) sowie die Testemotion (fröhlich/neutral) einen Einfluss auf den Own-Race Bias haben, wurde in diesem Experiment untersucht.

## 2. Methode

An dem Computerexperiment nahmen insgesamt 32 gesunde, weiße Versuchspersonen teil. Alle Teilnehmer waren Studierende der Friedrich-Schiller-Universität in Jena (davon 20 weiblich,  $M = 22,16$ ;  $SD = 2,67$ ). Unter allen Versuchspersonen waren 29 Rechtshänder und 3 Linkshänder. In der Studie wurden insgesamt 96 Stimuli verwendet. Diese beinhalteten 48 Schwarze und 48 Weiße, welche jeweils zur Hälfte aus Männern und Frauen bestanden. Darin waren 48 fröhliche und 48 neutrale Gesichter enthalten. Außerdem wurden bei den Probanden durch das Zeigen von 2 Videos neutrale oder fröhliche Emotionen evoziert. Um die evozierten Emotionen und Bildstimuli auszubalancieren, wurden acht verschiedene Bedingungen des Experiments durchgeführt.

Vor der Lernphase, in welcher jeweils 28 Same- und Other-Race Gesichter gelernt wurden, wurde ein neutrales/fröhliches Video abgespielt, um die entsprechende Emotion zu evozieren. Während des Lernens der Gesichter bestand die Aufgabe darin, die Gesichter als männlich oder weiblich durch die Tasten X und M zu klassifizieren. Ebenfalls wurden die Versuchspersonen instruiert, sich die Gesichter so präzise wie möglich einzuprägen, da sie am Ende abgefragt wurden. Um sicherzugehen, dass die Aufgabe verstanden wurde, erfolgte ein Übungsblock. Danach bestand die Aufgabe darin, so schnell und korrekt wie möglich über Tastendruck (X/M; gelernt/unbekannt) zu entscheiden, ob die Gesichter bereits gelernt wurden oder nicht. Vor der anschließenden Testphase, in welcher 48 gelernte Stimuli aus der Lernphase und zusätzlich 48 neue Stimuli gezeigt wurden, sahen die Probanden das zweite Video (neutral/fröhlich). Für die Durchführung der Testphase blieb die Aufgabeninstruktion die gleiche.

Es wurden die Akkuratheit (Treffer und korrekte Zurückweisungen) und die Reaktionszeit aufgezeichnet

### 3. Ergebnisse

Eine Analyse des dPrimes und KritC wurde mithilfe einer mehrfaktoriellen ANOVA mit Messwiederholung berechnet. Dabei wurden als Innersubjektfaktoren die Testemotion (fröhlich/neutral) und Ethnie (schwarz/weiß) sowie evozierte Emotion (fröhlich/neutral) als Zwischensubjektfaktor berücksichtigt. Für die Analyse der Reaktionszeit (RT) kam der Faktor Bekanntheit (neu/gelernt) dazu. Bei der Analyse des dPrimes ( $d'$ ), welches ein Maß dafür ist, wie gut eine Versuchsperson zwischen den Stimuli unterscheiden kann, wurde kein signifikanter Haupteffekt Ethnie ( $F(1,31) = 0.964, p = .334, \eta^2 = .031$ ) gefunden. Für die *Interaktion Testemotion x Ethnie* kam jedoch ein signifikanter  $d'$  ( $F(1,31) = 4.849, p = .035, \eta^2 = .139$ ) heraus. Bei der Betrachtung von KritC, welches ein Maß für die Antworttendenz ist, wurde ein signifikanter Haupteffekt Ethnie ( $F(1,31) = 12.797, p = .001, \eta^2 = .299$ ) deutlich. Auch für die *Interaktion Testemotion x Ethnie* konnte ein signifikanter KritC-Wert ( $F(1,31) = 6.199, p = .019, \eta^2 = .171$ ) ermittelt werden. Die Reaktionszeit wurde nur für die *Interaktion Ethnie x Bekanntheit* signifikant ( $F(1,31) = 6.841, p = .014, \eta^2 = .186$ ), für den Haupteffekt Ethnie ( $F(1,31) = 1.374, p = .250, \eta^2 = .004$ ) und die Interaktion zwischen Testemotion und Ethnie ( $F(1,31) = 0.179, p = .675, \eta^2 = .006$ ) konnte keine Signifikanz festgestellt werden. Bei der Analyse der evozierten Emotion ließ sich kein signifikanter Effekt feststellen.

### 4. Diskussion

Aus den erhobenen Daten ergibt sich, entgegen unserer Hypothese, kein Own-Race-Bias. Das heißt, weder bei der Akkuratheit noch bei der Reaktionszeit konnte gezeigt werden, dass Same-Race Gesichter besser wiedererkannt werden als Other-Race Gesichter. Es kann ausgeschlossen werden, dass dieser Effekt aufgetreten ist, weil die Probanden regelmäßigen Kontakt zu schwarzen Menschen hatten, da dies durch ein Kontaktformular erhoben wurde. Da sich jedoch keine Korrelation mit dem ORB ergab, kann der Kontakt als Erklärung ausgeschlossen werden.

Es findet sich dennoch eine signifikante Interaktion zwischen Ethnie und Bekanntheit. Für die schwarzen Stimuli gab es also mehr Hits (Treffer) als für die weißen Stimuli. Dies könnte durch den eher konservativen Antwortstil erklärt werden, wobei für weiße Stimuli noch konservativer geantwortet wurde als für schwarze Stimuli.

Dass fröhliche Emotionen den Own-Race Bias verringern, konnte auch nicht bestätigt werden, da kein ORB in diesem Experiment gefunden wurde.

Es konnte nicht bestätigt werden, dass die evozierte Emotion einen Einfluss auf den ORB hat und dass fröhlich gestimmte Menschen Stimuli mit fröhlichem Ausdruck besser wiedererkennen. Die Stimmungsabfrage vor und nach den Videos konnte aber auch bestätigen, dass die Videos keinen Einfluss auf die Stimmung hatten.

Da insgesamt kein ORB gefunden wurde, können die anderen Befunde nur unter Vorbehalt betrachtet werden. Dadurch ist es schwer, einen Einfluss durch die Emotionen der Stimuli sowie durch die evozierte Emotion zu belegen. Ein möglicher Grund für den nicht gefundenen ORB könnte sein, dass Stimuli aus unterschiedlichen Datenbanken verwendet wurden.

### 5. Literatur

- Chen, J. (2014). Face recognition as a predictor of social cognitive ability: Effects of emotion and race on face processing. *Asian Journal of Social Psychology, 17*(1), 61-69.
- Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & emotion, 19*(3), 313-332.
- Hu, C. S., Wang, Q., Han, T., Weare, E., & Fu, G. (2017). Differential emotion attribution to neutral faces of own and other races. *Cognition and Emotion, 31*(2), 360-368.
- Johnson, K. J., & Fredrickson, B. L. (2005). "We All Look the Same to Me" Positive Emotions Eliminate the Own-Race Bias in Face Recognition. *Psychological science, 16*(11), 875-881.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta analytic review

# Ein Zeuge, der lügt, kennt die Wahrheit nicht – oder doch? Einfluss der Gedächtnisstärke auf den Lügen-Effekt

Theresa Epperlein, Nils Frisius, Stefanie Hunger, Sarah Jakob, Laura Sterba

Leitung: Franziska Schreckenbach

## 1. Einleitung

In vielen Kriminalfilmen werden nach einer Tat die Opfer zum Tathergang, Täterprofil, Umgebung und anderen wichtigen Details befragt. Doch wie hilfreich sind Zeugen, die aufgrund mangelnder Erinnerungen unsicher in ihren Aussagen sind? Besteht eine Tendenz, sich die eigenen Lügen irgendwann tatsächlich selbst zu glauben? Was ist mit jenen Zeugen, die stets sozial erwünscht antworten? Einige Forscher gehen davon aus, dass ein implizites Gedächtnis für Lügen existiert (Koranyi, Schreckenbach & Rothermund, 2015). Doch lässt sich der Lügeneffekt bei Unsicherheit über Aussagen zu einer Situation überhaupt noch erkennen? Für unsere Untersuchung stellen wir die Hypothese auf, dass der Gedächtniseffekt für Lügen besonders stark auftritt, wenn bereits gefestigte Erinnerungen zu einem Ereignis vorhanden sind. Zusätzlich wollen wir überprüfen, inwiefern die Tendenz, sich sozial erwünscht darzustellen, Einfluss auf den Gedächtniseffekt nimmt.

