



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Die Auswirkungen des Betrugsfalls „Diederik Stapel“
auf die Wissenschaft

Verfasserin

Lisa Meisner

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Ao. Univ.-Prof. Mag. DDr. Andreas Hergovich Bakk.

Danksagung

Ich bedanke mich ganz herzlich bei meinem Diplomarbeitsbetreuer Professor ao. Univ.-Prof. Mag. DDr. Andreas Hergovich, Bakk., der mir die Bearbeitung dieses Themas vorgeschlagen hat, welches mich aufgrund seiner Aktualität und spannenden Hintergrundgeschichte während meiner gesamten Recherche immer wieder aufs Neue überrascht und begeistert hat.

Ein ganz besonderer Dank gilt meiner Familie, allen voran meinen Eltern Hannes und Sabine, meinen kürzlich verstorbenen Großeltern Heinz und Inge und meiner Großmutter Veronika, ohne deren finanzieller Unterstützung mein Psychologiestudium nicht möglich gewesen wäre. Danke für eure bedingungslose Hilfsbereitschaft in allen Belangen, für euer Lob, eure Anerkennung und Ermutigung und euren jahrelangen emotionalen Halt. Ihr habt mir die Sicherheit, das Vertrauen und die Wurzeln geschenkt, aus denen meine Flügel wachsen konnten.

Danke, Michael, dass du mir immer mit Rat und Tat zur Seite gestanden hast. Du hast mich nicht nur während meiner Diplomarbeitszeit unterstützt, sondern mir mit deiner Liebe, deiner Geduld und deiner Ruhe während meines gesamten Studiums eine starke Schulter zum Anlehnen geboten.

Des Weiteren möchte ich mich bei all meinen Freunden bedanken, die mich immer wieder zu dringend benötigten Ruhepausen verleitet und mir dadurch viel neue Energie und Kraft geschenkt haben. Danke für eure motivierenden Worte und dass ihr immer für mich da seid.

Zuletzt möchte ich mich noch bei allen sozialpsychologischen Experten¹ bedanken, die bereitwillig an meiner Umfrage teilgenommen haben. Danke ebenfalls an alle Autoren, die mir auf so freundliche und hilfsbereite Art und Weise ihre auf der Universität Wien nicht frei verfügbaren Artikel zur Verfügung gestellt haben.

¹ In dieser Diplomarbeit wird, aufgrund der besseren Lesbarkeit, das generische Maskulinum verwendet. Obwohl das mögliche oder das tatsächlich natürliche Geschlecht nicht jedes Mal explizit angeführt wird, möchte ich ausdrücklich darauf hinweisen, dass – im Sinne des generischen Maskulinums – auch Personen weiblichen Geschlechts gleichermaßen mit gemeint sind.

**Erfolg beruht nicht auf der Anerkennung durch andere.
Er ist die Frucht dessen, was du mit Liebe gepflanzt hast.**

Ist die Zeit der Ernte gekommen, kannst du dir sagen:

„Ich habe es geschafft.“

Paulo Coelho

Eidesstattliche Erklärung

Ich erkläre an Eides Statt, dass ich die Diplomarbeit mit dem Titel „Die Auswirkungen des Betrugsfalls *Diederik Stapel* auf die Wissenschaft“ selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und alle den benutzten Quellen wörtlich oder sinngemäß entnommenen Stellen als solches kenntlich gemacht habe.

Zusammenfassung

Im November 2011 kam an der Universität Tilburg der Forschungsbetrug des niederländischen Sozialpsychologen Diederik A. Stapel ans Tageslicht. Ein eigens gegründetes Untersuchungskomitee stellte in mindestens 54 publizierten Artikeln Hinweise auf die Manipulation und Fabrikation von Daten fest. Durch eine Literaturrecherche wurden in der vorliegenden Diplomarbeit die Forschungsfragen beantwortet, welche Auswirkungen Stapels Betrug im Sinne verschwendeter Ressourcen durch den Aufbau auf seinen betrugshaften Erkenntnissen, auf die Rufschädigung der (sozial-) psychologischen Disziplin und auf das bestehende Wissenschaftssystem hat. Zusätzlich wurde eine Expertenumfrage mit der Fachgruppe für Sozialpsychologie der Deutschen Gesellschaft für Psychologie durchgeführt, um Reaktionen und Stellungnahmen von Experten zum Fall Stapel zu erheben.

Abstract

In November 2011, the research misconduct of the Dutch social psychologist, Diederik A. Stapel, was revealed at Tilburg University. A specially formed investigation committee confirmed the allegations of data manipulation and fabrication for at least 54 published articles. In the diploma thesis at hand, a literature review answered the research questions, namely which consequences Stapel's fraud has in the sense of wasted resources by building on his fraudulent findings, for the reputation of the discipline of social psychology and for the existing science system. In addition, an expert survey with the section of Social Psychology of the German Society for Psychology was conducted to collect reactions and opinions on the Stapel case.

Inhaltsverzeichnis

1	Einleitung	1
2	Biographie	2
3	Die Aufdeckung des Betrugs.....	4
4	Das Untersuchungskomitee	6
5	Definitionen.....	7
6	Die Auswirkungen des Betrugs	8
6.1	Forschungsfrage 1. Welche externen Auswirkungen hat der Betrugsfall Stapel nach den von Stroebe et. al (2012) genannten Typen?	9
6.1.1	Schädigung beruflicher Karrieren.....	9
6.1.2	Schäden für Patienten	9
6.1.3	Behinderung des wissenschaftlichen Fortschritts	9
6.1.3.1	Forschungsfrage 1.1. Wie oft wurden Diederik Stapels Publikationen zitiert?	10
6.1.3.2	Forschungsfrage 1.2.: Dienten Stapels manipulierte Ergebnisse als Grundlage für weitere Forschungsarbeiten beziehungsweise wurden durch Replikation Forschungsressourcen verschwendet?	13
6.1.4	Forschungsfrage 1.3.: Welche von Stapel verfassten Buchkapitel basieren auf den Erkenntnissen manipulierter Studien?	112
6.1.5	Forschungsfrage 1.4.: Welche (Lehr-) Bücher stützen ihre Erkenntnisse auf Stapels manipulierte Studien?	113
6.1.5.1	Inhaltliche Darstellung der Stapel Literatur in den (Lehr-) Büchern.....	116
6.1.6	Rufschädigung der (Sozial-) Psychologie	135
6.2	Forschungsfrage 2: Wie weit verbreitet ist wissenschaftlicher Betrug?	136
6.3	Forschungsfrage 3: Ist die Psychologie tatsächlich anfälliger für Betrug als andere Wissenschaften?	138
6.4	Forschungsfrage 4: Welche Rolle spielt die bestehende Forschungskultur der Sozialpsychologie im Betrugsfall Diederik Stapel?	140
6.4.1	Der Mythos der sich selbst korrigierenden Natur der Wissenschaft	144
6.4.1.1	Forschungsfrage 4.1.: Wie gut funktioniert der „Peer-Review“ Prozess um Betrugsfälle in der Praxis aufzudecken?	145
6.4.1.2	Forschungsfrage 4.2.: Welche Rolle spielt die Replikation in der Aufdeckung von Betrug in der Praxis?	147
7	Forschungsfrage 5: Welche Maßnahmen können getroffen werden, um das Risiko für wissenschaftlichen Betrug in Zukunft zu reduzieren?	150
7.1	Institutionelle Veränderungen.....	150
7.2	Konkrete Maßnahmen nach Van Lange et al. 2012	152
7.2.1	Niederschwellige Ethikabteilung.....	152
7.2.2	Datenspeicherung und –veröffentlichung.....	153
7.2.3	Namentliche Kennzeichnung zuständiger Bereiche.....	153
7.2.4	Kommunikation mit den Medien	153
7.2.5	Deskriptiver Statistik höhere Aufmerksamkeit schenken	154
7.2.6	Replikationsförderung	154
7.3	Präventive Maßnahmen nach Stroebe et al. (2012)	155
7.3.1	Belohnungen	155

