

Programm 2022

17. Jenaer
Empiriepraktikumskongress

Psychologie
FSU Jena

Foyer, EAP
15.07.2022

Danksagung

Es ist wieder soweit: Heute, am 15. Juli 2022., findet bereits zum 17. Mal der Jenaer Empiriepraktikumskongress statt. Nachdem der Kongress aufgrund der Corona-Pandemie die letzten zwei Jahre virtuell stattfinden musste, freuen wir uns besonders, alle Beteiligten heute wieder von Angesicht zu Angesicht und vor Ort willkommen heißen zu dürfen. Dieser Kongress kommt jedes Jahr nur durch die Unterstützung und das Engagement zahlreicher Personen zustande, denen an dieser Stelle Dank gebührt.

Zu allererst möchten wir den Juroren und Jurorinnen aus der Professorenschaft, dem Mittelbau und der Studierendenschaft danken, die sich die Zeit nehmen, die einzelnen Arbeiten mit viel Sorgfalt, Expertise und Eifer zu begutachten.

Darüber hinaus gilt unser Dank allen teilnehmenden Gruppen und ihren BetreuerInnen, die durch die bunte Themenvielfalt, die kreativen Forschungsideen und nicht zuletzt durch die Freude am Austausch und dem Generieren neuer Ideen den EmPra-Kongress prägen und unvergesslich machen.

Für den Gastvortrag zum Thema *Bored to death? Motivations underlying self-infliction of pain during thinking for pleasure* haben wir Prof. Dr. Andreas Eder (Julius-Maximilians-Universität Würzburg) gewinnen können, der außerdem als Juror die Begutachtung der Arbeiten unterstützt. Für diese Bereicherung des Kongressprogramms danken wir ihm herzlich.

Einen besonderen Programmpunkt des EmPra-Kongresses bildet jedes Jahr die Auszeichnung und Prämierung der besten Arbeiten. Für die finanzielle Unterstützung durch Bereitstellung der Preisgelder 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 Christopher Schnell für die langjährige Unterstützung.

Des Weiteren danken wir den Studierenden aus der Fachschaft Psychologie für die Hilfe bei der Suche studentischer Juroren. Ebenso danken wir den Mitgliedern der CampusCouch, die den Sektverkauf organisieren, welcher einen maßgeblichen Beitrag zu einer angenehmen Kongressatmosphäre leistet.

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,

Adrian Jusepeitis und Merlin Urban

Inhaltsverzeichnis

Gruppe 1	8
<i>Typing Skills im universitären Alltag – Ein Leistungsprädiktor?</i>	
Alwine Bischoff, Anne Frotscher, Tim Karl, Dominic Kühnlein, Lea Späte	
Gruppe 2	10
<i>Du traust keinen Fake News – und dein Unterbewusstsein? Der Effekt diskreditierter Informationen auf direkt und indirekt gemessene Einstellungen</i>	
Sarah Akman, Cornelia Hartl, Lioba Mans, Finn Schröder, Lea Waloschcyk	
Gruppe 3	12
Manipulation der Schwierigkeit durch den Kontext am Beispiel vegetarischer Ernährung zur Erhöhung der prädiktiven Power des IATs	
Philine Bäzel, Wedeke Ernst, Valerie Gleisner, Luisa Haag, Sophie Schiemanz	
Gruppe 4	14
<i>Ich mag dich nicht – ich verstehe dich nicht. Kulturelle und persönliche Distanz und empathische Akkuratheit</i>	
Tim Bongardt, Sophie Kattein, Annemarie Schwanstecher, Paula Sill	
Gruppe 5	16
<i>Altersbezogene Relevanzunterschiede bei der Bewertung von Gesundheitszuständen</i>	
Kristin Wagner, Carolin Geißler, Lieselotte Langer, Lena Bernstein, Celine Wöhmann	
Gruppe 6	18
<i>Der Einfluss von persönlicher Relevanz auf das Bewertungsverhalten</i>	
Antonia Mesgarha, Aya Franke, Jannes Kolbig, Pauline Kamm, Theodore Wohlfarth	
Gruppe 7	20
<i>Exploring indirect response transfer from native language to a new language using multimodal Associations</i>	
Sarah Bischoff, Maria Galbas, Magdalena Krautwurst, Barbara Riedißer, Vanessa Wolf	
Gruppe 8	22
<i>Ich sehe was, was du nicht fühlst – Zwei Studien zum Zusammenhang zwischen Emotionserkennung, Masken und anderen Faktoren</i>	
Henriette Kleindienst, Paula Gut, Lara Schäfer, Julia Lehnert, Jakob Körner, Ben Erdeli, Josefine Tews, Aleksandra Spajic, Miriam Findeis, Helene Krämer	
Gruppe 9	24
<i>You got the job! Bevorzugen wir Mitarbeiter:innen, die eine ähnliche Persönlichkeit haben?</i>	
Theresa Splitt, Alina Klettke, Carolina von Zanthier, Nele Pieper	
Gruppe 10	26
<i>Backfire-Effekte verstehen – wie psychologische Reaktanz visuelle Aufmerksamkeit beeinflusst</i>	
Bennet Antonik, Matti Keßler, Jennifer Kroker, Katharina Rehling, Esther Stemmler	
Gruppe 11	28
<i>Wir verstehen uns auch ohne Worte – oder doch nicht? Empathische Akkuratheit gegenüber bilingualen Personen</i>	
C. Halcour, E. Joerges, C. Schmitz, Mi. Schwarz, L. Stock, N. Vernaldi	

Gruppe 12	30
<i>Auditive Szenenanalyse auf der Cocktailparty</i>	
Luise Bock, Merija Gühning, Luise Hillebrand, Alina Schönherr, M.S., Vera Wolferstetter	
Gruppe 13	32
<i>Aus den Augen aus dem Sinn - Einfluss der wahrgenommenen Seriosität einer Quelle auf die Entstehung von False Memories</i>	
Annalena Utecht, Kaya Lorenz, Antonia Nestler, Roya Haji Hosseini, Elisa Marski	
Gruppe 14	34
<i>Fragen in Augenzeugeninterviews - Welche Schuld tragen Single-Choice-Antworten? - Psychopol ermittelt - Gedächtnisverzerrung aufgrund suggestiver Fragestellungen</i>	
Helene Heilig, Franz Lehmann, Nele Reisser, Max Wagner, Ferdinand Werner	
Gruppe 15	36
<i>„Gute Nacht Siri, schlaf gut!“ – Wahrnehmung von virtuellen Assistenzsystemen</i>	
Odelia Herman, Sophie-Luise Krüger, Julius Protte, Ina Strugale, Carolin Zimmer	
Gruppe 16	38
<i>Siri, sei meine Tutorin! Der Einfluss einer Virtuellen Assistenz auf den Abruf beobachtungsbasierter Stimulus-Reaktions-Bindungen</i>	
Lina Flindt, Anna Mieske, Sarah Mühr, Ann-Christin Winter, Jennifer Wolf	
Gruppe 17	40
<i>Wie unterscheiden sich Normen und Vorurteile zwischen rechts- und linksorientierten Personen?</i>	
Miriam Augustin, Carolin Beier, Sophie Pimpl, Sophie Schäfer, Verena Schneider	
Gruppe 18	42
<i>I trusted you :` (... - Manipulation des Vertrauens durch Selbstoffenbarung im Repeated-Trust-Game</i>	
Laura Beuthien, Cedric Friedel, Pia Garcke, Ronja Uhlig, Tina Wenzel	
Gruppe 19	44
<i>Illusory Truth Effect - Liegt die Wahrheit im Ohr des Betrachters?</i>	
Julian Kunkel, Sabrina Bittel, Farida Gasimova, Gina Neff	
Gruppe 20	46
<i>Eine Pilotstudie zur Replikation des Illusory Truth Effekts unter Einfluss der Modalität</i>	
Luna Mahler, Alina Meier, Carolin Knorre, Sophie Stockmann	
Gruppe 21	48
<i>Escaping the Habit Hole: The Relationship between Habit Strength and Motivation Orientations in Breaking Habits</i>	
Friederike Kühr, M.L., Robin Lux, Jule Schubert, Sabiene Strömsdörfer	
Gruppe 22	50
<i>Wie ein Ei dem anderen: Was heißt Vorurteile sind sozial geteilt?</i>	
Antonia Boddenberg, A.B., Tuba Korkmaz-Walther, Claudia Lehmann, Hauke Lickfett, Paulin Otto-Schneider	
Gruppe 23	52
<i>Smart or not? - Understanding the formation of context-dependent stereotypes</i>	
Berenike Buschermöhle, Maret Weimann, Lara Krickow, Lena Huss, Eliza Arzuman	

Gruppe 24	54
<i>Investigating the Formation of Contextualized vs. Generalized Stereotypes Using Propositional Evaluation Paradigm</i>	
Johanna Fechner, Elisabeth Hefner, Anja Monstadt, Antonia Reinhardt, Suzanne Van de Moosdijk	
Gruppe 25	56
<i>Du gehörst zu uns, deshalb bestrafen wir dich!</i>	
Lina Marie Detzner, Rebeka Edinger, Teresa Urban, Lara Wöhmann, Yunqi Yang	
Gruppe 26	58
<i>Face to Face: Effekte von Ähnlichkeit und Sympathie auf Gesichtswiedererkennung</i>	
Jona John, Lina Kleemann, Isabel Kiefer, Laura Ramirez	
Gruppe 27	60
<i>Gesellschaftliche Normen als Fundament für Vorurteile</i>	
Felix Karl Konschak, Julia Ueberschär, Kim Nathalie Ciesla, Natalie Zidek, Sinja Groß	

Typing Skills im universitären Alltag – Ein Leistungsprädiktor?

Alwine Bischoff, Anne Frotscher, Tim Karl, Dominic Kühnlein, Lea Späte

Leitung: Dr. Laura Sperl

1. Einleitung

In vielen Lebensbereichen brachte die Corona-Pandemie Herausforderungen mit sich, so auch im universitären Kontext. Die Lehre musste innerhalb kürzester Zeit online gestaltet werden und daher wurden auch die meisten Klausuren nicht mehr handschriftlich, sondern getippt absolviert. Untersuchungen fanden unter Zeitdruck einen signifikanten Zusammenhang zwischen der individuellen Schreibleistung (d.h. wie schnell und akkurat jemand handschreiben/tippen kann) und der Textqualität (Christensen, 2004; Hughes & Babski-Reeves, 2005). Dieser Effekt wird mutmaßlich über die Arbeitsgedächtniskapazität vermittelt: Durch weniger automatisierte Typing Skills wird mehr kognitive Kapazität während des Tippens benötigt. Daher sind weniger kognitive Ressourcen für andere Schreibprozesse verfügbar und es wird eine schlechtere Textqualität beobachtet (van Waes, 2021). Folglich liegt nahe, dass Typing Skills entscheidend beim Absolvieren von Onlineklausuren mit Textproduktion sind. Daher führten wir zunächst eine Online-Studie zu den Erfahrungen und Präferenzen von Studierenden mit Onlineklausuren durch. Untersucht wurden diese Fragestellungen mittels eines Fragebogens mit integriertem Typing- und Arbeitsgedächtniskapazitätstest. Es zeigten sich u.a. signifikant negative Zusammenhänge der selbst eingeschätzten Typing Skills mit dem subjektiv empfundenen Stress während der Onlineklausuren. Außerdem gingen hohe Typing Skills mit einer Präferenz für Onlineklausuren einher. Daneben waren entgegen unserer Erwartung die objektiven Typing Skills jedoch kaum prädiktiv und auch Interaktionseffekte mit der Arbeitsgedächtniskapazität wurden nicht signifikant. Eine Limitation war allerdings, dass wir die Textqualität nur retrospektiv über eine Einschätzung der Note erfassten. Daher schlossen wir eine Laborstudie an, in der aktiv ein argumentativer Text produziert werden musste. Dieser wurde je nach Bedingung getippt oder mit Hand geschrieben, wobei bei einer Vermittlung des Effekts der Typing Skills auf die Textqualität über die Arbeitsgedächtniskapazität der gleiche Effekt auch bei handgeschriebenen Texten zu erwarten ist (Christensen, 2004). Zusätzlich erfassten wir Arbeitsgedächtniskapazität, Stressempfinden, Präferenzen und Tipp- sowie Handschriftgeschwindigkeit. Dabei untersuchten wir die Frage, inwieweit die individuellen Typing Skills bzw. die Handschriftflüssigkeit mittels der Arbeitsgedächtniskapazität die Präferenz der Schreibmodalität sowie die Textqualität in einer argumentativen Schreibaufgabe beeinflussen.

2. Methode

Stichprobe. In die Analysen gingen die Ergebnisse von 27 Proband:innen ein (20 weiblich, Alter: $M = 21$, $SD = 2$, davon 19 Psychologiestudierende).

Material und Prozedur. In einem Computerexperiment wurden demographische Daten, Arbeitsgedächtniskapazität (Backwards-Corsi-Test), Tipp- und Handschriftgeschwindigkeit (1-minütiger Tipp- bzw. Schreibtest), selbst eingeschätzte Typing Skills und Stressempfinden mittels PsyToolkit erhoben. Final wurde ein argumentativer Text mit randomisiert zugewiesener Schreibmodalität (getippt vs. handschriftlich) und Zeitdruckbedingung (mit vs. ohne) produziert.

Design. Das Experiment impliziert ein 2 (Schreibmodalität: Handschrift vs. Tippen) x 2 (Zeitdruck: mit vs. ohne Zeitdruck) Between-Subject-Design.

Datenanalyse. Mittels einer aus der Literatur adaptierten Kriterienliste wurden alle argumentativen Texte hinsichtlich der Textqualität und -quantität von je zwei unabhängigen Rater:innen bewertet. Die weitere Datenaufbereitung und -auswertung erfolgte mit Excel und R/RStudio.

3. Ergebnisse

Objektive Typing Skills reichten von 131 bis 331 ($M = 214$), Handschriftflüssigkeiten von 117 bis 197 ($M = 145$) korrekter Zeichen pro Minute. Die Textqualität reichte von 8.5 bis 15.5 von 16 Punkten ($M = 11.7$), Textlängen von 138 bis 303 Wörtern ($M = 212$). Versuchspersonen, die Tippen präferierten, unterschieden sich in ihren selbsteingeschätzten Typing Skills nicht signifikant von denen, die Handschrift bevorzugten ($p = .094$, $d = .71$). Objektive und selbst eingeschätzte Typing Skills hingen in dieser Studie signifikant zusammen ($p = .008$, $r = .50$). Während sich in der Gesamtstichprobe zwischen den Zeitbedingungen kein signifikanter Unterschied in der Textqualität zeigte, waren die Texte in der Zeitdruckbedingung im Mittel signifikant kürzer ($t(24.35) = 2.079$, $p = .048$, $d = -.84$). Nach Schreibmodalität getrennt ergaben sich keine Unterschiede in Textqualität bzw. -länge. Es konnte dabei in keinem Fall gezeigt werden, dass ein Effekt über den empfundenen Stress vermittelt wurde (alle $p \geq .245$). Über beide Zeitbedingungen hinweg konnten die Typing Skills 35.7% der Varianz in der Qualität ($p = 0.005$, $\beta = .598$) sowie 30.8% der Varianz in der Länge ($p = 0.013$, $\beta = .563$) der getippten Texte erklären, während die Handschriftflüssigkeit nicht signifikant Varianz in Qualität und Länge der handgeschriebenen Texte aufklärte. Keiner dieser Effekte wurde signifikant über die Arbeitsgedächtniskapazität mediiert (alle $p \geq .601$).

4. Diskussion

Entgegen unserer Erfahrung aus dem Wintersemester zeigte sich diesmal ein signifikanter Zusammenhang von subjektiv eingeschätzten und objektiven Typing Skills, was womöglich darauf zurückzuführen ist, dass der Tipptest diesmal beaufsichtigt werden und etwaige Probleme daher ausgeschlossen werden konnten. Unsere Erwartung, dass Personen mit besseren selbst eingeschätzten Typing Skills das Tippen präferierten, blieb unbestätigt. Dies könnte auf die kleine Stichprobe oder unberücksichtigte Faktoren, die die Präferenz beeinflussen, zurückgehen, so etwa die Angst vor technischen Problemen. Auch fanden wir in keiner unserer beiden Studien einen vermittelnden Effekt über das Arbeitsgedächtnis, was neben der Stichprobengröße an der Wahl des Corsi-Tests liegen könnte, welcher auf die visuell-spatiale Merkfähigkeit abzielt. Eine wichtigere Komponente für den Schreibprozess ist vermutlich hingegen die phonologische Merkfähigkeit (Yan et al., 2011). Während nicht bestätigt werden konnte, dass Zeitdruck die Textqualität beeinflusst, zeigte sich, dass unter Zeitdruck kürzere Texte produziert wurden. Im universitären Alltag bedeutet jedoch ein längerer Text oft die bessere Note (Fleckenstein et al., 2020). Zusätzlich zum Zeitdruck ist im Klausurkontext daneben meist ein starker Anreiz vorhanden, sich auch qualitativ anzustrengen, was Thema weiterer Studien sein könnte. Im Gegensatz zur Handschriftflüssigkeit hatten die Typing Skills einen Einfluss auf Textlänge und -qualität, es gibt demnach einen Zusammenhang zwischen individuellen Typing Skills und getippten Leistungen in argumentativer Textform. Sollte sich dies bestätigen, wäre das ein deutliches Argument für Präsenzklausuren und lässt für zukünftige Forschung die Frage offen, inwieweit Klausuren in Hinblick auf die individuell vorliegenden Schreibfähigkeiten fair gestaltet werden können.

5. Literatur

- Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written texts. *British Journal of Educational Psychology*, 74, 551-564.
- Fleckenstein, J., Meyer, J., Jansen, T., Keller, S., & Köller, O. (2020). Is a Long Essay Always a Good Essay? The Effect of Text Length on Writing Assessment. *Frontiers in psychology*, 11, 562462.
- Hughes, L. E., Babski M. S. & K (2005). Time Pressure and Mental Workload Effects on Perceived Workload and Key Strike Force during Typing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.
- van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2021). Measuring and Assessing Typing Skills in Writing Research. *In Journal of Writing Research*, 13(13), 107–153.
- Yan, C. M. W., Wong, A. M. Y., Shu, H., Wagner, R. K., McBride-Chang, C., Zhang, J. (2011). Writing Quality in Chinese Children: Speed and Fluency Matter. *Reading and Writing*.

Du traust keinen Fake News – und dein Unterbewusstsein? Der Effekt diskreditierter Informationen auf direkt und indirekt gemessene Einstellungen

Sarah Akman, Cornelia Hartl, Lioba Mans, Finn Schröder, Lea Waloschcyk

Leitung: M.Sc. Adrian Jusepeitis

1. Einleitung

In Zeiten der zunehmenden Digitalisierung kommt es vermehrt zu strategischer Desinformation und Verbreitung versehentlich fehlerbehafteter Information (Hanley & Munoriyarwa, 2021). Welchen Effekt verarbeitete, aber später diskreditierte Information auf uns hat, ist also ein Themengebiet von wachsender Bedeutung. Peters und Gawronski (2011) zeigten bereits, dass indirekt gemessene evaluative Assoziationen sich – anders als direkt im Fragebogen gemessene Bewertungen – unabhängig vom deklarierten Wahrheitsgehalt konsumierter Informationen, also unabhängig von Metainformationen, bilden. Im Unterschied zu dieser Studie werden wir indirekt gemessene propositionale Bewertungen und direkt gemessene Bewertungen kontrastieren. Der für diese Studie relevante Vorteil von propositionalen Maßen gegenüber assoziativen ist die Möglichkeit, den beigemessenen Wahrheitsgehalt von propositionaler Verknüpfung (i.e. inhaltlicher Verbindung) verschiedener Konzepte zu erfassen (De Houwer & Moors, 2010). Vor diesem Hintergrund beschäftigt sich diese Studie mit der Fragestellung, inwiefern sich ein diskreditierter Artikel anders auf implizite (i.e. unter Bedingung der Automtizität, z.B. unbewusst; De Houwer et al., 2015) und explizite Bewertungen von Personen auswirkt als ein als wahr präsentierter Artikel. In der folgenden Studie werden den Probanden zwei Artikel über fiktive Personen präsentiert, von denen einer unmittelbar nach dem Lesen als wahr und einer als falsch deklariert wird. Hierauf werden die expliziten und impliziten Bewertungen der Personen erfasst.

Hypothese: Die Diskreditierung von Informationen wirkt sich unterschiedlich auf direkte und indirekte Maße aus. Die Differenz in den indirekt erfassten Einstellungen zwischen den als wahr und falsch deklarierten Artikeln ist kleiner als in den direkt erfassten Einstellungen, da die Metainformation der Informationsvalidität weniger Einfluss auf die indirekte Messung ausübt als auf die direkte.

2. Methode

Die Stichprobe umfasst nach Probandenausschluss 90 Probanden, von denen 70 weibliche Psychologiestudentinnen sind. Jeder Versuchsperson werden zwei Zeitungsartikel zu zwei Bürgermeisterkandidaten präsentiert. In diesen Artikeln werden negative Informationen (Steuergeldveruntreuung, Vetternwirtschaft) über je einen Bürgermeisterkandidaten präsentiert. Einer der Artikel wird im Nachhinein als Falschinformation im Rahmen einer Kampagne eines Konkurrenten, der andere als wahr deklariert. Die Erfassung der expliziten Einstellung gegenüber den Kandidaten erfolgt durch einen Fragebogen nach der Manipulation. Den Versuchspersonen werden sechs Aussagen zu Charaktereigenschaften (z. B. Vertrauenswürdigkeit) der dargestellten Bürgermeisterkandidaten vorgelegt, die sie auf einer fünfstufigen Likert-Skala beurteilen sollen. Das Propositional Evaluation Paradigm (PEP) wird als indirekte Messmethode angewandt. Bei dieser handelt es sich um ein priming-basiertes, propositionales Maß zur Messung impliziter Bewertungen anhand von Reaktionszeiten (Müller & Rothermund, 2019). Wir nutzen ein $2 \times 2 \times 2 \times 2 \times 2$ gemischtes Design: Diskreditierung (within) \times Messmethode (within) \times Reihenfolge der Messmethoden (between) \times Reihenfolge der Artikel (between) \times Zuordnung von Artikel und Diskreditierung (between).

3. Ergebnisse

Es finden sich für den PEP Split-Half-Reliabilitäten zwischen .34 (Thema A) und .55 (Thema B). Beim Fragebogen ergeben sich Split-Half-Reliabilitäten zwischen .93 (Thema A) und .95 (Thema B). Zur Testung unserer Hypothese wurde eine zweifaktorielle ANOVA mit den Faktoren Messmethode (within) und Diskreditierung (between) eingesetzt, hierbei findet sich eine signifikante Interaktion von Messmethode und Diskreditierung, $F(1, 89) = 32.68, p < .001$. Die Diskreditierung wirkt sich weniger stark auf die indirekt gemessenen Einstellungen aus, $d = 0.95, t(89) = 6.86, p < .001$, als auf die direkt gemessenen Einstellungen, $d = 2.51, t(89) = 16.29, p < .001$. Außerdem zeigt sich ein Haupteffekt der Messmethode, $F(1, 89) = 19.33, p < .001$, sowie ein Haupteffekt der Diskreditierung, $F(1, 89) = 170.7, p < .001$. Zusätzliche T-Tests gegen 0 wurden getrennt nach Messmethode und Diskreditierung durchgeführt. Es finden sich für beide Maße und beide Bedingungen signifikant Abweichungen von einer neutralen Bewertung, siehe Abbildung 1, maximaler p-Wert $< .001$.

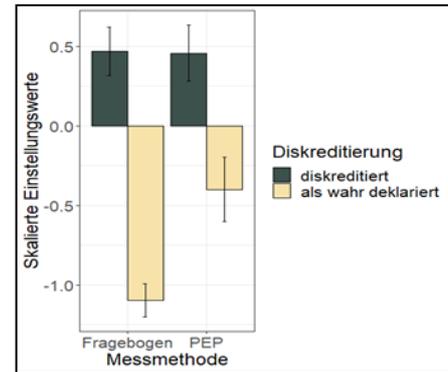


Abbildung 1

4. Diskussion

Ziel unserer Studie war es, den Effekt von diskreditierten Informationen auf direkt gemessene (explizite) und indirekt gemessene (automatische) Bewertungen zu untersuchen. Es zeigt sich eine signifikante Interaktion zwischen der Messmethode und der Diskreditierung des Artikels. Des Weiteren weisen die t-Tests pro Messinstrument auf signifikante Differenzen in den Einstellungen in Abhängigkeit von der Diskreditierung der Artikel hin. Die Ergebnisse sind nur teilweise kompatibel mit unserer Hypothese. Einerseits hatte die Diskreditierung je nach Messmethode tatsächlich einen unterschiedlichen Effekt auf die Einstellungen. Entgegen den Erwartungen schlägt sich die Diskreditierung von negativen Informationen über eine Person jedoch beim PEP genau wie beim Fragebogen in positiven Einstellungen der Person gegenüber nieder. Dies widerspricht der Annahme, dass sich die Metainformation des Wahrheitsgehalts der Quelle in einer indirekten Messung nur abgeschwächt abrufen lässt. Folglich lässt sich sagen, dass Falschinformationen, die auch als solche deklariert werden, den Leser auch auf impliziter Ebene nicht in die Richtung des Artikelinhalts beeinflussen. Als Stärke der Studie kann man insbesondere hervorheben, dass diese eine der wenigen Studien ist, die experimentell testet, wofür indirekte Maße, speziell das PEP, anfällig sind. Dennoch ist zu erwähnen, dass die verwendeten Messinstrumente unterschiedliche Reliabilitäten aufweisen, was die Interpretation der Ergebnisse in ihrer Gültigkeit limitiert. Um dem entgegenzuwirken, könnten in zukünftigen Studien weitere PEP-Blöcke hinzugefügt werden. Zusätzlich ist die Generalisierbarkeit eingeschränkt, da den Probanden in der Studie im Gegensatz zu vielen Alltagssituationen eindeutig vermittelt wird, welcher Artikel wahr oder falsch ist.

5. Literatur

- De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176–198). Guilford Press.
- De Houwer, Jan & Heider, Niclas & Spruyt, Adriaan & Roets, Arne & Hughes, Sean. (2015). The Relational Responding Task: Toward a New Implicit Measure of Beliefs. *Frontiers in Psychology*. 6.
- Hanley, Monika and Munoriyarwa, Allen. "Fake News: Tracing the Genesis of a New Term and Old Practices". *Digital Roots: Historicizing Media and Communication Concepts of the Digital Age*, edited by Gabriele Balbi, Nelson Ribeiro, Valérie Schafer and Christian Schwarzenegger, Berlin, Boston: De Gruyter Oldenbourg, 2021, pp. 157-176
- Müller, Florian & Rothermund, Klaus. (2019). The Propositional Evaluation Paradigm: Indirect Assessment of Personal Beliefs and Attitudes. *Frontiers in Psychology*. 10. 2385.
- Peters, Kurt & Gawronski, Bertram. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality & social psychology bulletin*. 37. 557-69.

Manipulation der Schwierigkeit durch den Kontext am Beispiel vegetarischer Ernährung zur Erhöhung der prädiktiven Power des IATs

Philine Bäzel, Wedeke Ernst, Valerie Gleisner, Luisa Haag, Sophia Schiemanz

Leitung: M. Sc. Merlin Urban

1. Einleitung

Das Ziel unserer Studie ist die Erhöhung der prädiktiven Power des IATs durch die Übertragung des Konzepts der Schwierigkeit aus der Klassischen Testtheorie (KTT) auf den IAT. Aus der KTT ist bekannt, dass mittelschwere Tests aufgrund höherer Varianz bessere Vorhersagen treffen als Tests mit extremen Schwierigkeiten (Bühner, 2011). Die Schwierigkeit des IATs lässt sich entsprechend der KTT durch den mittleren IAT-Effekt definieren. Ein IAT ist leicht bei einem stark positiven mittleren IAT Effekt (Versuchspersonen (VP) beurteilen relevante Targetkategorie positiver), schwierig bei einem stark negativen mittleren IAT Effekt (VP beurteilen relevante Targetkategorie negativer) und mittelschwer bei einem mittleren IAT-Effekt von 0 (VP beurteilen relevante Targetkategorie weder positiver noch negativer bzw. gleich viele VP beurteilen sie als positiver oder negativer). Forschung zur *malleability* des IATs weist darauf hin, dass sich der mittlere IAT-Effekt (i.e., die IAT-Schwierigkeit) durch den Kontext manipulieren lässt und dass der Kontext v.a. Menschen mit niedrigen Einstellungen beeinflusst (Park & Lee, 2007). Diesen Kontexteffekt wollen wir nutzen, um aus einem leichten IAT (relevante Targetkategorie: „vegetarische Ernährung“) einen mittelschweren IAT (mit höherer Varianz und Vorhersagekraft) zu machen. Es lassen sich folgende Hypothesen ableiten: *H1*. Die Schwierigkeit des IATs lässt sich durch den Kontext manipulieren. Je positiver die Darstellung von vegetarischer Ernährung, desto leichter der IAT; je negativer die Darstellung, desto schwieriger der IAT. *H2*. Der Effekt der Manipulation ist stärker für VP mit neutraleren Einstellungen als für VP mit extremen pos./neg. Einstellungen gegenüber vegetarischer Ernährung. *H3*. Je näher der IAT an einer mittleren Schwierigkeit ist, umso höher ist die Korrelation mit dem Kriterium.