## 2. Methode

*Stichprobe:* Nach Ausschluss von 11 Versuchspersonen (VPN) aufgrund einer zu geringen Accuracy bzw. der Missachtung der Aufgabenstellung setzte sich die studentische Stichprobe aus 12 männlichen und 30 weiblichen Probanden zusammen ( $M = 21.9$  Jahre,  $SD = 3.0$ ).

*Ablauf und Materialien:* Zu Beginn unseres Experiments bearbeiteten die VPN den Fragebogen „Balanced Inventory of Desirable Responding“ (BIDR), welcher 20 Items zur Erfassung der sozialen Erwünschtheit beinhaltet. Nach den Instruktionen zum Ablauf des weiteren Experiments am PC wurde eine fünfminütige Videosequenz abgespielt, wobei es sich um ein verbrecherisches Szenario im Rahmen eines Raubüberfalls handelte. Danach bekamen die Probanden acht Fragen präsentiert, von denen sich vier auf die Täter bezogen und vier auf die Opfer. Entsprechend ihrer Versuchspersonennummer gehörten die VPN der kurzen (Gruppe 1) oder der langen Bedingung (Gruppe 2) an. Während die VPN der Gruppe 1 direkt nach Absolvieren des ersten Teils zum Interview gebracht wurden, sahen die anderen erneut das Video. Im Interview sollten sie zu einem der beiden vorab präsentierten Themenkomplexe, Täter oder Opfer, lügen. Die Anweisungen, zu welchem Thema gelogen werden sollte, wurden ebenfalls über den PC gezeigt. Der Interviewer wies lediglich darauf hin, das Video selbst nicht zu kennen und demnach nicht direkt nachvollziehen zu können, zu welchen Fragen gelogen bzw. die Wahrheit gesagt wurde. Zurück am PC galt es für beide Versuchsgruppen, den zweiten Teil, das Reaktionszeitexperiment, zu durchlaufen. Die Täter- bzw. Opferfragen dienten hierbei als Prime und die Wörter „ehrlich“ und „gelogen“ als Target (degradiert durch Sonderzeichen). Per Tastendruck (D oder K) sollte dann entschieden werden, ob es sich bei der Antwort um „ehrlich“ oder „gelogen“ handelte. Jeder Durchgang begann mit einem Fixationskreuz (750ms), gefolgt von einer Prime-Frage. Nach weiteren 300ms erschien auf dem Bildschirm das Target-Wort und verblieb bis zum Tastendruck der VPN. Vor dem eigentlichen Experimentaldurchgang erfolgte ein kurzer Übungsblock. Der Versuchsaufbau spiegelte demnach ein 2 (Prime: ehrlich vs. gelogen) x 2 (Target: ehrlich vs. gelogen) x 2 (Bedingung: kurz vs. lang)-Design wider, wobei Prime und Target Within-Subject-Faktoren darstellten, während die Bedingung einen Between-Subject-Faktor repräsentierte. Zum Abschluss des Experiments folgte, ebenfalls am PC, ein Fragebogen mit 10 Aussagen zum Video, welche von der VPN als „richtig oder „falsch“ eingeschätzt werden sollten. Diese dienten als

Gedächtnisfragen, um zu testen, wie genau sich die Teilnehmer noch an das gezeigte Video erinnerten.

### 3. Ergebnisse

Um unsere Hypothesen zu prüfen, rechneten wir zunächst eine 2 (Prime: ehrlich vs. gelogen) x 2 (Target: ehrlich vs. gelogen) x 2 (Bedingung: kurz vs. lang) ANOVA. Schnellere Reaktionen auf kongruente Prime-Target-Kombinationen wären dabei ein Beleg für den Gedächtniseffekt. Die Dreifachinteraktion wurde jedoch nicht signifikant, ebenso zeigte sich keine Interaktion von Prime und Target über alle Probanden hinweg (alle  $F < 2.81$ ). Bei einem Vergleich der Erinnerungsleistung an das Video (mittlere Accuracy im Gedächtnistest am Ende des Experiments) stellte sich zwischen den beiden Gruppen ebenfalls kein signifikanter Unterschied ( $t < |1.16|$ ) heraus, die Manipulation der Bedingungen schien demnach keinen Effekt zu erzielen. Um dennoch zu testen, ob die Stärke der Erinnerung einen Einfluss auf den Gedächtniseffekt hat, berechneten wir ein lineares Regressionsmodell mit den Prädiktoren Erinnerungsleistung und Gruppe, sowie der abhängigen Variable Kongruenzeffekt (RT in inkongruenten Trials – RT in kongruenten Trials). Dieses Modell wurde signifikant,  $R^2 = 0.28$ ,  $F(2,35) = 7.70$ ,  $p = 0.002$ . Die Gedächtnisleistung erwies sich als signifikanter Prädiktor für den Kongruenzeffekt ( $b = 354.36$ ,  $t(39) = 3.44$ ,  $p = 0.001$ ), während die Gruppenzugehörigkeit keine zusätzliche Varianz aufklärte ( $t = -1.26$ ). So wiesen diejenigen VPN den stärksten Effekt auf, deren Gedächtnisleistung hinsichtlich des Videos am besten ausfiel. Infolgedessen rechneten wir eine zusätzliche 2 (Prime: ehrlich vs. gelogen) x 2 (Target: ehrlich vs. gelogen) ANOVA über 36 VPN mit sehr guten Leistungen in der Gedächtnisaufgabe (*mittlere Accuracy = 1*). Für diese Probandengruppe zeigte sich der Kongruenzeffekt durch eine signifikante Interaktion ( $p = 0.048$ ,  $F(1,35) = 4.19$ ). In einer weiteren Regression mit dem BIDR-Score als Prädiktor ließ sich kein Einfluss auf die Stärke des Kongruenzeffekts erkennen, ( $F < 1$ ).

### 4. Diskussion

In unserer Studie untersuchten wir, inwiefern sich das Erinnerungsvermögen an eine Situation und bestimmte Persönlichkeitsmerkmale auf den Gedächtniseffekt für Lügen auswirken. So versuchten wir zunächst nachzuweisen, dass uns bekannte, vertrautere Thematiken eine gefestigtere Gedächtnisspur erzeugen und für diese somit ein höherer Kongruenzeffekt auftritt als für Situationen, die uns weniger vertraut sind. Die dahingehende Manipulation mittels unserer zwei Videobedingungen ließ den erwarteten Effekt allerdings nicht zu. Legen wir unser Analyse hingegen die gruppenunabhängige und individuell verschiedene Gedächtnisleistung der VPN zugrunde, zeigen sich signifikante Ergebnisse: Je besser die Erinnerungen der VPN an den tatsächlichen Tathergang des Raubüberfalls ausfallen, desto stärkere Prime-Target-Kongruenzen traten auf. Dies könnte darauf hinweisen, dass geringere Gedächtnisleistungen die Grenze zwischen subjektiv und objektiv wahrgenommener Wahrheit und Lüge verschwimmen lassen. Dieser Umstand könnte speziell für die Interpretation von Zeugenaussagen ein Problem mit sich bringen: Wie kann der wahre Tathergang rekonstruiert werden, wenn ein Zeuge seine etwaige Lüge womöglich irgendwann selbst als Wahrheit ansieht bzw. nicht mehr in der Lage ist, zwischen bewusst verzerrten und wahren Aspekten zu trennen? Ferner untersuchten wir am Beispiel des BIDRs, welchen Einfluss bestimmte Persönlichkeitsmerkmale auf den Gedächtniseffekt nehmen. Dabei zeigten sich keine signifikanten Unterschiede zwischen Personen, die im Alltag eher zur Lüge tendieren als andere, die weniger sozial erwünschte Antworten geben.

### 5. Literatur

- Koranyi, N., Schreckenbach, F., Rothermund K., (2015): The Implicit Cognition of Lying: Knowledge about Having Lied to a Question is Retrieved Automatically. *Social Cognition*
- Hommel, B. (1998). Event Files: Evidence for Automatic Integration of Stimulus-Response Episodes. *Visual Cognition*, 5(1/2), 183-216. doi: 10.1016/j.tics.2004.08.00

## **Need for Speed – wie Risikokognitionen unsere Bestrafungstendenz beeinflussen**

**Paul Otto Löbl, Katharina Rinke, Rica Volmerich, Clara Ritter, Sophie Gräßer-Dietsch**

Leitung: David Urschler

### 1. Einleitung

In den letzten Jahren hat das Spielen von PC- und Videospiele zugenommen. Besonders unter männlichen Jugendlichen sind Autorennspiele beliebt (Gentile, Lynch, Linder, Walsh, 2004). Oft beinhalten diese Spiele rücksichtsloses Fahren, Geschwindigkeitsüberschreitungen und Kollisionen mit anderen Fahrzeugen. Aus vergangener Forschung weiß man bereits, dass gewaltbeinhaltende Videospiele die aggressionsbezogenen Risikokognitionen erhöhen (Anderson & Bushman, 2001) und ebenso der Affekt und das Verhalten beeinflusst wird (Fischer, Kubitzki, Guter, Frey, 2007). Doch was passiert, wenn man nach dem Spielen das Strafmaß von Delikten beispielsweise im Straßenverkehr beurteilen soll? Gehen wir dann wohlwollender mit den Tätern um, wenn wir gerade selbst am PC ein Autorennen gefahren sind? Bereits im vorherigen Semester erforschten wir verschiedene Einflussfaktoren, welche auf die Bestrafungstendenz wirken. Dieses Semester beschäftigten wir uns damit, inwiefern PC-Spiele vermittelt über aktivierte Risikokognitionen unsere Bestrafungstendenz beeinflussen.