7.3.2	Kosten	156
7.3.3	Erhöhung der Entdeckungswahrscheinlichkeit	157
7.4	Alternative Maßnahmen	158
7.4.1	„Archives of Scientific Policy“	158
7.4.2	Statistische Analyse	158
7.4.3	Replikation.....	159
7.4.4	Sicherung guter wissenschaftlicher Praxis.....	159
8	Umfrage	160
8.1	Ergebnisse.....	160
8.2	Kritik an der Umfrage	178
9	Diskussion	179
9.1	Behinderung des wissenschaftlichen Fortschritts	179
9.1.1	Verschwendung von finanziellen Ressourcen	179
9.1.2	Verschwendung von zeitlichen Ressourcen	179
9.1.3	Persönlicher Schaden	181
9.1.4	Rufschädigung.....	181
10	Literaturverzeichnis	184
11	Tabellenverzeichnis.....	188
12	Abbildungsverzeichnis.....	189
13	Anhang	190
13.1	Anhang 1: (Lehr-) Bücher Sozialpsychologie.....	190
13.2	Anhang 2: E-Mail Margot Custers (2013).....	193
13.3	Anhang 3: Fragebogen – Umfrage: Deutsche Gesellschaft für Psychologie.....	195
14	Lebenslauf	199

1 Einleitung

Der Betrugsfall Diederik Stapel, der im November 2011 an der Universität Tilburg ans Tageslicht kam, traf die Welt der Wissenschaft und vor allem die Sozialpsychologie wie einen Schlag ins Gesicht und schädigte sowohl ihr Image in den Medien als auch ihr kollektives Selbstwertgefühl massiv. Stapel veröffentlichte im Laufe seiner Karriere insgesamt 137 Publikationen in den renommiertesten Fachzeitschriften mit hohem Impact Faktor, wobei sich herausstellte, dass er bei mindestens 54 Artikeln, in denen er Erst- oder Co-Autor war, manipuliertes Datenmaterial verwendete.

Der Schaden für das Feld der Sozialpsychologie, vor allem für ihren Ruf, ist immens hoch. Stapel wurde in den Niederlanden jahrelang als akademischer Star gefeiert und gerade deswegen aufgrund seines Skandals von der internationalen Presse zerrissen. Er zierte monatelang die Schlagzeilen in Zeitungen, Internet-Foren, Blogs und Fachzeitschriften und gab Kritikern genug Anlass, um die zugrundeliegende wissenschaftliche Disziplin, die Psychologie, und deren Forschungsmethoden in Frage zu stellen.

Auf das bestehende Kreuzfeuer der Kritik an der (Sozial-) Psychologie als Wissenschaft und ihren Methoden antworteten betroffene Wissenschaftler einerseits mit Erklärungsversuchen, wie es dazu kommen konnte, dass ein solch immenses Ausmaß an wissenschaftlichem Fehlverhalten jahrelang unentdeckt blieb und welche Rolle dabei das bestehende Wissenschaftssystem spielt, andererseits gibt es massive Bestrebungen, Maßnahmen zu entwickeln und zu ergreifen, um Betrug in der Wissenschaft zukünftig zu reduzieren.

In der folgenden Diplomarbeit werden die direkten Auswirkungen des konkreten Betrugsskandals um Diederik Stapel für das Feld der Sozialpsychologie und in Folge die längerfristigen Konsequenzen für das gesamte System der Wissenschaft erarbeitet und dabei auftretende Forschungsfragen einerseits durch eine systematische Literaturrecherche und andererseits durch Stellungnahmen eines sozialpsychologischen Expertenkreises im Zuge einer Umfrage beantwortet.

Bevor auf diese Fragen eingegangen wird, sollen zunächst die Umstände des Betrugsfalls dargestellt werden, um die daraus resultierenden Konsequenzen besser nachvollziehen zu können.

2 Biographie

Diederik Alexander Stapel wurde am 19.10.1966 in Oegstgeest, in den westlichen Niederlanden als jüngstes von vier Kindern geboren. Sein Vater war Bauingenieur und arbeitete als Senior Manager am Schiphol Flughafen. Er habe ihm bereits als Kind die Hingabe zu seinem Beruf und die Ansicht vermittelt, dass sich jeder Mensch über die erzielten Erfolge im Berufsleben definiert (Bhattacharjee, 2013). Diederik sei ein exzellenter Schüler und Sportler gewesen, in seiner Freizeit schrieb er Theaterstücke und agierte als Schauspieler. Dabei lernte er auch seine zukünftige Ehefrau Marcelle kennen, die er 1997 heiratete. Nach seiner High School Zeit begann Stapel an der Universität in Pennsylvania sogar kurz ein Schauspielstudium, das er wenig später jedoch abbrach, um daraufhin in die Niederlande zurückzukehren um mit seinem Psychologie Studium zu beginnen. 1991 schloss er an der Universität Amsterdam die beiden Studiengänge Psychologie und Kommunikationswissenschaften „cum laude“ ab und unterhielt danach Forschungsk Kooperationen mit der Universität Chicago (1991-1992) und der Universität Michigan (1996), wo er zeitweilig als Berater tätig war. Er promovierte 1997, ebenfalls mit Auszeichnung, im Fach Sozialpsychologie an der Universität von Amsterdam und bekam für seine Doktorarbeit, unter der Betreuung von Willem Koomen, den ASPO (*Association for Social Psychologists*) Dissertationspreis verliehen. 2002 wurde er selbst zum Präsidenten der ASPO ernannt und hielt dieses Amt fünf Jahre lang inne. Ab 1998 wurde seine Forschung durch ein Stipendium der *Royal Netherlands Academy of Arts and Sciences* (KNAW) unterstützt, das ihm die Möglichkeit gab, fünf Jahre lang unabhängige Forschungsprojekte zu leiten. 1999 wurde Stapel mit dem *Jos Jaspers Award* von der *European Association of Experimental Social Psychology* (EAESP) ausgezeichnet. Im Jahr 2000 wurde Diederik die Professur für den „Van der Leeuw“ Lehrstuhl für das Fach der *Cognitive Social Psychology* an der Universität Groningen angeboten. Er nahm das Angebot an und erhielt zwei Jahre später Forschungsfördermittel aus dem „Pionier-Fond“ der *Netherlands Organization for Scientific Research* (NWO). Obwohl er gerne an der Universität Groningen arbeitete, wechselte er seinen Arbeitsplatz im September 2006 zur Tilburg Universität, wo er das *Tilburg Institute for Behavioral Economics Research* (TIBER) gründete. Stapel wurde 2009 der *Career Trajectory Award* von der *Society of Experimental Social Psychology* verliehen und 2010 wurde er zum Dekan der *Tilburg School for Social and Behavioral Sciences* ernannt, welches insgesamt neun Institute umfasst, unter anderem auch eines für Sozialpsychologie (Newsletter Studenten, 2010).