2. Methode

Stichprobe. In die Analyse gingen 63 VP ein (85.1% weiblich, 3% divers; Alter: $M = 21.9$, $SD = 3.5$). *Design*. Es wurde ein $2 \times 2 \times 2$ -mixed-Design mit den Between-Faktoren Kontextmanipulation (pos./neg. soziale Norm) und Einstellung (neutrale/extreme Einstellung, eingeteilt über den Mittelwert des IAT-Effekts zum ersten Messzeitpunkt) sowie dem Within-Faktor Messzeitpunkt (Prä/Post-IAT-Messung) verwendet. *Ablauf*. Innerhalb eines Onlineexperiments durchlief jede VP zuerst den IAT. Danach erfolgte die Kontextmanipulation sowie ein Manipulationscheck. Dann durchlief jede VP erneut denselben IAT. Abschließend beantwortete die VP verschiedene Fragebogenmaße und wurde über die Studie aufgeklärt. *Messinstrumente*. Es wurde ein Single-Target-IAT verwendet, der aus einem Attributübungsblock, gefolgt von je einem kurzen und langen kompatiblen sowie einem kurzen und langen inkompatiblen Testblock, bestand. Die Reihenfolge der (in)kompatiblen Blöcke wurde counterbalanciert. Es wurden jeweils 6 Bild-Stimuli für die Targetkategorie und die Attributkategorien (gut/schlecht) verwendet. Als Manipulation wurden je nach Bedingung vier positive oder negative soziale Normen zu vegetarischer Ernährung präsentiert (z.B. pos. (/neg.): „75,3% der deutschen Studierenden stimmten der Aussage, man sollte sich vegetarisch ernähren, (nicht) zu“). Die Normen wurden im Rahmen einer Coverstory, laut derer es sich um Studienergebnisse vom Umweltbundesamt handelte, präsentiert. Mit dem Manipulationscheck wurde geprüft, ob die Normen aufmerksam gelesen wurden und glaubwürdig waren. Die direkte Einstellung wurde über das Bauch- sowie das tatsächliche Gefühl zu vegetarischer Ernährung (1 = sehr neg. bis 10 = sehr pos.) erfragt. Das Verhalten wurde mit zwei dichotomen Items gemessen (z.B. „Ich ernähre mich vegetarisch“).

3. Ergebnisse

Zur Analyse der H1 und H2 wurde eine dreifaktorielle ANOVA im 2x2x2-Design mit den o.g. Faktoren durchgeführt. Entgegen H1 gab es keinen signifikanten Interaktionseffekt zwischen Manipulation und Messzeitpunkt, $F(1,60) = 1.75$, $p=0.19$, $Eta^2p = 0.03$, KI [0; 0.15]. Die Kontextmanipulation führte also nicht zu einer Veränderung der mittleren IAT-Effekte. Entgegen H2 gab es auch keine signifikante Dreifachinteraktion zwischen Manipulation, Messzeitpunkt und Einstellung, $F(1, 60) = 3.13$, $p=0.08$, $Eta^2p = 0.05$, KI [0, 0.18]. Die Manipulation führte also auch nicht zu einer unterschiedlichen Veränderung der IAT-Effekte in Abhängigkeit der Einstellungen der VP. Signifikant wurden hingegen der Haupteffekt der Einstellung, $F(1, 60)= 19.99$, $p < 0.01$, $Eta^2p= 0.25$, KI [0.08, 0.41], sowie der Interaktionseffekt zwischen Einstellung und Messzeitpunkt, $F(1, 60)=30.51$, $p<0.01$, KI [0.15, 0.49]. Ersteres ist durch unsere Einteilung in neutrale/extreme Einstellungen zu erklären, Letzteres bedeutet, dass eine Veränderung der IAT-Effekte je nach Einstellung stattfand. Alle weiteren Effekte wurden nicht signifikant, alle $ps > 0.19$. Bezüglich der H3 waren aufgrund der fehlgeschlagenen Manipulation keine signifikanten Unterschiede zwischen den Korrelationen zu erwarten, was laut der Z-Tests auch zutraf, alle $ps > .2$, bis auf den Unterschied zwischen den Korrelationen innerhalb der negativen Bedingung zu Messzeitpunkt 1 und 2 für die direkte Einstellung als Kriterium, $p = .03$.

Tabelle 1. IAT-Mittelwerte und -Varianzen sowie Korrelationen zwischen den IATs und den Kriteriumsmaßen nach Messzeitpunkt und Kontextbedingung

Variable	Pos. Kontextmanipulation			Neg. Kontextmanipulation				
	M (SD)	1	2	3	M (SD)	1	2	3
1. Vegetarismus-IAT T1	0.34 (0.29)				0.41 (0.36)			
2. Vegetarismus-IAT T2	0.31 (0.24)	.33			0.35 (0.34)	.27		
3. Vegetarismus-Einstellung	8.58 (1.47)	.23	.25		8.16 (1.52)	.06	.45*	
4. Vegetarismus-Verhalten	0.4 (0.48)	.3	.26	.64*	0.51 (0.48)	-.12	.19	.68*

Anmerkung. T1: Messzeitpunkt 1; T2: Messzeitpunkt 2. * $p < .05$.

4. Diskussion

Entgegen H1 ließ sich die IAT-Schwierigkeit nicht durch den Kontext beeinflussen. Da bisherige Forschung dafür spricht, dass sich der IAT-Effekt über den Kontext manipulieren lässt, liegen die Gründe dafür vermutlich im Manipulationsdesign. So fanden die VP der positiven Bedingung die Studie zwar glaubwürdig, die der negativen jedoch nicht. Dies erklärt allerdings nur die fehlende Veränderung der IAT-Schwierigkeit in der neg. Bedingung, nicht aber in der pos. Bedingung. Möglicherweise stellen soziale Normen als Kontextmanipulation bei vegetarischer Ernährung die falsche Wahl dar, da Studierende gefestigtes Wissen und Überzeugungen haben und sich nicht einfach durch andere beeinflussen lassen. Weiterhin veränderten sich die mittleren IAT-Effekte entgegen H2 für VP mit neutralen sowie extrem pos./neg. Einstellungen: die mittleren IAT-Effekte Ersterer stiegen, die Letzterer sanken. Dies könnte auf Boden- und Deckeneffekte zurückzuführen sein. Bezüglich H3 lassen sich aufgrund der fehlgeschlagenen Manipulation Unterschiede in den Korrelationen nicht auf die Schwierigkeits-Überlegung zurückführen, sondern stellen, wenn vorhanden, vermutlich lediglich Schwankungen aufgrund einer zu niedrigen VP-Anzahl dar.

5. Literatur

- Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion (3. Aufl.). Always learning. Pearson Deutschland.
- Park, J., Lee, K. F. & Lee, G. (2007). Implicit Attitudes Toward Arab-Muslims and the Moderating Effects of Social Information. *Basic and applied social psychology*, 29(1), 35-45. <https://doi.org/10>

Ich mag dich nicht – ich verstehe dich nicht. Kulturelle und persönliche Distanz und empathische Akkuratheit

Tim Bongardt, Sophie Kattein, Annemarie Schwanstecher, Paula Sill

Leitung: Dr. phil. Andrea Schlesier-Michel

1. Einleitung

Kulturen unterscheiden sich nicht nur in ihren Werten und Normen, sondern auch in ihren emotionalen Erfahrungen und in der Art und Weise, wie diese untereinander kommuniziert werden (Bänzinger, 2016). Die *empathische Akkuratheit* (EA) ist ein wichtiger Bestandteil einer erfolgreichen Kommunikation und stellt die Fähigkeit dar, die Gedanken und Gefühle unserer Gegenüber akkurat „lesen“ zu können. Daher wollten wir untersuchen, wie gut die Gedanken und Gefühle bilingualer Personen erfasst werden und welche Rolle die kulturelle und persönliche Distanz (in Form von Ähnlichkeit, Stereotypen und Sympathie) zwischen Target und Perceiver spielen. Während es zur Ähnlichkeit und Sympathie Vorbefunde gibt, ist die Bedeutung von Stereotypen kontroverser diskutiert (Ekerim-Akbulut et al., 2019). Unsere erste Hypothese war, dass die EA höher ist, je ähnlicher sich Target und Perceiver sind. Unsere zweite Hypothese besagt, dass negative Stereotype negativ mit EA korrelieren, während positive Stereotype positiv mit EA korrelieren. Dazu erwarteten wir, dass die EA umso schlechter sei, je weniger S, woraus sich die dritte Hypothese ergibt.

2. Methode

Perceiver. Die Stichprobe umfasste insgesamt 122 deutschsprachige Teilnehmende (1 diverse, 100 weibliche, 21 männliche Personen). Das Alter variierte von 18 bis 74 Jahren und betrug im Durchschnitt 29 Jahre. Der überwiegende Teil der Probanden setzte sich aus Personen mit Abitur (95%) zusammen, wobei sich insgesamt 55% der Teilnehmenden noch in Ausbildung befanden, meist als Psychologie-Studierende.

Prozedur, Design und Material: Die Erhebung der Daten erfolgte über einen Fragebogen auf der Plattform SoSci.de. In einer ersten vorbereitenden Phase wurden von uns Videos für diesen Fragebogen erstellt. Dafür rekrutierten wir drei bilinguale 18-20jährige weibliche Personen („Targets“), welche sowohl in Deutsch als auch in ihrer Zweitsprache (Chinesisch, Russisch, Rumänisch) über verschiedene Erlebnisse berichteten. Zu festgelegten Zeitpunkten wurden diese Videos gestoppt und die Targets sollten auf einer neunstufigen Skala von -4 (extrem negativ) bis +4 (extrem positiv) ihre generelle sowie auf einer fünfstufigen Skala von 0 (gar nicht) bis 4 (stark) ihre spezifische Gefühlslage für diesen Zeitpunkt angeben. Hierbei handelte es sich um die Ausprägung vier positiver (froh, stolz, zufrieden, ausgeglichen; Cronbach's $\alpha = .90$) und drei negativer Emotionen (traurig, verärgert, verängstigt; Cronbach's $\alpha = .72$). In drei unterschiedlichen Bedingungen (Video mit Ton, Video ohne Ton und eine reine Tonaufnahme) mussten in der darauffolgenden Erhebungsphase die Perceiver nun die Gefühle der Target-Personen auf denselben Skalen für dieselben Zeitpunkte einschätzen. Sie sahen dabei randomisiert entweder nur die Aufnahmen, die auf Deutsch oder die, die auf der Zweitsprache basierten.

Maße: Die jeweilige empathische Akkuratheit wurde als Korrelation (Fishers-z-standardisiert) zwischen Selbstbericht des Targets und Fremdbbericht der Perceiver über die Stopp-Punkte errechnet. Dabei fassten wir die drei EA-Bereiche (generelle Gefühlslage, positive und negative Emotionen) zu einem Maß zusammen (Cronbach's $\alpha_{\text{Video}} = .92$, $\alpha_{\text{Audio}} = .84$, $\alpha_{\text{Voll}} = .85$). Für jedes Target schätzten die Perceiver die wahrgenommene kulturelle Ähnlichkeit (1 Item), die wahrgenommene persönliche Ähnlichkeit (8 Items), positive und negative Stereotype (jeweils 5 Items), sowie Sympathie (6 Items) ein. Alle verwendeten Variablen waren ausreichend reliabel (Cronbach's α von .72 bis .95) und annähernd normalverteilt. Ausreißer wurden (sofern vorhanden) mittels Winsorizing korrigiert.

3. Ergebnisse

Hypothese 1. Unsere erste Hypothese konnten wir teilweise bestätigen. In der ersten Bedingung „nur Audio“ korrelierte die EA signifikant mit der wahrgenommenen Kulturähnlichkeit (siehe Tabelle 1). Dabei handelte es sich um einen mittelgroßen Effekt. Auch in der vollständigen Video-Bedingung war die Korrelation signifikant und positiv (kleiner Effekt), jedoch nicht in der Bedingung „nur Video“ (ohne Ton). Dieses Ergebnismuster zeigte sich ebenfalls für die persönliche Ähnlichkeit.

Hypothese 2. Die negativen Stereotype korrelierten in keiner Bedingung signifikant mit der EA. Die positiven Stereotype korrelierten jedoch signifikant in der Bedingung „nur Audio“ mit der EA, wobei es sich um einen kleinen Effekt handelte. Die Korrelationen in der „nur Video“ und „volles Video“ waren nicht signifikant. Somit ließ sich Hypothese 2 nicht bestätigen.

Hypothese 3. Die verwendete Sympathie-Skala korrelierte signifikant mit der EA in der reinen Audio-Bedingung, ebenso wie in der Bedingung „volles Video“. Beides waren kleine Effekte. Die Korrelation in der „nur Video“-Bedingung war nicht signifikant. Dies bestätigt unsere Hypothese nur zum Teil.

Bivariate Korrelation EA mit ...

	kultureller Ähnlichkeit	persönliche Ähnlichkeit	negativen Stereotypen	positiven Stereotypen	Sympathie	Alter
nurAudio	.46***	.35***	.04	.23**	.29**	-.08
nurVideo	.08	-.12	.06	-.06	-.05	-.08
Video mit Ton	.16*	.18*	-.01	.04	.21**	-.14*

Tabelle 1. Bivariate Korrelationen der Empathischen Akkuratheit (EA) mit den verschiedenen Prädiktoren kultureller und persönlicher Distanz. * $p < .05$, ** $p < .01$, *** $p < .001$.

4. Diskussion

Unsere Ergebnisse erweitern die bisherige Forschung, da sie zeigen, dass zwischen Maßen persönlicher und kultureller Distanz und Empathischer Akkuratheit scheinbar nur dann bedeutsame Zusammenhänge existieren, wenn verbale und/oder paralinguistische Hinweisreize und Informationen vorliegen. Die auf rein bildlichem Material (Mimik) basierende EA zeigte keinerlei Zusammenhänge mit den untersuchten Prädiktoren. Dazu zeigte sich in unserer Untersuchung, dass die Teilnehmenden kaum negative Einschätzungen (Stereotype) zu den bilingualen Target-Personen angaben. Im Hinblick auf die etwas einseitige Stichproben-Zusammensetzung könnte dies darauf hindeuten, dass die Versuchspersonen häufig sozial erwünschte Antwortmuster abgaben. Tatsächlich genutzte Stereotype könnten sich implizit im Item der subjektiv wahrgenommenen Kulturähnlichkeit ausdrücken, welches sich in partiellen Korrelationen über die persönliche Ähnlichkeit hinaus bedeutsam mit der EA verbunden zeigte.

5. Literatur

- Bänziger, T. (2016). Accuracy of judging emotions. In J. A. Hall, M. Schmid Mast, & T. West (Eds.), *The social psychology of perceiving others accurately* (pp. 23-51). Cambridge University Press. <https://doi.org/10.1017/CBO9781316181959.002>
- Ekerim-Akbulut, M., Selçuk, B., Slaughter, V., Hunter, J. A. & Ruffman, T. (2019). In two minds: Similarity, threat, and prejudice contribute to worse mindreading of outgroups compared with an ingroup. *Journal of Cross-Cultural Psychology*, 51(1), 25–48. <https://doi.org/10.1177/0022022119883699>

Altersbezogene Relevanzunterschiede bei der Bewertung von Gesundheitszuständen

Kristin Wagner, Carolin Geißler, Lieselotte Langer, Lena Bernstein, Celine Wöhmann

Leitung: Dr. Karoline Bading

1. Einleitung

Durch voranschreitende Forschung werden immer mehr Medikamente entwickelt und auf den Markt gebracht. Ob diese Medikamente genommen werden, hängt auch von ihrer Bewertung ab. Wie werden nun diese Bewertungen solcher Einstellungsobjekten (CS) erworben? In der aktuellen Forschung findet sich unter anderem die propositionale Theorie. Hiernach werden Beziehungen von den Ereignissen mit den Objekten (US) abgeleitet [De Houwer, 2018]. Wir vermuten, dass die Variation in der Stärke dabei auch von der persönlichen Relevanz assoziierter Ereignisse bzw. Objekte abhängt. Zur Überprüfung adaptieren wir die Lernprozedur von Hu et al. (2017). Hierbei werden Teilnehmenden Medikamente präsentiert, die positive/negative Gesundheitszustände verursachen/verhindern. Der Befund zeigte eine disordinale Interaktion von US-Valenz und CS-US-Relation (verursachen: CS-Bewertungen in Richtung US-Valenz; verhindern: CS-Bewertung entgegen US-Valenz). Unserer Studie verwendet unterschiedliches Bildmaterial, welches normativ relevant für „junge“ vs. „ältere“ Leute sein sollte. Weiterhin rekrutieren wir zwei Stichproben unterschiedlicher Altersgruppen („Jüngere“: 18-30 Jahre; „Ältere“: >45 Jahre). Mit diesem Design untersuchen wir zwei Hypothesen. Erstens überprüften wir den Material-Relevanz-Effekt (extremere Bewertungen für „junges“ („altes“) als für „altes“ („junges“) Material bei Jüngeren (Älteren)). Zweitens vermuten wir stärkere Effekte in der „verhindern“ als in der „verursachen“ Bedingung. Als Begründung hierfür sehen wir, dass die subjektive Wahrscheinlichkeit des Gesundheitszustandes wichtiger für „verhindern“ sein sollte als für „verursachen“.

2. Methode

Design und Teilnehmer. Die Studie folgte einem 2 (US-Valenz: positiv vs. negativ, within) x 2 (CS-US Relation: verursachen vs. verhindern, within) x 2 (Material: jung vs. alt, within) x 2 (Altersgruppe: jünger vs. älter, between) Design. Teilnehmende der „jüngeren“ Altersgruppe wurden über einen E-Mail-Verteiler der FSU Jena rekrutiert und mit 0.5 VP-Stunden vergütet. Teilnehmende der „älteren“ Altersgruppe wurden aus unserem persönlichen Bekanntenkreis akquiriert. Unvollständige Datensätze und Personen, die angaben, die Studienteilnahme nicht ernst genommen zu haben, wurden ausgeschlossen. Die finale Stichprobe umfasste 62 Personen ($N_{\text{jünger}} = 51$, $N_{\text{älter}} = 11$).

Material und Prozedur. In der Onlinestudie wurden die Teilnehmenden begrüßt und instruiert. Danach folgte die an Hu et al. angelehnte Lernprozedur, in welcher Medikamente (CS) zusammen mit Gesundheitszuständen (US) präsentiert wurden. Außerdem wurde für jedes CS-US Paar angegeben, ob der CS den US verursacht oder verhindert. Als CS verwendeten wir die fiktiven Medikamente von Hu et al. (2017). Als US fungierten Bilder von 4 positiven und 4 negativen Gesundheitszuständen. Pro Valenz waren jeweils 2 Gesundheitszustände wahrscheinlich und normativ relevant für jüngere Personen (z.B. Akne) und 2 weitere für ältere Personen (z.B. Hyperpigmentierung). Die CS-US-Relation Triplets wurden jeweils für 5 Sekunden präsentiert (5 Wiederholungen pro Triplet). Nach der Lernprozedur wurden die CS Bewertungen auf einer Skala von 1 (sehr negativ) bis 10 (sehr positiv) erfasst. Das Experiment endeten mit einem kurzen Debriefing sowie einer Verabschiedung.

3. Ergebnisse

Datenaggregation: Zur Überprüfung der Hypothesen berechneten wir individuelle Differenzscores, welche den Material-Relevanz-Effekt in den vier US- Valenz x CS-US Relation Bedingungen quantifizieren. Für jede Bedingung wurde die Bewertung des mit „alten“ Material assoziierten Medikaments von der Bewertung des mit „jungem“ Material assoziierten Medikaments abgezogen. Um die Scores einheitlich zu polen, wurden die Differenzscores in der „positiv - verhindern“ und in der „negativ-verursachen“ Bedingung mit -1 multipliziert. Entsprechend zeigt ein positiver (negativer) Score an, dass das mit „jungem“ („altem“) Material assoziierte Medikament extremer bewertet wurde als das mit „altem“ („jungem“) Material assoziierte Medikament.

Datenauswertung: Mit einer 2 (US-Valenz) x 2 (CS-US-Relation) x 2 (Altersgruppe) gemischten Varianzanalyse wurden die Differenzscores analysiert. Der Haupteffekt Altersgruppe wurde in der ANOVA nicht signifikant ($F(1,60)=0.02$, $p=.879$, $\eta^2<.001$). Wie erwartet war der mittlere Score in der „jungem“ Stichprobe ($M = 0.51$, $SD = 3.42$) signifikant, $t(50) = 1.824$, $p = .0377$. Entgegen unserer Erwartung war der mittlere Differenzscore in der „alten“ Stichprobe jedoch nicht kleiner sondern größer als 0 ($M = 0.61$, $SD = 3.91$), wich allerdings nicht signifikant davon ab, $t(10) = 0.908$, $p = .385$. Die durch Hypothese 2 implizierte Interaktion „Altersgruppe“ x „CS-US Relation“ wurde zwar signifikant, $F(1,60)=5.79$, $p=.019$, $\eta^2=.021$, entsprach aber nicht dem vorhergesagten Muster. In der „jungem“ Stichprobe zeigten sich entgegen unserer Erwartung deskriptiv positivere Differenzscores in der „verursachen“ Bedingung ($M=0.62$, $SD=3.36$) als in der „verhindern“ Bedingung ($M=0.40$, $SD=3.50$), $t(50)=, -0.523$, $p = .699$. In der „alten“ Stichprobe fand sich (statt eines negativeren) ein positiverer Score in der „verhindern“ Bedingung ($M=1.82$, $SD=4.46$) als in der „verursachen“ Bedingung ($M=-0.59$, $SD=2.89$), $t(10)=-1.729$, $p=.114$. In der Varianzanalyse wurde überraschenderweise der Haupteffekt für die CS-US Relationen signifikant ($F(1,60)=4.04$, $p=.049$, $\eta^2 =.015$). Dieser beruhte auf durchschnittlich höheren Differenzscores in der „verhindern“ Bedingung ($M = 0.65$, $SD = 3.71$) als in der „verursachen“ Bedingung ($M=0.40$, $SD=3.30$). Alle anderen Haupteffekte und Interaktionen wurden nicht signifikant, alle $p \geq .180$.

4. Diskussion

Unsere erste Hypothese geht davon aus, dass jüngere (ältere) Teilnehmenden ihre entsprechenden Medikamente extremer bewerten. Dieser Effekt wurde nicht signifikant, bestätigte sich zwar im Muster für die jünger Stichprobe, jedoch nicht für die ältere Stichprobe. Eine Alternativerklärung für diesen Befund könnte die allgemein höhere Attraktivität jüngerer Gesundheitszustände über die Altersgruppe hinweg oder die empfundene Ungleichwertigkeit der verschiedenen Gesundheitszustände sein. In unserer zweiten Hypothese erwarteten wir einen stärkeren Material-Relevanz-Effekt in der „verhindern“ Bedingung. Auch diese Hypothese bestätigte sich nicht, da sich weder in der jüngeren, noch in der älteren Stichprobe stärkere Effekte für die „verhindern“ Bedingung fanden. In der Annahme, älteres Material sei relevanter für die höhere Altersgruppe, kann es womöglich hinderlich gewesen sein, dass das Durchschnittsalter der höhere Altersgruppe und ihre Teilnehmendenanzahl zu niedrig war um ihre Gesundheitszustände signifikant stärker zu bewerten als ihre Vergleichsgruppe. Das jüngere Material wurde daher eventuell als relevanter eingestuft, da ein Alter von 45 Jahren mental noch zu weit vom Rentenalter und den damit verbunden Gesundheitszuständen entfernt ist. Die Teilnehmenden der „älteren“ Gruppe könnten sich dementsprechend besser mit dem jüngeren Material identifiziert und es dadurch als relevanter erachtet haben.

5. Literatur

- De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin*, 13(3). <https://doi.org/10.5964/spb.v13i3.28046>
- Hu, X., Gawronski, B., & Balas, R. (2017). Propositional Versus Dual-Process Accounts of Evaluative Conditioning. *Social Psychological and Personality Science*, 8(8), 858–866. <https://doi.org/10.1177/1948550617691094>

Der Einfluss von persönlicher Relevanz auf das Bewertungsverhalten

Antonia Mesgarha, Aya Franke, Jannes Kolbig, Pauline Kamm, Theodore Wohlfarth

Leitung: Dr. Karoline Bading

1. Einleitung

Bewertungen sind eine grundlegende Determinante menschlichen Verhaltens und daher ein wichtiges Forschungsthema. Propositionale Theorien des evaluativen Lernens besagen, dass die Bewertung eines Einstellungsobjektes (CS) aus seiner Beziehung zu anderen (positiven oder negativen) Objekten und Ereignissen (US) abgeleitet wird (De Houwer, 2018). Bisherige Forschung konnte diese Annahme bestätigen: so zeigten Hu, Gawronski und Balas (2017), dass Medikamente (CSs), welche positive oder negative Gesundheitszustände (GZ) (USs) verursachen (verhindern), entsprechend (entgegen) der Valenz des verursachten (verhinderten) USs bewertet werden. Mit unserer Forschung im Rahmen des Empirie-Praktikums wollten wir untersuchen, ob neben US-Valenz und CS-US Beziehung, auch die persönliche US-Relevanz einen Einfluss auf die CS Bewertung hat. Um einen möglichen Effekt der US-Relevanz zu untersuchen, haben wir die Lernprozedur von Hu et al. (2017) adaptiert: die Versuchspersonen (VP) lernten über Medikamente, welche positive oder negative GZ verursachten oder verhinderten. Neben den Faktoren US-Valenz der CS-US Beziehung manipulierten wir auch die US- Relevanz. Hierfür verwendeten wir sowohl „junges“, als auch „altes“ US-Material, d.h. (positive und negative) GZ, welche wahrscheinlich und normativ relevant für Personen im Studentenalter vs. Rentenalter sein sollten. Des Weiteren manipulierten wir den Altersfokus der VP (Jetzt-Fokus-Gruppe = JFG, Rentenalter-Fokus-Gruppe = RFG). Unsere Annahme bestand darin, dass der Fokus auf das Rentenalter, junge Personen in ältere Menschen hineinversetzen lässt und dadurch mit dem Rentenalter assoziierte GZ relevanter erscheinen. In diesem Kontext wurden drei Hypothesen aufgestellt: (1) Junge VP halten Material, das mit jungem Alter assoziiert ist („junges Material), für normativ relevanter und geben betragsmäßig extremere CS Bewertungen für Medikamente ab, die mit „jungem“ (im Vergleich zu „altem“) US-Material assoziiert sind. (2) Der Material-Relevanzeffekt ist stärker für „verhindernde“ Medikamente, als für „Verursachende“. Dies wird begründet, dass die subjektive Wahrscheinlichkeit des GZ wichtiger ist, wenn dieser verhindert (vs. verursacht) werden soll. (3) Die Manipulation des Altersfokus schwächt bzw. hebt die Material-Relevanzeffekte auf, da sich der Fokus der jungen VP durch die Manipulation in der RFG auf das hohe Alter ausweitet und somit „altes Material“ an Relevanz gewinnt. Jedoch wird dadurch die Relevanz der jungen GZ nicht abnehmen. Demnach sollten die Bewertungen der Medikamente, die mit „jungem Material“ in Beziehung stehen und die Medikamente, die mit „altem Material“ assoziiert sind, ähnlich extrem ausfallen.

2. Methode

Teilnehmende und Design. 98 VP nahmen über den Emailverteiler der FSU Jena am eigenen PC teil. Die Onlinestudie umfasste ein 2 (US-Valenz: positiv vs. negativ, *within*) x 2 (CS-US Relation: verursachen vs. verhindern, *within*) x 2 (Material: jung vs. alt, *within*) x 2 (Altersfokus: JFG vs. RFG)-Design.

Ablauf. Nach einer Begrüßung und Coverstory wurde der Altersfokus der VP mithilfe von vier Fragen zur aktuellen Lebenssituation vs. zum Rentenalter manipuliert. Danach begann die Lernprozedur. Dabei wurden CS, CS-US Relation und US zeitgleich präsentiert. Als CSs wurden acht fiktive Medikamente von Hu, Gawronski & Balas (2017) gezeigt. Als USs dienten acht Bilder von GZ, welche sich hinsichtlich ihrer Valenz (positiv vs. negativ) und ihres Altersbezugs (jung vs. alt) unterschieden. Nach der Lernprozedur wurden die Medikamente auf einer Ratingskala von 0 (sehr negativ) bis 10 (sehr positiv) bewertet. Die Durchführung schloss mit einem Debriefing ab.