### 2. Methode

Unsere erste Hypothese (H1) lautete: *Risikorelevante PC-Spiele verstärken Risikokognitionen.* Unsere zweite Hypothese (H2) lautete: *Risikorelevante PC-Spiele verringern das Ausmaß der Bestrafungstendenz.*

Bei der Studie nutzten wir ein einfaktorielles Design mit drei Stufen zur Klärung unserer Fragestellung. Die Studie bestand aus zwei Testteilen - einem 10 minütigen PC-Spiel und einem anschließenden Online-Fragebogen. Zunächst lasen die Probanden eine Coverstory, die das eigentliche Ziel unseres Experimentes verdecken sollte.

Die Kontrollgruppe (KG) spielte das Spiel „*RollerCoaster Tycoon 2*“.

Hierbei sollte ein Freizeit-Erlebnis-Park nach eigenen Vorstellungen gestaltet werden. Das Ziel des Spiels ist es, den Gewinn des Parks zu steigern und Gäste anzulocken. Um das zu erreichen, konnten verschiedene Attraktionen gebaut werden. Dieses Spiel sollte die Risikokognitionen nicht verstärken und die Bestrafungstendenz nicht beeinflussen.

Die Ergebnisgruppe (EG2) spielte das Spiel „*Need for Speed Most Wanted*“.

Dieses ist ein Arcade-Rennspiel, bei dem man möglichst schnell im Ziel ankommen soll. Dabei gab es konkurrierende Fahrer und andere Verkehrsteilnehmer, mit denen man im öffentlichen Straßenverkehr kollidieren konnte. Um die Ziele des Spiels zu erreichen, müssen die Straßenverkehrsregeln bewusst missachtet werden. Laut H1 ist hier eine Verstärkung der Risikokognitionen zu erwarten. Das sollte laut H2 wiederum einen negativen Einfluss auf die Bestrafungstendenz haben.

Die Vergleichsgruppe (EG) spielte das Spiel „*BMW M3 Simulator*“.

Dabei handelt es sich um eine Rennspiel-Simulation. Auch hier sollte man möglichst schnell ins Ziel gelangen. Der Unterschied zu „*Need for Speed Most Wanted*“ besteht darin, dass der Spieler allein und auf einer abgegrenzten Rennstrecke und unter kontrollierten Bedingungen fährt. Außerdem sind die Fahreigenschaften viel realistischer. Hier ist ein geringerer Einfluss auf die Risikokognitionen und die Bestrafungstendenz als bei der Ergebnisgruppe zu erwarten.

Nach Ende der Spielzeit, wurden in allen Gruppen die Risikokognitionen in einem Online-Fragebogen abgefragt über Homonyme, die entweder risikobezogen oder neutral interpretiert werden konnten z.B. „*RENNEN*“. Als Nächstes wurden Fragen gestellt, die der Ablenkung der

Probanden dienen sollten und nicht relevant für unsere Fragestellung waren (z.B. „Wie interessant fanden Sie das Spiel?“).

Im Hauptteil des Fragebogens wurden fünf Szenarien in Textform dargelegt: eine Schlägerei, Fahrerflucht, Vandalismus, Diebstahl mit Sachschaden und ein illegales Autorennen. Im Anschluss daran sollten die Probanden angeben, wie wichtig ihnen jeweils eine Bestrafung ist. Außerdem sollten sie festlegen, wie stark der Täter bestraft werden soll.

Zum Schluss wurden noch demographische Fakten der Versuchsteilnehmer erfasst.

### 3. Ergebnisse

Am Experiment nahmen 28 männliche und 26 weibliche Versuchspersonen teil (N=55, Durchschnittsalter = 21,24), die randomisiert auf die drei Versuchsbedingungen verteilt wurden (EG2: N=18, EG: N=18, KG: N=19). Das Strafmaß wurde mit einer 8-stufigen Ordinalskala erfasst. Für die EG2 fiel das durchschnittliche Strafmaß am geringsten aus, für die EG am höchsten. (KG: M= 4.11, EG: M= 4,2, EG2: M= 3,99). Die Gruppenunterschiede des verkehrsbezogenen Szenarios waren noch deutlicher. Hier war die Bestrafungstendenz der EG2 auch am geringsten. (KG: M= 6.1, EG: M= 6, EG2: M= 5.67). Bezogen auf die Wichtigkeit der Bestrafung (7-stufige Intervallskala) fielen die Unterschiede ähnlich aus (KG: M=6.67, EG: M=6.89, EG2: M=6.53). Laut Varianzanalysen (Anova) war jedoch keiner dieser Gruppenunterschiede signifikant. *Hypothese H2 kann also widerlegt werden.* Die EG2 hatte die meisten Risikokognitionen, die KG die geringsten. Allerdings zeigten auch hier die einschlägigen Varianzanalysen keinen signifikanten Zusammenhang auf. *Demnach kann die Hypothese H1 ebenfalls widerlegt werden.*

Ein signifikantes Ergebnis war, dass die Personen die angegeben haben bereits selbst im Straßenverkehr auffällig gewesen zu sein, ein geringeres Strafmaß bei dem verkehrsbezogenen Szenario ansetzten und ihnen die Strafe weniger wichtig ist.

### 4. Diskussion

Zwar waren unsere Ergebnisse nicht signifikant, allerdings ließen sich einige Tendenzen in die von uns vermutete Richtung feststellen. Das wird besonders dadurch bestärkt, dass die Unterschiede im Bestrafungsverhalten dann deutlicher ausfielen, wenn man sie separat für das verkehrsbezogene Szenario betrachtet. Möglicherweise lag es an der verhältnismäßig kleinen Stichprobe, dass diese Ergebnisse nicht signifikant waren. Außerdem könnte man bei zukünftigen Studien die kurze Spielzeit von 10 Minuten erhöhen, um dadurch gegebenenfalls stärkere Effekte zu erzielen.

Es blieb unklar, ob die tendenziellen Unterschiede im Bestrafungsverhalten tatsächlich über die verstärkten Risikokognitionen vermittelt worden sind oder von anderen Effekten der PC- Spiele, wie z.B. einer erhöhten körperlichen Erregung, verursacht wurden. Besonders, wenn man versucht ein virtuelles Rennen zu gewinnen könnte die körperliche Anspannung deutlich erhöht sein, da derartige Spiele sehr aufregend sind.

### 5. Literatur

- The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. D.A. Gentile, P.J. Lynch, J.R. Linder, D.A. Walsh. *Journal of Adolescence: Applied* 2004. Vol. 27, 5-22.
- Effects of violent video games on aggressive behaviour, aggressive cognition, aggressive affect, physiological arousal, and prosocial behaviour: A meta-analytic review of the scientific literature. C.A. Anderson, B.J. Bushman. 2001. *Psychological Science: Applied* 2001, Vol. 12, 353-359.
- Virtual Driving and Risk Taking: Do Racing Games Increase Risk-Taking Cognitions, Affect, and Behaviors?. P. Fischer, J. Kubitzki, S. Guter, D. Frey. *Journal of Experimental Psychology: Applied* 2007. Vol. 13, No. 1, 22-31.