Im September 2011 wurde Diederik Stapel vom Rektor der Universität Tilburg, Prof. P. Eijlander, gekündigt, nachdem ihm drei seiner Studenten der Manipulation und Fabrikation von Datenmaterialien beschuldigt hatten. Prof. P. Eijlander bildete daraufhin

ein Untersuchungskomitee, welches dem Ausmaß des Betrugs und der Art und Weise der Verletzung der wissenschaftlichen Integrität Stapels nachgehen sollte.

3 Die Aufdeckung des Betrugs

Gerüchte um Verletzungen der wissenschaftlichen Integrität verfolgten Stapel bereits von Groningen nach Tilburg, doch der Verdacht einzelner Personen weckte nie genug Aufmerksamkeit, um eine Untersuchung einzuleiten. Ein Forschungskollege der Tilburg Universität, Ad Vingerhoets, fragte Stapel beispielsweise, ob er ihm bei einer Idee zu einer Studie helfen wolle, in der es darum gehen sollte, dass Menschen mehr Empathie zeigen, wenn sie eine andere Person weinen sehen. Stapel war begeistert von dieser Forschungsidee und bot seinem Kollegen an, für ihn die Daten an Grundschulen zu sammeln, da er dafür seine Kontakte einsetzen könnte. Die Versuchsgruppe der Kinder sollte einen Comic ausmalen, auf dem ein weinendes Kind zu sehen war, die Kontrollgruppe dagegen einen ausdruckslosen Comic. Unmittelbar danach hätten die Kinder Süßigkeiten erhalten und entscheiden sollen, wie viel sie bereit sind, von ihren Süßigkeiten mit anderen Kindern zu teilen – eine Messung des prosozialen Verhaltens. Die beiden Forscher bereiteten gemeinsam mit einem Assistenten die Comics und Fragebögen vor und einige Wochen später meldete sich Stapel bei Ad Vingerhoets, dass er bereits alle Daten erfolgreich gesammelt hätte. Ad Vingerhoets war begeistert, als er sah, dass die Analysen einen signifikanten Unterschied zwischen den beiden Versuchsgruppen zeigten und Kinder, die den Empathie-Comic ausmalten tatsächlich mehrere Süßigkeiten verteilten als die Kinder in der Kontrollgruppe. Aufgrund der hoch signifikanten Effekte waren sich die beiden sicher, ihr Experiment in einem Journal mit hohem Impact-Faktor publizieren zu können. Als Vingerhoets mit der Verschriftlichung des Manuskripts begann, kam ihm die Frage in den Sinn, ob es bei ihrem Experiment nicht vielleicht auch Geschlechterunterschiede bezüglich des gezeigten prosozialen Verhaltens gäbe. Er fragte Stapel nach den Daten, um weitere Analysen durchführen zu können und dieser antwortete ihm, dass sie noch nicht in den Computer eingegeben worden wären. Vingerhoets war geschockt, da Stapel ihm zuvor Mittelwerte, Standardabweichungen und die Höhe der Reliabilität des Fragebogens präsentiert hatte, für deren Berechnungen er mit Sicherheit einen eingegebenen Datensatz benötigt hätte. In seiner Unsicherheit, ob er den Dekan der Universität tatsächlich des Betrugs beschuldigen sollte, wandte sich Vingerhoets an einen älteren Kollegen im Ruhestand, der ihm davon abriet, seine Vermutung öffentlich zu machen. Er folgte dieser Empfehlung und entschied, den Vorfall dem Rektor nicht zu melden (Bhattacharjee, 2013).

Im Sommer 2009 schöpften drei seiner Studenten erstmals Verdacht, als sie ein Manuskript Stapels lasen, bei dessen Stichprobe es sich angeblich um Schulkinder handeln sollte, deren durchschnittliches Alter bei 19 Jahren lag, was für die bedachte Zielpopulation eindeutig zu hoch war. Die drei Studenten baten Stapel um eine Erklärung,

woraufhin er entgegnete, dass es sich dabei um einen Fehler handle. Die Studenten nahmen seine Antwort als einmaligen Fehler hin und dachten nicht weiter über die Sache nach, bis es kurz darauf zu einem weiteren Meeting des Forschungsteams kam, in dem Stapel die Ergebnisse einer neuen Studie präsentierte, wobei die gewonnenen Daten seinen Hypothesen so perfekt entsprachen, dass ein Mitglied der Zuhörerschaft scherzend sagte: „It is as if he made up these data himself“ (Keulemans, 2012a, zitiert nach Stroebe, Postmes & Spears, 2012).

Daraufhin begannen seine Diplomanden die rohen Datenmasken Stapels zu untersuchen, die ihnen für diverse Forschungsprojekte zur Verfügung gestellt worden waren, und fanden in ihren Analysen verdächtig hohe Effektgrößen. Am auffälligsten war jedoch, dass Stapel eine Reihe von Daten einer Studie gleichzeitig für eine andere Studie verwendet hatte.