3. Ergebnisse

Datenaggregation: Um unsere Hypothesen über den Effekt des Materials (jung vs. alt) auf die CS Bewertungsstärke zu prüfen, berechneten wir individuelle Differenz-Scores (DSs). Dazu wurde für jede US-Valenz x CS-US Relation Bedingung die Bewertung des mit „altem“ US-Material assoziierten Medikaments von der Bewertung des mit „jungem“ Material assoziierten Medikament abgezogen. Für eine einheitliche Polung für beide US-Valenz Bedingungen, wurden die DSs für „positiv verhindern“ und „negativ verursachen“ mit -1 multipliziert. Positive DSs zeigen also an, dass mit „jungem“ US-Material assoziierte Medikamente extremer bewertet wurden, als Medikamente, die mit „altem“ US Material in Verbindung standen.

Datenauswertung: Die DSs wurden mit einer 2 (US-Valenz) x 2 (CS-US Relation) x 2 (Altersfokus) gemischten Varianzanalyse ausgewertet. Wir erwarteten einen HE „Altersfokus“ (JFG vs. RFG), einen HE „CS-US Relation“ (verursachen vs. verhindern), und eine Interaktion von „Altersfokus“ und „Relation“. Keiner der drei vorhergesagten Effekte wurde signifikant, alle $ps \geq .33$. Trotz der nicht-signifikanten Effekte schauten wir, ob das Datenmuster deskriptiv in die vorhergesagte Richtung ging. Entsprechend unserer Vorhersage zeigte sich für den HE „Altersfokus“, dass die DSs in der JFG durchschnittlich höher waren, als in der RFG. Außerdem war der mittlere DS in der JFG signifikant größer 0, $t(50) = 1,824$, $p = .04$. Dagegen wich der DS in der RFG nicht signifikant von 0 ab, $t(40) = 0,89$, $p = .38$. Bezüglich des HE „CS-US Relation“ zeigte sich entgegen unserer Erwartung, dass der durchschnittliche DS in der „verursachen“ Bedingung größer ausfiel als in der „verhindern“ Bedingung. Der mittlere DS in der „verursachen“ Bedingung war signifikant größer 0, $t(91) = 2,1692$, $p = 0,03$, wohingegen der mittlere DS in der „verhindern“ Bedingung nicht signifikant von 0 abwich, $t(91) = 0,84$, $p = 0,40$. Auch das Muster der Interaktion von „Altersfokus“ und „CS-US Relation“ entsprach nicht unserer Vorhersagen. Zwar fand sich in der RFG Bedingung kein Unterschied der DSs für „verursachende“ vs. „verhindernde“ Medikamente, $t(40) = -0,78$, $p = 0,44$, jedoch in der JFG Bedingung höheren DSs für „verhindernde“ (vs. „verursachende“) Medikamente, $t(50) = -0,52$, $p = 0,70$.

4. Diskussion

Unsere Hypothesen zum Effekt der US-Relevanz haben sich nur teils bestätigt. Es zeigte sich, dass junge VP tatsächlich extremere Bewertungen für Medikamente abgeben, die mit „jungem“ US-Material assoziiert sind. Deskriptiv zeigte sich der Material-Relevanz-Effekt abgeschwächt, wenn der situative Altersfokus der jungen VP auf das Rentenalter gerichtet wird. Dies unterstützt die Annahme, dass VP in der RFG das „alte“ US-Material im Vergleich relevanter finden und betragsmäßig extremere CS Bewertungen abgeben. Die Richtung des (nicht-signifikanten) HE „Relation“ verlief nicht wie angenommen, da sich stärkere Material-Relevanz-Effekte für „verursachende“, als für „verhindernde“ Medikamente fanden. Demnach scheint die subjektive Wahrscheinlichkeit der GZ in der „verhindern“ Bedingung nicht wichtiger zu sein. Gründe für die nichtbestätigten Hypothesen könnten eine zu schwache Manipulation des Altersfokus sein und das die ausgewählten GZ nicht altersspezifisch genug waren und somit für alle Altersgruppen relevant. Denkbar wäre auch eine Konfundierung des Einflusses von Relevanz mit den verschiedenen Schweregraden der GZ, die stark auf das Bewertungsverhalten wirkt. Für zukünftige Forschung sollte altersspezifischere Material generiert und eine größere Stichprobe erhoben werden.

5. Literatur

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, 13(3), 1-21.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, 43(1), 17-32.

Exploring indirect response transfer from native language to a new language using multimodal Associations

Sarah Bischoff, Maria Galbas, Magdalena Krautwurst, Barbara Riedißer, Vanessa Wolf

Supervision: M.Sc. Mrudula Arunkumar

1. Introduction

Understanding how humans learn has been a main goal in psychology for long. From animal literature, we know that two neutral stimuli (S1 and S2) can be linked when presented together, and even if only one of the stimuli (S2) is often paired with an unconditioned response, the unpaired stimulus (S1) associated with S2 can also lead to the response. This is known as sensory pre-conditioning. (Brogden, 1939). In humans, it has been shown that amount of attentional control (Bejjani, Zhang & Egnér, 2018) or even rewards (Wimmer & Shohamy, 2012) can be transferred across associated stimuli. However, it is yet unclear whether this is specific to rewards and cognitive states and if the modality of the stimuli is important. In an unpublished previous study, we found that a response can be transferred across associated stimuli with different modalities. In this study, we aim to replicate the finding using more complex and applicable stimuli in the form of a language learning scenario. We will investigate if responses can transfer from a visual German word to a “new language” word (pseudoword) presented in the auditory modality.

2. Method

Based on *a-priori* power calculation, we recruited 71 participants. Students of FSU Jena were compensated by credit hours. As auditory stimuli (S1) we recorded the two pseudo words *mank* and *dels* portrayed as a new language words that participants are unfamiliar with. As visual stimuli (S2) we used the written German words *Haus* (written in blue letters on black screen) and *Wald* (yellow letters). As response we used a key press which either was 4 (left hand) or 8 (right hand) based on the number displayed on the screen. All stimuli and the response were intended to be neutral without any existing associations between them. The experiment was constructed using PsychoPy (Peirce et al., 2019) and run online on Pavlovia (<https://pavlovia.org/>).

The experiment consisted of *three* phases within participant. In Phase 1, S1 was linked with S2. Using a contingency of 100% *mank* was always paired with *Haus*, *dels* always with *Wald* (40 trials each pair). The setup was as follows: First, there was the inter-trial-screen display (800ms), followed by the auditory pseudoword (800ms), followed by the written German word (800ms). The task was to observe and say the German word aloud. In 100 trials of Phase 2, only S2 was linked with a response by using a number identification task (forced choice paradigm) with the following setup: First a fixation cross (500ms), followed by the German word (200ms), followed by a number (presented until the key was pressed). There was a contingency ratio of 90:10 for the S2 and response, so 90% of the trials were contingent/valid (*Haus* before 4, *Wald* before 8) and 10% not contingent/invalid. During 80 trials in Phase 3, the task was to guess what number would have appeared (free choice paradigm). Half of the trials showed S2 (manipulation check for learning from Phase 2) and half showed S1 (test for transfer of response from associated S2). The setup started with a fixation cross (500ms), followed by the stimulus, visual S2 or auditory S1 (about 200ms), and a white question mark (until guessing the response). The choice made was recorded and we hypothesized that if transfer is successful, then participants would guess 4 to follow *mank* and 8 to follow *dels*.

The experiment ended with questions regarding awareness of the associations and their prediction. The study was preregistered using the AsPredicted template on OSF (10.17605/OSF.IO/FC5U3)

3. Results

Data from the forced choice task, namely reaction time (RT) and error rate (ER) for valid vs. invalid trials, along with the number of guesses made in the free choice task were analysed according to the preregistered analyses. Outliers and very fast RTs from the forced choice tasks were removed. In the free choice task, if participants guess the number contingent with S2 (when S2 or associated S1 is displayed) they were coded as valid guesses. No data was analysed in Phase 1. In Phase 2 we tested whether the S2-response association was learnt by comparing the performance on valid and invalid trials using t-tests. Participants react significantly faster $t(70) = 3.733, p < .001$ and also significantly more accurate, $t(70) = 6.307, p < .001$ in valid trials than in invalid trials.

In Phase 3, we checked the learning for S2-R and more importantly, checked for transfer of response to S1 using the proportion of valid guesses as a dependent measure. Using t-tests we tested whether the proportion of valid keypresses was significantly above 50% (level of chance). For the S2 trials participants guessed significantly above chance, $t(70) = 7.287, p < .001$. Crucially we also see that participants transfer the response to S1 by performing significantly better than chance $t(70) = 6.615, p < .001$.

4. Discussion

This study showed that a transfer of response from visual to auditory modality is possible. Contrary to previous study we did not use an arbitrary audio stimulus but a word (pseudoword) that represents a translation from a devised foreign language to a common German word. Participants were able to learn the associations between the German word and the number response and more importantly transferred this response to the auditory pseudoword by guessing based on the response contingent with the associated German word in the visual modality. Our experiment design had no explicit indication of contingencies which shows that transfer occurs in an incidental learning set-up for our experiment. However, a point in our design that raises questions is the fact that the two German words *Haus* and *Wald* were presented in two different colours, and we can't disentangle whether the learned connection to the pseudoword was with the German word or the colour it was displayed in. Nevertheless, it can still be concluded that response could be transferred to the auditory pseudoword from the associated visual word.

To sum up we discovered that a successful transfer is possible even in less ambiguous set-ups using more complex stimuli involving transfer from familiar native language to unfamiliar language.

This could give further implications in multimodal language learning.

5. References

- Bejjani, C., Zhang, Z. & Egner, T. (2018). Control by association: Transfer of implicitly primed attentional states across linked stimuli. *Psychonomic Bulletin Review* 25, 617–626. <https://doi.org/10.3758/s13423-018-1445-6>
- Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of experimental psychology*, 25(4), 323–332. <https://doi.org/10.1037/h0058944>
- Wimmer, G. E., & Shohamy, D. (2012). Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. *Science*, 338(6104), 270–273. <https://doi.org/10.1126/science.1223252>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

Ich sehe was, was du nicht fühlst – Zwei Studien zum Zusammenhang zwischen Emotionserkennung, Masken und anderen Faktoren

Henriette Kleindienst, Paula Gut, Lara Schäfer, Julia Lehnert, Jakob Körner, Ben Erdeli, Josefine Tews, Aleksandra Spajic, Miriam Findeis, Helene Krämer

Leitung: Dr. Jürgen M. Kaufmann

1. Einleitung

Die Emotionserkennung unseres Gegenübers ist in alltäglichen Situationen wichtig, um die Interaktion zu erleichtern und Missverständnissen vorzubeugen. Seit Beginn der COVID-19-Pandemie ist das Tragen einer medizinischen Maske üblich geworden, um die Ansteckung zu reduzieren. Allerdings ist das Tragen der Masken auch mit negativen Effekten verbunden, so zeigte Carbon (2020), dass die Akkuratheit der Emotionserkennung darunter leidet. Deshalb war das erste Ziel unserer Studien, zu überprüfen, ob das Tragen einer medizinischen Maske die Emotionserkennung in Gesichtern statistisch signifikant verschlechtert. McCrackin et al. (2021) untersuchten neben dem Einfluss von Masken den Einfluss individueller Unterschiede in den Big-Five-Persönlichkeitseigenschaften auf die Emotionserkennung. Sie fanden positive Zusammenhänge der Emotionserkennung mit hoher Verträglichkeit und niedriger Extraversion. Wir stellten deshalb die Hypothese auf, dass sich die Emotionserkennungsleistung von Personen in Abhängigkeit ihrer Ausprägungen in den Big Five unterscheidet. Zusätzlich interessierten wir uns dafür, ob es einen Unterschied zwischen bekannten und unbekanntem Gesichtern gibt, wenn deren Emotionen erkannt werden sollen. Bislang ist der Einfluss der Bekanntheit von Gesichtern auf die Emotionserkennung noch nicht weitgehend erforscht. Einzelne Studien wie die von Barzut und Markovic (2017) deuten allerdings an, dass die Bekanntheit eine Rolle spielen könnte, sei es auch nur bei einzelnen Emotionen. In einer Folgestudie wurde deshalb geprüft, ob sich die Bekanntheit eines Gesichts positiv auf die Emotionserkennung auswirkt. Die Hypothesen lauteten, dass sowohl für Gesichter mit als auch Gesichter ohne Maske die Emotionserkennungsleistung für bekannte Gesichter besser ist als für unbekannte Gesichter.

2. Methode

Die Online-Experimente wurden mit PsyToolkit Version 3.4.2 (Stoet, 2010; 2017) erstellt und durchgeführt. In der ersten Studie beantworteten die VPn (insgesamt 92: 77 Frauen, 15 Männer; 18-59 Jahre) zunächst Items des Big-Five-Inventory (Rammstedt et al., 2014), um die Ausprägungen ihrer Big Five zu messen, und bearbeiteten danach den Experimentaltel, in dem sie die Emotionen von Gesichtern aus der Radboud Faces Database (Langner et al., 2010) kategorisierten. Dies fand in einem 2 (Maske / keine Maske) x 7 (6 Basisemotionen und neutraler Ausdruck) Messwiederholungsdesign statt. In der zweiten Studie sollten ebenfalls Emotionen benannt werden. Das Stimulusmaterial, das hierbei von uns selbst generiert wurde, enthielt Fotos von Dozierenden der Bachelormodule der Psychologie der FSU Jena und Fotos von einer für Bachelorstudierende der Psychologie der FSU Jena unbekanntem Vergleichsgruppe. Jede VP wurde außerdem zum subjektiven Bekanntheitsgrad der ihr gezeigten Personen befragt. Es nahmen 98 FSU-Bachelorstudierende der Psychologie (86 Frauen, 11 Männer, 1 diverse Person; 18-32 Jahre) teil. Zudem wurden bis jetzt und auch über den 06.07.2022 hinaus Nicht-Psychologiestudierende rekrutiert, um mögliche Konfundierungen durch das selbst erhobene Stimulusmaterial zu kontrollieren. Die Berechnung der a-priori Poweranalysen ergab eine gewünschte Stichprobengröße von 95 VPn für jede der Erhebungen.

3. Ergebnisse

Die Analyse der Daten erfolgte mit R-Studio (Version 1.3.1093). Um den Maskeneffekt auf die Emotionserkennung zu überprüfen, wurde für die Akkuratheiten eine ANOVA im 2 (Maske) x 7 (Emotionen) Messwiederholungsdesign berechnet. Die Ergebnisse $F(1, 92) = 768.47, p < .001, \eta^2 = .893$, 95 % KI [.852, .917] belegen deutlich einen Haupteffekt der Maske mit niedrigeren Akkuratheiten für Gesichter mit Masken (Bedingung „Maske“: $M = .696$; Bedingung „keine Maske“: $M = .873$). Es konnte kein Zusammenhang zwischen der Emotionserkennungsleistung der VPn und ihren Big-Five-Ausprägungen nachgewiesen werden, da die Pearson-Produkt-Moment-Korrelationen für den Zusammenhang zwischen Akkuratheit (jeweils mit und ohne Maske) und den einzelnen Big-Five-Dimensionen nicht signifikant wurden. Um zu überprüfen, ob sich die Emotionserkennung bei bekannten und unbekanntem Gesichtern unterscheidet, wurde für die Akkuratheiten in Studie 2 eine 2 (Bekanntheit) x 7 (Emotionen) x 2 (Maske) ANOVA im Messwiederholungsdesign berechnet. Es wurde ein Haupteffekt der Bekanntheit ($F(1, 97) = 12.86, p = .001, \eta^2 = .117$, 95 % KI [.024, .241]) nachgewiesen, der auf einen Vorteil für bekannte Gesichter zurückzuführen war (bekannt: $M = .561$, unbekannt: $M = .542$). Weiterhin wurde ein Haupteffekt für die Maske ($F(1, 97) = 1057.32, p < .001, \eta^2 = .916$, 95 % KI [.884, .934]) und eine Zweifach-Interaktion von Bekanntheit und Maske ($F(1, 97) = 53.74, p < .001, \eta^2 = .357$, 95 % KI [.208, .478]) festgestellt. Durch das Betrachten der Mittelwerte von Bekanntheit und Maske, welche pro VP und über alle Emotionen hinweg gemittelt wurden, zeigte sich die Richtung des Interaktionseffekts: Der Vorteil in der Emotionserkennung für bekannte Gesichter zeigte sich in der Bedingung mit Maske ($M = .497$ für bekannte vs. $M = .441$ für unbekannte Gesichter), aber nicht in der ohne ($M = .625$ für unbekannte vs. $M = .643$ für unbekannte).

4. Diskussion

Die Ergebnisse beider Studien zeigen eine erhebliche Beeinträchtigung der Emotionserkennung durch das Tragen einer medizinischen Maske. Dieser nach Carbon (2020) replizierte Effekt hat in pandemischen Zeiten eine hohe Alltagsrelevanz. Ein Zusammenhang zwischen den individuellen Ausprägungen der Big Five und der Emotionserkennung konnte nicht bestätigt werden. Falls es diesen Zusammenhang auf Populationsebene geben sollte, dann hatte die Studie, aufgrund der homogenen Stichprobe von weiblichen Psychologiestudierenden, eventuell eine zu geringe Variabilität der Persönlichkeitseigenschaften. In der Folgestudie zeigte sich, dass die Emotionen in maskierten Gesichtern von Bekannten im Mittel besser erkannt wurden als von Unbekannten. Ob dieser Effekt wirklich auf der Bekanntheit beruht oder nur ein Stimuluseffekt war, können wir zum jetzigen Zeitpunkt nicht abschließend sagen, da die Datenerhebung noch läuft, aber wenn wir Ihr Interesse geweckt haben, dann können Sie dies gern am 15.07.2022 zum 17. Jenaer Empiriepraktikumskongress bei uns erfragen.

5. Literatur

- Barzut, V. & Marković, S. (2017). Facial expression categorization of unfamiliar and familiar faces. Conference: Current trends in Psychology.
- Carbon, C. C. (2020). Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions. *Frontiers in Psychology*, 11.
- McCrackin, S., Capozzi, F., Mayrand, F., & Ristic, J. (2021). The influence of face masks on emotion recognition and the role of individual differences.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS-GESIS Leibniz Institute for the Social Sciences.
- Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096-1104
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.

You got the job! Bevorzugen wir Mitarbeiter:innen, die eine ähnliche Persönlichkeit haben?

Theresa Splitt, Alina Klettke, Carolina von Zanthier, Nele Pieper

Leitung: M.Sc. Hannah Klink

1. Einleitung

Menschen haben die natürliche Neigung, sich zu vergleichen (Festinger, 1954). Dabei wird sich sozial insbesondere mit ähnlichen Personen auf wesentlichen Dimensionen verglichen (Gilbert et al., 1995). Die Wahrnehmung eines Gesichts und damit auch Persönlichkeitseigenschaften einer fremden Person, wirken sich auf unser Urteil über diese aus (Wero, 1993). Dabei braucht das Gehirn nur eine Zehntelsekunde, um ein Urteil über einen unbekanntem Menschen zu fällen (Jimenez, 2013). Sobald eine positive Eigenschaft eines Menschen wahrgenommen wird, überträgt man dies auch auf andere Bereiche (Dion et al., 1972). Dieser sogenannte Halo-Effekt wurde z.B. für die Tendenz belegt, dass attraktive Menschen als intelligenter wahrgenommen werden (Eagly et al., 1991). Die eigenen Eigenschaften dienen grundsätzlich als Anhaltspunkt für die Bewertung einer anderen Person. Charakterlich ähnliche Personen werden als sympathischer wahrgenommen und durch den Halo-Effekt auch insgesamt positiver bewertet. Dies bedeutet, dass die wahrgenommene Ähnlichkeit einen Einfluss auf die Gesamtbeurteilung auch in Leistungssituationen hat (Rausch et al., 2015). In unserer Studie untersuchten wir daher ob trotz ähnlicher fachlicher Qualifikationen Bewerber:innen, die einem selbst ähnlicher sind, eher als passend ausgesucht werden (H1) und ob eine stärkere wahrgenommene Ähnlichkeit zu mehr Sympathie, Vertrauen und damit auch zu gefühlt besserer Eignung für den Job führt (H2).

2. Methode

Stichprobe. Unsere Stichprobe umfasste 122 Personen (91 weiblich, 27 männlich, 3 nicht-binär, 1 Person keine Angabe). Der Altersdurchschnitt betrug 27.93 Jahre (*SD*: 13.41). Die Proband:innen setzten sich aus 93 Studierenden und 29 Berufstätigen zusammen.

Stimuli. Es wurden acht computergenerierte, weibliche Gesichter von der Website „thispersondoesnotexist.com“ verwendet. die den Proband:innen als Bewerberinnen für eine zu besetzende Stelle in einem Unternehmen präsentiert wurden.

Ablauf. Zunächst wurden alle Gesichter der Bewerberinnen hintereinander für einen ersten Überblick präsentiert. Anschließend wurden die Gesichter zusammen mit einem kurzen Kontext (Name, Alter und Wohnort) ein zweites Mal gezeigt, wobei die Proband:innen anschließend die Sympathie und Kompetenz der Gesichter sowie die wahrgenommene Ähnlichkeit zwischen sich und den Gesichtern einschätzen sollten (Skala 1-100). Die dritte Präsentation der Gesichter folgte demselben Ablauf wie die Zweite, nur diesmal mit einer zusätzlichen Beschreibung der Gesichter bezüglich der drei Big Five Dimensionen Gewissenhaftigkeit, Offenheit und Extraversion (langer Kontext). Danach sollte aus allen Gesichtern die geeignetste Bewerberin ausgewählt und dies kurz begründet werden. Zuletzt wurden die selbsteingeschätzten Ausprägungen auf den drei Big Five Dimensionen (Rammstedt et al., 2014) sowie demografische Daten erfasst.

3. Ergebnisse

Erste Hypothese: Wir berechneten zunächst einen Ähnlichkeitsscore mit jedem Gesicht, der sich aus der Übereinstimmung der Versuchsperson mit dem Gesicht bezüglich der Ausprägung auf den Big Five Dimensionen, der wahrgenommenen Ähnlichkeit und dem Altersunterschied (niedrig = hohe Ähnlichkeit) zusammensetzte. Bei vier von sieben Gesichtern ergab ein t-Test einen signifikanten Unterschied im Ähnlichkeitsscore zwischen dem ausgewählten und den nicht

ausgewählten Gesichtern, wobei bei allen vier Gesichtern die Ähnlichkeit zum ausgewählten Gesicht höher war (Gesicht 1: $t(24)=8.56, p<.001$; Gesicht 4: $t(14)=-5.04, p<.0001$; Gesicht 5: $t(36)=7.57, p<.001$; Gesicht 7: $t(16)=3.04, p=0.008$). Für die anderen drei Gesichter zeigte sich kein signifikanter Effekt (Gesicht 2: $t(3)=-0.12, p=0.91$; Gesicht 3: $t(15)=0.09, p=0.92$; Gesicht 6: $t(3)=0.87, p=0.45$). Für Gesicht 8 war kein t-Test möglich, da dieses nur einmal ausgewählt wurde.

Zweite Hypothese: Wir analysierten zunächst die Bewertung der wahrgenommenen Ähnlichkeit, der eingeschätzten Sympathie und der eingeschätzten. Dabei zeigte sich mithilfe von t-Tests signifikante Unterschiede der Bewertungen zwischen dem kurzen und langen Kontext, weshalb wir die anschließende Analyse getrennt nach Kontexten durchführten. Die wahrgenommene Ähnlichkeit konnte als Prädiktor für die Einschätzung der Sympathie identifiziert werden, sowohl innerhalb der kurzen ($b = 0.33, t(878) = 12.26, p < .001$) als auch innerhalb der langen Kontexte ($b = 0.44, t(919) = 18.82, p < .001$). Dabei erklärte die wahrgenommene Ähnlichkeit einen signifikanten Anteil der Varianz der Sympathie-Einschätzung bei den kurzen ($R^2 = 0.15, F(1,878) = 150.2, p < 0.001$) und den langen Kontexten ($R^2 = 0.28, F(1,919) = 354.1, p < 0.001$). Auch für die eingeschätzte Kompetenz konnte die wahrgenommene Ähnlichkeit als Prädiktor identifiziert werden, und zwar ebenfalls innerhalb der kurzen ($b = 0.25, t(871) = 9.75, p < .001$) und der langen Kontexte ($b = 0.33, t(921) = 14.24, p < .001$). Die wahrgenommene Ähnlichkeit erklärte auch hier einen signifikanten Anteil der Varianz der Kompetenz-Einschätzung bei den kurzen ($R^2 = 0.10, F(1,871) = 95.06, p < 0.001$) und den langen Kontexten ($R^2 = 0.18, F(1,921) = 202.90, p < 0.001$). Es ergaben sich also in allen Regressionen signifikante Ergebnisse.

4. Diskussion

Objektive Urteile von Menschen unterliegen Verzerrungen, wie Rausch und Kollegen (2015) mit Bewertungen von Schüler:innen zeigen konnten. Aus diesen Erkenntnissen entstand unsere erste Forschungsfrage, ob Bewerberinnen trotz ähnlicher fachlicher Qualifikationen eher als passender ausgesucht werden, wenn sie einem selbst ähnlich sind. Bei vier von sieben Gesichtern erhielten wir signifikante Ergebnisse. Da dies der Mehrheit der Gesichter entspricht, konnte unsere erste Hypothese insgesamt bestätigt werden. Eine mögliche Erklärung für die signifikanten Ergebnisse bei nur vier von sieben Gesichtern könnte die Berechnung des Ähnlichkeitsscores sein, der mit zusätzlichen Variablen (Wohnort der Proband:innen, Tätigkeit der Bewerberinnen) aussagekräftiger sein könnte. Unsere zweite Hypothese besagt, dass stärkere Ähnlichkeit zu mehr Sympathie, Vertrauen und damit zu einer gefühlt besseren Eignung für den Job führt. Diese konnte mit ausschließlich signifikanten Effekten bestätigt werden. Allerdings meldeten uns viele Versuchspersonen zurück, dass sie gerade sich unähnliche Personen ausgewählt haben. In den nachfolgenden Studien müsste daher die Auswahl-situation und ihre Anforderungen näher beschrieben werden. Insgesamt konnten wir so die Studie von Rausch und Kollegen (2015) jedoch bestätigen und die Ergebnisse auf einen anderen Kontext ausweiten. Vermeintlich objektive Urteile sind so auch unbewussten Einflüssen unterzogen, die diese Urteile maßgeblich verzerren können. Darum muss verstärkt Bewusstsein für derartige Verzerrungen geschaffen werden.

5. Literatur

- Dion, K.; Berscheid, E. & Walster E. (1972). What is beautiful is good. *J. Pers. Soc. Psychol.*, 24, 285-290.
- Eagly, A.H.; Ashmore, R.D.; Makhijani, M.G. & Longo, L.C. (1991). What is beautiful is good, but ...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109-128.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140.
- Gilbert, D.T.; Giesler, R.B. & Morris, K.A. (1995). When comparison arise. *J. Pers. Soc. Psychol.*, 69(2), 227.
- Jimenez, F. (2013). Der erste Eindruck bleibt – weil er stimmt. *Welt*.
<https://www.welt.de/gesundheit/psychologie/article114446800/Der-erste-Eindruck-bleibt-weil-er-stimmt.html>
- Rausch, T.; Karing, C.; Dörfler, T. & Artelt C. (2015): Personality similarity between teachers and their students influences teacher judgement of student achievement, *Educational Psychology*, 2015, 1-16.
- Wero, B. (1993). Accuracy in Face Perception: View from Ecological Psychology. *Journal of Personality*, 61(4)

Backfire-Effekte verstehen – wie psychologische Reaktanz visuelle Aufmerksamkeit beeinflusst

Bennet Antonik, Matti Keßler, Jennifer Kroker, Katharina Rehling, Esther Stemmler

Leitung: Dr. Philipp Sprengholz

1. Einleitung

Häufig werden Menschen mit Forderungen und Verboten konfrontiert, die nicht Adhärenz sondern gegenteiliges Verhalten auslösen (sog. Backfire Effekt). Die Theorie der psychologischen Reaktanz geht davon aus, dass Menschen, die sich in ihren Freiheiten verletzt fühlen, negative Kognitionen und Affekte wie beispielsweise Wut erleben, welche dazu motivieren, die eingeschränkte Freiheit wiederherzustellen (Brehm, 1966; Dillard & Shen, 2005). Bisher wurde nicht untersucht, welchen Einfluss Reaktanz auf Aufmerksamkeitsprozesse hat. Damit bedrohte oder tatsächlich eingeschränkte Freiheiten wiederhergestellt werden können, muss Reaktanz zu einer Ausrichtung der Aufmerksamkeit auf solche Stimuli führen, die etwas mit der Freiheitseinschränkung zu tun haben. Diese Annahme wurde im Rahmen von zwei Studien überprüft.