# **“Ebony and Ivory live together in holistic harmony”: Der Einfluss von Ethnie und Bekanntheit auf die Composite Face Illusion**

**Gesine Franke, Bianka Knorr, Alexandra Kruppa, Katharina Kühne, Nandiyin  
Nandiyin**

Leitung: Dr.J.M. Kaufmann

## 1. Einleitung

Die Tendenz Gesichter der eigenen Ethnie besser wiedererkennen zu können, der sogenannte „Own-Race Bias“ (ORB), wurde bereits in vielen Studien nachgewiesen (Meissner & Bringham, 2001). Rossion und Michel (2011) schlugen als Erklärung vor, dass Gesichter der eigenen ethnischen Gruppe (Same-Race, SR) generell holistischer verarbeitet werden, als Gesichter anderer Ethnien (Other-Race, OR). Evidenz hierfür lieferte eine Studie, welche die holistische Wahrnehmung mit Hilfe der „Composite-face illusion“ operationalisierte (Michel, Rossion, Han, Chung, & Caldara, 2006): Wird die obere Hälfte eines Gesichts mit der unteren Hälfte eines anderen Gesichts kombiniert, wird dieses „Composite Face“ eher als ein neues Gesicht wahrgenommen, und es fällt schwer, die einzelnen Identitäten zu erkennen, aus denen es zusammengesetzt ist. Die Autoren fanden, dass dieser Effekt stärker ausgeprägt war, wenn beide Gesichtshälften der ethnischen Gruppe der Versuchspersonen entsprachen, und schlussfolgerten daraus, dass OR-Gesichter weniger holistisch wahrgenommen werden als SR-Gesichter. Diese Ergebnisse sind bisher allerdings auf unbekannte Gesichter beschränkt. Wenn, wie häufig postuliert (siehe z.B. Rossion & Michel, 2011), SR und OR Gesichter generell qualitativ anders verarbeitet werden, sollte sich auch für bekannte OR Gesichter eine schwächer ausgeprägte „Composite Face Illusion“ zeigen. Sollte der Effekt aber spezifisch für unbekannte Gesichter sein, würde dies die Notwendigkeit einer teilweisen Rekonzeptualisierung des ORB nahelegen. In der folgenden Studie wurde daher untersucht, ob die Bekanntheit der präsentierten SR und OR Gesichter die holistische Verarbeitung ihrer Identität -gemessen mit der „Composite-face illusion“- beeinflusst. Basierend auf theoretischen Überlegungen und einer Pilotstudie aus dem Wintersemester erwarteten wir, im Widerspruch zu Rossion & Michel (2011), dass sich auch für OR Gesichter das Phänomen holistischer Verarbeitung der Identität zeigen sollte, wenn sie den Versuchspersonen bekannt sind.

## 2. Methode

An der Studie nahmen insgesamt 28 Versuchspersonen mit einem Durchschnittsalter von  $M = 20$  Jahren (Range: 18-27) teil. Davon waren 20 Personen weiblich, 7 männlich, eine Versuchsperson gab kein Geschlecht an.

Für das Experiment wurden Graustufen-Bilder der Gesichter von insgesamt 60 Individuen verwendet. Die Hälfte waren jeweils bekannte (d.h. berühmte) und unbekannte Personen, und jeweils die Hälfte von diesen war afrikanisch, bzw. mitteleuropäisch aussehend. Die Bekanntheit wurde vorher in einer Pilotstudie getestet, um sicherzustellen, dass der Bekanntheitsgrad der schwarzen (v.a. afro-amerikanischen Künstler, Sportler und Politiker) und weißen Berühmtheiten vergleichbar war. Zusätzlich zu den Originalgesichtern wurden „Composite Faces“ hergestellt. Hierfür wurden die Gesichter mit Hilfe von Photoshop in eine obere und untere Hälfte geteilt und dann jeweils mit der Gesichtshälfte eines anderen Gesichts gleicher Ethnie, Bekanntheit und Geschlechts zusammengesetzt. Von jeder Kombination wurden zwei Versionen hergestellt: In der „Aligned-Bedingung“ waren beide Gesichtshälften genau übereinander angeordnet, während sie in der „Misaligned-Bedingung“ horizontal verschoben waren. Die Trails der Testdurchgänge setzten sich wie folgt zusammen: Nach einem Fixationskreuz in der Mitte des Bildschirms (300ms) folgte ein schwarzer Bildschirm (200ms), gefolgt von einem Originalgesicht (600ms), auf welches erneut

ein schwarzer Bildschirm (300ms) folgte. Dann wurde ein Composite-Teststimulus präsentiert (1000ms), auf den reagiert werden musste. Die Aufgabe bestand darin, jeweils die untere Hälfte der gezeigten Composite Stimuli zu ignorieren und per Tastendruck zu entscheiden, ob die obere Hälfte des Composites dem vorher gezeigten Originalgesicht entsprach, oder nicht. Die Versuchspersonen erhielten genaue Instruktionen zum Ablauf, die vor Beginn des Experiments auf dem Bildschirm gezeigt wurden. Das Experiment bestand aus insgesamt 480 Trials á 3,6 Sekunden und dauerte insgesamt ca. 50 Minuten.

### 3. Ergebnisse

Zunächst wurde sowohl für die Antwortgenauigkeiten, als auch für die Reaktionszeiten für alle Bedingungskombinationen der Composite Effekt berechnet („Misaligned-Bedingung“ minus „Aligned-Bedingung“). Für diese Werte wurden dann Varianzanalysen mit Messwiederholung auf den Faktoren Bekanntheit (bekannt vs. unbekannt), Ethnie (afrikanisch vs. europäisch) und Matchingbedingung (Originalgesicht im Composite enthalten vs. Originalgesicht nicht enthalten) berechnet. Es zeigte sich, dass der Composite Effekt in den Antwortgenauigkeiten nur durch den Faktor Matchingbedingung moduliert wurde,  $F(1,27) = 17.001$ ,  $p < .001$ ,  $\eta_p^2 = .386$ , d.h. der beschriebene Effekt war für Match-Trials größer als für Mismatch-Trials, (Matching:  $M = .040$ ; Mismatch:  $M = -.012$ ).

Bei den Reaktionszeiten fand sich ein Haupteffekt für Bekanntheit,  $F(1,27) = 6.571$ ,  $p = .016$ ,  $\eta_p^2 = .196$ , welcher einen größeren Composite Effekt für unbekannte Gesichter im Vergleich zu bekannten zeigte. Des Weiteren fand sich auch hier, wie bei den Genauigkeiten, ein Haupteffekt für den Faktor Matchingbedingung,  $F(1,27) = 50.326$ ,  $p < .001$ ,  $\eta_p^2 = .651$  mit stärker ausgeprägtem Composite Effekt für Matchtrials.

### 4. Diskussion

Wie postuliert fand sich in dieser Studie ein Composite Effekt. Dieser wurde allerdings nicht signifikant durch die Ethnie der Gesichter moduliert. Diese Ergebnisse lassen darauf schließen, dass der ORB, der sich typischerweise in Gesichterlern- und Gedächtnisaufgaben zeigt, nicht durch Unterschiede in der holistischen Verarbeitung von same-race und other-race Gesichtern zu Stande kommt. Die Beobachtung, dass bei der Verwendung bekannter Stimuli vergleichbar große Composite Effekte für Gesichter der eigenen und einer anderen Ethnie entstehen, legt die Vermutung nahe, dass SR und OR Gesichter nicht generell qualitativ anders verarbeitet werden. Sollten sich diese Ergebnisse auch für andere Ethnien, idealerweise unter Verwendung eines gekreuzten Designs (also auch mit Stichproben anderer Ethnien) replizieren lassen, sollte das Konzept des ORB neu überdacht werden. Für weiterführende Forschung wäre es interessant zu prüfen, ob sich der Einfluss von Bekanntheit auf den ORB durch intensivere oder reale Bekanntschaft verstärkt.

### 5. Literatur

- Michel C., Rossion B., Han J., Chung C-S., Caldara R. (2006). Holistic Processing Is Finely Tuned for Faces of One's Own Race. *Psychological Science*, 17, 608-615.
- Meissner C.A., Brigham J.C. (2001). Psychology, Thirty Years of Investigation the Own-Race Bias in Memory for Faces a Meta-Analytic Review. *Public Policy, and Law*, 7, 3- 35.
- Rossion, B., & Michel, C. (2011). An Experience-Based Holistic Account of the Other-Race Face Effect. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 215-243). Oxford New York: Oxford University Press

# “I’ve been looking for freedom” – Effekte des IAT auf den freien Willen, Kontrollüberzeugungen und aggressives Verhalten

Paula Kiefert, Wilma Middendorf, Stella Pfund, Deborah Löschner, Alexander Fuchs

Leitung: Franziska Meißner

## 1. Einleitung

Welche unbekanntes Tücken liegen im Schatten eines impliziten Assoziationstests (IAT) verborgen? Der IAT ist ein sehr bekanntes und häufig verwendetes Maß zur indirekten Erfassung von Einstellungen. Inzwischen wurde allerdings gezeigt, dass der IAT, über seine Funktion zur Messung von Einstellungen hinaus, nachfolgendes Verhalten beeinflusst. Vorauer (2012) zeigte in einer Studie, dass sich das Durchführen des IATs negativ auf Intergruppenkontakte auswirkt und vermutete reduzierte Verhaltenskontrolle und Selbstwirksamkeit als mögliche vermittelnde Prozesse. Jedoch wurden potentielle Mediatoren nur indirekt und methodisch unsauber erfasst.