Einer der Studenten wandte sich mit seinem Verdacht an einen jungen Professor der Universität, der auch an Stapels Forschungsmeetings teilnahm. Der Professor war geschockt über die perfekten Datensätze und hohen Effekte, die Stapel immer wieder in seinen Meetings präsentierte und meinte dazu:

I don't know that I ever saw that a study failed, which is highly unusual. Even the best people, in my experience, have studies that failed constantly. Usually, half don't work. (Bhattacharjee, 2013)

Der junge Professor konsultierte einen Kollegen in den USA, der wiederum meinte, dass er sich nicht verpflichtet fühlen solle, seinen Verdacht zu melden. Doch die Studenten ließen sich damit nicht abspeisen und kontaktierten im August 2011 schließlich den Leiter ihres Forschungsinstituts, welcher sich daraufhin mit dem Rektor der Universität Tilburg in Verbindung setzte. Stapel hätte sogar davon kommen können, hätte er sich „klüger“ verhalten und die gefundenen Fehler auf seine Schlampigkeit zurückgeführt. Doch er behauptete, dass seine Daten von Hilfskräften in verschiedenen Schulen gesammelt worden wären. Als der Rektor diese Schulen kontaktieren wollte, um die Sache aufzuklären, musste Stapel zugeben, dass diese Schulen gar nicht existierten. Diese Tatsache war schließlich der unwiderlegbare Beweis („smoking gun“) seines Betrugs und der Auslöser für alles, was danach folgen sollte.

4 Das Untersuchungskomitee

Vom Rektor der Universität Tilburg, Prof. P. Eijlander, wurde kurz nach der Entlassung Stapels, am 9. September 2011, ein Untersuchungskomitee, unter der Leitung von Willem J.M. Levelt gegründet. Stapel gestand dem Rektor, dass er nicht erst in Tilburg gefälschtes Datenmaterial für seine Untersuchungen verwendet hatte, sondern bereits seit Beginn seiner Karriere in Amsterdam auf die Manipulation von Daten zurückgriff. Um den vollen Umfang des durch das wissenschaftliche Fehlverhalten Stapels entstandenen Schadens erfassen zu können, hatte das Komitee zum Ziel, die gesamte Forschungsarbeit Stapels von Beginn seiner Karriere in Amsterdam, über Groningen, bis zu hin zu den Publikationen, die er während seiner Amtszeit in Tilburg veröffentlichte, zu untersuchen. Daher wurde vom Rektor der Universität Amsterdam, Prof. D. C. van den Boom, das Drenth-Komitee und vom Rektor der Universität Groningen, Prof. E. Sterken, das Noorth-Komitee gegründet.

Das Untersuchungskomitee verfolgte während ihrer akribischen Analyse zwei Fragestellungen (Stapel Investigation, 2012):

1. The Committee should examine which publications are based on fictitious data or fictitious scientific studies, and during which period the misconduct took place.
2. The Committee should offer a view on the methods and the research culture that facilitated this infringement, and make recommendations as to how a recurrence may be prevented.

Die Ergebnisse der Untersuchung können im Endbericht der Stapel Investigation (2012) nachgelesen werden.

5 Definitionen

Die National Science Foundation (2001) definiert „Scientific Misconduct“ (=wissenschaftliches Fehlverhalten) wie folgt:

- A significant departure from accepted practices of the scientific community for maintaining the integrity of the research record.
- The misconduct must be committed intentionally or knowingly, or in reckless disregard of accepted practices.

Weiter definiert sie den engeren Begriff „Research Misconduct“ (=Forschungsbetrug):

- Research Misconduct is defined as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research or in reporting research results.

Unter den Begriffen der Fabrikation, der Falsifikation und des Plagiats versteht man:

- Fabrikation: Das Produzieren und Berichten eines kompletten Datensatzes.
- Falsifikation: Das Manipulieren von Untersuchungsmaterialien, Untersuchungsgegenständen, Untersuchungsbedingungen, Untersuchungsabläufen oder das Verändern, Hinzufügen oder Weglassen von Daten und Ergebnissen, so dass die betriebene Forschung im abgebildeten Forschungsergebnis nicht mehr akkurat abgebildet wird.
- Plagiat: Die Inbesitznahme von Ideen, Forschungsprozessen, Ergebnissen und Wörtern, ohne auf deren Ursprung zu vermerken.

6 Die Auswirkungen des Betrugs

Nach Stroebe, Postmes und Spears (2012) kann wissenschaftliches Fehlverhalten vier Typen von externen Auswirkungen haben:

a) Die Schädigung der beruflichen Karrieren von Studenten, Mitarbeitern und Co-Autoren, die oftmals unwissend an der betrugshaften Forschungsarbeit und den Publikationen beteiligt waren.

b) Die Schäden für Patienten, die in Folge fälschlich gewonnener Erkenntnisse über Behandlungsmöglichkeiten oder Wirksamkeitsnachweisen von Medikamenten und Therapien verheerend sein können. Zum Beispiel der als Betrug identifizierte Artikel von Wakefield et al. (1999, zitiert nach Stroebe et al., 2012) in der Zeitschrift *Lancet*, welcher den Impfstoff gegen Masern, Mumps und Röteln mit einer erhöhten Auftretensrate von Autismus bei Kindern in Zusammenhang brachte, resultierte in einem erheblichen Rückgang von Impfungen, welcher für den Tod unzähliger ungeschützter Kinder verantwortlich hätte sein können. Oder die Studie von Bezwoda, Seymour & Dansey (1995, zitiert nach Stroebe et al., 2012), in der behauptet wurde, dass bei Patienten mit Metastasen bildendem Brustkrebs, der hoch dosierten Chemotherapie und einer Knochenmarktransplantation der Vorzug gegenüber anderen Behandlungsmethoden gegeben werden soll, resultierte darin, dass sich tausende Brustkrebspatienten einer teuren, schmerzhaften und teilweise fatalen Operation unterzogen.

c) Betrug kann den wissenschaftlichen Fortschritt behindern, da Forscher wertvolle Ressourcen (Forschungsgelder und Zeit) damit verschwenden, weitere Untersuchungen auf fälschlichen Theorien und Ergebnissen aufzubauen. Beziehungsweise wird durch die Unterstützung durch begrenzte Forschungsgelder, anderen Wissenschaftlern die Möglichkeit genommen, in ihrer Forschungsarbeit, und damit verlorenen Erkenntnissen zu anderen bestehenden Problematiken, voranzukommen. Der Betrugsfall von Jan Hendrik Schön, einem Physiker, der behauptete, einen Hochleistungstransistor (ein elektronisches Bauelement zum Schalten und Verstärken von elektrischen Signalen) aus Plastik und anderen Materialien gebaut zu haben, führte beispielsweise dazu, dass Forschungsteams jahrelang ihre Zeit damit verschwendeten, seine Versuche zu replizieren (Stroebe et al., 2012).

d) Das öffentliche Bekanntwerden eines Betrugsfalls schädigt den Ruf der zugrunde liegenden Forschungsdisziplin und reduziert das grundsätzliche Vertrauen in die Wissenschaft.

6.1 Forschungsfrage 1. Welche externen Auswirkungen hat der Betrugsfall Stapel nach den von Stroebe et. al (2012) genannten Typen?