2. Studie I

Methode: In der ersten Studie wurden studentische Teilnehmende ($N = 90$) in zwei Gruppen eingeteilt. Während die Kontrollgruppe keine weiteren Informationen erhielt, las die Experimentalgruppe einen kurzen Text, in dem die negativen Auswirkungen des Fleischkonsums auf Gesundheit und Umwelt dargelegt und die Teilnehmenden dazu aufgefordert wurden, in Zukunft auf Fleisch zu verzichten. In einem anschließenden Dot Probe Task wurden sowohl kritische als auch unkritische Durchgänge präsentiert. In einem unkritischen Durchgang wurden nebeneinander zwei Bilder gezeigt, die nichts mit Nahrung im Allgemeinen und Fleisch im Speziellen zu tun hatten. In den kritischen Durchgängen zeigte eines der beiden Bilder eine fleischreiche Mahlzeit (z.B. einen Hamburger). In beiden Arten von Durchgängen entstammten die Bilder dem IAPS 2008. Sie wurden nur kurz gezeigt (500ms), anschließend erschien an der Position eines der beiden Bilder ein schwarzer Punkt. Je nach Position des Punktes sollte so schnell wie möglich eine bestimmte Taste gedrückt werden.

Hypothese: Es wurde angenommen, dass Reaktanz zu einem Attentional Bias führt, d.h., dass Personen in der Experimentalgruppe ihre Aufmerksamkeit in kritischen Durchgängen eher auf fleischbezogene Bilder lenken und deshalb besonders schnell (langsam) reagieren, wenn der Punkt anschließend (nicht) an der Position erscheint, an der die fleischreiche Mahlzeit gezeigt wurde.

Ergebnis: Im Vergleich zur Kontrollgruppe konnte in der Experimentalgruppe entgegen unserer Hypothese kein Attentional Bias gefunden werden. Die Differenz aus kongruenten (Punkt an gleicher Position wie fleischreiche Mahlzeit) und inkongruenten kritischen Durchgängen (Punkt an anderer Position als fleischreiche Mahlzeit) unterschied sich nicht signifikant zwischen Experimental- und Kontrollgruppe, Welch's $t(65.67) = 0.38, p = .700$.

Diskussion: Der fehlende Nachweis eines Attentional Bias konnte mit verschiedenen methodischen Faktoren zusammenhängen. Womöglich löste die Aufforderung zum Verzicht auf Fleisch nicht genügend Reaktanz aus, zudem waren die Bilder teilweise schwer zu erkennen und unterschieden sich in ihrer Valenz. Diese potenziellen Störfaktoren sollten in einer zweiten Studie behoben werden.

3. Pretest für Studie II

Methode: In einer Umfrage wurde jeweils 30 Personen ein Text gezeigt, der dazu aufforderte, gendergerechte Sprache zu verwenden, auf selbige zu verzichten, keine Pornofilme zu schauen, Maske im Supermarkt zu tragen oder kein Fleisch mehr zu essen (analog zu Studie 1). In einer weiteren Bedingung sollten sich 30 Teilnehmende vorstellen, dass die Wehrpflicht wieder eingeführt wird. Anschließend wurde das Ausmaß an Reaktanz mit vier Items der Salzburger State Reactance Scale (Sittenthaler et al., 2015) gemessen (Bsp.: *Es ärgert mich, dass die Wehrpflicht wieder eingeführt werden soll*) und zu einem Score gemittelt (Cronbach's α jeweils $> .90$).

Ergebnis: Im Mittel löste das Wehrpflicht-Szenario am meisten Reaktanz aus, sodass es als Basis für Studie II ausgewählt wurde.

4. Studie II

Methode: Analog zu Studie I wurden $N = 48$ Teilnehmende eines deutschen Online-Panels zufällig einer Kontrollgruppe (keine Informationen) oder einer Experimentalgruppe (Lesen des Wehrpflichtszenarios) zugeordnet. Anschließend folgte ein Dot Probe Task, der statt Bildern Wörter verwendete. In kritischen Durchgängen hatte eines der beiden jeweils gezeigten Wörter militärischen Bezug (z.B. Truppe, Helm), alle anderen Wörter waren neutral formuliert (z.B. Team, Mütze). Wie in Studie 1 erschien nach Präsentation der Stimuli ein Punkt an einer der beiden Wortpositionen und durch Drücken einer von zwei Tasten musste darauf so schnell wie möglich reagiert werden.

Hypothese: Es wurde wiederum angenommen, dass Reaktanz zu einem Attentional Bias führt, d.h., dass Personen in der Experimentalgruppe ihre Aufmerksamkeit in kritischen Durchgängen eher auf militärbezogene Wörter lenken und deshalb besonders schnell (langsam) reagieren, wenn der Punkt anschließend (nicht) an der Position erscheint, an der das militärbezogene Wort gezeigt wurde.

Ergebnis: In der Experimentalgruppe konnte ein Attentional Bias von 9.30 ms (95% KI = [-15.92, -2.67]) gefunden werden. Wie vorhergesagt, unterschied sich die Differenz aus kongruenten (Punkt an gleicher Position wie militärbezogenes Wort) und inkongruenten kritischen Durchgängen (Punkt an anderer Position als militärbezogenes Wort) signifikant zwischen Experimental- und Kontrollgruppe, Welch's $t(45.12) = -2.83, p = .007$.

Diskussion: Die Androhung einer Wehrpflicht führte zu Reaktanz und diese wiederum zu einer Ausrichtung der visuellen Aufmerksamkeit auf Stimuli, die mit der Freiheitseinschränkung in Zusammenhang standen (in diesem Fall Wörter mit Militärbezug). Das Ergebnis zeigt, dass Zusammenhänge zwischen psychologischer Reaktanz und Verhaltensänderungen durch die Anpassung von Aufmerksamkeitsprozessen ermöglicht werden.

5. Literatur

Brehm, JW (1966) *A Theory of Psychological Reactance*. Oxford: Academic Press.

Dillard, J. P., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, 72(2), 144-168.

Sittenthaler, S., Traut-Mattausch, E., Steindl, C., & Jonas, E. (2015). Salzburger State Reactance Scale (SSR Scale): Validation of a scale measuring state reactance. *Zeitschrift für Psychologie*, 223(4), 257.

Wir verstehen uns auch ohne Worte – oder doch nicht? Empathische Akkuratheit gegenüber bilingualen Personen

C. Halcour, E. Joerges, C. Schmitz, Mi. Schwarz, L. Stock, N. Vernaldi

Leitung: Dr. phil. Andrea Schlesier-Michel

1. Einleitung

Das Aufwachsen in verschiedenen Kulturen geht nicht nur mit der Sozialisation durch Werte und Normen einher, sondern auch mit unterschiedlichen emotionalen Erfahrungen und einer verschiedenen Art und Weise diese zu kommunizieren und wahrzunehmen (Elfenbein & Ambady, 2002). Das Konstrukt der *empathischen Akkuratheit (EA)* umfasst die Fähigkeit die Gedanken und Gefühle unseres Gegenübers richtig zu erkennen. Aufbauend auf Faktoren wie Sprache, Paralinguistik und nonverbalen Hinweisreizen, die je nach Kultur unterscheiden können, untersuchten wir, inwiefern verschiedene Informationskanäle (Ton und/ oder Bild und/ oder Verständnis der verwendeten Sprache) von Bedeutung für die *empathische Akkuratheit* gegenüber bilingualen Personen sind. Für unsere Untersuchung stellten wir folgende Hypothesen auf: (1) Die EA ist am besten, wenn alle Kanäle genutzt werden können. (2) Da positive Emotionen *nonverbal* vor allem über die Mimik, negative jedoch über die Paralinguistik transportiert werden (Elfenbein & Ambady, 2002), erwarteten wir für die *nicht-deutschen Aufnahmen* eine höhere EA für positive Gefühle, wenn man nur das Video (Mimik) hat als wenn man nur das Audio (Paralinguistik) hat, für negative Gefühle jedoch eine höhere EA, wenn man nur das Audio (vs. nur das Video) hat. (3) Die EA ist bei den *Videos ohne Ton* höher, wenn ursprünglich Deutsch als Erzählsprache verwendet wurde, als wenn die Muttersprache verwendet wurde.

2. Methode

Target-Videos: Wir erstellten Stimulus-Videos mit bilingualen Personen, die Deutsch und mind. eine andere weniger geläufige Sprache fließend sprechen. Aus sieben geführten Interviews wählten wir drei möglichst vergleichbare Personen aus: alle weiblich, 18 bis 20 Jahre alt mit den Zweitsprachen Russisch, Rumänisch und Chinesisch.

Ablauf und Design. Die *Target-Personen* führten ein ca. einstündiges Zoom-Interview mit Fragen zu positiven und negativen Erlebnissen, wobei die Personen einmal auf Deutsch und einmal auf der jeweils anderen Sprache berichten sollten. Anschließend ging die Versuchsleitung die Videoaufnahme mit der Person durch und stoppte jenes zu bestimmten Zeitpunkten, an denen diese einschätzen sollte, wie sie sich zum Zeitpunkt des Erzählens gefühlt hatte. Auf einer neunstufigen Skala von -4 (*extrem negativ*) bis +4 (*extrem positiv*) sollte dabei die generelle Gefühlslage eingeschätzt werden, und auf einer fünfstufigen Skala von 0 (*gar nicht*) bis 4 (*stark*) vier positive (froh, stolz, zufrieden, ausgeglichen; Cronbach's $\alpha = .90$) und drei negative Emotionen (traurig, verärgert, verängstigt; Cronbach's $\alpha = .72$). Es ergaben sich so jeweils 9-10 Kurzvideos, wovon sich die Hälfte auf positive (vs. negative) Erlebnisse bezog. Diese dienten als Grundlage für eine *online-Studie* (SoSci-Survey). Die Studienteilnehmenden (*Perceiver*) dort wurden zufällig einer von zwölf Bedingungen zugeordnet, in der sie entweder die deutsch- oder die fremdsprachigen Videos sahen („between-factor: Sprache“), und zwar in rotierter Reihenfolge als Audioaufnahme, als Videoaufnahme ohne Ton, und als Video mit Ton („within-factor: Bedingung“), wobei sie pro Bedingung jeweils ein unterschiedliches Target sahen/hörten. Sie sollten dabei an den gleichen Stellen wie die Targets deren Emotionen einschätzen. Die jeweilige *empathische Akkuratheit* wurde als Korrelation (Fishers-z-standardisiert) zwischen Selbstbericht des Targets und Fremdbbericht der Perceiver über die Stopp-Punkte errechnet. Für H1 und H3 fassten wir die drei EA-Bereiche (generelle Gefühlslage, positive und negative Emotionen) zu einem Maß zusammen (Cronbach's $\alpha_{\text{Video}} = .92$, $\alpha_{\text{Audio}} = .84$, $\alpha_{\text{Voll}} = .85$).

Stichprobe. Die Rekrutierung der durch Power-Analyse angestrebten 160 Perceiver erfolgte über persönliche Kontakte (Schneeballsystem), sowie verschiedene Fachschaften in Deutschland. 184 Personen begannen den Fragebogen auszufüllen, 131 (71.2%) Personen beendeten ihn vollständig. Weitere neun Perceiver schlossen wir wegen zu kurzer Bearbeitungszeit, Auffälligkeiten im Aufmerksamkeitscheck oder basaler Sprachkenntnisse in mind. einer der verwendeten Zweitsprachen aus. So ergab sich eine letztendliche Stichprobe von 122 Personen (davon 21 Männer, 1 divers; $M_{\text{Alter}} = 29$ Jahre, $SD_{\text{Alter}} = 14$ Jahre) mit einem eher hohen SÖS (95% Abitur). Die deutsche wurde etwas häufiger beendet als die Zweitsprachenversion ($n = 68$ vs. $n = 54$).

3. Ergebnisse

Vorbereitende Analysen: Ein Shapiro-Wilk-Test auf Normalverteilung bestätigte diese für alle Variablen. Es gab zudem keine Ausreißer außerhalb des 99%-Konfidenzintervalls.

Hypothese 1: Eine Messwiederholungs-ANOVA mit Interaktion (within-Faktor Bedingung nur Video, nur Audio, Video mit Ton; between-Faktor Deutsch vs. Zweitsprache), zeigte signifikante Unterschiede in der EA sowohl zwischen den Bedingungen, $F(1.9, 224.6^1) = 52.68, p < .001$, partielles $\eta^2 = .31$, zwischen den Sprachversionen, $F(1, 117) = 257.23, p < .001$, partielles $\eta^2 = .69$, sowie eine Interaktion, $F(1.9, 224.6^1) = 43.83, p < .001$, partielles $\eta^2 = .27$. Die erwartete bessere EA für das deutsche Video mit Ton im Vergleich allen anderen Bedingungen, wurde jedoch post-hoc nicht durch unsere Ergebnisse bestätigt. Dieser Mittelwert ($M_{EA} = .86, SD_{EA} = 0.07$) unterschied sich nicht signifikant von der Bedingung „deutsch - nur Audio“ ($M_{EA} = .84, SD_{EA} = 0.11$).

Hypothese 2: In *t*-Tests für abhängige Stichproben zeigten sich weder Unterschiede in der EA für positive Gefühle, $t(53) = 0.755, p = .227$, Cohen's $d_z = 0.10$, noch für negative Gefühle, $t(51) = 0.944, p = 0.175$, Cohen's $d_z = 0.13$, zwischen „nur Audio“ ($M_{EA+} = .30, SD_{EA+} = 0.39; M_{EA-} = .23, SD_{EA-} = 0.45$) und „nur Video“ ($M_{EA+} = .34, SD_{EA+} = 0.43; M_{EA-} = .33, SD_{EA-} = 0.35$) in der Zweitsprache.

Hypothese 3: Der *t*-Test für unabhängige Stichproben zeigte keine signifikanten Unterschiede zwischen den beiden „nur Video“-Bedingungen, $t(119) = 0.233, p = .408$, Cohen's $d = 0.04$ ($M_{EA\text{deutsch}} = .40, SD_{EA\text{deutsch}} = 0.28; M_{EA\text{Zweitsprache}} = .37, SD_{EA\text{Zweitsprache}} = 0.33$).

4. Diskussion

Die Daten unterstützten unsere spezifischen Hypothesen nicht. Bei den deutschen Videos war die EA in beiden Bedingungen, in denen man den Ton hören konnte, am höchsten, das zusätzliche Vorhandensein mimischer Information brachte keine zusätzliche Erhöhung. Bei der Zweitsprache machte es keinen Unterschied, ob man mimische oder paralinguistische Informationen zur Verfügung hatte, die EA für positive und negative Gefühle war jeweils gleich niedrig. Zudem unterschied sich die EA für die „nur Videos“ nicht, egal ob sie auf Deutsch oder der Zweitsprache basierten. Insgesamt zeigen unserer Ergebnisse, dass verständliche verbale Informationen die wichtigste Informationsquelle für das Erkennen der emotionalen Befindlichkeit des Gegenübers sind. Stehen einem diese nicht zur Verfügung (unbekannte Zweitsprache) zeigten post-hoc-Tests die höchste EA wenn sowohl paralinguistische als auch mimische Informationen vorliegen (Video mit Ton). Unsere Stichprobe bestand zu einem großen Teil aus Studierenden und zu einem noch größeren Anteil aus Personen mit hohem Bildungsniveau, was zu Verzerrungen geführt haben könnte, ebenso wie, dass es sich um Videoaufnahmen und keine echten Interaktionen handelte.

5. Literatur

Elfenbein, H. A. & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203-235.

¹ Huynh-Feldt- Korrektur der Freiheitsgrade wegen sign. Mauchly-Test auf Sphärizität (Huynh-Feldt- $\epsilon = .933$).

Auditive Szenenanalyse auf der Cocktailparty

**Luise Bock, Merija Gühring, Luise Hillebrand, Alina Schönherr, M.S., Vera
Wolferstetter**

Leitung: PD Dr. Romi Zäske

1. Einleitung

Unser alltägliches Leben setzt sich aus einer Vielzahl von Geräuschen zusammen. In sogenannten Cocktailparty-Situationen sind mehrere Menschen in verschiedene Gespräche aus unterschiedlichen Richtungen involviert und gleichzeitig versucht man seinem eigenen Gesprächspartner zu folgen. Auditive Szenenanalyse wurde bislang v.a. für Sprachverstehen untersucht¹, selten jedoch für die Segregation von Sprecheridentitäten innerhalb sog. Multitalker-Stimuli. Hier wurde gezeigt, dass es Hörern mit steigender Sprecheranzahl schwerer fällt, die Sprecher zu zählen und damit zu unterscheiden^{2,3}. Wie können wir die Segregation von mehreren Sprechern in Cocktailparty-Situationen unterstützen, wenn eine zusätzliche aufgabenirrelevante Stimme im Hintergrund spricht? Ausgehend von Studien zum Sprachverstehen sollte eine Beeinflussung höherer Verarbeitungsprozesse bei der Segregation der Sprecher helfen¹. Dazu zählt u.a. die Lenkung des Aufmerksamkeitsfokus auf das relevante Sprachsignal (Multitalkers) und die Entfernung aufgabenirrelevanter semantischer Information aus der Hintergrundsprache. Hierfür hinterlegten wir die aufgabenrelevanten Multitalker-Stimuli, welche einen Vokal im linken oder rechten Ohr sprachen, mit einer binaural präsentierten Stimme. Diese äußerte einen irrelevanten Satz. Manipuliert wurde (1) die Sinnhaftigkeit des irrelevanten Satzes (Vorwärts- vs. Rückwärtssprache) und (2) das Vorhandensein eines ortsbasierten Hinweisreizes (Cue) für den Multitalker-Stimulus (mit vs. ohne Cue auf das entsprechende Ohr). Wir erwarteten eine größere Genauigkeit in der Zähltaufgabe, d.h. eine erleichterte Segregierbarkeit der Multitalker-Stimuli, wenn (1) die Hintergrundsprache rückwärts im Vgl. zu vorwärts abgespielt wird und (2) der Präsentationsort des Multitalker-Stimulus durch einen Cue angezeigt wird im Vgl. zu keinem Cue.

2. Methode

In die Auswertung dieses Online-Experiments^{4,5} gingen nach Ausschluss unvollständiger Datensätze Daten von $N = 23$ Personen ein (15 weiblich, 7 männlich, 1 binär); Alter: $M = 22,1$ $SD = 3,13$, range: 19-31 Jahre). Eine a priori Poweranalyse⁶ hatte eine Stichprobengröße von $N = 34$ ergeben, um mittelstarke Haupteffekte mit einer Power von .80 und $\alpha = .05$ zu finden. Alle Proband*innen verfügten über ein normales Hörvermögen, waren deutsche Muttersprachler, kannten vorher keinen der Sprecher und verwendeten Kopfhörer. Als Multitalker-Stimuli dienten gehaltene Vokale (1500 ms) von 31 Sprecher*innen (14♀) aus der JESS-Datenbank⁷. Bis zu 5 Sprecher*innen wurden gleichzeitig entweder über den linken oder rechten Kanal gespielt, wobei per Tastendruck die Anzahl gleichzeitiger Sprecher*innen (1-5) geschätzt werden musste. Als irrelevante Hintergrundsprache dienten binaural präsentierte Sätze von 20 Einzelsprecher*innen (10♀), welche vorwärts oder rückwärts abgespielt wurden (Bedingung Hintergrundsprache). Wir variierten zudem blockweise, ob der Ort der aufgabenrelevanten Multitalker-Stimuli zuvor von einem Signalton (Cue) angezeigt wurde oder nicht (Cueing-Bedingung). Daraus ergibt sich ein 2 x 2 Innersubjekt-Design. Aus insg. 500 Stimuli wurden 80 Trials randomisiert gezogen.

3. Ergebnisse

Die Auswertung der mittleren Antwortgenauigkeiten mittels einer zweifaktoriellen ANOVA mit Messwiederholung auf beiden Faktoren ergab weder sign. Haupteffekte von Hintergrundsprache ($F(1, 22) = 1.308, p = .265, \eta_p^2 = .06$) und ortsbasiertem Cueing ($F(1, 22) < 1$), noch eine Interaktion der Faktoren ($F(1, 22) < 1$). Auch zeigte sich kein sign. Effekt der Faktoren auf die Reaktionszeiten ($0.08 \leq F(1, 22) \leq 1.84, .189 \leq ps \leq .777$). Deskriptiv fand sich eine Abnahme der Antwortgenauigkeit mit steigender Sprecheranzahl (Abb. 1).

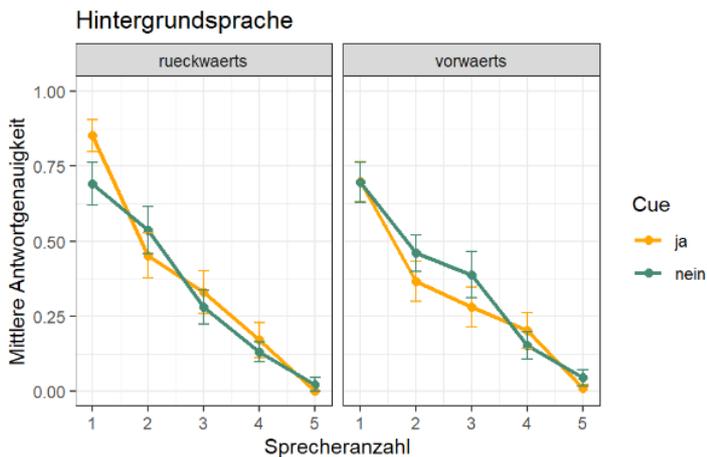


Abb. 1. Mittlere Antwortgenauigkeiten in der Zähl Aufgabe in Abh. der experimentellen Bedingungen (Hintergrundsprache und Cueing), sowie getrennt nach Sprecheranzahl. Fehlerbalken sind Standardfehler des Mittelwerts (SEM).

4. Diskussion

Entgegen unserer Hypothesen und Befunden zur Sprachverarbeitung¹ wurde die Segregierbarkeit von Multitalker-Stimuli weder durch die Sinnhaftigkeit irrelevanter Hintergrundsprache noch durch ortsbasiertes Cueing moduliert. Für den Nulleffekt von irrelevanter Hintergrundsprache könnte eine Erklärung sein, dass das Zählen von Sprechern im Ggs. zu linguistischer Verarbeitung auf relativ frühen Verarbeitungsstufen stattfindet. Eine Interferenz durch Hintergrundsprache fand möglicherweise deshalb nicht statt, weil die relevante Zähl Aufgabe keiner Sprachverarbeitung bedarf. Überprüfen ließe sich dies in einer Folgestudie mit einer Änderung der Aufgabenstellung, wobei nicht die Sprecher gezählt, sondern die gesprochenen Vokale identifiziert werden. Das Ausbleiben eines Cueing-Effekts könnte damit zusammenhängen, dass die Aufgabenschwierigkeit unabhängig von Aufmerksamkeitsprozessen insgesamt sehr hoch war, ein Cue hier also keine Hilfe darstellt. Alternativ könnten unsere Ergebnisse durch die eingeschränkte Kontrollierbarkeit von Störvariablen in Online-Studien verfälscht worden sein oder mit einer geringen Power unserer relativ kleinen Stichprobe zusammenhängen, was in Folgestudien berücksichtigt werden sollte.

5. Literatur

- [1] Wang, X. & Xu, L. (2021). Speech perception in noise: Masking and unmasking. *Journal of Otology*, 16, 109-119.
- [2] Bock, L., Gühring, M., Hillebrand, L., Schönherr, A., Schumacher, M., Wolferstetter, V. (2022). Forschungsbericht zum Empiriepraktikum „Auditive Szenenanalyse“ im WiSe. Lehrstuhl für Allgemeine Psychologie 1, Friedrich-Schiller-University Jena.
- [3] Kwak, C., Han, W. (2020). Towards Size of Scene in Auditory Scene Analysis: A Systematic Review, *Journal of Audiology & Otology*, 24(1), 1-9.
- [4] Stoet, G. (2010). PsyToolkit – A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096-1104
- [5] Stoet, G. (2010). PsyToolkit – A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31
- [6] Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191
- [7] Zäske, R., Skuk, V.G., Golle, J., Schweineberger, S.R. (2019, 14.Juni). The Jena Speaker Set (JESS) – A database of voice stimuli from unfamiliar young and old adult speakers. <https://osf.io/m5zdf/>

Aus den Augen aus dem Sinn - Einfluss der wahrgenommenen Seriosität einer Quelle auf die Entstehung von False Memories

Annalena Utecht, Kaya Lorenz, Antonia Nestler, Roya Haji Hosseini, Elisa Marski

Leitung: Dr. Jürgen M. Kaufmann

1. Einleitung

Ist auf unser Gedächtnis immer Verlass? Dieser Annahme zu vertrauen kann trügerisch sein, denn es ist möglich, sich an Dinge zu erinnern, die man nie erlebt hat. So kann es passieren, dass Erinnerungen durch den Inhalt anderer Quellen wie Texte oder Erzählungen ergänzt oder ersetzt werden. Wie leicht derartige „False Memories“ entstehen können, zeigten bereits Loftus und Palmer (1974). Die Quellen, aus der die (Fehl-)Informationen zum kritischen Ereignis stammen, sind dabei ebenfalls von Bedeutung. Ausgehend davon stellten wir uns in Studie 1 die Frage, inwiefern die wahrgenommene Seriosität einer Quelle, mit der falsche Informationen präsentiert werden, die Entstehung von „False Memories“ beeinflusst. Wir nahmen an, dass nach Präsentation falscher Informationen in einer seriös erscheinenden Quelle im Vergleich zu einer unseriös wirkenden häufiger falsche Erinnerungen entstehen. Darauf aufbauend untersuchten wir in Studie 2 zusätzlich, ob die Ausprägung des Selbstwertgefühls die Übernahme von Falschinformationen aus einer seriösen oder unseriösen Quelle beeinflusst. Dabei erwarteten wir, dass sich Personen in Abhängigkeit ihres Selbstwertgefühls in der Anfälligkeit gegenüber „False Memories“ unterscheiden.

2. Methode

Stichprobe. In Studie 1 nahmen 128 Personen teil. In die Analyse gingen die Daten von 93 Personen im Alter vom 18 bis 64 Jahren ein. In Studie 2 gingen die Daten von 74 Versuchspersonen im Alter von 18 bis 56 Jahren ein. In beiden Studien wurden die Übrigen aufgrund technischer Probleme und unvollständiger Bearbeitung des Fragebogens ausgeschlossen.

Material und Prozedur. In beiden Studien war zunächst ein Video zu sehen, das jeweils einmalig präsentiert wurde. In Studie 1 zeigte dieses einen Autounfall an einer Londoner Kreuzung. Anschließend folgte in Abhängigkeit der Gruppenzugehörigkeit die Lektüre eines seriösen bzw. unseriösen Zeitungsartikels, der ein fiktives Interview zum Unfall beinhaltete. Die zusätzliche Kontrollgruppe las einen neutralen Text. Das Video aus Studie 2 zeigte einen Raubüberfall in einem Supermarkt in Los Angeles. Darauf folgten seriös bzw. unseriös wirkende fiktive Twitter-Beiträge, welche den Raubüberfall thematisierten. Anschließend wurde in beiden Studien ein Fragebogen mit sechs kritischen und 15 neutralen Items präsentiert. Die kritischen Items fragten jeweils die Falschinformationen ab. In Studie 2 wurde nach Abschluss des Fragebogens mithilfe der Rosenberg Self-Esteem Scale (1989) das selbsteingeschätzte Selbstwertgefühl ermittelt. Jede Versuchsperson wurde randomisiert einer Seriositätsbedingung zugeordnet. Die Datenerhebung erfolgte online über die Software PsyToolkit PsyToolkit (Version 3.4.2; Stoet, 2010, 2017).

3. Ergebnisse

In Studie 1 konnte ein Manipulationscheck mittels einer einfaktoriellen ANOVA mit dem Zwischen-Gruppen-Faktor experimentelle Bedingung belegen, dass die Seriosität der drei Artikel signifikant unterschiedlich eingeschätzt wurde, $F(2,90) = 52.16$, $p < .001$, $\eta^2 = .537$. Ein Welch's t -Test für zwei unabhängige Stichproben zeigte einen signifikanten Unterschied für die Einschätzung der Seriosität zwischen den Bedingungen seriös und unseriös von 1.41 mit 95% KI [1.03, 1.79] zugunsten der seriösen Bedingung, $t(59.76) = 7.465$, $p < .001$. Der Chi-Quadrat-Test für den Vergleich der Häufigkeiten korrekter vs. falscher Antworten auf die kritischen Items zwischen den drei experimentellen Bedingungen wurde signifikant, $\chi^2(2) = 30.089$, $p < .001$. Es zeigte sich ferner

ein signifikanter Unterschied in der Verteilung der Häufigkeit richtiger und falscher Antworten für die kritischen Items zwischen der seriösen und unseriösen Bedingung, $\chi^2(1) = 13.724$, $p < .001$. Auch in Studie 2 wurde der Chi-Quadrat-Test für den Vergleich der Häufigkeiten korrekter vs. falscher Antworten auf die kritischen Items zwischen den Bedingungen signifikant, $\chi^2(1) = 19.082$, $p < .001$. Die Versuchspersonen der Gruppe mit dem seriösen Twitter-Beitrag gaben mehr falsche Antworten im Verhältnis zu richtigen Antworten bei den kritischen Items, als die Gruppe mit der unseriösen Quelle. Eine Pearson-Moment-Korrelation legte einen positiven Zusammenhang zwischen der Ausprägung des Selbstwertgefühls und der mittleren Akkuratheit für die kritischen Items nahe, $r(74) = .27$, $p = 0.017$.