In der aktuellen Studie soll die Fragestellung auf einen anderen Kontext erweitert werden: Es soll untersucht werden, ob der IAT auch nachfolgendes *aggressives* Verhalten beeinflusst und welche Wirkmechanismen zugrunde liegen. Angenommen wurde, dass die wahrgenommene Verhaltenskontrolle (Selbstwirksamkeitserwartung [SWE] und Locus of Control [LOC]; vgl. Vorauer, 2012), sowie der Glaube an Determinismus medierende Variablen sind. Letzteres geht zurück auf Dalski, Knitter, Kölbel, Middendorf und Vogel (2017), die nahelegten, dass implizite Maße Determinismusglauben erhöhen. Ein ausgeprägter Glaube an Determinismus steigert wiederum aggressive Verhaltensweisen (Baumeister, Masicampo & DeWall, 2009).

Darauf aufbauend sollten sich Probanden nach einem Aggressivitäts-IAT aggressiver verhalten als Probanden, die einen Fragebogen zur selben Thematik ausgefüllt haben (H1). SWE und LOC als vermittelnde Prozesse sollten durch den IAT geschwächt werden, da das Antwortverhalten weniger kontrollierbar erscheint; dadurch wird im Anschluss aggressiveres Verhalten gezeigt (H2). Der Glaube an Determinismus sollte dagegen durch das Bearbeiten des IATs gestärkt werden, wodurch wiederum ebenfalls aggressiveres Verhalten gezeigt wird (H3).

## 2. Methode

Am Experiment haben 41 Probanden zwischen 19 und 30 teilgenommen. Der Kontrollgruppe (KG,  $N = 21$ ) wurde ein Aggressivitäts-Fragebogen dargeboten. Die Experimentalgruppe (EG,  $N = 20$ ) bekam einen Aggressivitäts-IAT. Im Anschluss wurde in beiden Gruppen SWE (10 Items,  $\alpha = .86$ ), LOC (24 Items,  $\alpha = .65$ ) sowie der Glaube an Determinismus (9 Items,  $\alpha = .72$ ) gemessen.

Anschließend spielten die Probanden in der Taylor Competitive Reaction Time Task (TCRTT, 25 Items,  $\alpha = .97$ ) gegen einen fiktiven Gegner. Dabei wurde der für den Gegner eingestellte „Bestrafungston“ über 25 Durchgänge hindurch gemittelt und als Maß für Aggressivität verwendet. Als weiteres Mittel, um Aggressivität zu messen, wurde hypothetisches Verhalten in Szenarien erfasst (selbst generiert, 17 Items,  $\alpha = .57$ ). Durch einen anfänglichen Fehler im Programm liegen hierbei nur für 29 Probanden Daten vor. Am Ende bearbeiteten die Versuchspersonen einen Manipulationscheck, von dem drei Items die wahrgenommene Kontrolle bzgl. des eingangs durchgeführten Aggressivitätstests (IAT/Fragebogen) erfassten. Diese wurden zur für die Auswertung relevanten Variable „handlungsspezifische Kontrolle“ zusammengefasst ( $\alpha = .62$ ).

## 3. Ergebnisse

Die Überprüfung des IAT-Einflusses auf das gezeigte Verhalten zeigte zunächst keine bedeutsamen Unterschiede zwischen KG und EG, weder für TCRTT,  $t(39) = 0.24$ ,  $p = .810$ ,  $d = 0.08$ , noch für die Szenarien,  $t(27) = -0.27$ ,  $p = .787$ ,  $d = -0.10$ ). Diese Ergebnisse schließen jedoch nicht aus, dass

dieser Nulleffekt durch gegensätzliche Wirkmechanismen entstand. Die Effekte der Manipulation auf die potentiellen Mediatorvariablen stützen diese post-hoc generierte Annahme: Der Glaube an Determinismus war im Vergleich zur EG ( $M = 2.88$ ,  $SD = 0.56$ ) bei der KG ( $M = 3.20$ ,  $SD = 0.55$ ) marginal signifikant stärker ausgeprägt,  $t(39) = 1.85$ ,  $p = .072$ ,  $d = 0.58$ . Sowohl auf SWE, ( $t(39) = 0.02$ ,  $p = .983$ ,  $d = 0.01$ ), als auch auf LOC, ( $t(39) = -1.06$ ,  $p = .297$ ,  $d = -0.33$ ), zeigte die Manipulation keinen Effekt. Bei beiden Konstrukten ist jedoch davon auszugehen, dass sie tiefer in der Person verankerte *traits* erfassen. Die Untersuchung der situationsbedingten, handlungsspezifischen Kontrolle hingegen verdeutlichte, dass die EG ( $M = 3.53$ ,  $SD = 1.00$ ) wie erwartet weniger Handlungskontrolle erlebte als die KG ( $M = 4.25$ ,  $SD = 0.86$ ),  $t(39) = 2.48$ ,  $p = .018$ ,  $d = 0.77$ .

Auf diesem Wissen baute die Betrachtung des Zusammenhangs der potentiellen Mediatoren und dem aggressiven Verhalten auf. Während sich bzgl. TCRTT keine bedeutsamen Zusammenhänge zeigten, korrelierten die Antworten auf die Szenarien positiv mit dem Glauben an Determinismus ( $r = .32$ ,  $p = .089$ ) und negativ mit der handlungsspezifischen Kontrolle ( $r = -.34$ ,  $p = .068$ ). Eine Mediationsanalyse bestätigte letztlich die ergebnisimplizierte Vermutung eines gegensätzlichen Wirkmechanismus zwischen der Manipulation und dem hypothetischem Verhalten in den Szenarien: Einerseits zeigte sich ein signifikanter *positiver* indirekter Effekt der Manipulation auf das Verhalten vermittelt über handlungsspezifische Kontrolle (*bootstrapped effect* = .14,  $SE = .10$ ; 95% Bias corrected and accelerated CI [.003, .370]). Andererseits war auch ein signifikanter *negativer* indirekter Effekt, vermittelt über den Glauben an Determinismus (*bootstrapped effect* = -.15,  $SE = .08$ ; 95% Bias corrected and accelerated CI [-.348, -.025]) zu beobachten.

#### 4. Diskussion

Obwohl sich unsere Hypothese, dass der IAT Auswirkungen auf aggressives Verhalten zeigt, in den ersten Analysen nicht bestätigt hat, konnten wir doch zwei Mediatorvariablen identifizieren. Somit wirkte sich die Manipulation durch den IAT auf zwei konträre Arten auf das hypothetische Verhalten in den Szenarien aus. Während Probanden in der EG weniger handlungsspezifische Kontrolle wahrnahmen und sich dies in aggressiverem Verhalten zeigte, schwächte die Bearbeitung des IAT den Glauben an Determinismus, was zu weniger aggressivem Verhalten führte. Die Steigerung des aggressiven Verhaltens durch mangelndes (spezifisches) Kontrollgefühl ist hypothesenkonform. Der über den Glauben an Determinismus vermittelte Effekt weist jedoch in eine erwartungskonträre Richtung. Es ist jedoch vorstellbar, dass dies auf die gewählte Kontrollgruppe zurückgeht: Vielleicht sorgte die zur Bearbeitung des Aggressivitäts-Fragebogens nötige Introspektion bei den Probanden dafür, dass das eigene Verhalten stärker hinterfragt wurde, was sich dann in höherem Glauben an Determinismus äußerte. In zukünftigen Studien sollten die berichteten Ergebnisse daher mit einer anderen Kontrollgruppe überprüft werden.

Problematische Aspekte der Studie sind sowohl die kleine Stichprobengröße als auch die mangelnde Glaubwürdigkeit der TCRTT-Coverstory. Außerdem entstammten die Szenarien keiner validierten Skala und erfassten lediglich hypothetisches Verhalten. Eine Aussage über tatsächliches aggressives Verhalten lässt sich somit schwerlich treffen. Dennoch liefern die Ergebnisse interessante Hinweise auf mögliche vermittelnde Effekte zwischen dem IAT und aggressivem Verhalten. Eine genauere Analyse der beiden vermittelnden Pfade wäre wünschenswert und notwendig, um konkretere Aussagen treffen zu können.

#### 5. Literatur

- Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, 35, 260-268.
- Dalski, A., Knitter, L. A., Köbel, O., Middendorf, W. M., Vogel, A. (2017). Der Glaube an einen freien Willen und sein Einfluss auf unsere Eigen- und Fremdbewertung. *Empira Bericht, Jena: Friedrich Schiller Universität*.
- Vorauer, J. D. (2012). Completing the Implicit Association Test Reduces Positive Intergroup Interaction Behavior. *Psychological Science*, 23, 1168-1175.