6.1.1 Schädigung beruflicher Karrieren

Die Konsequenzen für Stapels Doktoranden, deren Dissertationen teilweise auf seinem gefälschten Datenmaterial beruhen, sind vermutlich am schwersten. Die meisten durften zwar ihren Dokortitel behalten, solange ihnen dieser bereits verliehen wurde (in einem Fall wäre dies zwei Wochen nach Bekanntwerden des Betrugsfalls gewesen), jedoch muss es herzerreißend sein, jahrelang für einen Abschluss und ein wenig Ansehen in der wissenschaftlichen Gemeinde gekämpft zu haben um danach für lange Zeit den Stempel eines „Mitgehilfen“ aufgedrückt zu bekommen, obwohl der endgültige Untersuchungsbericht ergeben hat, dass sowohl den Studenten als auch den Co-Autoren keine Schuld und Mitwissen am gefälschten Datenmaterial zugesprochen werden kann.

6.1.2 Schäden für Patienten

Da Stapel keine klinische Forschung betrieb, hat sein Betrugsfall zumindest keine Auswirkungen auf eine Patientenpopulation. Er arbeitete hauptsächlich an Problemen im Bereich der sozialen Kognitionsforschung, so dass sich der gesundheitliche Schaden für die Nachwelt in Grenzen hält.

6.1.3 Behinderung des wissenschaftlichen Fortschritts

Es ist schwierig zu erfassen, welche Auswirkungen Stapels Fehlverhalten für die Sozialpsychologie bezüglich verschwendeter Ressourcen hat. Sicher ist, dass einige Doktors-Anwärter ihre Forschungsarbeit aufgeben bzw. neu gestalten mussten, da sich Stapels Ergebnisse nicht replizieren ließen. Jedoch sind sich die Autoren des Endberichts (Stapel Investigation, 2012) einig, dass sich der Schaden Stapels Betrugs für die Weiterentwicklung sozialpsychologischer Theorien in Grenzen hält, wofür sie auch maßgeblich das bestehende Wissenschaftssystem verantwortlich machen (worauf zu späterem Zeitpunkt noch näher eingegangen werden soll). Die Autoren meinen:

The preference of journals for elegant, concise, attractive findings is consistent with the assessment of various of the Committees' interviewees, that Mr Stapel has made no significant contribution to the social psychological theory. His main contributions were in inventive experimental applications and variants, and in 'interesting' findings. He had a good feeling for what was 'in', we were told. His

citation score is no higher, and is rather lower than the average level of his Tilburg research group [...]. (S. 54)

Dennoch soll in der vorliegenden Diplomarbeit recherchiert werden, welches Echo Stapels Arbeiten in der Wissenschaft ausgelöst haben.

6.1.3.1 Forschungsfrage 1.1. Wie oft wurden Diederik Stapels Publikationen zitiert?

Um den wissenschaftlichen Einfluss Diederik Stapels überhaupt einschätzen zu können, wurde zunächst eine Literaturrecherche in der Thomson Reuters (früher: *ISI Web of Knowledge*) Online-Zitationsdatenbank *Web of Science* durchgeführt. Die Eingabe des Autors „Diederik A. Stapel“ ergab 168 Ergebnisse, wovon sich 122 auf Publikationen in Fachzeitschriften oder Beiträgen in Büchern beziehen. 46 Suchergebnisse sind „Retractions“, also die offizielle Rücknahme eines Artikels der jeweiligen Fachzeitschrift, aufgrund des bestätigten Verdachts auf Datenmanipulation seitens der Stapel Untersuchungskommission (Stapel Investigation, 2012). Im Jahr 2011 wurde ein Artikel zurückgezogen, 2012 31 Artikel und 2013 weitere 14. Diederik Stapels Publikationen wurden insgesamt 2169 Mal zitiert, davon stammen 390 Zitierungen von einer Selbstzitation ab. Durchschnittlich wurde eine Publikation Stapels 12.91 Mal zitiert.

In Tabelle 1 sind die Ergebnisse der Literaturrecherche dargestellt, in Abbildung 1 die graphische Darstellung Stapels Publikationsanzahl von 1994 bis 2013 und in Abbildung 2 die graphische Darstellung der Zitierungen Stapels durch andere Autoren im Zeitraum von 1996 bis 2013.

Tabelle 1: Zitationsreport des Autors Diederik A. Stapel

Ergebnisse	169
Ergebnisse ohne <i>Retraction Statements</i>	122
Summe Zitierung durch andere	2232
Summe Zitierung durch andere ohne Selbstzitation	1840
Zitierende Artikel	1451
Zitierende Artikel ohne Selbstzitation	1313
Durchschnittliche Zitierungen pro Artikel	13.21
h-Index	27