4. Diskussion

Studie 1 konnte zunächst den Befund von Loftus und Pickrell (1995) replizieren: Das Induzieren von Falschinformationen kann die Erinnerung an ein kritisches Ereignis verändern. Darüber hinaus wurden die präsentierten Artikel unterschiedlich seriös wahrgenommen, was den Erfolg unserer Manipulation belegt. Der Vergleich korrekter Antworten zwischen den drei Artikeln auf die kritischen Items zeigte, dass diese häufiger fehlerhaft beantwortet wurden, wenn ein seriös statt unseriös erscheinender Artikel gelesen wurde. Dies ist ein bedeutsamer Befund, da wir entsprechend unserer Hypothesen wohl eher dazu tendieren, Falschinformationen zu übernehmen, wenn deren Quelle vertrauenswürdig gestaltet ist. Hiermit übereinstimmend konnten auch Zhu et al. (2010, S.1200) anhand anderer Studiendesigns einen sogenannten „Glaubwürdigkeitseffekt der Quelle“ zeigen. Eine mögliche Erklärung könnte die Behauptung von Hannah und Sternthal (1984) sein, bei einer Informationsaufnahme würde nicht nur die Information selbst im Gedächtnis gespeichert, sondern auch die Glaubwürdigkeit ihrer Quelle.

In Studie 2 gelang es uns ebenfalls, mit einer seriös in Vergleich zu einer unseriös gestalteten Quelle mehr False Memories zu erzeugen. Weiterhin fanden wir, dass ein hohes Selbstbewusstsein seltener mit der Übernahme von Falschinformationen einhergeht als ein niedriges. Personen mit höherem Selbstbewusstsein scheinen weniger anfällig für die Übernahme von Falschinformationen zu sein.

5. Literatur

- Hannah, D. B., & Sternthal, B. (1984). Detecting and Explaining the Sleeper Effect. *Journal of Consumer Research*, 11(2), 632. <https://doi.org/10.1086/209000>
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589. [https://doi.org/10.1016/S0022-5371\(74\)80011-3](https://doi.org/10.1016/S0022-5371(74)80011-3)
- Loftus, E. F., & Pickrell, J. E. (1995). The Formation of False Memories. *Psychiatric Annals*, 25(12), 720–725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Zhu, B., Chen, C., F. Loftus, E., Lin, C., & Dong, Q. (2010). Treat and trick: A new way to increase false memory. *Applied Cognitive Psychology*, 24(9), 1199–1208. <https://doi.org/10.1002/acp.1637>
- Rosenberg, Morris. 1989. *Society and the Adolescent Self-Image*. Revised edition. Middletown, CT: Wesleyan University Press.
- Stoet, G. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42, 1096–1104 (2010). <https://doi.org/10.3758/BRM.42.4.1096>
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31

Fragen in Augenzeugeninterviews - Welche Schuld tragen Single-Choice-Antworten? - Psychopol ermittelt - Gedächtnisverzerrung aufgrund suggestiver Fragestellungen

Helene Heilig, Franz Lehmann, Nele Reisser, Max Wagner, Ferdinand Werner

Leitung: Dr. Jürgen M. Kaufmann

1. Einleitung

Erinnerungen prägen uns Menschen. Wir gehen davon aus, dass die Dinge so geschehen sind, wie wir sie erinnern. Doch ob unsere Erinnerungen der Realität entsprechen, können wir nicht wissen. Was passiert, wenn unser Gehirn zusätzliche Informationen hinzufügt oder vermeintlich unwichtige weglässt? Erinnerungsverzerrungen dieser Art werden als "False Memory" bezeichnet. Menschen können ganze Ereignisse erinnern, die ihnen nie widerfahren sind (Loftus et al., 1995). Dies ist im Bereich der Forensik von großer Bedeutung, wo Erinnerungen von Augenzeug*innen zu Verurteilungen führen können. Diese Untersuchungen beschäftigten sich mit sogenannten Suggestionseffekten. Suggestion kann durch Suggestivfragen (Fragen mit bereits eingebauter Antwort, z.B. Single-Choice-Fragen) erfolgen. In der Forensik wird Suggestion als eine Veränderung von Gedächtnisinhalten, bzw. Induktion von Pseudoerinnerungen, verstanden. In den vorliegenden Studien ging es darum, inwiefern im Bereich der Forensik Suggestionseffekte auftreten können. Was für einen Effekt hat die Art der Fragestellung bei der Befragung von Augenzeug*innen? Die Bereitstellung fehlerhafter Informationen nach einem Ereignis kann die Billigung dieser Informationen bei einem späteren Erinnerungstest erhöhen (Schacter et al., 2011). Daher behandelten die vorliegenden Studien im Besonderen die Frage, inwiefern das Präsentieren von Antwortmöglichkeiten in Form eines Single-Choice-Formats (Suggestion) die Erinnerung beeinflussen. Bisherige Forschung zeigte, dass Multiple-Choice-Tests zu einer Kreation von falschem Wissen führen können (Roediger et al., 2005). Aus den vorangegangenen Überlegungen bildete sich folgende Hypothese bezüglich der Untersuchungen: Die Verwendung des Single-Choice-Formats, verglichen mit einem offenen Antwortformat, führt zu einer höheren Fehlerrate bei der späteren Überprüfung des zu Erinnernden durch Ja/Nein-Fragen.

2. Methode

Anhand einer Poweranalyse ($f = 0.3$, $\alpha = 0.05$, $\beta = 0.8$) zeigte sich, dass zur Durchführung des vorliegenden Experiments eine Stichprobengröße (N) von 90 Versuchspersonen (VP), und somit je 45 Personen für Kontroll- und Experimentalgruppe (KG und EG), notwendig war, um einen existierenden Effekt aufzudecken. Die Stichprobe der Erhebung des Wintersemesters 2021/22 bestand aus 90 Personen ($M = 25.16$, $SD = 9.88$). Aufgabe der Personen war es, innerhalb eines Online-Experiments, das über „PsyToolkit“ gestaltet wurde, ein Video von einem Überfall auf einen Geldtransporter in Berlin, zu betrachten und Fragen zu den Inhalten dieses Videos zu beantworten (Stoet 2010, 2017). Die EG erhielt indes 10 Fragen im Single-Choice-Format, mit je 5 Antwortoptionen, wobei bei der Hälfte dieser Fragen keine korrekten Antwortoptionen gegeben wurden. Neben je drei Antwortmöglichkeiten, welche potenziell gezeigte Details des Videos thematisierten, konnten die VP weiterhin zwischen den Möglichkeiten „Ich weiß nicht“ und „Nichts davon“ wählen. Die KG hatte dieselben Fragen innerhalb eines offenen Antwortformats erhalten und konnte ebenfalls von der „Ich weiß nicht“-Option Gebrauch machen. Nach einer Zwischensequenz in Form des Knobelspiels „Die Türme von Hanoi“, sollten anschließend beide Gruppen dieselben Ja-Nein-Fragen beantworten, welche die Aspekte der bereits gestellten Fragen aufgriffen. Der Aufbau des Experiments, welches im Sommersemester 2022 durchgeführt wurde, entsprach dem des Wintersemesters, bis auf zwei Veränderungen: dass zum einen ein anderes Video, in diesem Fall eines Raubüberfalls auf einen Apple-Store in Kaiserslautern, ausgewählt und dass zum anderen

anstelle eines Knobelspiels ein Zeitintervall von einer Woche gesetzt wurde. Somit war das zweite Experiment in zwei Erhebungszeitpunkte unterteilt. Die Stichprobe bestand aus 80 Teilnehmer*innen ($M = 24.04$, $SD = 7.5$), aufgeteilt in 41 Personen innerhalb des Single-Choice-Formats und 39 Personen im offenen Antwortformat. In dem Online-Experiment war es die Aufgabe der VP nach der Präsentation des Videos Fragen in einem Single-Choice- oder offenen Antwortformat und nach einer Woche Ja-Nein-Fragen zu den zum ersten Erhebungszeitpunkt erfragten Inhalten zu beantworten. Die Gestaltung der Fragen und Antwortoptionen des ersten Erhebungszeitpunktes war zwar inhaltlich eine andere, entsprach aber konzeptionell der Gestaltung der Fragen aus der Erhebung des Wintersemesters.

3. Ergebnisse

Es wurde in der ersten Studie eine lineare Regression gerechnet, um die Akkuratheit im Ja/Nein-Antwortformat durch den dichotomen Prädiktor der Gruppenzugehörigkeit (SC_vs_Offen) vorherzusagen. Der Prädiktor Gruppenzugehörigkeit war in der ersten Studie mit $t(88) = 0.39$, $p = 0.7$, $b = 0.01$, 95% KI [-0.05, 0.07] statistisch nicht signifikant. Durch die Gruppenzugehörigkeit wurden hier 0.17%, 90% KI [0, 0.041], der Varianz der Akkuratheit erklärt ($R^2 = 0,0017$), $F(1, 88) = 0.15$, $p = 0.7$. Danach wurde das Antwortverhalten vor der Zwischenaufgabe durch mehrere Chi-Quadrat-Tests ausgewertet. Im ersten Chi-Quadrat-Test prüften wir einen bedeutsamen Unterschied bezüglich der sicheren und unsicheren Antworten zwischen den beiden Gruppen. Dieser Test wurde mit $\chi^2(89) = 1.53$, $p = 0.22$ nicht signifikant. Der zweite Test dieser Art, bei dem wir auf einen bedeutsamen Unterschied zwischen den beiden Gruppen im Hinblick auf richtig, falsche und unsichere Antworten testeten, wurde mit $\chi^2(88) = 22.05$, $p < 0.001$ statistisch signifikant. So auch der letzte Test, bei dem wir die unsicheren Antworten aus der Analyse ausschlossen und stattdessen nur die richtigen und falschen Antworten auf einen bedeutsamen Unterschied zwischen dem Single-Choice- und dem offenen Antwortformat testeten. Hier zeigte sich mit $\chi^2(89) = 19.64$, $p < 0.001$ ebenso ein statistisch signifikanter Unterschied.

Gleiches wiederholen wir für die Auswertung der Daten der zweiten Studie, deren Datensammlung zum aktuellen Zeitpunkt noch nicht vollständig abgeschlossen ist. Daher können wir diese Ergebnisse hier noch nicht, allerdings später auf dem Empiriepraktikumskongress, präsentieren.

4. Diskussion

In der ersten Studie konnte die Hypothese nicht bestätigt werden. Es zeigte sich kein bedeutsamer Effekt in der Art der Fragestellung, im Hinblick auf die Akkuratheit. Mit der Einführung bzw. Verlängerung des Zeitintervalls zwischen den Fragenformaten, knüpft die zweite Untersuchung an den Schwächen der ersten an. Ob diese Veränderung zu einem anderen Ergebnis führt, wird sich in der Auswertung zeigen. Zudem konnten nicht alle Schwächen der ersten Studie verbessert werden, wie bspw. das Onlineformat, sodass Folgestudien hieran ansetzen könnten.

5. Literatur

- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725. <https://doi.org/10.3928/0048-5713-19951201-07>.
- Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. In *Journal of Experimental Psychology: Learning Memory and Cognition* (Vol. 31, Issue 5, pp. 1155–1159). <https://doi.org/10.1037/0278-7393.31.5.1155>.
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL.
- Schacter DL, Guerin SA, St Jacques PL. (2011). Memory distortion: an adaptive perspective. *Trends Cogn Sci*. 2011;15(10):467-474. doi:10.1016/j.tics.2011.08.004.
- Stoet, Gijsbert (2010): PsyToolkit: a software package for programming psychological experiments using Linux. In: *Behavior research methods* 42 (4), S. 1096–1104. DOI: 10.3758/BRM.42.4.1096.
- Stoet, Gijsbert (2017): PsyToolkit. In: *Teaching of Psychology* 44 (1), S. 24–31. DOI: 10.1177/0098628316677643.

„Gute Nacht Siri, schlaf gut!“ – Wahrnehmung von virtuellen Assistenzsystemen

Odelia Herman, Sophie-Luise Krüger, Julius Protte, Ina Strugale, Carolin Zimmer

Leitung: Carina G. Giesen

1. Einleitung

In unserem Alltag verlassen wir uns heute immer mehr auf die Hilfe von virtueller Assistenz: Siri und Alexa schalten für uns das Licht ein, managen unsere Termine und beantworten emsig unsere Fragen. Führt das dazu, dass wir hinter ihnen nicht mehr einen maschinellen Algorithmus sehen, sondern eine wesenhafte Kommunikationspartnerin? Dieses Experiment hatte das Ziel herauszufinden, ob Menschen virtuelle AssistentInnen wie Siri oder Alexa stärker mit menschlichen Eigenschaften assoziieren als einen gewöhnlichen Computer. Hierzu wurde ein Impliziter Assoziationstest (IAT, Greenwald et al., 1998) durchgeführt, mit dem Einstellungen untersucht werden können. Die ProbandInnen hatten dabei die Aufgabe, Wortstimuli zu vier verschiedenen Kategorien zuzuordnen. Da sie dafür nur zwei Tasten nutzen konnten, musste bei zwei Kategorien jeweils dieselbe Taste gedrückt werden. Es wurde die Hypothese aufgestellt, dass Versuchspersonen (VP) virtuelle AssistentInnen eher als menschlich wahrnehmen und Computer als nicht menschlich. In diesem Fall sollten die Reaktionszeiten kürzer sein, wenn die Stimuli der Kategorie „virtuelle Assistenz“ und „menschlich“ bzw. „Computer“ und „nicht menschlich“ (kompatible Bedingungen) der gleichen Taste zugeordnet sind, als wenn es sich um eine inkompatible Zuordnung handelt. Die zu überprüfende Annahme war demzufolge, dass es eine signifikante Mittelwertsdifferenz zwischen der kompatiblen und der inkompatiblen Bedingung sowie in den Fehlerraten gibt.

2. Methode

Stichprobe. Die Stichprobe bestand aus 77 Versuchspersonen (VP; Alter: 18-69 Jahre; 38 Frauen, 37 Männer, 2 divers; Muttersprache deutsch). VP waren Psychologiestudierende der FSU Jena sowie Personen, die über Prolific erhoben wurden, welche mit 0.25 Versuchspersonenstunden bzw. 1.50 GBP entlohnt wurden. Die Studie wurde präregistriert: <https://aspredicted.org/y45ep.pdf>

Design. Den VP wurde suggeriert, dass sie an einer Studie zur Wahrnehmung von virtuellen AssistentInnen teilnehmen. Die Kompatibilität der Zuordnung der Kategorien zu den Tasten wurde als within-Faktor blockweise manipuliert. Der Faktor hat die Ausprägung kompatibel („Virtuelle Assistentin“ und „menschlich“ auf derselben Taste sowie „Computer“ und „nicht menschlich“) und inkompatibel („Virtuelle Assistentin“ und „nicht menschlich“ auf derselben Taste sowie „Computer“ und „menschlich“). Die Blockreihenfolge (zuerst kompatibler Block/ inkompatibler Block) wurde ausbalanciert. Als abhängige Variable dienten die Reaktionszeit und Fehlerraten.

Ablauf. Als Material dienten Stimuli der Target-Kategorien „Virtuelle Assistentin“ (Siri, Alexa, interaktiv, sprachgesteuert) bzw. „Computer“ (Maschine, Rechner, binär, systematisch) und der Attribut-Kategorien „menschlich“ (GEFÜHLE, BEWUSSTSEIN, LEBENDIG, DENKEND) und „nicht menschlich“ (GEGENSTAND, OBJEKT, UNBELEBT, STATISCH). Die ersten zwei Blöcke dienten der Übung der Attribute und anschließend der Targets. Im anschließenden ersten kombinierten Block wurden je eine Target- und Attribut-Dimension mit der gleichen Taste kategorisiert. Darauf folgte eine Übung der Targets, aber diesmal mit vertauschter Zuordnung der Dimensionen zu den Tasten. Zuletzt folgte ein zweiter kombinierter Block. Insgesamt durchlief jede VP 128 Durchgänge. Jeder Trial begann mit Anzeige der Kategorien am rechten und linken Rand des Bildschirms für 250 ms. Darauf erschien der Stimulus in der Bildschirmmitte, welchen die VP durch Tastendruck kategorisierten. Bei falscher Klassifikation wurde negatives Feedback angezeigt.

3. Ergebnisse

Aufbereitung. Die finale Auswertung wurde mit einem Datensatz von 77 Personen durchgeführt. Aufgrund von Fehlern und/oder Ausreißern mussten 15,17% der Werte ausgeschlossen werden.

Analyse. Um die oben genannte Hypothese zu überprüfen, wurde mit den Reaktionszeit- und Fehlerdaten der kompatiblen und der inkompatiblen Bedingung ein gerichteter t -Test gerechnet. Dieser wurde jeweils signifikant für RT $t(76) = 8.67, p < .001, d = -0.99$ (siehe Abbildung 1) und für Fehlerraten: $t(76) = -5.86, p < .001, d = -0.67$. VP reagierten langsamer bzw. machten mehr Fehler im inkompatiblen Block. Die anfängliche Hypothese konnte demnach bestätigt werden, da die RT

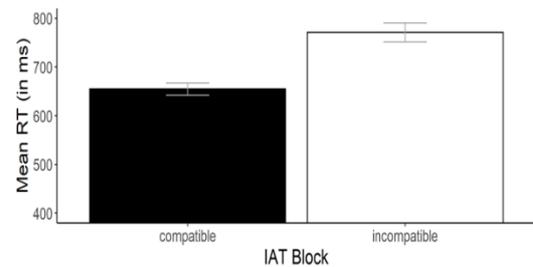


Abbildung 1

in der kompatiblen Bedingung kürzer ausfielen als in der inkompatiblen Bedingung. Abschließend wurde untersucht, ob es einen Einfluss von „Wechselkosten“ gab. Damit wird die Reduktion von RT bezeichnet, die auftritt, wenn VP in der kompatiblen Bedingung die Zuordnungen der Stimuli kognitiv vereinfachen („Rekodieren“). Bei einer Varianzanalyse ergab sich ein Interaktionseffekt der Faktoren IAT-Block und Aufgabe (Wechsel oder Wiederholung der Stimulusdimension zum Vorgängertrial), $F(1, 75) = 30.98, p < .001, \eta_p^2 = .29$ ($M_{\text{kompatibel}} = 26$ ms vs. $M_{\text{inkompatibel}} = 87$ ms). Diese Befunde sprechen für eine Vereinfachung der Aufgabe im kompatiblen Block durch die VP.

4. Diskussion

In unserer Studie konnte die Hypothese bestätigt werden, da die Reaktionszeiten in der kompatiblen Bedingung signifikant geringer waren als in der inkompatiblen Bedingung. Unsere Ergebnisse sprechen dafür, dass Virtuelle AssistentInnen menschlichere Eigenschaften zugeschrieben werden als Computern. Die zusätzliche Analyse der Wechselkosten legt nahe, dass dieser Effekt nicht durch Assoziationen entsteht, sondern durch Rekodierung im kompatiblen Block zustande kommt, durch die die Attributkategorien (menschlich/nicht menschlich) auch auf Stimuli der Targetkategorien angewendet wurde. Die Verwendung von Namen (Siri, Alexa) hat sich hier möglicherweise günstig ausgewirkt. Man kann spekulieren, ob diese Namen von Unternehmen wie Google und Apple so vorgegeben werden, weil sie wahrscheinlich genau das erreichen wollen.

Limitationen und Implikationen für weiterführende Forschung. Eine Verzerrung der Ergebnisse kann aufgrund der Vertrautheit der Begriffe entstehen, wie bei den Namen Siri und Alexa durch ihre Präsenz in der Werbung. Außerdem haben wir selbst die Stimuli festgelegt, nachdem wir unsere Hypothesen aufgestellt haben. Daher könnten wir die Auswahl implizit beeinflusst haben. In aufbauenden Studien sollte die Stimulenauswahl überarbeitet und verbessert werden, durch einen Test der impliziten Zuordnung vor dem eigentlichen IAT. Um Rekodierungseffekte zu verhindern, kann der IAT-Recoding-Free nach Rothermund et al. (2009) verwendet werden. Des Weiteren kann eine klare Unterscheidung der Kategorien Virtuelle AssistentIn und Computer schwierig sein, da Virtuelle AssistentInnen auf Computern programmiert sind. Aufgrund dieser Limitationen empfehlen wir weiterführende Forschung, welche unsere Implikationen beachtet.

5. Literatur

Greenwald AG, McGhee DE, Schwartz JL (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-80. doi: 10.1037//0022-3514.74.6.1464.

Rothermund K, Teige-Mocigemba S, Gast A, Wentura D (2009). Minimizing the influence of recoding in the Implicit Association Test: the Recoding-Free Implicit Association Test (IAT-RF). *Quarterly Journal of Experimental Psychology*, 62(1), 84-98. doi: 10.1080/17470210701822975.

Siri, sei meine Tutorin! Der Einfluss einer Virtuellen Assistenz auf den Abruf beobachtungsbasierter Stimulus-Reaktions-Bindungen

Lina Flindt, Anna Mieske, Sarah Mühr, Ann-Christin Winter, Jennifer Wolf

Leitung: Dr. Carina G. Giesen

1. Einleitung

Welchen Einfluss haben beobachtete Handlungen auf unser eigenes Handeln? In vorangegangenen Studien konnte gezeigt werden, dass beobachtete Reaktionen anderer Personen die eigene Handlungssteuerung beeinflussen können. Dieser Prozess wird durch den Abruf von *beobachtungsbasierten Stimulus-Reaktionen (SR)-Episoden* erklärt. Wird der Stimulus wiederholt, wird die Verknüpfung zur vorangegangenen Beobachtung abgerufen und beeinflusst dann die eigene Reaktion (Giesen et al., 2014). Bisherige Untersuchungen legen nahe, dass Abrufeffekte ausblieben, wenn die Teilnehmenden dachten, sie würden einen Computer bei einer Reaktion beobachten (Giesen & Rothermund, 2022). Befunde der Untersuchung unserer EMPRA-Gruppe zeigen jedoch, dass eine Verknüpfung von Stimulus und beobachteter Reaktion auch dann entsteht, wenn den Teilnehmer:innen suggeriert wird, mit einer virtuellen Assistenz (VA) zu interagieren (Winter et al., 2022). Diese Befunde legen nahe, dass VA menschenähnlicher wahrgenommen werden als Computer. Diesen Befund wollten wir replizieren: In der folgenden Studie wurde gezielt manipuliert, ob Teilnehmer:innen glauben mit einer VA vs. mit einem Computer zu interagieren und wie sich dies auf den Abruf beobachtungsbasierter SR-Episoden auswirkt.

2. Methode

Stichprobe. Die Studie wurde mit Prolific Academic erhoben. Daten von 82 Personen (43 Frauen; $M_{\text{Alter}} = 29.27$ Jahre; Muttersprache deutsch) wurden analysiert. Alle Versuchspersonen (VP) gaben Consent, wurden mit Geld (3.75£) entlohnt (Dauer 25 Minuten), und über die Studienziele aufgeklärt. *Ablauf.* Das Experiment ist eine interaktive Farbkategorisierungsaufgabe, die je nach Bedingung mit dem Computer oder einer VA („Interaktive Spiel-Assistentin, ISA) bearbeitet wird: VP wurden randomisiert einer Bedingung zugewiesen und bearbeiteten einen Übungsblock (32 Trials) und vier Experimentalblöcke (total: 128 Trials). Pro Durchgang erscheint ein Wort zentral in der Bildschirmmitte innerhalb eines weißen Quadrats. In der linken, oberen Ecke des Bildschirms der VP befindet sich ein roter Button und in der rechten, oberen Ecke ein grüner. Diese leuchten bei jeder Farbkategorisierung in der jeweils gewählten Farbe auf. Ist das Wort weiß, muss die VP die Reaktion des Interaktionspartners beobachten (Prime), die Reaktion des Interaktionspartners wird über das „Aufleuchten“ der Buttons angezeigt. Ist das Wort rot oder grün, muss die VP durch Tastendruck die Farbe klassifizieren (Probe). Nach 25% aller Probes erfolgt ein Gedächtnistest zur Abfrage der beobachteten Reaktion. Sind alle 128 Prime-Probe-Sequenzen der Studie absolviert, wird der Manipulationscheck durch abschließende Fragen überprüft, gefolgt vom Debriefing und der Entlohnung.

Design. Bei der Studie handelt es sich um ein 2x2x2 gemischt faktorielles Design mit dem between-subject-Faktor *Interaktionspartner* (ISA, n=45, oder Computer, n=37). Die within-subject-Faktoren sind Reaktionskompatibilität (geforderte Probereaktion ist kompatibel [50%] / inkompatibel [50%] mit beobachteter Primereaktion) und Stimulusrelation von Prime zu Probe (Wortwechsel [50%] / Wortwiederholung [50%]). Reaktionszeiten im Probe sind die abhängige Variable. Die Studie wurde auf [as.predicted.org](https://aspredicted.org) präregistriert (<https://aspredicted.org/qr9zx.pdf>).

3. Ergebnisse

Es werden nur Daten von VP ausgewertet, die sich im abschließenden Manipulationscheck: „Mit wem hast du zusammengearbeitet?“ korrekt an ihren Interaktionspartner erinnerten. Dadurch werden 12 VP aus der Analyse ausgeschlossen (15%). Ausgeschlossen wurden Reaktionszeiten, die schneller als 200 ms oder langsamer als 1.5 IQA über dem 75. Perzentil der individuellen RT Verteilung waren (5.6%; Tukey, 1977) und fehlerhafte Reaktionen im Gedächtnistest (0.86%). Wir führten eine

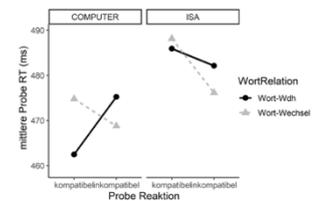


Abb. 2. Dreifach-Interaktion

2 (Reaktionskompatibilität: kompatibel vs. Inkompatibel) x 2 (Stimulusrelation: Wiederholung vs. Wechsel) x 2 (Interaktionspartner: Computer vs. ISA) ANOVA durch. Es zeigt sich mit $F(1,68) = 9.85, p = .003, \eta_p^2 = .13$ eine signifikante Interaktion zwischen der Stimulusrelation und Reaktionskompatibilität. Dies spricht für einen allgemeinen Abruf von beobachtungs-basierten SR Episoden. Die dreifach-Interaktion zwischen Interaktionspartner, Wortrelation und Reaktionsrelation ist mit $F(1,80) = 2.26, p = .138, \eta_p^2 = .03$ nicht signifikant. Dies spricht gegen unsere Hypothese, dass stärkere Abrufeffekte bei der Interaktion mit einer VA auftreten.

Explorative Analysen. Im Anschluss untersuchen wir, wie groß die einzelnen Abrufeffekte in den Interaktionspartnergruppen sind. Die Berechnung einer ANOVA zeigt, dass bei der „VA“-Bedingung die Stimulusrelation x Reaktionskompatibilität Interaktion mit $F(1,33) = 1.15, p = .291, \eta_p^2 = .03$ nicht signifikant ist. In der Computer-Bedingung ist die Stimulusrelation x Reaktionskompatibilität Interaktion mit $F(1,35) = 11.62, p = .002, \eta_p^2 = .25$ signifikant. Dieser Datentrend ist unserer Hypothese entgegengesetzt und widerspricht den Vorbefunden (vgl. Giesen & Rothermund, 2022).

4. Diskussion

Entgegen unserer Hypothese, stärkere Abrufeffekte in der Versuchsbedingung „VA“ als in der Bedingung „Computer“ zu finden, zeigte sich in dieser Untersuchung ein gegensätzlicher Datentrend. Dies widerspricht nicht nur unserer Hypothese, sondern auch den Befunden der Vorstudien (Giesen & Rothermund, 2022; Winter et al., 2022). Dieses Ergebnis lässt sich unserer Meinung nach damit erklären, dass in den Vorstudien bei den Studienteilnehmenden die Erwartung geweckt wurde, mit einer anderen Person zu interagieren und ihnen dann mitgeteilt wurde, dass sie aus Ermangelung anderer verfügbarer Teilnehmer:innen der Computer-Bedingung zugeteilt wurden. Möglich ist, dass der beobachtete Nulleffekt in der Computer-Bedingung durch diese Erwartungsverletzung ausgelöst wurde. In unserer Studie wurde angekündigt, dass interaktive Spielpartner in Form von VA untersucht würden – d.h. man erwartet von vorneherein, mit einem Computer zu interagieren – insofern wird hier auch keine Erwartung verletzt. In Folgestudien muss deshalb untersucht werden, an welcher Stelle a) in VP die Erwartung auf Interaktion mit einem Menschen geweckt wird (die dann frustriert wird) und durch b) eine Interaktion mit Computer vs. VA ersetzt wird. Weiterhin ist der dargestellte Kontrast (Bedingung: Computer vs. VA) weniger prägnant als zwischen „Mensch und Computer“ oder „Mensch und VA“. Dies könnte zu weniger ausgeprägten Effekten geführt haben.