# **„Und führe uns nicht in Versuchung...“ - Implizites ‚wanting‘ und ‚liking‘ für Ziele und Versuchungen im Studienkontext**

**Luisa Borsche, Jennifer Graf, Fenna Plume, Maja Sindermann, Laura Wodara**

Leitung: Laura Anne Grigutsch

## 1. Einleitung

Wenn wir ein Ziel anvisieren, bewerten wir ziieldienliche Reize positiver, während Reize, die dem Ziel entgegenstehen, negativer bewertet werden (Fishbach, Zhang, & Trope, 2010). Ebenso entsteht eine Annäherungsreaktion für ziieldienliche Reize, für Versuchungsreize dagegen eine Vermeidungsreaktion (Fishbach & Shah, 2006). Trotz dieser kognitiven Schutzmechanismen ist es oft nicht leicht Versuchungen standzuhalten und ein Ziel konsequent zu verfolgen. Die aktuelle Studie versucht zu erklären, warum das so ist.

Ideen hierzu stammen aus der Incentive-Sensitization-Theorie (Robinson & Berridge, 1993), welche Mögen (Liking) und Begehren (Wanting) als zwei voneinander unabhängige, auf unterschiedlichen Gehirnmechanismen basierende Prozesse ansieht. Demnach könnten Versuchungsreize zwar, wie bereits experimentell gezeigt, evaluativ abgewertet werden, jedoch weiterhin starkes ‚wanting‘ hervorrufen. In dieser Studie wird versucht die bereits existierenden Befunde zu replizieren und um die Erfassung von ‚wanting‘ zu ergänzen. Dafür wird die Wahrnehmung von Freizeit als mit Studienerfolg konfligierende Versuchung manipuliert. Es wird erwartet, dass diese Manipulation ‚liking‘, sowie die Annäherungs-Tendenz für Freizeit-Reize reduziert, ‚wanting‘ für diese Reize jedoch unverändert lässt.

Wir möchten in unserer Studie die bisherigen Befunde mittels eines auf unsere Fragestellung angepassten Wanting- und Liking-Impliziten-Assoziations-Test (W-IAT und L-IAT) von Koranyi, Grigutsch, Algermissen und Rothermund (2017) erweitern, sowie eine Approach-Avoidance-Task von Fishbach und Shah (2006) in unsere Ergebnisse integrieren.

## 2. Methode

An der ca. 40-minütigen PC-Studie nahmen 40 Studierende der FSU Jena (26 weiblich, Alter:  $M=22.38$ ,  $SD=3.10$ ) teil. Davon wurden 21 Personen der Kontrollgruppe (KG) und 19 Personen der Experimentalgruppe (EG) zugewiesen.

Zu Beginn wurde durch Fragen zum Studienfach das akademische Ziel aktiviert. Anschließend wurde durch einen manipulierten Zeitungsartikel die Wahrnehmung von „Freizeit“ als damit konfligierendem Konzept manipuliert. Während das Konzept ‚Freizeit‘ in der Experimentalgruppe als für Studienerfolg hinderliche Versuchung dargestellt wurde, wurde es in der Kontrollgruppe unter dem Blickwinkel eines notwendigen Ausgleichs behandelt. Nach einer Fülleraufgabe wurden der L-IAT und der W-IAT durchgeführt, deren Reihenfolge between-subject balanciert wurde. Der L-IAT bestand aus fünf Blöcken, wobei in zwei kritischen Testblöcken Bilder per Tastendruck in die Kategorien ‚Studium‘ und ‚Freizeit‘ (Target-Dimension), sowie ‚positiv‘ und ‚negativ‘ (Attribut-Dimension) sortiert werden sollten. Dies geschah über zwei Reaktionstasten, sodass jeweils zwei Kategorien (eine jeder Dimension) auf der gleichen Taste lagen. Die Tastenzuordnung unterschied sich zwischen den Testblöcken. Der W-IAT glich dem L-IAT mit folgenden Ausnahmen: (1) Die Kategorien der Attribut-Dimension waren „will ich“ und „will ich nicht“, (2) vor der Bearbeitung wurde ein physiologisches Bedürfnis (Durst) aktiviert, (3) die der „will ich“ Kategorie zuzuordnenden Stimuli (Getränke) waren bedürfnisrelevant und (4) Ausführung der „will ich“-Reaktion führte zur tatsächlichen Bedürfnisreduktion (Gewinn von Wasser zum späteren Konsum).“ Nun folgten zwei Blöcke einer Approach-Avoidance-Task. Hierbei sollten Bilder je nach Kategoriezugehörigkeit (Studium vs Freizeit) mittels eines Joysticks herangezogen oder

weggeschoben werden. Die Zuordnung der Bewegung zu den Kategorien unterschied sich zwischen den Blöcken. Abschließend wurde ein Fragebogen zu expliziten Einstellung zum Studium und den Aussagen des Artikels bearbeitet.

### 3. Ergebnisse

IAT-Effekte für beide IATs wurden auf Basis des D-Scores (Greenwald, Nosek & Banaji, 2003) gebildet, wobei höhere Werte eine Präferenz von Freizeit anzeigen. Eine 2 x 2 mixed model ANOVA mit dem between-subject Faktor ‚Bedingung‘ (KG vs EG) und dem within-subject Faktor ‚IAT‘ (L-IAT vs W-IAT) ergab einen Haupteffekt des Faktors IAT-Typ  $F(1, 38)=6.12, p=.018, \eta^2_p=.139$ , da im L-IAT generell höhere Werte erzielt wurden, als im W-IAT. Der Haupteffekt des Faktors Bedingung war marginal signifikant  $F(1, 38)=2.53, p=.060$  (einseitig),  $\eta^2_p=.062$ , da die IAT-Effekte in der Experimentalgruppe geringer ausfielen, als in der Kontrollgruppe. Die erwartete Interaktion Bedingung x IAT war nicht signifikant ( $F<1$ ).

Zur Auswertung der Approach-Avoidance-Task wurden Inverse-Efficiency-Scores berechnet. Es wurde eine 2 x 2 mixed model ANOVA mit dem between-subject Faktor ‚Bedingung‘ (KG vs EG) und dem within-subject Faktor ‚Block‘ (Studium-wegschieben vs. Freizeit-wegschieben) berechnet. Diese ergab keine Haupteffekte (beide  $Fs<1$ ), jedoch eine signifikante Interaktion beider Faktoren,  $F(1, 37)=3.62, p=.033$  (einseitig),  $\eta^2_p=.089$ , welche daher rührte, dass Personen der Kontrollgruppe der „Studium-wegschieben“-Block leichter fiel, als der „Freizeit-wegschieben“-Block, während Probanden der Experimentalgruppe das umgekehrte Muster zeigten.

Es gab keine signifikanten Gruppenunterschiede in den Vorstellungen über das Studium, der aktuellen Studienleistung und der Einstellung zum gelesenen Zeitungsartikel (alle  $ps>.261$ ).

### 4. Diskussion

Mit der Approach-Avoidance-Task und dem L-IAT ist uns eine konzeptuelle Replikation der Befunde von Fishbach und Kollegen gelungen. Die Aktivierung eines Ziels (Studienerfolg) und damit konfligierender Versuchungen (Freizeit) führen also zu zieldienlichen Annäherungs- und Vermeidungsreaktionen gegenüber Ziel- und Versuchungsreizen sowie zu einer Verschiebung evaluativer Bewertungen (liking) zugunsten der Ziel-Reize. Überraschend zeigte sich das gleiche Muster auch für Begehren (wanting). Für zukünftige Forschung könnte es interessant sein, Zusammenhänge der untersuchten Prozesse mit tatsächlichen Performanzindikatoren erfolgreicher Zielverfolgung zu betrachten, besonders in Situationen, in welchen die Selbstkontrolle versagt und der Versuchung nachgegeben wird. Spannend bleibt also nach wie vor die Frage was die Versuchung zur Versuchung macht.

### 5. Literatur

- Fishbach, A., & Shah, J. Y. (2006). Self-control in action: Implicit dispositions toward goals and away from temptations. *Journal of Personality and Social Psychology*, 90(5), 820–832. <https://doi.org/10.1037/0022-3514.90.5.820>
- Fishbach, A., Zhang, Y., & Trope, Y. (2010). Counteractive evaluation: Asymmetric shifts in the implicit value of conflicting motivations. *Journal of Experimental Social Psychology*, 46(1), 29–38. <https://doi.org/10.1016/j.jesp.2009.09.008>
- Koranyi, N., Grigutsch, L. A., Algermissen, J., & Rothermund, K. (2017). Dissociating implicit wanting from implicit liking: Development and validation of the Wanting Implicit Association Test (W-IAT). *Journal of Behavior Therapy and Experimental Psychiatry*, 54, 165–169. <https://doi.org/10.1016/j.jbtep.2016.08.008>
- Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction, 18(3), 247–291.