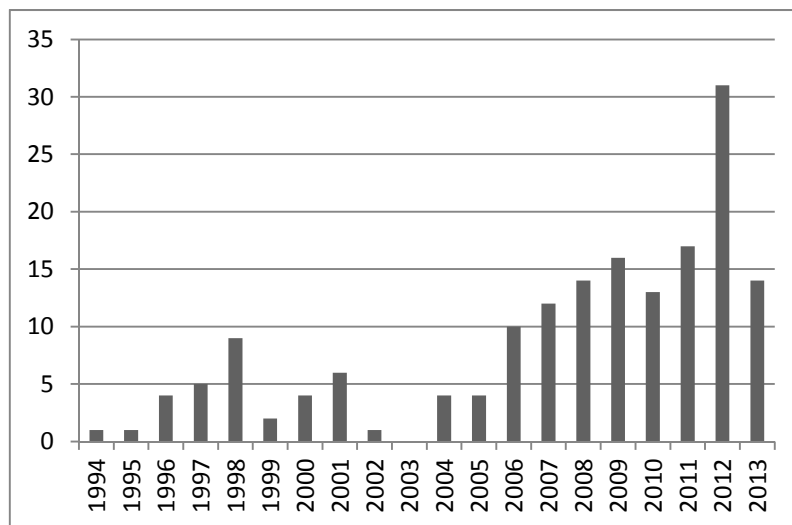


Abbildung 1: Anzahl der Publikationen von 1994 – 2013

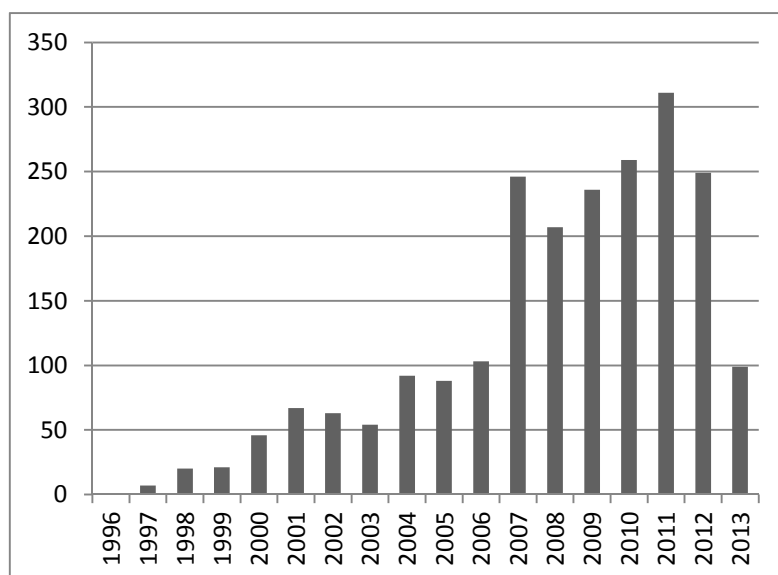


Abbildung 2: Anzahl der Zitierungen durch andere von 1996 – 2013

Der „h-Index“ in Tabelle 1, auch Hirsch-Index, Hirschfaktor oder Hirsch-Koeffizient genannt, ist ein bibliometrisches Maß, das auf Zitationen der Publikationen eines Autors zu einem Zeitpunkt basiert. Ein hoher h-Index spricht für einen großen wissenschaftlichen Einfluss des Autors. Ein Wissenschaftler hat einen Hirsch-Index h , wenn h von seinen insgesamt N Publikationen mindestens h -mal, die restlichen $(N - h)$ Publikationen höchstens h -mal zitiert wurden. Zur Ermittlung kann man alle Veröffentlichungen des Autors nach Zitathäufigkeiten absteigend aufreihen. Danach wird durchgezählt, bis die r -te Veröffentlichung weniger als r Zitierungen hat. h ist dann $r - 1$.

Zum Beispiel:

Bei den Zitathäufigkeiten 65, 58, 55, 17, 2, 2, 1, 1, 0, 0 ist der Hirschfaktor 4, weil vier Veröffentlichungen mindestens 4 Mal, die restlichen höchstens 4 Mal zitiert wurden (Wikipedia, 2013).

Der h-Index zur Bewertung wissenschaftlicher Leistungen wurde 2005 von dem argentinischen Physiker Jorge E. Hirsch entwickelt und in der Zeitschrift *Proceedings of the National Academy of Sciences* veröffentlicht. Sehr einflussreiche Physiker erreichen laut Hirsch einen Index von 62 bis 107. Zur Zeit der Publikation des h-Indexes (Hirsch, 2005) hatte Edward Witten mit 120 den höchsten Hirsch-Index unter Physikern. In den Biowissenschaften erreichen berühmte Wissenschaftler sogar einen h-Index zwischen 120 und 191, welcher mit der grundsätzlich höheren Publikationsanzahl in diesem Fachgebiet zusammenhängt.

Diederik Stapel erzielt einen h-Index von 27 und schneidet damit im Vergleich zu anderen Wissenschaftlern im Fachbereich Psychologie relativ gut ab. Das amerikanische *Institute for Scientific Information* (ISI) versuchte mittels der Datenbank *Web of Science* die einflussreichsten Wissenschaftler der Welt zu ermitteln. Unter dem Titel „Highly Cited“ veröffentlichte das Institut Listen der einflussreichsten Forscher aus den verschiedensten wissenschaftlichen Fachgebieten, wobei die Häufigkeit der Zitierung durch andere Autoren als Messkriterium für einen hohen Einfluss herangezogen wurde (Highly Cited Liste, 2012). Laut ISI bietet „Highly Cited“ die Möglichkeit, in den einzelnen Wissenschaftsgebieten Schlüssel-Forscher und -Institute zu identifizieren, die einen fundamentalen Beitrag zur Entwicklung der Wissenschaft in der jüngsten Vergangenheit geleistet haben. Ausgewertet wurden Zitierungen von der Publikationsliste einzelner Forscher innerhalb der letzten 10 Jahre (Highly Cited, 2013). Im Jänner 2013 wurden die Listen das letzte Mal aktualisiert. Die Liste für das Fachgebiet Psychiatrie/Psychologie beinhaltet 308 Wissenschaftler, die jedoch nicht nach ihrem Einfluss gerangiert sind. Darin findet sich zum Beispiel der amerikanische Forscher Andrew L. Elliot des *Departments of Clinical and Social Science in Psychology* der *University of Rochester*, der über einen h-Index von 56 verfügt oder der niederländische Psychologe Eric-Jan Wagenmakers von der Universität Amsterdam mit einem h-Index von 24. Diederik Stapel findet sich natürlich nicht auf der aktuellen Liste, jedoch kann mithilfe des Hirsch-Koeffizienten, als Vergleichsmaß für wissenschaftlichen Einfluss, festgehalten werden, dass Stapel zu seiner Zeit durchaus einen wesentlichen Beitrag zu fundamentalen Theorien geleistet hat. Zur Veranschaulichung wurde in Tabelle 2 der h-Index von Stapel mit dem h-Index weiterer Forscher, die in der folgenden Arbeit mehrmals zitiert werden, gegenüber gestellt.

Tabelle 2: h-Index Vergleich

Autor	h-Index
Diederik A. Stapel	27
Daniel Kahneman	72
Brian Nosek	28
Elliot Aronson	56
John F. Dovidio	47
Daniele Fanelli	19
Herbert Bless	26
Uri Simonsohn	8
Wolfgang Stroebe	34
Ap Dijksterhuis	32

6.1.3.2 Forschungsfrage 1.2.: Dienten Stapels manipulierte Ergebnisse als Grundlage für weitere Forschungsarbeiten beziehungsweise wurden durch Replikation Forschungsressourcen verschwendet?

6.1.3.2.1 Verschwendung finanzieller Ressourcen

Tabelle 3: Anzahl durch Forschungsgelder geförderte betrügerische Publikationen

Komitee	Zeitraum	Betrügerische Dissertationen	Betrügerische Publikationen	Durch Forschungsgelder geförderte Publikationen
Levelt	2007 - 2011	3	34	14
Noort	2002 - 2011	7	24	23
Drenth	1996 - 1999	2	5	0 (2 durch KNAW Stipendium)
Summe	1996 - 2011	12	63	37 (39)

Diederik Stapel schloss sein Doktorat der Psychologie 1997 an der Universität Amsterdam ab und 1998 wurde seine Forschung durch ein Stipendium der *Royal Netherlands Academy of Arts and Sciences* (KNAW) unterstützt, das ihm die Möglichkeit gab, fünf Jahre lang unabhängige Forschungsprojekte zu leiten. In dieser Zeit publizierte er zwei als betrügerisch identifizierte Artikel, bei denen davon ausgegangen werden kann, dass sie durch diese Geldquelle finanziert wurden.