5. Literatur

- Giesen, C. & Rothermund, K. (2022). Reluctance against the machine: Retrieval of observational stimulus–response episodes in online settings emerges when interacting with a human, but not with a computer partner. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-022-02058-4>
- Giesen, C., Herrmann, J. & Rothermund, K. (2014). Copying competitors? Interdependency modulates stimulus-based retrieval of observed responses. *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1978–1991. <https://doi.org/10.1037/a0037614>
- Winter, A., Krüger, S., Zimmer, C., Wolf, J., Mieske, A. (2022). Lernen von der künstlichen Intelligenz: Abruf beobachtbarer Reiz-Reaktions-Episoden einer KI als Maßstab eigener Handlung. *Unpublished manuscript, Friedrich-Schiller-Universität Jena, Germany*.

Wie unterscheiden sich Normen und Vorurteile zwischen rechts- und linksorientierten Personen?

Miriam Augustin, Carolin Beier, Sophie Pimpl, Sophie Schäfer, Verena Schneider

Leitung: Clemens Lindner

1. Einleitung

„Das größte Problem sind die Vorurteile der anderen!“ Solch ein Gedanke mag uns nicht unbekannt vorkommen, aber wie kommen wir und andere Menschen zu dieser Einschätzung? In der Sozialpsychologie werden Vorurteile als negative Bewertung einer Gruppe oder eines Individuums aufgrund dessen Gruppenzugehörigkeit definiert. Vorurteile können dabei als sozial geteilt verstanden werden (Crandall et al. 2002), wobei die Normativität hinsichtlich der Einschätzung einer Einstellung als Vorurteil innerhalb der Gruppe konsistent ist, sich zwischen Gruppenkontexten jedoch unterscheiden kann. Crandall et al. (2013) postulieren in diesem Zusammenhang das ‚Normative Fenster‘, in dem sich jene Vorurteilsgruppen befinden, bei denen kein Konsens besteht, ob negative Einstellungen als Vorurteile zu werten sind (z.B., Trans*-Personen). An den beiden äußeren Rändern des Fensters befinden sich einerseits Gruppen, gegenüber denen Konsens besteht, dass negative Einstellungen hochgradig inakzeptabel (‘true lows‘ z.B., Kindergärtner), bzw. akzeptabel (‘unsuppressed‘ z.B. Vergewaltiger) sind. Aber wie kann man das ‚normative Fenster‘ adressieren? Können wir empirische Befunde dafür finden, dass man Gruppen anderer normativer Kontexte mehr Vorurteile zuschreibt als sich selbst? Unterscheiden sich die der Fremdgruppe zugeschriebenen Vorurteile von denen, die die Fremdgruppe selbst berichtet? Welche Rolle spielen dabei prominente Prädiktoren für Vorurteile, wie Autoritarismus, Soziale Dominanz und politische Ideologie?

2. Methode

Diese Fragestellung haben wir mit einem Online-Fragebogen (Plattform: SoSci-Survey) mit vier Bedingungen between-subjects untersucht. Genauer wurden die Proband*innen nach Normen hinsichtlich der Akzeptanz negativer Äußerungen versus Vorurteile in der Eigengruppe versus der Fremdgruppe gefragt. Die Zuweisung zu den Bedingungen erfolgte randomisiert. Da das links-rechts Spektrum eine bedeutsame Differenzierung normativer Kontexte zu sein scheint (Brandt et al., 2014), wurde eine forced-choice links-rechts-Skala von 1 (links) bis 6 (rechts) zur Zuteilung zur Eigen- versus Fremdgruppenbedingung verwendet (bspw. linksorientierte VPN beantwortet rechte Vorurteile in Fremd-Bedingung). Der Aufbau des Fragebogens war in allen Bedingungen identisch: nach der Einverständniserklärung bearbeiteten die VPN die Soziodemografie mit der politischen Orientierung für die Zuteilung zum Gruppenkontext (s.o.). Anschließend füllten die Teilnehmenden etablierte Skalen bekannter Prädiktoren für Vorurteile aus, nämlich Autoritarismus (rechtsgerichteter Autoritarismus, RWA; nationaler Gruppenautoritarismus, NGA), soziale Dominanzorientierung (SDO), sowie eine politische Ideologieskala (Konservatismus, Ökonomischer Libertarismus, Liberalismus, Sozialismus). Schließlich wurde allen Teilnehmenden dieselbe Liste von 40 Targetgruppen präsentiert (gegenderte Präsentation balanciert between-subjects). Die Auswahl der Targetgruppen entstand aus Material von Crandall et al., 2002. Die Vorurteile wurden auf einem Gefühlsthermometer 0 kalt – 100 warm erfasst, die deskriptive Norm wurde als Akzeptanz negativer Einstellungen auf einer 5-stufigen Likert-Skala von 1 „überhaupt nicht akzeptiert“ bis 5 „vollkommen akzeptiert“ erfasst.

3. Ergebnisse

Insgesamt nahmen 121 Proband*innen an der Studie teil. Im Mittel waren die Proband*innen $M = 27.8$ Jahre alt ($SD = 12,9$), überwiegend weiblich identifiziert ($n = 86, 71\%$), gut ausgebildet ($n =$

111; 92% Abitur und höher) und links orientiert ($M = 2,43$; links [1-3] $n = 111$; 92%). Bei unseren Prädiktoren finden wir wie erwartet, positive Zusammenhänge zwischen RWA, NGA ($r = .53$) und SDO ($r_s = .44$; $=.23$). Autoritarismus (RWA, NGA) ist weiterhin stark mit Konservatismus konfundiert ($r_s = .63$; $=.42$). Liberalismus und Sozialismus sind hingegen negativ mit Autoritarismus und SDO korreliert. Hinsichtlich unserer Fragestellung finden wir, dass innerhalb der Kontexte die Eigengruppennorm hoch mit den Eigengruppenvorurteilen korreliert ($r = .87$). Ebenso sind zugeschriebene Fremdgruppennorm und Fremdgruppenvorurteile hoch korreliert ($r = .78$). Interessanterweise finden wir ebenso eine hohe Korrelation zwischen Eigengruppenvorurteilen und zugeschriebenen Fremdgruppenvorurteilen ($r = .74$), wohingegen Eigengruppennorm und Fremdgruppennorm nur moderat korreliert sind ($r = .42$). Betrachten wir die normativen Kontexte zwischen rechts und links finden wir auch hier hohe positive Korrelationen. Die Mittelwerte der einzelnen Vorurteilstargets zeigen eine breite Varianz im Ausdruck von Vorurteilen von ‚true lows‘ (z.B. Familienmenschen), über das ‚normative Fenster‘ (z.B. Soldaten) bis ‚unsupressed‘ Vorurteilen (z.B. Vergewaltiger). Um die Differenzielle Zuschreibung von Vorurteilen zu untersuchen, können wir die Differenz zwischen zugeschriebenen Fremdgruppenvorurteilen und Eigengruppenvorurteilen berechnen. Wir können zeigen, dass links-identifizierte Personen in unserer Stichprobe linke Vorurteilstargets im Mittel $M = 9.88$ ($SD = 12.1$) Skalenpunkte schlechter bewerten, als sie das bei rechten Personen erwarten würden. Umgekehrt werden bei rechten Targetgruppen den rechten Personen im Mittel eine $M = 17.60$ ($SD = 13.6$) Skalenpunkte schlechtere Bewertung der Gruppe zugeschrieben.

4. Diskussion

In den Korrelationen der Prädiktoren (SDO, RWA, NGA) mit den Bewertungen der Vorurteilstargets zeigten sich inkonsistent positive und negative Werte, wodurch sich diese nicht zur Vorhersage der Ausprägung von Vorurteilen eignen. Die hohe Korrelation zwischen Eigengruppenvorurteilen und Eigengruppennormen unterstützt die Annahme, dass Vorurteile als soziale Norm angesehen werden können. Die hohe Korrelation zwischen Eigengruppenvorurteilen und Fremdgruppenvorurteilen lässt vermuten, dass die Gruppen sich in ihren Vorurteilen gar nicht so stark unterscheiden, wie man es annehmen könnte. Durch die Differenz zwischen Eigengruppenvorurteilen und zugeschriebenen Vorurteilen in der linken Gruppe erkennt man, dass die linksorientierten Proband*innen bei einigen Targets, der Auffassung waren, dass es große Unterschiede in den Vorurteilen gibt. In unserer Studie wurden nur zehn Personen aufgrund ihrer Selbstzuschreibung auf der Rechts-Links-Skala als rechtsorientierte Menschen eingestuft. Die Personen, die zur rechten Gruppe gezählt wurden, gaben auch nur eine leichte rechte Neigung an. Die rechte Stichprobe ist zu klein, um Aussagen über die Frage zu geben, ob es signifikante Unterschiede zwischen den zugeschriebenen und den eigentlichen Vorurteilen in den Gruppen gibt. Die Ergebnisse unserer Studie lassen sich nicht auf die Allgemeinbevölkerung übertragen, da unsere Stichprobe überwiegend linksorientiert und weiblich ist und einen hohen Bildungsabschluss besitzt. Interessant für weiterführende Forschung wäre es herauszufinden, wie die Differenz zwischen zugeschriebenen und selbsteingeschätzten Vorurteilen zustande kommt und warum die Unterschiede zwischen den Gruppen von den Proband*innen hervorgehoben wurden.

5. Literatur

- Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science*, 23(1), 27–34. <https://doi.org/10.1177/0963721413510932>
- Crandall, C. & Eshleman, A. & O'Brien, L. (2002). Social Norms and the Expression and Suppression of Prejudice: The Struggle for Internalization. *Journal of personality and social psychology*. 82. 359-78. 10.1037//0022-3514.82.3.359.
- Crandall, C. & Ferguson, M. & Bahns, A. (2013). When we see prejudice: The normative window and social change. 10.4324/9780203567708.

I trusted you :` (... - Manipulation des Vertrauens durch Selbstoffenbarung im Repeated-Trust-Game

Laura Beuthien, Cedric Friedel, Pia Garcke, Ronja Uhlig, Tina Wenzel

Leitung: M.Sc. Jost Blasberg & M.Sc. Fabian Münch

1. Einleitung

Vertrauen ist eines der Grundgerüste sozialer Beziehungen. Fühlen wir uns Personen nahe, geben wir preis, was uns bewegt, wie wir denken und wie wir fühlen. Dieser dynamische Prozess des Informationsaustauschs wird Selbstoffenbarung genannt. Teilen wir negativ konnotierte Informationen, sprechen wir von negativer Selbstoffenbarung. Doch welche Auswirkungen hat negative Selbstoffenbarung auf das Vertrauen?

Vertrauen wird nach Rousseau et al. (1998) definiert als die Bereitschaft, „Risiken und Schwachstellen auf der Grundlage positiver Erwartungen in Bezug auf die Absichten oder Verhaltensweisen eines anderen zu akzeptieren“ (S.395). Besonders im Falle eines Interessenkonflikts ist Vertrauen von großer Bedeutung (Weiss et al., 2021). In diesem Kontext wird Vertrauen experimentell durch *Repeated-Trust-Games* gemessen (Berg et. al.,1995). Die Protagonisten des Spiels sind *Sender* und *Empfänger*. Während der Sender dem Empfänger Geld zusendet, kann dieser entscheiden, ob er das Geld aufteilen oder gänzlich für sich behalten will. Vertrauen wird hier durch die gesendete Menge des Senders operationalisiert. Das Spiel wird über mehrere Runden wiederholt, sodass man den Verlauf der Vertrauensentwicklung beobachten kann. Nachdem in der Mitte des Spiels ein Vertrauensbruch durch den Empfänger stattfindet, falls er den gesamten Betrag für sich behält, kann in den darauffolgenden Runden der Vertrauenswiederaufbau (*Vertrauensrecovery*) beobachtet werden (Haselhuhn et al.,2010). In dieser Studie soll untersucht werden, ob verschiedene Arten der Selbstoffenbarung das Vertrauensverhältnis zwischen den Spielenden beeinflussen. Insgesamt sollte Selbstoffenbarung zu mehr Vertrauen beim Gegenüber führen. Weil negative Selbstoffenbarung insbesondere im dyadischen Kontext zu mehr Vertrauen führen kann (Liu & Kang, 2017), erwarten wir außerdem, dass diese den stärksten Effekt auf die Vertrauensentwicklung und die Vertrauensrecovery im Vergleich zu positiver oder neutraler Selbstoffenbarung hat.

2. Methode

In unserer Online-Studie ließen wir 46 Proband:innen (72 % Frauen, 2% Divers, 18-31 Jahre) ein Repeated-Trust-Game spielen. Das Spiel beruhte auf dem Versenden (0-100 Sekunden pro Runde) und Empfangen von Versuchspersonen-Sekunden (VpS). Die Einteilung in Sender und Empfänger geschah scheinbar randomisiert, Proband:innen waren jedoch immer Sender. Das Trust-Game war in 7 Runden unterteilt. Runden konnten als gültig oder ungültig gewertet werden. In ungültigen Runden (Runde 1 und 4) konnten die Proband:innen Nachrichten verschicken und erhalten, jedoch konnte nicht auf die andere Nachricht reagiert werden. In gültigen Runden wurden die gesendeten VpS verdreifacht und vom Empfänger (Computer) in einem bestimmten Verhältnis zurückgeschickt. Den Versuchspersonen wurde durch Wartebildschirme und Terminvergabe suggeriert, dass sie gegen eine andere Person spielen. Nach der 1. Runde erfolgte über die Nachricht des Computers die Manipulation der Selbstoffenbarung (3 Gruppen: positiv, negativ, keine). In Runde 2 spielte der Computer fair (Vertrauensaufbau). In Runde 3 erfolgte der Vertrauensbruch, in welcher der Computer die gesamten VpS behielt. Daraufhin konnten in der ungültigen 4. Runde wieder Nachrichten gesendet werden. Runden 4 bis 7 dienten der Vertrauensrecovery.

3. Ergebnisse

In einer Repeated-Measures-Anova wurde ein signifikanter Effekt auf Vertrauen durch die Runden gefunden ($F(6, 258) = 6.139, p < 0.001$), allerdings konnte kein signifikanter Effekt auf Vertrauen für die unterschiedlichen Selbstoffenbarungsbedingungen festgestellt werden ($F(2,43) = 1.145, p = 0.328$). Auch die Interaktion zwischen Runden und Selbstoffenbarungsbedingungen war nicht statistisch bedeutsam ($F(12,258) = 1.205, p = 0.279$).

Zur weiteren Analyse der Kontraste nutzten wir lineare Modelle der einzelnen Runden. Diese Analysen wurden Bonferroni-korrigiert (adjustiertes $\alpha = .0125$). Das lineare Modell direkt nach der Selbstoffenbarung (Runde 2) zeigte keinen Unterschied von keiner und positiver Selbstoffenbarung zu negativer Selbstoffenbarung (Keine: $\beta = 14.82, t = 1.95, p = 0.058$, Pos: $\beta = 15.29, t = 1.91, p = 0.062$).

Das lineare Modell zur Untersuchung der Trust-Recovery zeigte einen marginalen Rundeneffekt ($\beta = 7.13, t = 2.22, p = 0.028$). Mit steigender Rundenzahl nahm die gesendete Menge an VpS daher wieder signifikant zu. Die Unterschiede in den Experimentalbedingungen waren in dieser Phase nicht bedeutsam (Keine: $\beta = 17.79, t = 0.72, p = 0.473$, Pos: $\beta = 13.95, t = 0.53, p = 0.591$).

4. Diskussion

In dieser Studie wurde untersucht, inwiefern sich Selbstoffenbarung allgemein und besonders negative Selbstoffenbarung auf das Vertrauen auswirkt. Insgesamt konnte keine unserer Hypothesen bestätigt werden. Negative, positive und keine Selbstoffenbarung führten zum gleichen Spielverhalten im Repeated-Trust-Game.

Zum einen könnte ein Grund für die nicht gefundenen Effekte sein, dass es sich um eine Online-Studie handelte und jenseits der unidirektionalen Nachrichten kaum Interaktion mit dem/der Gegenspieler:in möglich war. Chetty et al. (2021) weisen darauf hin, dass die Unterschiede in den Beträgen zumindest teilweise auf die Risikoeinstellung des/der Spielers/Spieler:in zurückzuführen sind und nicht unbedingt auf Unterschiede im Vertrauen. Bei einer Konfundierung mit der persönlichen Risikoeinstellung wäre so auch eine rationale Spieltaktik bei einigen Teilnehmer:innen möglich gewesen. Zudem muss es nicht nur Vorteile haben, seine Geheimnisse offen zu legen (Vrij et al., 2003). Die Offenlegung von negativen Geheimnissen kann auch zu negativen Reaktionen auf Seiten anderer führen (Afifi & Caughlin, 2006). So könnte negative Selbstoffenbarung auch Irritation zur Folge haben, weil eine fremde Person unaufgefordert etwas Negatives über sich preisgibt.

Abschließend lässt sich anhand der Ergebnisse der Studie sagen, dass weder Selbstoffenbarung noch speziell negative Selbstoffenbarung einen Effekt auf das Vertrauen und auf die Vertrauensrecovery im Repeated-Trust-Game haben.

5. Literatur

- Afifi, W. A. & Caughlin, J. P. (2006). A Close Look at Revealing Secrets and Some Consequences That Follow. *Communication Research, 33*(6), 467–488.
- Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior, 10*(1), 122–142.
- Chetty, R., Hofmeyr, A., Kincaid, H. & Monroe, B. (2021). The Trust Game Does Not (Only) Measure Trust: The Risk-Trust Confound Revisited. *Journal of Behavioral and Experimental Economics, 90*, 101520.
- Haselhuhn, M. P., Schweitzer, M. E. & Wood, A. M. (2010). How Implicit Beliefs Influence Trust Recovery. *Psychological Science, 21*(5), 645–648.
- Liu, B., & Kang, J. (2017). Publicness and directedness: Effects of social media affordances on attributions and social perceptions. *Computers in Human Behavior, 75*, 70-80.
- Vrij, A., Paterson, B., Nunkoosing, K., Soukara, S. & Oosterwegel, A. (2003). Perceived advantages and disadvantages of secrets disclosure. *Personality and Individual Differences, 35*(3), 593–602.
- Weiss, A., Michels, C., Burgmer, P., Mussweiler, T., Ockenfels, A. & Hofmann, W. (2021). Trust in everyday life. *Journal of Personality and Social Psychology, 121*(1), 95–114. h

Illusory Truth Effect - Liegt die Wahrheit im Ohr des Betrachters?

Julian Kunkel, Sabrina Bittel, Farida Gasimova, Gina Neff

Leitung: PD Dr. Romi Zäske

1. Einleitung

Wann schenken wir einer Aussage Glauben? Der *Illusory truth effect* (i.F. ITE) ist ein Phänomen, bei dem die Wiederholung einer Aussage deren wahrgenommene Glaubwürdigkeit erhöht [1]. Wir untersuchten den ITE mittels wahrheitsambiger Aussagen zum Allgemeinwissen, welche zuvor weder als eindeutig wahr noch falsch klassifiziert worden waren [2]. Um die Testpower im Vergleich zur Pilotstudie im Wintersemester 2021/22 zu erhöhen, untersuchten wir den ITE mittels eines reinen Innersubjektdesigns, bei dem unsere Probanden 24 Aussagen zweimal hinsichtlich Glaubwürdigkeit bewerteten. Die Hälfte der Aussagen wurde schriftlich bzw. auditiv präsentiert. Um eine bewusste Erinnerung an das erste Wahrheitsurteil zu minimieren, wurde zwischen den Wiederholungen zudem eine ‚filler task‘ eingeführt. Eine Hypothese, gestützt auf einen deskriptiven Befund nach Souza [3], war, dass schriftliche Stimuli leichter zu verarbeiten sind und daher der ITE in dieser Modalität größer ist im Vergleich zur auditiven. Dies wäre im Einklang mit Modellen, wonach der ITE auf einer erhöhten *fluency*, d.h. einer erleichterten Verarbeitung wiederholter Aussagen [4], basiert. V.a. mit Blick auf Anwendungsbereiche wie Werbung oder Politik ist wichtig herauszufinden, ob der ITE mit der Präsentationsmodalität interagiert. Diese Studie wurde auf aspredicted.org präregistriert.

2. Methoden

Teilnehmer. Auf Basis einer Effektstärke von $d = .39$ [5] ermittelten wir mittels G*Power [6] eine Stichprobengröße von 54. Zur gleichmäßigen Verteilung von Probanden auf acht Experimentversionen umfasste die finale Stichprobe 56 Probanden (11 Männer, 42 Frauen, 3 ohne Angabe, M Alter = 22,0, $SD = 6,9$ von 18 bis 70 Jahren). Drei Personen gaben Lese- oder Hörstörungen an, wurden aber mangels starker Einschränkung im Datensatz belassen. Die Teilnahme an der ca. 20-minütigen Studie war freiwillig und wurde bei Bedarf mit einer halben Versuchspersonenstunde vergütet.

Material, Vorgehen und Design. Für das Experiment wurden 24 wahrheitsambige Aussagen in deutscher Sprache verwendet [2]. Nach Zuschneiden mit der Software Praat [7] umfassten die auditiven Stimuli 48 einzelne Audiospuren von 24 Sprechern. Die 48 visuellen Stimuli umfassten die gleichen Aussagen und wurden in 24 Schriftarten, zentriert in weißer Schrift auf schwarzem Hintergrund, erstellt. Die Probanden durchliefen vier Phasen, in denen die Stimuli bewertet wurden: 1. Wahrheitsrating (sechsstufig *wahr* bis *falsch*) für 24 Aussagen (je zwölf auditiv/schriftlich), 2. Angenehmheitsrating (sechsstufig *unangenehm* bis *angenehm*) für Sprecher und Schriftarten von zwölf zusätzlichen Aussagen (Filler-Task), 3. Wahrheitsrating (Wdh. Phase 1), 4. visueller Wissenscheck, ob die Aussagen bereits vorher bekannt waren. Das Experiment wurde online mit PsyToolkit [8, 9] durchgeführt. Nach Zustimmung zur Teilnahme wurden die Probanden zufällig und gleichmäßig einer der acht Experimentversionen zugeordnet. Dies diente der Vermeidung von Positionseffekten durch Balancierung von Blockreihenfolgen (schriftlich, auditiv) sowie der Zuordnung verschiedener Sprecher und Schriftarten zu experimentellen Bedingungen.

3. Ergebnisse

Wir untersuchten die Wahrheitsratings in einer 2 (Modalität: visuell-auditiv) x 2 (Wiederholung: erste-zweite Präsentation) ANOVA im Innersubjektdesign. Der Haupteffekt von Modalität war nicht signifikant, $F(1,55) < 1$. Der Haupteffekt von Wiederholung (ITE) war nicht signifikant, $F(1,55) =$

1.426, $p = .237$, $\eta^2_p = .002$. Der Interaktionseffekt beider Faktoren war ebenfalls nicht signifikant, $F(1,55) = 1.028$, $p = .315$, $\eta^2_p < .001$. Explorativ untersuchten wir analog die Reaktionszeiten: Der Haupteffekt von Modalität war signifikant, $F(1,55) = 57.978$, $p < .001$, $\eta^2_p = .122$. Der Haupteffekt von Wiederholung (Priming-Effekt) war signifikant, $F(1,55) = 88.628$, $p < .001$, $\eta^2_p = .271$. Der Interaktionseffekt beider Faktoren war ebenfalls signifikant, $F(1,55) = 16.738$, $p_{\text{unkorr.}} < .001$, $\eta^2_p = .020$, mit größerem Priming-Effekt für die auditive Präsentation. Zwei t-Tests für abhängige Stichproben ergaben in beiden Modalitäten einen signifikanten Priming-Effekt (auditiv: $t(55) = 9.118$, $p_{\text{unkorr.}} < .001$; visuell: $t(55) = 7.326$, $p_{\text{unkorr.}} < .001$). Explorativ korrelierten wir den ITE (Differenz der mittleren Wahrheitsratings der ersten und zweiten Präsentation) mit den Angenehmheitsratings aus der Filler-Task und fanden einen Trend für eine positive Korrelation für die auditive Bedingung: $r(10) = .5546$, $p_{\text{unkorr.}} = .066$.

4. Diskussion

Trotz Anpassung der vorausgehenden Studie im Wintersemester konnten wir den ITE nicht replizieren. Zunächst könnte ein Grund dafür im Stimulusmaterial liegen: hierbei handelt es sich um nicht-alltagsrelevante Fakten, sodass möglicherweise aus mangelndem persönlichen Bezug keine hinreichend tiefe Verarbeitung stattfand. Es ist jedoch anzumerken, dass auch in früheren Studien der ITE mit nicht-alltagsrelevanten Stimuli nachgewiesen wurde [10]. Zudem ist nicht auszuschließen, dass trotz ablenkender Filler-Task Erinnerungseffekte an vorangegangene Ratings auftraten. In Folgestudien könnte man Erinnerungseffekte weiter reduzieren, z.B. durch eine längere Pause zwischen erstem und zweitem Wahrheitsrating oder durch eine höhere Anzahl von Aussagen. Wir empfehlen zudem die Untersuchung unter standardisierten Laborbedingungen, um Störvariablen, die mit Online-Testungen einhergehen können, zu minimieren. Die Manipulation der wiederholten Präsentation erwies sich insofern als erfolgreich, als dass wir in beiden Modalitäten Priming-Effekte und damit einen Hinweis auf erhöhte *fluency*, d.h. effizientere Verarbeitung wiederholter Stimuli, fanden [4]. Weiterhin zeigte sich ein Trend für eine Zunahme des ITE mit höherer Angenehmheit der Stimmen. Dieser mögliche Zusammenhang sollte in Folgestudien näher untersucht werden, um die Rolle von Stimmeigenschaften für die Glaubwürdigkeit von Aussagen besser zu verstehen.

5. Literatur

- [1] Hasher, L., Goldstein, D. & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)
- [2] Keller, S. (2020). Zweifelhafte Aussagen – Was die Stimme über die tatsächliche Überzeugtheit verrät. Bachelorarbeit, Friedrich-Schiller-Universität Jena, Deutschland.
- [3] Souza, A. L. E. (2012, Mai). *Modality-specific effects of processing fluency on cognitive judgments* (Dissertation). University of Texas at Austin. <http://hdl.handle.net/2152/ETD-UT-2012-05-5235>
- [4] Unkelbach, C. & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, 160, 110–126. <https://doi.org/10.1016/j.cognition.2016.12.016>
- [5] Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, 14(2), 238–257. <https://doi.org/10.1177/1088868309352251>
- [6] Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/brm.41.4.1149>
- [7] Boersma, P., Weenink, D. (1992–2022): Praat: doing phonetics by computer. (Version 6.2.09) [Software]. <https://www.praat.org>.
- [8] Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(2), 1096–1104.
- [9] Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31.
- [10] Fazio, L. K., Brashier, N. M., Payne, B. K. & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology. General*, 144(5), 993–1002. <https://doi.org/10.1037/xge0000098>

Eine Pilotstudie zur Replikation des Illusory Truth Effekts unter Einfluss der Modalität

Luna Mahler, Alina Meier, Carolin Knorre, Sophie Stockmann

Leitung: PD Dr. Romi Zäske

1. Einleitung

Täglich werden wir privat, in der Werbung oder in der Politik mit Aussagen und Botschaften konfrontiert, deren Wahrheitsgehalt wir bewusst oder unbewusst beurteilen. Bisherige Forschung beschreibt den robusten Effekt, dass Aussagen nach wiederholter Darbietung als wahrer eingestuft werden als neue Informationen [1]. Ein Erklärungsansatz für diesen Illusory Truth Effects (i.F. ITE) ist das fluency-cognition-model [2]: Danach beruhen Wahrheitsurteile maßgeblich auf fluency, also auf der erleichterten Verarbeitung bei Wiederholung. Der ITE entsteht demnach aus der Lernerfahrung, dass eine erleichterte Verarbeitung positiv mit Wahrheit korreliert ist. Vorherige Studien untersuchten den ITE meist in nur einer Modalität, schriftlich oder auditiv [3], [4]. In dieser Pilotstudie adressierten wir daher neben der Replikation des ITE erstmalig, ob der Effekt durch die Modalität der Stimuluspräsentation (schriftlich vs. auditiv) moderiert wird. Sollte fluency eine Rolle spielen, müsste diejenige Modalität einen größeren ITE bedingen, die „flüssiger“ zu verarbeiten ist. Nach Souzas [5] deskriptiven Ergebnissen scheinen schriftliche Stimuli einfacher zu verarbeiten sein. Für unsere Studie würde das bedeuten, dass der ITE in der schriftlichen Bedingung größer sein sollte.