# Du bist nicht du, wenn du gestresst bist! Wie Stress das Entscheidungsverhalten bei Männern und Frauen beeinflusst

Franziska Sophia Roller, Rico Stecher, Heide Wenzel

Leitung: Dr. Ewald Naumann

## 1. Einleitung

Auf der Grundlage der Literatur zum Thema „Stress und Entscheidungsverhalten“ (Starcke & Brand, 2016) wird folgende Hypothese getestet:

*Akuter Stress beeinflusst das Entscheidungsverhalten in unsicheren Entscheidungssituationen negativ, da akuter Stress Erfahrungslernen verändert.*

Eine unsichere Entscheidungssituation ist eine, bei der die Ergebniswahrscheinlichkeiten der Entscheidungsalternativen dem Entscheider unbekannt sind. Die *balloon analogue risk task* (BART; Lejuez et al. 2002) wurde als valide Operationalisierung dieser Situation verwendet. In unsicheren Entscheidungssituation kann man nur aus Feedback in ähnlichen, zeitlich vorgeordneten Situationen lernen. Bei akutem Stress wird positives Feedback verstärkt gesucht, negatives Feedback nur unzureichend verarbeitet. Dies führt dazu, dass gestresste Personen riskanter entscheiden. Dieser Prozess ist dem Bewusstsein eher nicht zugänglich, sondern automatisch und kann durch die *feedback-related negativity* (FRN; Miltner et al. 1997) indiziert werden. Riskantere Entscheidungen nach akutem Stress werden nur bei Personen erwartet, die nach dem Stressor das Stresshormon *Cortisol* ausschütten. Als akuter Stressor wird der *socially evaluated cold-pressor test* (SECPT, Schwabe et al. 2008) verwendet. In diesem Test werden zwar alle Personen subjektiv gestresst, aber nur die Hälfte reagiert mit einem substantziellen Cortisolanstieg auf den Stressor.

## 2. Methode

Es wurden 32 Probanden (16 Frauen; u.V. Geschlecht), die nach solchen Kriterien selektiert waren, dass eine unverfälschte Cortisolmessung gewährleistet werden konnte, ausgewertet. 10 Probanden durchliefen die Warmwasserbedingung (KG), 22 wurden mit dem SECPT gestresst. Letztere wurden mit einem geschlechtsspezifischen Mediansplit des Cortisolanstieges in Cortisol-high- und low-responder unterteilt (u.V. Stressgruppe). Das Experiment wurde von der Ethikkommission der Universität Jena genehmigt. Alle erspielten Beträge wurden den Versuchspersonen ausgezahlt.

Die Probanden wurden zunächst dem SECPT ausgesetzt. Sie mussten die rechte Hand 3 Minuten lang in kaltes (in KG: handwarmes) Wasser halten, während sie von einer Versuchsleiterin analysiert wurden. Danach wurden sie dem BART unterzogen, der zunächst aus einer Übung und dann 4x50 Trials bestand. Zu Beginn eines Trials des BART erhält der Proband einen Geldbetrag, den er entweder behalten oder erhöhen konnte, indem er einen virtuellen Ballon weiter aufblies. Je weiter er den Ballon aufblies, desto größer der Gewinn und desto höher die Platzwahrscheinlichkeit des Ballons. Platzte er, ging der Betrag des Trials verloren. Ein Ballon konnte maximal 6mal aufgepumpt werden (u.V. Stufe). Ein Daumen nach oben (positives Feedback) bzw. ein Daumen nach unten (negatives Feedback) signalisierte, ob der Ballon geplatzt war (u.V. Feedback). Die abhängige Variable war die geschätzte bedingte Wahrscheinlichkeit, nach positivem bzw. negativem Feedback auf einer Stufe, beim nächsten Erreichen der gleichen Stufe, den Ballon aufzupumpen.

Speichelentnahmen zur Cortisolanalyse erfolgten vor dem Anbringen des EEG, vor dem SECPT, nach der BART-Übung und danach nach jedem Block des BART (insg. 7mal; Messzeitpunkte C<sub>1</sub>-C<sub>7</sub>). Der Cortisolanstieg wurde wie folgt berechnet:  $\Delta C = (C_3 + C_4) - (C_1 + C_2)$ . Die affektive Stressreaktion wurde mithilfe der State-Version der *Positive and Negative Affect Schedule* (Krohne et al. 1997) erfasst, die immer sofort nach C<sub>2</sub>-C<sub>7</sub> ausgefüllt werden musste (insg. 6mal).

Die abhängigen Variablen wurden mit mehrfaktoriellen ANOVAs ausgewertet, die immer die Gruppenfaktoren Geschlecht und Stressgruppe und die jeweiligen relevanten messwiederholten Faktoren beinhalteten. Signifikante Wechselwirkungen wurden mit geeigneten Folgetests aufgelöst.

### 3. Ergebnisse

Bei der Cortisolausschüttung zeigte sich ein signifikanter Stressgruppe x Geschlecht x Messzeitpunkt - Interaktionseffekt [ $F(12, 156)=6,55$ ;  $p<0,001$ ;  $\eta^2=0,335$ ]. (Abbildung 1).

Wider Erwarten änderte sich der positive Affekt während des Experimentes nicht; der negative Affekt verringerte sich über die Zeit.

Bei den geschätzten bedingten Wahrscheinlichkeiten wurde ein Stressgruppe x Geschlecht x Feedback -Interaktionseffekt signifikant [ $F(2,26)=3,63$ ;  $p<0,05$ ;  $\eta^2=0,22$ ]. (Abbildung 2) Gestresste Frauen reagierten im BART nach negativem Feedback vorsichtiger, nur gestresste Männer, die auch eine Cortisolreaktion zeigten, agierten riskanter. Allerdings war das Risikoverhalten der gestressten Probanden in keiner Bedingung höher als in der Warmwasserbedingung. Bei der FRN-Amplitude zeigte sich der erwartete Haupteffekt Feedback [ $F(1,29)=4,98$ ;  $p<0,05$ ;  $\eta^2=0,16$ ] mit negativeren Amplituden nach negativem Feedback. Kein Effekt, der Stressgruppe beinhaltete, wurde signifikant.

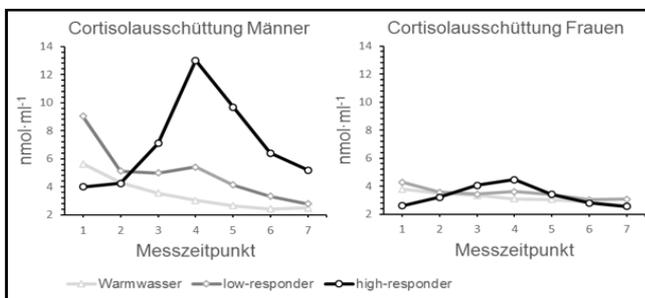


Abbildung 1. Cortisolkonzentration im Speichel bei Männern und Frauen

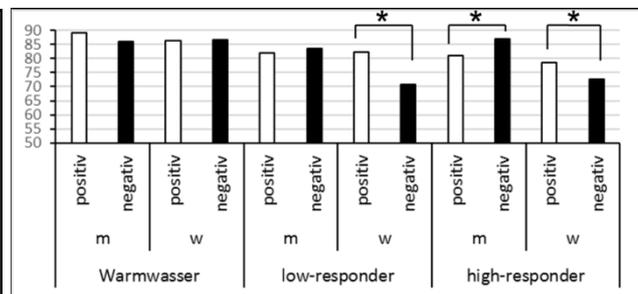


Abbildung 2. geschätzte bedingte Wahrscheinlichkeiten, den Ballon aufzupumpen, in Prozent

### 4. Diskussion

Die oben formulierte Hypothese muss verworfen werden. Der Stressor führte weder bei Cortisol-low-respondern, noch bei -high-respondern zu erhöhtem Risikoverhalten. Zwar verändert sich das Risikoverhalten bei gestressten Männern und Frauen unterschiedlich in Abhängigkeit vom Feedback, dies spiegelt sich aber nicht in der FRN wieder. Allerdings könnte der Effekt, dass Frauen keinen substanziellen Cortisolanstieg zeigten, daran liegen, dass nur weibliche Versuchsleiter den SECPT durchgeführt haben. Zudem erscheint die *Positive and Negative Affect Schedule* ungeeignet, die affektive Stressreaktion zuverlässig zu messen.