In Tabelle 3 sind die Untersuchungsergebnisse der Stapel Investigation (2012) dargestellt. Das *Levelt Komitee* stellte für den Zeitraum von 2007 bis 2011 Betrug in 3 Dissertationen und 34 Publikationen Stapels fest, wovon 13 Publikationen durch einen „Pioneer Grant“ der *Netherlands Organisation for Scientific Research* (NWO) und eine Publikation durch einen „Transforum Research Grant“ finanziert wurde. Die NWO wird durch das niederländische Ministerium für Bildung, Kultur und Wissenschaft finanziert und verfügt jährlich über ein Budget von 625 Millionen Euro, das sie in die herausragendsten wissenschaftlichen Talente und die besten Forschungsprojektanträge investiert (Netherlands Organisation for Scientific Research, 2013). Das *Noort Komitee* konnte in 7 Dissertationen und 24 Publikationen Stapels Betrug feststellen, wovon 18 Publikationen durch einen NWO „Pioneer Grant“, drei durch einen NWO „Subsidy T32 MH19728 Grant“ und zwei durch seine KNAW-Mitgliedschaft unterstützt wurden. Aus Stapels Forschungszeit in Amsterdam gingen 2 betrügerische Dissertationen und 5 Publikationen hervor (*Drenth Komitee*), die nach den Autoren der Stapel Investigation (2012) offiziell nicht von einer Organisation gesponsert wurden, jedoch erhielt Stapel, wie bereits erwähnt, von 1998 bis 2002 ein Stipendium der KNAW, welches höchst wahrscheinlich auch zu Finanzierungszwecken dieser Forschungsprojekte verwendet wurde. Das Untersuchungskomitee bestätigte insgesamt in 63 Publikationen Hinweise auf betrügerische Forschungsarbeit, wobei insgesamt 37 dieser Publikationen durch eine Forschungsförderungsquelle finanziert wurden. Allein die NWO förderte davon 34 Publikationen mit einer Geldsumme von 2.2 Millionen Euro, weswegen sie 2012 noch verkündete, Diederik Stapel aufgrund der Veruntreuung von Forschungsgeldern anklagen zu wollen (Stroebe et al., 2012).

Die niederländische Staatsanwaltschaft (*Dutch Public Prosecution Service*) und das FIOD (*Fiscal Information and Investigation Service*), welches für die Verfolgung von finanziellen Betrugsfällen zuständig ist, prüfte den Fall auf staatlicher Ebene und entschied, dass sich Stapel wegen Missbrauchs öffentlicher Forschungsgelder in keinem gerichtlichen Prozess weiter verantworten müsse. In einer außergerichtlichen Einigung wurde ein Strafmaß von 120 Stunden gemeinnütziger Arbeit und der 18-monatige Verzicht des Anspruchs auf ein halbes Monatsgehalt, das Stapel nach niederländischem Recht aufgrund von Krankheit während seiner Entlassung von der Tilburg Universität 2011, zugestanden hätte,

festgelegt. Laut zuständigen Staatsanwälten habe Stapel keine Forschungsgelder veruntreut, da die Finanzmittel hauptsächlich für die Gehälter von Mitarbeitern und Doktoranden verwendet wurden und er das Geld nicht zur persönlichen Bereicherung missbrauchte. In die außergerichtliche Entscheidung sei ebenfalls miteingeflossen, dass Stapel bereits freiwillig seinen Dokortitel abgegeben und bei der Untersuchung des Schadensausmaßes durch die drei Universitäts-Komitees bereitwillig zur Mitarbeit und Aussage zur Verfügung stand. Außerdem sagten Stapels ehemalige Doktoranden aus, welche zu den Hauptleidtragenden zählen, dass sie froh darüber wären, wenn der Fall endlich abgeschlossen und die ganze Geschichte nicht noch einmal vor Gericht aufgerollt werden würde (Van Noorden, 2013).

Um zu recherchieren, welche Konsequenzen die NWO nun aus dem Betrugsfall zog, wurde die Organisation per E-Mail kontaktiert und mit folgenden Fragen konfrontiert:

1. Gibt es seitens der NWO eine offizielle Stellungnahme zu den Vorfällen des entlarvten Forschungsbetrügers Diederik Stapel?
2. Wie viel Euro der NWO missbrauchte Diederik Stapel exakt für seine jahrelange Forschung?
3. Wie viele Forscher mussten anstelle Diederik Stapels auf eine Förderung der NWO verzichten?
4. Welche (rechtlichen) Konsequenzen zogen die Verantwortlichen der NWO schlussendlich? Kam es, wie 2012 angekündigt, zu einer Anklage aufgrund einer Veruntreuung von Forschungsgeldern?

Eine Pressesprecherin der NWO, Margot Custers, teilte in ihrer Antwort mit, dass ihre Organisation den Empfehlungen der niederländischen Staatsanwälte folgte und sich dafür entschied, keine rechtlichen Schritte gegen Stapel einzuleiten. Er habe für seine Forschung insgesamt 1.4 Millionen Euro von der NWO erhalten. Seit Bekanntwerden dieses massiven Betrugsfalles, habe die NWO weitere Schritte unternommen, um die wissenschaftliche Integrität zukünftig zu sichern (eine Kopie der E-Mail findet sich im Anhang):

NWO asks all researchers who receive funding from NWO to commit to the principles of the Netherlands Code of Conduct on Scientific Conduct and to act accordingly.

NWO has opened a research integrity desk where complaints about research integrity can be filed. Anyone can file a complaint about research integrity against a researcher who is or was funded by NWO or about an applicant who is still in the procedure. NWO believes that such a research integrity desk will be a more

accessible way for someone for filing a complaint than at their own university, since most of the researchers funded by NWO are employed at the Dutch universities.

Auf weitere Fragen ging die zuständige Pressesprecherin nicht ein.

6.1.3.2.2 Verschwendung von Zeit durch den Aufbau auf fälschlichen Ergebnissen

Schwieriger als die Verschwendung von finanziellen Ressourcen ist zu erfassen, welche Auswirkungen Stapels Fehlverhalten für die sozialpsychologische Wissenschaft hat. Sicher ist, dass einige Doktorats-Anwärter ihrer Forschungsarbeit aufgeben bzw. neu gestalten mussten, da sich Stapels Ergebnisse nicht replizieren ließen. Jedoch sind sich die Autoren des Endberichts (Stapel Investigation, 2012) einig, dass sich der Schaden Stapels Betrugs für die Weiterentwicklung sozialpsychologischer Theorien in Grenzen hält. Die Autoren meinen:

The preference of journals for elegant, concise, attractive findings is consistent with the assessment of various of the Committees' interviewees, that Mr Stapel has made no significant contribution to the social psychological theory. His main contributions were in inventive experimental applications and variants, and in 'interesting' findings. He had a good feeling for what was 'in', we were told. His citation score is no higher, and is rather lower than the average level of his Tilburg research group [...]. (S. 54)

Vergleicht man Stapels Hirsch-Index als Maß für wissenschaftlichen Einfluss mit dem anderer anerkannter Psychologen (siehe Forschungsfrage 1.1.), so lässt sich darüber diskutieren, inwiefern er tatsächlich zur Entwicklung sozialpsychologischer Theorien beigetragen hat.

a. Fehlgeschlagene Replikation von Priming-Experimenten

Eine Möglichkeit zeitliche Ressourcen zu verschwenden, besteht darin, die Forschungsergebnisse eines Autors direkt zu replizieren und daran zu scheitern, da der gesamte Versuchsaufbau erfunden bzw. die berichteten signifikanten Ergebnisse auf manipulierten Daten beruhen.