2. Methode

Unter Annahme eines mittelstarken Effekts ($d = .39$) [4] ermittelten wir eine notwendige Stichprobengröße von $N = 98$ für den HE Wiederholung (ITE) [6]. Aus Zeitgründen und zum Einsparen von Ressourcen entschieden wir, das Design mit $N = 32$ Probanden aus unserem Bekanntenkreis zunächst zu pilotieren und die exakte Power a-posteriori zu bestimmen. In die Auswertung gingen damit insgesamt 17 Datensätze aus Gruppe 1 ein (8 weibl., 8 männl., 1 divers; Alter $M = 31,1$; $SD = 16,3$), sowie 15 Datensätze von Gruppe 2 (11 weibl., 4 männl., Alter $M = 22,4$; $SD = 3,5$). Weitere 8 Datensätze wurden ausgeschlossen wegen Unvollständigkeit. Wir nutzten 24 Aussagen zum Allgemeinwissen [7], deren Wahrheitsgehalt im Durchschnitt unbekannt ist (z.B. „Shakespeares „Hamlet“ wurde in die Star-Trek-Sprache Klingonisch übersetzt.“). Die Aussagen wurden sowohl als Audiostimuli vorbereitet, eingesprochen von je 2 verschiedenen Sprechern, sowie als schriftliche Stimuli in je 2 verschiedenen Schriftarten (zentriert weiß auf schwarz). Die insg. je 48 auditive und visuelle Stimuli unterteilten wir in vier Sets à 12 Aussagen (A1, A2, B1, B2), die wir in 8 Block-Kombinationen (Experimentversionen) balancierten, um Reihenfolgeeffekte zu kontrollieren. Unser 2x2 gemischtes Design bestand aus dem Gruppenfaktor „Wiederholung“ (einmalig vs. wiederholt) und dem Innersubjektfaktor „Modalität“ (schriftlich vs. auditiv). Jede Gruppe bewertete je 12 schriftliche und auditive Aussagen nach vermutetem Wahrheitsgehalt (4-stufige Skala). Die Gruppe mit wiederholter Präsentation nahm die Wahrheitsratings vor, nachdem alle Stimuli zunächst nach Interessanz bewertet worden waren. Das Online-Experiment wurde mit PsyToolkit durchgeführt [8], wobei die Probanden randomisiert einer Gruppe und Experimentversion zugeteilt wurden.

3. Ergebnisse

Die 2x2 ANOVA für mittlere Wahrheitsratings mit dem Zwischensubjektfaktor Wiederholung (einmalig vs. wiederholt) und Innersubjektfaktor Modalität (auditiv vs. visuell) ergab weder signifikante Haupteffekte für Modalität ($F(1, 30) < 1$, $p = .41$, $\eta^2 = .02$) und Wiederholung ($F(1, 30) = 1.632$, $p = .211$, $\eta^2 = .05$), noch eine signifikante Interaktion beider Faktoren ($F(1, 30) = 1.394$,

$p = .247$, $\eta^2p = .04$). Explorativ führten wir eine analoge ANOVA für die Reaktionszeiten durch. Diese ergab einen Priming-Effekt mit schnelleren Reaktionen auf wiederholte vs. einmalig präsentierte Items ($F(1, 30) = 22.229$, $p < .001$, $\eta^2p = .43$). Zudem ergab eine a-posteriori-Poweranalyse für den HE Modalität eine benötigte Stichprobengröße von $N = 100$, für den HE Wiederholung ein $N = 118$ und für die Interaktion $N = 50$.

4. Diskussion

Mit dem vorliegenden Stimulusmaterial konnten wir den ITE nicht replizieren, weder für schriftliche noch für auditive Aussagen. Ursächlich hierfür könnte die geringe Stichprobengröße bzw. das gewählte Versuchsdesign sein. Wir manipulierten die Wiederholungsbedingung zwischen Gruppen, um Erinnerungseffekte an die eigenen Wahrheitsratings zu vermeiden. Für Folgestudien empfehlen wir jedoch ein reines Innersubjekt-Design, um die Power zu erhöhen, zumal sich der Zwischensubjektfaktor „Wiederholung“ in der a-posteriori Poweranalyse sehr anfällig für die genaue Stichprobenszusammensetzung zeigt. Die online Datenerhebung erschien unter ökonomischen Gesichtspunkten zwar günstig, hat praktisch aber den Erhebungszeitraum verlängert und unvollständige Datensätze bei einigen Probanden verursacht. Da Störvariablen bei Online-Studien nicht vollständig kontrollierbar sind und daher Konfundierungen mit bspw. Umgebungs- oder tageszeitlichen Faktoren auftreten können, sollte die Studie unter kontrollierten Laborbedingungen wiederholt werden. Da der ITE nicht repliziert wurde, lässt sich keine fundierte Aussage über einen möglichen Modalitätseffekt treffen. Jedoch weisen unsere generellen Priming-Effekte auf eine effizientere Verarbeitung wiederholter Aussagen hin, was im Einklang mit dem fluency-cognition Model steht [2].

5. Literatur

- [1] Hasher, L., Goldstein, D. & Toppino, T. (1977). Frequency and the confidence of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)
- [2] Fazio, L. K., Brashier, N. M., Payne, B. K. & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of experimental psychology. General*, 144(5), 993–1002. <https://doi.org/10.1037/xge0000098>
- [3] Begg, I. M., Anas, A. & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121(4), 446–458. <https://doi.org/10.1037/0096-3445.121.4.446>
- [4] Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, 14(2), 238–257. <https://doi.org/10.1177/1088868309352251>
- [5] Souza, A. L. E. (2012, Mai). *Modality-specific effects of processing fluency on cognitive judgments* (Dissertation). University of Texas at Austin. <http://hdl.handle.net/2152/ETD-UT-2012-05-5235>
- [6] Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- [7] Keller, S. (2020). Zweifelhafte Aussagen – Was die Stimme über die tatsächliche Überzeugtheit verrät. (B.Sc.), Friedrich-Schiller-Universität Jena, Deutschland
- [8] Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31.

Escaping the Habit Hole: The Relationship between Habit Strength and Motivation Orientations in Breaking Habits

Friederike Kühr, M.L., Robin Lux, Jule Schubert, Sabiene Strömsdörfer

Leitung: Juhi Jayesh Parmar

1. Introduction

Research suggests that habit strength moderates the intention-behaviour relationship, indicating a moderation of weak habits - but not strong ones - by intentions (Wood & Neal, 2007). Additionally, it has been found that the motivation orientations of intentions - broadly approach or avoidance-based - further moderate behaviour change, with studies reporting mixed results regarding which orientation is the better predictor for breaking habits (Fuglestad et al., 2008; Gomez et al., 2013). Therefore, we wish to investigate how both habit strength and motivational orientations combine to influence breaking habits. Given the contradictory findings, we used two-tailed tests to test all possibilities. We chose the habit “*Instagram* use”, and created an intervention to support participants in achieving either approach (e. g. more time for friends and family) or avoidance goals (e. g. less social pressure or insomnia).

Our hypotheses are: If there is a difference between the two motivation orientations, and the approach-based orientation is a better predictor than avoidance-based, we expect a larger reduction of *Instagram* use during post-test – and vice versa. Additionally, we expect a main effect of habit strength, meaning participants with strong habits will show a smaller reduction in habit strength than participants with weak habits. Lastly, we expect a main effect of time, meaning participants – regardless of their habit strength and motivation orientation – will show a decrease in their *Instagram* use because of participating in our study.

2. Methods

Participants were tested online at two time-points seven days apart (pre and post intervention, “t1” & “t2”) using SoSci Survey. To measure subjective habit strength of *Instagram* use, we administered the Self Report Habit Index (SRHI) – a 12-point scale measuring behaviour frequency and automaticity (Verplanken & Orbel, 2003). For an objective measure, participants were asked to upload a screenshot of their daily average of *Instagram* use in minutes (“IU-mins”). At t1, participants filled out the SRHI, then uploaded a screenshot of their IU-mins. Next, participants were randomly assigned to a written motivation task (on a piece of paper) that would either be approach-based, avoidance-based, or the control group, which had to be put up somewhere in their daily environment (to ensure interaction with the orientation assigned).

After a seven day break, participants completed the SRHI again, and uploaded their IU-mins screenshot. Additionally, we administered the Regulatory Focus Questionnaire (Higgins et al., 2001) at t2, to explore how “orientation match” moderates habit change.

Thus, we used a mixed within-between-subjects design with three independent variables: habit strength (between: strong vs. weak), motivation orientation (between: control, promotion, prevention), and time (within: pre, post) resulting in a 2 x 3 x 2 design. However, due to non-compliance and possible misinterpretation of our instructions, data were not based on the assigned group, but on participants’ interpretations. Therefore, two changes were made post-hoc: 1) the control group was excluded, resulting in a 2 x 2 x 2 design. 2) instead of dichotomising “habit strength” based on pre-intervention SRHI scores as planned, we used participants’ pre-intervention IU-mins.

3. Results

After excluding participants who did not complete both tests or uploaded false data, we had a total sample size of $n = 36$. A 3-way mixed ANOVA produced a significant main effect of habit strength in IU-mins $F(1, 32) = 19.25, p < 0.0001, ges = 0.33$. Additionally, there was a significant interaction between habit strength and time $F(1,32) = 6.54, p = 0.016, ges = 0.03$. In the SRHI, the main effect of time was significant $F(1, 32) = 5.87, p = 0.01, ges = 0.03$. No other main effects or interactions were significant.

4. Discussion

Interpretation: Overall, participating in our study reduced subjective habit strength (SRHI) across conditions. Moreover, whereas there was a reduction of IU-mins for participants with strong habits between t1 and t2, participants with weak habits showed an increase. This might be due to a higher intrinsic motivation to change the habit, in the “strong group”, given the high negative association attached to high social media use. In the weak group, we noticed that some participants started using *Instagram* with zero IU-mins at pre-test, creating noise in our data. Additionally, participating in the study may have led to heightened awareness of the app, resulting in increased use for the “weak group”.

Limitations: Since “motivation orientation” was not significant in both DVs, it is either the case that both orientations were equally effective in promoting habit change, or that the manipulation was not effectively administered (misinterpreting instructions) or engaged with (no tracking during 7-day period). This led to dissolving the control group, allowing only speculations about the role of motivation orientation. Future studies should ensure instruction comprehension, motivation engagement, and add manipulation checks at t2. Further, we had to change our dichotomization criteria after data collection, making our analysis post-hoc. Additionally, our study was underpowered (Gpower: $n = 62$) which could explain why we did not find significant effects. Lastly, presumably participants dropped out at t2 when there was no reduction of *Instagram* usage, making the data skewed.

5. Literature

- Fuglestad, P. T., Rothman, A. J., & Jeffery, R. W. (2008). Getting there and hanging on: the effect of regulatory focus on performance in smoking and weight loss interventions. *Health Psychology, 27*(3S), S260. doi: [https://doi.org/10.1037/0278-6133.27.3\(Suppl.\).S260](https://doi.org/10.1037/0278-6133.27.3(Suppl.).S260)
- Gomez, P., Borges, A., & Pechmann, C. C. (2013). Avoiding poor health or approaching good health: Does it matter? The conceptualization, measurement, and consequences of health regulatory focus. *Journal of Consumer Psychology, 23*(4), 451-463. doi: <https://doi.org/10.1016/j.jcps.2013.02.001>
- Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology, 31*(1), 3-23. <http://dx.doi.org/10.1002/ejsp.27>
- Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology, 33*(6), 1313–1330. <https://doi.org/10.1111/j.1559-1816.2003.tb01951.x>
- Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological review, 114*(4), 843–863. <https://doi.org/10.1037/0033-295X.114.4.843>

Wie ein Ei dem anderen: Was heißt Vorurteile sind sozial geteilt?

Antonia Boddenberg, A.B., Tuba Korkmaz-Walther, Claudia Lehmann, Hauke Lickfett, Paulin Otto-Schneider

Leitung: Prof. Dr. Thomas Kessler

1. Einleitung

Woher stammen eigentlich unsere Vorurteile? Laut der Ideologie-Konflikt-Hypothese bewerte man solche Gruppen als negativ, deren Wertesystem als inkonsistent mit der eigenen Ideologie wahrgenommen wird (Chambers, Schlenker & Collisson 2013). Darüber hinaus seien Vertreter*innen beider Seiten des politischen Spektrums, also sowohl rechts-konservativ als auch links-liberal orientierte Personen, gleichermaßen vorurteilsbehaftet (Chambers et al. 2014).

Wir möchten untersuchen, ob Vorurteile gegenüber bestimmten Gruppierungen tatsächlich in der bloßen Ablehnung der ihr zugeschriebenen Weltanschauung begründet sind, oder darin, dass wir sie als fremd bzw. im Gegensatz zur Eigengruppe als weniger prototypisch für eine gemeinsame übergeordnete Kategorie betrachten. Das Eigengruppen-Projektions-Modell erklärt, dass wir eine Subgruppe umso positiver bewerten, je mehr diese mit den Eigenschaften einer umfassenderen Kategorie übereinstimmt. Dabei empfinden wir die Prototypikalität und somit auch Normativität der Eigengruppe im Vergleich zur referentiellen Kategorie i.d.R. höher als die der Fremdgruppe, was zu einer Abwertung dieser Gruppe führt (Mummendey, Kessler & Otten, 2009).

2. Methode

Die auf der Online-Plattform SoSciSurvey durchgeführte Studie war eine einfaktorielle between-subject Studie mit 3 randomisiert zugewiesenen Bedingungen: der Typikalität, dem Konservatismus und den Vorurteilen. Die Typikalität wurde von 46 Versuchspersonen auf einer Skala von 1 („untypisch“) bis 7 („typisch“) eingeschätzt, der Konservatismus von 43 Personen auf einer Skala von 1 („links-liberal“) bis 7 („rechts-konservativ“) und die Vorurteile von 45 Personen, indem erfragt wurde wie die Gefühle der Proband*innen gegenüber Menschen der verschiedenen Herkünfte sind von 1 („kalt-negativ“) bis 7 („warm-positiv“). Diese Bedingungen wurden für 55 soziale Gruppen eingeschätzt, darunter bekannte europäische Länder wie Frankreich, weniger bekannte wie San Marino und ethnische Minoritäten, z.B. die Samen.

Insgesamt nahmen 134 Versuchspersonen daran teil, davon sind 85 weiblich, 47 männlich und 2 divers mit einer Altersspanne von 18 bis 79 ($M = 36.7$, $SD = 15.53$). Alle Teilnehmenden sind deutsche Staatsbürger und Muttersprachler. Zusätzlich wurden die Teilnehmenden gebeten, ihre politische Orientierung auf einer Skala von 1 („links-liberal“) bis 101 („rechts-konservativ“) zu verordnen ($M = 34.33$). Daraus ergaben sich 98 Personen, die sich als politisch links-liberal (1-50) einschätzten und 36 Personen, die sich als politisch rechts-konservativ (51-101) einschätzten.

3. Ergebnisse

Die Typikalität korreliert um $r = .87$ mit den Vorurteilen, der Konservatismus mit den Vorurteilen um $r = .70$ und Typikalität und Konservatismus korrelieren um $r = .53$.

Weiterhin wurden vier Subsets erstellt, die Typikalitätseinschätzung der Links-Liberalen und der Rechts-Konservativen sowie die Konservatismusbeurteilung der Links-Liberalen und der Rechts-Konservativen. Diese Gruppenwerte wurden mit den Vorurteilen korreliert. Der Konservatismus korreliert in der Gruppe der Links-Liberalen um $r = .70$ mit den Vorurteilen, in der Gruppe der Rechts-Konservativen beträgt die Korrelation $r = .55$. Der Unterschied ist hierbei nicht signifikant ($z = 1.07$). Die Typikalität korreliert in der Gruppe der Links-Liberalen um $r = .85$ mit den Vorurteilen. In der Gruppe der Rechts-Konservativen beträgt die Korrelation $r = .87$.

Anschließend wurde die Tabelle der Vorurteilsbedingung so codiert, dass die Targets in den Zeilen stehen und die Teilnehmenden die Spalten bilden. So konnte die Item-Total-Korrelation für die einzelnen Proband*innen ermittelt werden, welche angibt wie gut eine Versuchsperson zu den Restlichen passt. Diese wurde für die 3 Bedingungen über die Versuchspersonen gemittelt, die mittlere Item-Total-Korrelation für die Typikalität beträgt $r = .60$, für den Konservatismus $r = .40$ und für die Vorurteile $r = .47$. Die mittlere Item-Total-Korrelation über die Typikalität, den Konservatismus und die Vorurteile hinweg beträgt $r = .47$.

4. Diskussion

Die hohen Item-Total-Korrelationen innerhalb der Bedingungen unterstützen die Hypothese, dass sich Personen in der Einschätzung verschiedener Gruppen sehr ähnlich sind. Auch innerhalb der politischen Lager sowie über alle Versuchspersonen hinweg sind die Item-Total-Korrelationen hoch. Dies zeigt, dass einzelne Personen im Mittel nicht von der Gesamtheit abweichen.

Grundsätzlich entsprechen unsere Ergebnisse nicht der Ideologie-Konflikt-Hypothese, da sich die ermittelten Vorurteile der beiden politischen Lager gegenüber Ländern, welche als konservativ eingeschätzt wurden, nicht signifikant voneinander unterscheiden. Hierbei ist jedoch zu erwähnen, dass sich eine Mehrheit der Versuchspersonen als links-liberal einordnete und somit ein Ungleichgewicht in der politischen Orientierung der Studienteilnehmer*innen besteht. Jedoch zeigen die Ergebnisse, dass interindividuelle Unterschiede in den persönlichen Erfahrungen mit bestimmten Gruppen kaum ins Gewicht fallen.

Die Theorie des Projektionsmodells konnte mithilfe der Korrelationen zwischen Typikalität und den Vorurteilen bestätigt werden. Je weniger typisch wir eine Gruppe also für eine umfassendere übergeordnete Kategorie ("Europa") empfinden, desto schlechter bewerten wir sie. Unsere Erhebung zeigt, dass wir ein geteiltes Verständnis über Gruppen besitzen und sie in ihrer Typikalität ähnlich einschätzen. Diese Annahme wird zusätzlich dadurch bekräftigt, dass die Typikalität der Gruppierungen, entsprechend des between-subject-designs, von anderen Versuchspersonen beurteilt wurde, als von denen, deren Vorurteile erhoben wurden. Dies bedeutet, dass die individuelle Bewertung einer Gruppierung auf Vorurteilen beruht, die einem gewissen Konsens entsprechen, welcher auf beiden Seiten des politischen Spektrums vorherrscht. Man könnte also annehmen, dass Vorurteile sozial geteilt sind.

5. Literatur

- Brandt, M., Reyna, C., Chambers, J., Crawford, J., & Wetherell, G. (2014). The Ideological-Conflict Hypothesis: Intolerance Among Both Liberals and Conservatives. *Current Directions in Psychological Science*, 20, 27-34.
- Chambers, J., Schlenker, B., & Collisson, B. (2013). Ideology and Prejudice: The Role of Value Conflicts. *Psychological Science OnlineFirst*, 23, 1-10.
- Mummendey, A., Kessler, T., & Otten, S. (2009). Sozialpsychologische Determinanten - Gruppenzugehörigkeit und soziale Kategorisierung. In A. Beelmann, & K. Jonas, *Diskriminierung und Toleranz - Psychologische Grundlagen und Anwendungsperspektiven* (1. Aufl., 43-60). Wiesbaden: VS Verlag für Sozialwissenschaften - GWV Fachverlage GmbH.

Smart or not? - Understanding the formation of context-dependent stereotypes

Berenike Buschermöhle, Maret Weimann, Lara Krickow, Lena Huss, Eliza Arzuman

Leitung: M. Sc. Tingting Huang

1. Introduction

Stereotypes as cognitive attitudes are undoubtedly necessary for us to deal with the high number of stimuli in our daily lives. According to textbook, stereotypes are formed as rigid and over generalized associations between social category and certain traits. However, more and more research suggest that stereotypes are context-dependent (Casper, Rothermund & Wentura, 2011). Taken them together, stereotypes can be highly robust and context-free or flexible and context-dependent. Research on the formation of attitudes found that the formation of context-dependent vs. context-free attitudes is determined by the heterogeneity of the previous experiences with the attitude object. (Rydell & Gawronski, 2009).

The present study aimed to examine whether the formation of context-dependent vs. context-independent stereotypes is influenced by the heterogeneity of the characteristics of the social category. We assume that newly formed stereotypes are context-free when the characteristics of the social category are homogenous across contexts whereas if the characteristics of the social category are heterogenous, the stereotypes should be context-dependent.

2. Methods

We had a total of 46 participants, $N = 46$ ($N_{\text{homo}} = 19$, $N_{\text{heter}} = 27$), 32 female, $M_{\text{age}} = 21.57 \pm 2.22$ years, who were randomly assigned to either a heterogenous or a homogenous condition.

In the first part of the study, participants were asked to form impressions about two unknown groups called “Fenian” and “Lerian” by reading behavioral descriptions of their members in two contexts, which were represented by two different background colors (i.e., turquoise and purple). In the heterogeneous condition, both groups behaved differently across the two contexts. In the homogeneous condition, both groups behaved the same across the two contexts.

To identify implicit stereotypes, we used the propositional evaluation paradigm (Müller & Rothermund, 2019). We displayed primes which varied in category and context. Also, we included a third background color (i.e., grey) to examine whether the stereotypes would transfer to the novel context. After presenting the prime, participants had to categorize a target (true or false) by pressing the according key. The reaction time was measured, and the endorsement effect indicating the smartness of the group was calculated ($RT_{\text{false}} - RT_{\text{true}}$). During the explicit testing, we again presented statements about the groups in the different contexts. Participants had to rate each statement on a scale from 1 (“do not agree”) to 9 (“totally agree”).

For both measures, we had the same 2Heterogeneity (homogenous vs. heterogenous) * 2Category (Lerian vs. Fenian) * 3Context (turquoise vs. purple vs. grey) design.

3. Results

To analyze the *implicit* data, we conducted a 2Heterogeneity (heterogeneous vs. homogenous) * 2Category (Lerian vs. Fenian) * 3Context (turquoise vs. purple vs. grey) repeated measures ANOVA of the endorsement effects. The most important finding was, that the three-way interaction of heterogeneity, category and context turned out to be significant, $F(2, 88) = 8.01$, $p < .01$, partial $\eta^2 = .16$. As expected, there was no context difference in the endorsement effects for both categories in homogeneous condition; for heterogeneous condition, we found context differences that for both categories, the endorsement effects for the smart-related context were stronger than the other two

contexts which did not differ from each other. The only exception was the Lerian category in the heterogeneous condition that there was a stronger endorsement effect in the grey context than in the smart-unrelated context, $t(44) = -3.86, p < .01$.

For the *explicit* ratings, we again ran a 2Heterogeneity (heterogeneous vs. homogenous) * 2Category (Lerian vs. Fenian) * 3Context (turquoise vs. purple vs. grey) repeated measures ANOVA. Most importantly, the three-way interaction of heterogeneity, category and context turned out to be significant, $F(2, 88) = 8.41, p < .01$, partial $\eta^2 = .16$. The pairwise contrasts between the contexts for each combination of heterogeneity and category yielded the pattern as predicted with one unexpected result that in the heterogeneous condition, the Fenian category was rated to be equally smart in the smart-related and in the grey context, $t(44) = 1.86, p > .05$.

4. Discussion

Taking a look at the *implicit results*, we can conclude, that in the homogenous condition our learning session was successful insofar as Lerians were perceived as smarter than Fenians in both contexts. This stereotype transfers to the novel context, confirming our hypothesis that stereotypes that are formed under homogenous conditions are context-independent. In the heterogeneous condition both groups were believed to be smarter in the smart-related contexts, indicating that our learning session was successful. We found no transfer effect of this stereotype to the novel context, which supports our hypothesis that stereotypes that are formed under heterogeneous conditions are context-dependent. The significant difference between smart-unrelated context and grey context for Lerians is not necessarily inconsistent with our expectations, since the effect was mainly driven by the unexpected negative effect in the smart-unrelated context.

Now focusing on the *explicit results*, our pattern in the homogenous condition again indicates that the learning session was successful. The acquired stereotype transfers to the novel context, which supports our hypothesis that stereotypes that are formed under homogenous conditions are context-independent. In the heterogeneous condition, Lerians were rated smarter in the smart-related context than in the smart-unrelated context. Since there was no transfer effect to the novel context, we can conclude that participants formed a context-dependent stereotype. Fenians were also rated smarter in the smart-related context than in the smart-unrelated context. In the pairwise comparison, however, they were rated equally smart in the grey context and in the smart-related context. This suggests a generalization rather than a contextualization effect across the contexts, which is inconsistent with our hypothesis. This inconsistency might be due to particularly high generalizability of the smart behaviors presented for Fenian group members in the smart-related context.

5. References

- Casper, C., Rothermund, K. & Wentura, D. (2011). The Activation of Specific Facets of Age Stereotypes Depends on Individuating Information. *Social Cognition*, 29(4), 393–414. <https://doi.org/10.1521/soco.2011.29.4.393>
- Müller, F. & Rothermund, K. (2019). The Propositional Evaluation Paradigm: Indirekt Assessment of Personal Beliefs and Attitudes. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02385>
- Rydell, R. J. & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion*, 23(6), 1118–1152. <https://doi.org/10.1080/02699930802355255>

Investigating the Formation of Contextualized vs. Generalized Stereotypes Using Propositional Evaluation Paradigm

Johanna Fechner, Elisabeth Hefner, Anja Monstadt, Antonia Reinhardt, Suzanne Van de Moosdijk

Leitung: M.Sc. Tingting Huang

1. Introduction

Psychology professors – smart, omniscient, intimidating. But what happens to that stereotype after a few canapés and glasses of champagne at the Empra-Congress? Will experiencing a professor in a convivial setting, possibly showing non-smart behavior, change the way students perceive him or her in next morning's course? Or will these students develop contextualized stereotypes, expecting their lecturer to show smart behavior inside and normal behavior outside the auditorium?

In the present study we aim to investigate the principles that guides the formation of *contextualized versus generalized* stereotypes. Based on research by Rydell and Gawronski (2009), our hypothesis is that the *heterogeneity of stereotypical associations* and the *salience of different contexts* (Gawronski et al., 2010) leads to the kind of stereotype formed. More precisely, we suggest that heterogeneous stereotypical associations lead to context-dependent stereotypes, if the context is made highly salient, whereas they become less context-dependent if the context salience is low. Homogeneous stereotypical associations lead to highly generalized stereotypes, if the context salience is low and stereotypes will generalize less, if the context is made highly salient. To test this proposal, we invented the novel group "Zinian", which we manipulated to be either smart in every context (homogeneous condition) or only smart in some contexts and average in the other contexts (heterogeneous condition). Afterwards we examined the newly learned stereotypes in both implicit and explicit measures.

2. Methods

The final sample of the present study consisted of 19 students (3 males, 16 females), $M_{\text{age}} = 22,11 \pm 3.52$ years old, age range 19-32 years. They were recruited via e-mails through the psychology mail distributor and rewarded with experimental hours. All participants were German native speakers with no uncorrected visual defects. One participant was excluded due to high error rate in the PEP task (more than 40% errors). Each participant was randomly assigned to one of the four conditions: homogeneous – high salience, homogeneous – low salience, heterogeneous – high salience and heterogeneous – low salience.

The first phase of the study was the impression formation paradigm in which participants were asked to form impression of the Zinian group by reading brief descriptions of behaviors of Zinian members against certain colored backgrounds which represent context. In the heterogeneous condition, Zinian act differently across different contexts while in the homogeneous condition, they always act smart. participants are presented with descriptions about Zinian in two contexts for high salience condition (i.e., turquoise and purple). For heterogeneous condition, Zinian are smart in the turquoise context and ordinary in the purple context while for homogenous condition, Zinian are smart in both contexts. To reduce the attention to context in low context salience conditions, participants are presented with descriptions about Zinian in four contexts (i.e., turquoise, purple, orange-red and yellow). Thus for heterogeneous condition, Zinian are smart in turquoise and orange-red contexts and ordinary in the other two contexts; For homogenous condition, Zinian are smart in all contexts.

The second phase of the experiment consisted of the experimental phase. A propositional evaluation paradigm (PEP, Müller & Rothermund, 2019) was used, it is an implicit measure to identify propositional beliefs and attitudes. Each PEP trial had the following temporal sequence of events: Firstly, the phrase (e.g., "In this context, Zinian are smart") was presented against a certain colored

background (i.e., turquoise, purple or grey). Afterwards, the target word “True” or “False” was presented and participants were asked to press one of the two keys (i.e., “D” or “L”) to indicate whether the target was “True” or “False”. In catch trials, a question “?? True or False ??” appeared after the prime presentation and participants had to indicate whether the statement presented before was true or false based on their own evaluation. The target word or the question remained on the screen until a response was registered.