### 5. Literatur

- Krohne, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS). *Diagnostica-Göttingen*, 42, 139-156
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75-84.
- Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, 9(6), 788-798.
- Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, 33(6), 890-895.
- Starcke, K., & Brand, M. (2016). Effects of Stress on Decisions under Uncertainty: A Meta-Analysis. *Psychological Bulletin*, 142(9), 909-933.

## **Das aufgePEPte Reaktionszeitexperiment, does sex really sell? Einfluss von Stereotypen auf die Bewerberauswahl**

**Julian Kern, Doreen Krall, Tobias Kűfner, Laura Luxen, Christiane Petzold, Delina  
Ubben**

Leitung: Florian Műller

### 1. Einleitung

Soziale Erwűnschtheit kann explizite Einstellungsmessungen (z.B. durch Fragebűgen) verfűlschen. Implizite Maűe, die spontane und automatische Reaktionen erfassen, sollten dafűr weniger anfűllig sein. Wir nutzen in unserem Experiment das Propositional Evaluation Paradigm (PEP) zur Messung automatischer Bewertungen komplexer Aussagen. Um den prognostischen Wert des PEP zu testen, untersuchen wir die Auswirkung von Geschlechtsstereotypen auf die Bewerberauswahl. Aus der Stereotypenforschung (*Eagly & Karau, 2002*) gibt es den Befund, dass auf stereotyp műnnlich oder weibliche Stellenbeschreibungen eher stereotyp passende Bewerber gewűhlt werden. Folglich soll der PEP die Ausprűgung von Geschlechtsstereotypen erfassen. Zusűtzlich wurden Geschlechtsstereotype explizit via Fragebogen erfasst. Durch unsere Studie kann herausgefunden werden, inwiefern Stereotype die Bewerberauswahl beeinflussen, um in der Praxis Anpassungen treffen zu kűnnen.

### 2. Methode

*Stichprobe:* Es wurden 74 Probanden (51% weiblich, Alter:  $M=23.40$ ,  $SD=3.85$ , Spanne =19-36, 11% Psychologiestudenten) an der Ernst-Abbe-Hochschule und der Friedrich-Schiller-Universitűt Jena angeworben und mit Versuchspersonenstunden oder Sűbigkeiten entlohnt.

*Stereotypmessung mit PEP:* Zunűchst wurden den Probanden Wort fűr Wort (Prűsentationszeit 500ms + 25ms x Buchstabenanzahl) Sűtze gezeigt, die Geschlechtsstereotype (műnnlich / weiblich) beschreiben (angelehnt an den „Bem Sex-Role-Inventory“, *Schneider-Dűker & Kohler, 1988*). Die Paarung jedes Satzes mit einem műnnlichen, als auch einem weiblichen Namen ergab kongruente (z.B. „Lisa ist sehr hilfsbereit.“) und inkongruente Sűtze (z.B. „Jan ist sehr fűrsorglich.“). Ein unabhűngig vom Satzinhalt folgender Reaktionsprompt („richtig“, „falsch“) forderte die Probanden auf entweder die "richtig" oder die "falsch" Taste zu drűcken. Um sicherzustellen, dass die Probanden die Sűtze lesen, gab es zusűtzliche „Catch“-Trials (Prompt: „? richtig – falsch ?“) in denen bewertet werden sollte, ob ein Rechtschreibfehler im Satz vorlag. Geschlechterstereotype zeigen sich im PEP durch kurze RT beim Drűcken von „richtig“ („falsch“) auf kongruente (inkongruente) Trials. Fűr die Analyse wurde eine Effektvariable erstellt, welche sich aus der Differenz dieser beiden RTs ergab.

*Color-Matching-Aufgabe:* Als Pufferaufgabe zwischen Einstellungs- und Verhaltensmessung sollten die Probanden fűr insgesamt 2 Minuten die Sűttigung eines auf dem Bildschirm gezeigten Rechtecks, kontinuierlich an eine zufűllig schwankende Zielvorgabe via Mausbewegung anpassen.

*Verhalten:* Nach jeder von insgesamt 6 Stellenausschreibung, die stereotyp-műnnliche oder stereotyp-weibliche Eigenschaften betonten, sollten die Teilnehmer fűr insgesamt 10 Fotos von Bewerbern (50% weiblich) schnell und spontan via Tastendruck entscheiden, ob sie diese fűr die Stelle geeignet hielten (ja/nein). Stereotypkongruentes Verhalten zeigt sich hier in der Bevorzugung stereotypkongruenter bzw. der Ablehnung inkongruenter Bewerber.

*Expliziter Fragebogen:* Hier wurden die geschlechtsspezifischen Eigenschaften aus dem PEP auf einer Rating -Skala von 1 – „typisch-műnnlich“ bis 7- „typisch-weiblich“ bewertet.

### 3. Ergebnisse

Bei PEP und Bewerberauswahl wurden Reaktionszeiten größer oder gleich 3s entfernt. Zusätzlich wurden im PEP Reaktionszeiten entfernt die den individuellen Mittelwert einer Person um mehr als 2SD überschritten. Für PEP und expliziten Fragebogen wurde je eine Variable berechnet, bei der hohe Werte für eine Zustimmung zu stereotypen Aussagen stehen. Analog wurde für die Bewerberauswahlaufgabe das Verhältnis stereotypkongruenter zu -inkongruenten Entscheidungen gebildet, so dass auch hier hohe Werte für stereotypkongruentes Verhalten stehen. Wider Erwarten zeigte sich kein Zusammenhang von PEP und Auswahlverhalten ( $r = 0.06$ ,  $p = 0.6473$ ). Einzig der Zusammenhang von expliziten Stereotypen und Auswahlverhalten war bedeutsam, allerdings war dieser entgegen der Erwartung negativ ( $r = -0.33$ ,  $p = 0.008$ ).

Um die Reliabilität des PEP zu schätzen wurde die Odd-Even Methode angewandt. Diese lag für männliche Stereotype bei  $r_{sb} = -.22$ , für weibliche bei  $r_{sb} = .04$  und ist damit ungenügend.

### 4. Diskussion

Die Studie diente zum einen der Validierung des PEP zur Messung impliziter Bewertungen und zum anderen der Untersuchung des Zusammenhangs von Stereotypen und Bewerberauswahl. Allerdings fand sich kein Zusammenhang zwischen via PEP gemessenen Vorurteilen und dem Wahlverhalten. Einerseits ist dies für die Validität des PEP als implizites Messinstrument problematisch. Andererseits zeigten sich in der Studie eine Reihe von Auffälligkeiten. Der PEP schien viele Probanden zu überfordern, so erreichten nur 53% eine Genauigkeit über 80% in den Catch Trials, was die Verarbeitung der Sätze in Frage stellt. Es würde sich hier also lohnen nur Personen zum Hauptexperiment zuzulassen, die eine bestimmte Genauigkeitsrate erfüllen. Außerdem ist es denkbar, dass die Verwendung individueller Vornamen (anstelle Gruppenbezeichnungen, wie "Männer" / "Frauen"), zur Messung individuumspezifischer (anstelle gruppenspezifischer) Einstellungen führte, was auch die schlechte Reliabilität des PEP erklären könnte. Trotz des Ergebnisses der Studie würde es sich lohnen die Probleme anzugehen und weitere Studien zum PEP zu machen, da der PEP, wenn er valide und reliabel misst, ein flexibles Messinstrument ist.

### 5. Literatur

- Bem, S. (1974). The Measurement of Psychological Androgyny. *Journal of Consulting and Clinical Psychology, Vol. 42, No. 2, 155 - 162.*
- Eagly, A., & Steffen, V. (1986). Gender Stereotypes, Occupational Roles, and Beliefs about Part - Time Employees. *Psychology of Women Quarterly, 252 - 262.*
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4. Aufl.).* Sage.
- Gawronski, B., & De Houwer, J. (in press). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology (2nd edition).* New York: Cambridge University Press.
- Müller, F., & Rothermund, K. (2016). The Propositional Evaluation Paradigm (PEP): Indirect Assessment of Personal Beliefs and Attitudes.
- Ward, C., Thorn, B., Clements, K., Dixon, K., Sanford, S. (2006). Measurement of Agency, Communion, and Emotional Vulnerability with the Personal Attributes Questionnaire. *Journal of Personality Assessment, 206 - 219.*









## Programm

14.00 Uhr	Foyer	Eröffnung der Posterausstellung
14.00 – 16.00 Uhr	Foyer	Präsentation der Poster
16.15 Uhr	HS 4	Festvortrag von Dr. Veronika Job (Universität Zürich) zum Thema: „Is it All in Your Head? Der Einfluss subjektiver Theorien auf die Selbstregulation“
17.00 Uhr	HS 4	Auszeichnung der besten Poster
anschließend	Foyer	Ausklang der Veranstaltung