After completing the PEP task, participants were asked in an explicit rating task to rate to which degree they agree with the stereotypic statements (i.e., “In this context, Zinian are smart”) being shown in one of the three contexts (i.e., turquoise, purple or grey). Ratings were given on a Likert scale ranging from 1 (totally disagree) to 9 (totally agree).

3. Results

Explicit Results

A 2 (context salience: high vs. low) x 2 (heterogeneity: homogeneous vs. heterogeneous) x 3 (context type: turquoise vs. purple vs. grey) repeated measures ANOVA was conducted with explicit ratings which revealed a significant two-way interaction effect between heterogeneity and context type, $F(2, 30) = 5.13, p < .05$, partial $\eta^2 = .25$. The pairwise comparison revealed no significant difference between the contexts in homogeneous condition. In the heterogeneous condition, the rating in the turquoise context was significantly higher than in the purple, $t(15) = 6.816, p < .001$, Cohen’s $d = 1.671$, and in the grey, $t(15) = 4.562, p = .001$, Cohen’s $d = 1.684$, context.

The three-way interaction between heterogeneity, salience and context did not reach significance $F(2,30) = .96, p > .05$.

Implicit Results

Erroneous responses and outlier values were excluded from the analyses. We calculated the endorsement effects as follows: $RT_{\text{false}} - RT_{\text{true}}$.

For the implicit task, a 2 (context salience: high vs. low) x 2 (heterogeneity: homogeneous vs. heterogeneous) x 3 (context type: turquoise vs. purple vs. grey) repeated measures ANOVA was performed. Neither the three-way interaction between heterogeneity, salience and context $F(2,30) = .57, p > .05$, nor any of the two-way interaction effects turned out to be significant.

4. Discussion

Due to the very limited sample size, all our results need to be interpreted with cautiousness. When it comes to the implicit results, no meaningful interpretation can be made. For the explicit results, non-conforming with our hypothesis, the context salience seems to have no impact on the formation of stereotypes. Interestingly, we found that even heterogeneity seems to have no influence. Although the difference fails to reach significance, in homogeneous condition the group was always rated smarter in the novel context than in the learned contexts. In the heterogeneous condition, the context which was associated with smart behavior in the learning session has significantly higher smartness ratings in the testing session. Altogether, this supports the hypothesis that stereotypes are formed to be highly context-sensitive.

5. References

- Gawronski, B., Rydell, R. J., Vervliet, B. & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of experimental psychology. General*, 139(4), 683–701. <https://doi.org/10.1037/a0020315>
- Müller, F. & Rothermund, K. (2019). The Propositional Evaluation Paradigm: Indirect Assessment of Personal Beliefs and Attitudes. *Frontiers in Psychology*, 10(2385). <https://doi.org/10.3389/fpsyg.2019.02385>
- Rydell, R. J. & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion*, 23(6), 1118–1152. <https://doi.org/10.1080/02699930802355255>

Du gehörst zu uns, deshalb bestrafen wir dich!

Lina Marie Detzner, Rebeka Edinger, Teresa Urban, Lara Wöhmann, Yunqi Yang

Leitung: M.Sc. Clemens Lindner

1. Einleitung

Vorurteile haben in Gruppenkontexten eine normative Funktion, zu diesem Ergebnis kamen Crandall et al. (2002). Sie zeigten, dass geäußerte Vorurteile eine direkte Funktion ihrer sozialen Akzeptanz sind. Anders gesagt: Menschen verleihen Vorurteilen bereitwillig Ausdruck, wenn diese sozial geteilt sind. Definiert sind sie dabei als negative Bewertung einer Gruppe oder einer Person, basierend auf deren Gruppenzugehörigkeit. Wie wird eine solche Gruppennorm aufrechterhalten? Die Theorie des Autoritarismus konstatiert, dass bestimmte Individuen unterwürfig gegenüber Autoritäten sind und Abweichler stärker bestrafen (Adorno et al., 1950). Altemeyer (1981) definierte Autoritarismus als Kovariation der Persönlichkeitsdimensionen Konventionalismus, autoritäre Aggression und Unterwerfung und entwickelte als Maß die Skala für rechtsgerichteten Autoritarismus (RWA). Duckitt (1989) zufolge liegt dieser Kovariation eine emotionale Identifikation mit einer sozialen Gruppe zugrunde. Auf dieser Konzeptualisierung basiert das Modell von Stellmacher und Petzel (2005): Sie definieren Gruppenautoritarismus (GA) als die individuelle Überzeugung von der angemessenen Beziehung, die zwischen Gruppen und ihren Mitgliedern bestehen sollte. Er ist Ausdruck der situationsspezifischen Aktivierung einer autoritären Disposition im Gruppenkontext. Mit der generischen GA-Skala der Autoren liegt im Gegensatz zur RWA-Skala ein ideologiefreies Autoritarismus-Maß vor. In gruppenspezifischen Skalen schwankt die ideologische Konfundierung mit dem jeweiligen Gruppenkontext (hier: linke oder rechte Gruppenzugehörigkeit). Kessler und Cohrs (2008) postulierten, dass Autoritarismus ein Prozess ist, der auf die Eigengruppe und deren Normen gerichtet ist. Folglich sollte er nur die Bestrafungsneigung bei Normabweichung in der Eigengruppe vorhersagen, nicht in der Fremdgruppe. Dies untersucht unsere Studie. Normabweichungen sind dabei im Einklang mit Crandall et al. (2002) als Kontakt-Intention zu einer abgewerteten Fremdgruppe operationalisiert.

2. Methode

Die Fragestellungen wurden mit einem Online-Fragebogen über SoSci-Survey untersucht. Es wurden drei verschiedene Maße von Autoritarismus erhoben: RWA (deutsche Version), linksgerichteter Autoritarismus (LWA, selbst übersetzt), sowie der Gruppenautoritarismus entsprechend der zuvor angegebenen politischen Orientierung. Anschließend lasen die Proband:innen vier Szenarien, in denen Normabweichung als Kontaktintention zur Fremdgruppe dargestellt wurde: In zwei Szenarien wichen Linke von der linken Vorurteilsnorm ab, in zwei wichen Rechte von der rechten Vorurteilsnorm ab. Die Szenarien wurden balanciert. Die normativen Kontexte waren: Polizei - linke Demonstrant:innen, Burschenschaften - linke Hochschulgruppe, patriotischer Heimatverein - LGBTQIA+-Initiative und Pegida - junge Journalist:innen. Jede Person las jeweils *eine* Abweichung je normativem Kontext. Nach dem Lesen eines Szenarios gaben die Proband:innen das Ausmaß ihrer moralischen Empörung, eine Beurteilung der Person und deren Forderung, den gewünschten Umgang mit der Situation sowie die Intensität der Durchsetzung des Umgangs.

3. Ergebnisse

88 Personen füllten den Fragebogen vollständig aus. Im Durchschnitt waren die Proband:innen 24 ($M=24$, $sd=7.71$) Jahre alt, haben sich vorwiegend weiblich identifiziert ($n=64$, 73%) und besaßen die Hochschulreife als höchsten Bildungsabschluss. Die meisten Proband:innen verorteten sich politisch links ($n=81$, 92%, Skala 1-6). Es wurden Korrelationen zwischen den unabhängigen

Variablen (RWA, LWA, LGA, RGA) und der politischen Orientierung getestet. Wie erwartet ist die RWA-Skala mit rechter Ideologie konfundiert ($r = .56$) und die RWA-Skala korreliert negativ mit der LWA-Skala ($r = -.21$, n.s.). Die LGA-Skala ist positiv mit der LWA- ($r = .13$, n.s.) und RGA-Skala korreliert. Die Korrelationen für die RGA-Skala sind aufgrund der wenigen Fallzahlen nicht aussagekräftig. Die moralische Empörung bei linker Normabweichung ($M = 2.58$) ist höher als bei Rechter ($M = 1.68$) und korreliert negativ mit der RWA- ($r = -.37$), positiv mit der LWA- ($r = .13$, n.s.) und nicht mit der LGA-Skala ($r = .01$, n.s.). Die moralische Empörung bei rechter Normabweichung korreliert positiv mit der RWA- ($r = .11$, n.s.), nicht mit der LWA- ($r = .005$, n.s.) und negativ mit der LGA-Skala ($r = -.12$, n.s.). Bei der Intensität der Bestrafung sind die Ausprägungen bei linker Normabweichung ($M = 3.07$) durchschnittlich höher als bei der rechten ($M = 2.04$) und korrelieren negativ mit der RWA- ($r = -.21$, n.s.) und nicht mit der LWA- ($r = -.04$, n.s.) und LGA-Skala ($r = -.03$, n.s.). Die Ausprägungen bei rechter Normabweichung korrelieren positiv mit der RWA- ($r = .22$), negativ mit LWA- ($r = -.11$, n.s.) und nicht mit der LGA-Skala ($r = -.06$, n.s.). Bei der Evaluation der Normabweichler:innen sind die Ausprägungen bei linker Normabweichung ($M = 2.55$) durchschnittlich niedriger als bei der rechten ($M = 2.72$). Die Evaluation der linken Normabweichler:innen korreliert positiv mit der RWA- ($r = .20$, n.s.), negativ mit der LWA- ($r = -.16$, n.s.) und nicht mit der LGA-Skala ($r = .004$, n.s.). Die Evaluation der rechten Normabweichler:innen korreliert nicht mit der RWA- ($r = -.02$, n.s.), negativ mit der LWA- ($r = -.11$, n.s.) und positiv mit der LGA-Skala ($r = .11$, n.s.). RWA- ($r = .11$, n.s.) korreliert positiv mit der Bestrafungsintensität bei rechter Normabweichung und LWA-Skala ($r = .10$, n.s.) und negativ mit der LGA-Skala ($r = -.28$).

4. Diskussion

Die Studie untersucht, ob die Kontaktintention zu einer Fremdgruppe als Normabweichung wahrgenommen wird und wie stark es bestraft wird. Wir sind davon ausgegangen, dass autoritäre Persönlichkeiten bzw. autoritäre Gruppenmitglieder diese Normabweichung *besonders* bestrafen. Wir finden in unserer linken Stichprobe eine höhere moralische Empörung bei linker Normabweichung. Insgesamt zeigt sich in unserer Studie, dass die Kontaktintention zu einer abgewerteten Fremdgruppe *nicht* als klare Verletzung der Gruppennorm wahrgenommen wird und entsprechend keine starke Bestrafungsneigung vorliegt. Trotzdem finden wir bei linker Normabweichung konsistent eine negative Korrelation mit RWA, was zeigt, dass auch die wenig rechts-autoritären stärker bestrafen und schlechter bewerten. Dieser Befund stellt eine Herausforderung des RWA dar, wonach vor allem Individuen mit der autoritären Persönlichkeit eine generell stärkere Bestrafungsneigung haben sollten. Insgesamt werden die Kontaktintentionen aber eher nicht bestraft. Trotzdem finden wir keine positive Vorhersage unserer Prädiktoren für die Bestrafung, weshalb wir unsere zentrale Hypothese nicht mit den vorliegenden Daten bestätigen können. Es könnte sein, dass die beschriebenen Szenarien von den Proband:innen nicht eindeutig als Eigen- oder Fremdgruppe wahrgenommen wurden. Zudem haben nicht genügend rechte Personen an der Studie teilgenommen, um aussagekräftige Effekte zu erzielen. Darüber hinaus können wir keine Aussage über die Validität der LWA-Skala im deutschsprachigen Raum treffen, da wir Die Skala für diese Studie erstmals in deutscher Sprache erhoben haben.

5. Literatur

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Altemeyer, B. (1981). *Right-wing Authoritarianism*. Winnipeg, Canada: University of Manitoba Press.
- Crandall, C. S., Eshleman, A. & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–378.
- Duckitt, J. (1989). Authoritarianism and group identification: A new view of an old construct. *Political Psychology*, 10, 63–84.
- Kessler, T., & Cohrs, J. C. (2008). The evolution of authoritarian processes: Fostering cooperation in large-scale groups. *Group Dynamics: Theory, Research, and Practice*, 12(1), 73–84.
- Stellmacher, J., & Petzel, T. (2005). Authoritarianism as a group phenomenon. *Political Psychology*, 26, 245–274.

Face to Face: Effekte von Ähnlichkeit und Sympathie auf Gesichtswiedererkennung

Jona John, Lina Kleemann, Isabel Kiefer, Laura Ramirez

Leitung: M.Sc. Hannah Klink

1. Einleitung

Im Rahmen des empirischen Forschungsseminars „Kognitive Neurowissenschaften“ haben wir uns mit dem Forschungsgegenstand „Gesichtsverarbeitung“ auseinandergesetzt. Neben der Wiedererkennung ist das Herauslesen von sozial relevanten Informationen aus Gesichtern ein alltäglich bedeutsamer Prozess. Nicht nur Eigenschaften von anderen (Attraktivität, Sympathie) werden bewertet, auch die Ähnlichkeit zur eigenen Person wird (unbewusst) wahrgenommen (Gilbert et al., 1995). Bisherige Studien haben einen Zusammenhang zwischen Attraktivität, der visuellen Aufmerksamkeit und der Wiedererkennung (Wiese et al., 2014) sowie zwischen Sympathie und visueller Aufmerksamkeit (Goller et al., 2019) feststellen können. Es scheint plausibel, dass mit erhöhter Aufmerksamkeit eine tiefere Verarbeitung und somit auch besseres Gedächtnis einhergehen. Gleichzeitig werden ähnlichere Personen als sympathischer wahrgenommen (Rausch et al., 2015). Bisher gab es keine veröffentlichten Studien, die sich mit wahrgenommener Ähnlichkeit von Gesichtern in Bezug auf die Wiedererkennungsleistungen beschäftigen. Semantische Kontexte haben einen Einfluss auf die Wahrnehmung und Bewertung von Gesichtern (Tskhay & Rule, 2015), weswegen wir diese benutzt haben, um Ähnlichkeit zu induzieren. Wir haben folgende Forschungsfragen entwickelt: Menschen, die wir als uns ähnlich einschätzen, erkennen wir besser wieder (H1) und Menschen die wir als uns ähnlich einschätzen, finden wir sympathischer (H2).

2. Methode

Stichprobe: Insgesamt wurden 48 Versuchspersonen (37 weiblich, 11 männlich) im Alter von 18 bis 69 Jahren, welche beide Messzeitpunkte abgeschlossen haben, rekrutiert. Das Versuchspersonenprofil besteht aus Psychologiestudierenden, Studierenden aus anderen Fächern, arbeitende Personen und Schüler:innen. Eine a-priori Poweranalyse ($\alpha = .05$, Power = .8) unter Einbezug der im Wintersemester durchgeführten Pilotstudie zum gleichen Thema gab eine Stichprobe von mindestens 59 Versuchspersonen vor. Aufgrund der begrenzten Erhebungszeit konnte diese Zahl nicht erreicht werden.

Materialien: Es wurden 32 computergenerierte, männliche und weibliche Gesichter von der Website „thispersondoesnotexist.com“ verwendet, die den Versuchspersonen im Rahmen der Studie präsentiert wurden. Davon waren 16 Gesichter Teil der Bewertung im ersten Messzeitpunkt und 16 Gesichter (angepasstes Alter und Geschlecht) dienten als unbekannte Gesichter in der Wiedererkennungsaufgabe. Um ein Ähnlichkeitsempfinden zwischen präsentiertem Gesicht und Versuchsperson zu erzeugen, wurden zusätzlich 8 semantische Kontexte generiert. Darin waren Informationen über den Namen, Hobbies, die aktuelle Tätigkeit und drei Eigenschaften der Big 5 (Extraversion, Gewissenhaftigkeit, Offenheit) enthalten. Jeder Kontext wurde einmal für ein weibliches Gesicht und einmal für ein männliches Gesicht verwendet.

Durchführung: Die Studie wurde online auf SoSciSurvey mit zwei Messzeitpunkten durchgeführt. Nach der Begrüßung der Versuchspersonen und Abfrage der Einwilligung wurden zunächst relevante Persönlichkeitsmerkmale der Versuchspersonen erfasst (3 Dimensionen der Big 5, s.o.; Danner et al., 2016). Anschließend wurden die bereits beschriebenen Fotos von Personen mit den Kontexten dargeboten, welche die Versuchspersonen aufmerksam lesen sollten. Danach erfolgte eine Bewertung der dargestellten Person hinsichtlich Sympathie, Vertrauenswürdigkeit und wahrgenommener Ähnlichkeit, womit der erste MZP beendet war. Zwei Tage später folgte die Wiedererkennungsaufgabe: Versuchspersonen sollten aus jeweils 8 präsentierten Gesichtern (4 neu)

die bereits bekannten Gesichter auswählen. Zuletzt wurden demografische Daten erhoben (Alter, Geschlecht, Tätigkeit, Hobbies) und die Versuchspersonen verabschiedet. Insgesamt dauerte die Teilnahme an der Studie 25 Minuten, Studierende der Universität Jena konnten eine halbe Versuchspersonenstunde erhalten.

3. Ergebnisse

Zur Analyse der Wiedererkennungslleistung wurde ein t -Test zwischen erfolgreich wiedererkannten und nicht wiedererkannten Gesichtern in Bezug auf die wahrgenommene Ähnlichkeit durchgeführt. Dabei wurden keine signifikanten t -Werte ($t > 2.011$) erreicht; alle t -Werte waren kleiner als 1. Die Analyse der zweiten Hypothese ergab einen schwach positive Korrelation zwischen der Bewertung der Sympathie und der subjektiv wahrgenommenen Ähnlichkeit ($r < .3$), allerdings wurde dieser Zusammenhang nur bei 5 Gesichtern signifikant ($p < .05$).

4. Diskussion

Das Ziel unserer Studie war es zu untersuchen, ob die Wiedererkennungslleistung eines fremden Gesichts mit der Ähnlichkeit zu sich selbst zusammenhängt. Unsere erste Hypothese konnten wir nicht bestätigen, aufgrund unserer Daten lässt sich kein statistisch signifikanter Zusammenhang zwischen der wahrgenommenen Ähnlichkeit und der Wiedererkennung finden. Ein schwacher visueller Trend in den Daten ist zwar vorhanden, könnte aber dem Zufall geschuldet sein. Die zweite Hypothese konnte mit einem schwachen positiven Zusammenhang bestätigt werden. Die Wahrnehmung der Ähnlichkeit geht mit einem verstärkten Gefühl der Sympathie einher. Aufgrund der Daten können wir darüber keine kausalen Aussagen machen; es ist jedoch anzunehmen, dass eine (unbewusst) wahrgenommene Ähnlichkeit zu positiveren Bewertungen gegenüber einer anderen Person führen kann (Rausch et al., 2015). Eventuell könnte die erste Hypothese mit einer größeren Stichprobe noch bestätigt werden, dieser Zusammenhang sollte in weiteren Studien untersucht werden. Eine weitere Möglichkeit ist, dass unsere Operationalisierung der Ähnlichkeit über demografische Daten und Persönlichkeitseigenschaften nicht den gewünschten Effekt bei den Versuchspersonen auslöste. Die präsentierten Informationen könnten zu wenig individuell gewesen sein, um Identifikation auszulösen. In zukünftigen Studien könnte mit einer Vorerhebung ein genaueres Persönlichkeitsprofil der Versuchspersonen erstellt und so in die Erhebung eingebunden werden. Insgesamt können wir so nur begrenzte Aussagen über die Effekte von Ähnlichkeit zwischen Personen machen.

5. Literatur

- Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C. J., & John, O. P. (2016). Die deutsche Version des Big Five Inventory 2 (BFI-2). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Gilbert, D.T.; Giesler, R.B. & Morris, K.A. (1995). When comparison arise. *J. Pers. Soc. Psychol.*, 69(2), 227.
- Goller, J., Mitrovic, A., & Leder, H. (2019). Effects of liking on visual attention in faces and paintings. *Acta Psychologica*, 197(2019), 115-123.
- Rausch, T.; Karing, C.; Dörfler, T. & Artelt C. (2015): Personality similarity between teachers and their students influences teacher judgement of student achievement, *Educational Psychology*, 2015, 1-16.
- Tskhay, K. & Rule, N. (2015). Semantic Information Influences Race Categorization From Faces. *Personality and Social Psychology Bulletin*, 41(6), 769-778.
- Wiese, H., Altmann, C., & Schweinberger, S. (2014). Effects of attractiveness on face memory separated from distinctiveness: Evidence from event-related brain potentials. *Neuropsychologia*, 56, 26-36.

Gesellschaftliche Normen als Fundament für Vorurteile

Felix Karl Konschak, Julia Ueberschär, Kim Nathalie Ciesla,
Natalie Zidek, Sinja Groß

Leitung: Prof. Dr. Thomas Kessler

Einleitung

Vorurteile gelten in unserer, wie auch in den meisten anderen Gesellschaften als große Herausforderung und Problem. Crandall (2002) definiert Vorurteile als (negative) Bewertung einer Gruppe oder eines Individuums aufgrund der Gruppenzugehörigkeit.

Nach Sherif (1948) basieren Vorurteile allgemeingesprochen auf intergrupuale Einstellungen und Gruppenprozesse. Um ein „gutes Gruppenmitglied“ zu sein müssen Individuen die Vorurteile der anderen übernehmen und sich an diese anpassen. Wer von den Vorurteilen der eigenen Gruppe abweicht gilt hingegen als deviant. Einzelne Gruppen unterscheiden sich jedoch nicht nur in ihren vorherrschenden Normen, sondern ebenfalls in deren Vorurteilen (Crandall, Eshleman & O'Brien, 2002). Der Zusammenhang zwischen Vorurteilen und Normen ist bisher nur wenig erforscht. In drei amerikanischen Studien an einer studentischen Stichprobe konnten enge Zusammenhänge zwischen beiden nachgewiesen werden (Crandall et al., 2002). Chialdini, Kallgren und Reno (2008) nehmen an, dass Normen Verhalten allgemein beeinflussen könne, um aber deren Wirkmechanismen genauer zu untersuchen, muss zwischen den deskriptiven und präskriptiven Normen unterschieden werden. Die *deskriptive Norm* beschreibt, wie sich die meisten Menschen in einer sozialen Gruppe für gewöhnlich verhalten. Sofern man sich daran orientiert, was alle anderen denken und tun, scheint man mit seiner Meinung oder Handlung auf der sicheren Seite zu sein.

Die *präskriptive Norm* wiederum beschreibt, was der Großteil der Gruppe für angemessen oder richtig empfindet. Normkonformes Verhalten und Denken wird typischerweise gelobt, wohingegen abweichendes Handeln und Denken in der Regel zu Ablehnung und Sanktionen führt (Chialdini, Kallgren & Reno, 2008).

In unserer Studie untersuchten wir den systematischen Zusammenhang von Vorurteilen und den zwei oben genannten Normentypen. Unser Ziel war es ein genaueres Verständnis davon zu entwickeln, ob Normen einen Einfluss auf Vorurteile haben. Dabei interessiert uns auch, inwiefern die Untersuchungsteilnehmer einander in ihrer Einschätzung von Vorurteilen, deskriptiven und präskriptiven Normen ähneln.

2. Methode

Unsere Studie bestand aus einem between-Subject-Design mit den drei randomisiert zugeteilten Bedingungen „Vorurteile“, „deskriptive Normen“ und „präskriptive Normen“. Diese wurden für je 55 Targets erhoben, die aus Nationen des europäischen Raumes bestehen.

Unsere Stichprobe bestand aus 74 Frauen (63,25%), 52 Männern (35,9%) und einer Person mit neutralem Geschlecht (0,85%) im Alter von 18 bis 80 Jahren, Mittelwert 35,39 ($SD=15,29$).

Es wurden auch demografische Daten (Alter, Geschlecht, Muttersprache und Staatsbürgerschaft), die politische Einstellung der Teilnehmenden sowie deren Identifikation mit Deutschland und Europa erhoben.

3. Ergebnisse

Es wurden in jeder Bedingung Mittelwerte für die Targets berechnet. Die Mittelwerte der deskriptiven und präskriptiven Normen korrelierten leicht negativ, aber nicht signifikant miteinander ($r(53) = -.23$, $p = .09$). Der Zusammenhang der Mittelwerte der deskriptiven Normen und der Vorurteile ist jedoch

stark ($r(53) = .93, p < .01$) Wohingegen die präskriptiven Normen nicht signifikant mit Vorurteilen korrelieren ($r(53) = -.25, p = .07$).

Um die Ähnlichkeit zwischen den Teilnehmenden zu analysieren, berechneten wir Korrelationen zwischen den Profilen der einzelnen Untersuchungsteilnehmer und dem Mittelwert der restlichen Stichprobe (Total-Item-Korrelation). Für die deskriptiven Normen ergibt sich eine hohe Korrelation der Individuen mit dem Rest der Stichprobe, ($r_{\text{ind,group}} = .71$), für Vorurteile eine mittlere ($r_{\text{ind,group}} = .52$), und bei den präskriptiven Normen hingegen lediglich eine geringe Korrelation ($r_{\text{ind,group}} = .11$).

4. Diskussion

Ziel unserer Studie war es, herauszufinden wie die beiden Normentypen (deskriptiv und präskriptiv) mit Vorurteilen zusammenhängen. Die Ergebnisse zeigen einen starken Zusammenhang zwischen Vorurteilen und deskriptiven Normen, weshalb wir davon ausgehen, dass das Verhalten anderer Gruppenmitglieder ausschlaggebend für die Vorurteile eines Individuums innerhalb dieser Gruppe ist. Des Weiteren konnten wir feststellen, dass sich die einzelnen Mitglieder hinsichtlich ihrer deskriptiven Normen stark ähneln, was sich in der hohen Korrelation ($r_{\text{ind,group}} = 0.71$) ausdrückt. Ebenso ähneln sich die einzelnen Individuen in der Vorurteilsbedingung.

Problematisch sind die präskriptiven Normen. Hier fanden wir eine sehr geringe Total-Item-Korrelation, die belegt, dass sich die Probanden in ihren Angaben nicht einig sind. Das könnte daran liegen, dass die Frage nach der präskriptiven Norm unklar gestellt war und zu unterschiedlichen Intuitionen der Untersuchungsteilnehmer geführt hat. So haben 11 von 35 Probanden bei allen Targets gleiche Angaben gemacht. Wir vermuten hier sozial erwünschtes Antwortverhalten, da angegeben wurde, dass das Ausdrücken negativer Gefühle gegenüber allen Nationen maximal verboten ist. In die Berechnung wurden deshalb nur 24 Personen einbezogen, was für verlässliche Berechnungen nicht ausreicht.

Dennoch lässt sich zeigen, dass Mitglieder einer bestimmten Gruppe sehr ähnliche Intuitionen gegenüber einzelnen Beobachtungsobjekten besitzen. Außerdem liegt ein gemeinsames Verständnis über deren allgemeiner Bewertung vor. Individuelle Prozesse, wie zum Beispiel die Persönlichkeitseigenschaften einer Person, können nur die Variation innerhalb eines Targets, nicht aber die unterschiedliche Bewertung über Targets hinweg erklären. Aus diesem Grund ist es wichtig, die vorherrschenden gesellschaftlichen Normen in die Betrachtung von Vorurteilen mit einzubeziehen.

In Anbetracht unserer geringen Stichprobengröße können die gefundenen Effekte jedoch nicht zuverlässig interpretiert werden. Für aussagekräftige Korrelationen bedarf es Untersuchungen mit mindestens 100 Teilnehmenden pro Bedingung. Aufgrund unserer Ergebnisse sind wir dennoch der Ansicht, dass eine weitere Erforschung beschriebener Zusammenhänge unabdingbar ist, um die Entstehung von Vorurteilen besser zu verstehen.

5. Literatur

- Cialdini, R.B., Kallgren, C.A., Reno, R.R. (2008). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. *Advances in Experimental Social Psychology*, 24, 201-234. [https://doi.org/10.1016/S0065-2601\(08\)60330-5](https://doi.org/10.1016/S0065-2601(08)60330-5)
- Crandall, C.S., Eshleman, A., O'Brien, L. (2002). Social Norms and the Expression and Suppression of Prejudice: The Struggle for Internalization. *Journal of personality and social psychology*, 82 (3), 359-78. DOI:10.1037//0022-3514.82.3.359
- Kessler, T. & Fritzsche, I. (2018). *Sozialpsychologie*. Springer.
- Sherif, M. (1948). The necessity of considering current issues as part and parcel of persistent major problems: Illustrated by the problem of prejudice. *International Journal of Opinion and Attitude Research*, 2 (1), 63– 68.

Programm

14.00 Uhr	Foyer	Eröffnung der Posterausstellung
14.00 – 16.00	Foyer	Präsentation der Poster
16.15 Uhr	HS 4	Festvortrag von Prof. Dr. Andreas Eder (Julius-Maximilians-Universität Würzburg) zum Thema: <i>„ Bored to death? Motivations underlying self-infliction of pain during thinking for pleasure“</i>
17.00 Uhr	HS 4	Auszeichnung der besten Poster
anschließend	Foyer	Ausklang der Veranstaltung

Thalia Buchhandlung

»Neue Mitte Jena«

Leutragraben 1

07743 Jena

Tel. 03641 4546-0

www.thalia.de