Stapels Forschungsschwerpunkt lag in der sozialen Kognitionsforschung rund um das Thema des sogenannten „Primings“.

Unter dem Begriff „soziale Kognition“ versteht man allgemein die Wahrnehmung sowie die kognitive Verarbeitung von Stimuli aus der sozialen Umwelt, die dazu dient, das eigene und das Verhalten anderer zu verstehen. Soziale Kognition ist gleichermaßen bedeutsam für bewusstes und unbewusstes Handeln in sozialen Interaktionen und umfasst im Kern das Erkennen und Interpretieren von Emotionen in Gesichtsausdrücken, Körperhaltung und Sprachmelodie, das Reflektieren eigener Befindlichkeiten und Absichten sowie das Vergegenwärtigen von Intentionen, Gedanken und Gefühlen Anderer (Brüne, 2008).

Das Verstehen der Prinzipien, die maßgeblich an der menschlichen Verhaltensänderung beteiligt sind, sind für die (Sozial-) Psychologie elementar. Die Annahme der Priming Theorie, dass Verhalten durch unbewusste Einflüsse gesteuert werden kann, indem relevante Stereotype, Einstellungen, Eigenschaften, Ziele oder Konzepte (durch die Vorgabe von Wörtern oder wortbezogenen Aufgaben) aktiviert werden, stellt die ursprüngliche Theorie in Frage, dass der beste Weg, Verhalten zu ändern, über bewusste Meinungen und Bestrebungen abläuft (Shanks et al., 2013). Vertreter der Priming Theorie sind jedoch von deren unbewussten Beeinflussung überzeugt. Bargh und Huang (2009) schreiben beispielsweise:

Priming refers to the passive, subtle, and unobtrusive activation of relevant mental representations by external, environmental stimuli, such that people are not and do not become aware of the influence exerted by those stimuli. In harmony with the situationist tradition, this priming research has shown that the mere, passive perception of environmental events directly triggers higher mental processes in the absence of any involvement by conscious, intentional processes [...]. (S. 128)

Diese sehr einflussreiche Theorie, dass unser Verhalten durch unbewusste Aktivierung verändert werden kann, wurde vor kurzem heftig kritisiert.

David Shanks, ein Psychologe des *University Colleges* in London, publizierte im April 2013 einen Artikel in der Zeitschrift *PLoS ONE*, in dem er berichtete, dass die Replikationsversuche ganzer neun Experimente rund um das Thema „Intelligence Priming“ scheiterten, welche erstmals 1998 von dem niederländischen Sozialpsychologen der Universität Nijmegen, Ap Dijksterhuis und seinem Kollegen van Knippenberg, durchgeführt wurden. Diederik Stapel wirkte ebenfalls in einer der drei Original Intelligenz-Priming Studien mit (Dijksterhuis, Spears, Postmes, Stapel, Koomen et al., 1998), welche mit insgesamt 178 Zitierungen die Nummer eins in Stapels Zitationsreport belegt und daher auch vom Untersuchungskomitee in Augenschein genommen wurde – jedoch konnten keine Hinweise auf betrügerische Forschungsmethoden festgestellt werden.

Die Autoren Dijksterhuis und van Knippenberg (1998; 2000) wollten in ihren Experimenten den beiden Fragen nachgehen, ob sich die Effekte der Wahrnehmung auf das Verhalten

nur auf relativ simple Aktionen beschränken oder ob sie auch auf komplexere Verhaltensmuster angewandt werden können und welche Parameter einen Einfluss auf die Wahrnehmungs-Verhaltens-Verbindung haben – ob es beispielsweise einen Zusammenhang zwischen der Stärke des Primes und der Stärke des resultierenden Verhaltens gibt und wann dieser Effekt der Wahrnehmung auf das Verhalten wieder abklingt. Sie behaupteten, dass die Aktivierung der Eigenschaft „intelligent“ (durch das direkte Primen der Eigenschaft oder durch die Verwendung eines Stereotyps), zur Aktivierung eines Sets konkreter Verhaltensweisen führt, die mit dieser Repräsentation verbunden sind (z.B. sich auf ein Problem konzentrieren, einen analytischen Ansatz anwenden, systematisch über mögliche Lösungen nachdenken).

Die Unterscheidung zwischen Stereotypen, die Eigenschaften aktivieren, auf der einen Seite, und konkreten (Vor-) Bildern, die diese Eigenschaften veranschaulichen, auf der anderen Seite, sind dabei für die soziale Wahrnehmung und das resultierende automatische Verhalten ausschlaggebend.

Die Autoren des Intelligenz-Primings gehen davon aus, dass die Aktivierung durch ein Vorbild (z.B. Albert Einstein) zu einem unterschiedlichen Verhalten führt, als die Aktivierung durch eine Eigenschaft (intelligent) oder eines Stereotyps (Professor). Das Primen mit konkreten (Vor-) Bildern kann unter bestimmten Umständen zu sogenannten Kontrasteffekten führen. Die Aktivierung der Eigenschaft „Feindseligkeit“ führte beispielsweise in einem Experiment von Herr (1986) zu feindseligerem Verhalten der Versuchspersonen. Wenn diese jedoch mit dem Bild von Adolf Hitler geprimt wurden, verhielten sie sich weniger feindselig. Konkrete (Vor-) Bilder dienen als Vergleichsstandard und können durch eine Entfernung der Selbsteinschätzung von diesem, einen Kontrasteffekt auslösen, wohingegen die Annäherung an ein aktiviertes Konstrukt zu einem Assimilationseffekt führt. Das Primen mit dem Bild Hitlers aktiviert zwar das Konzept Feindseligkeit und damit einen potentiellen Assimilationseffekt, allerdings führt der sich aufzwingende Vergleich mit sich selbst zu einem Kontrasteffekt, der den Assimilationseffekt des aktivierten Konstrukts überschattet. Das vorgegebene Bild sollte jedoch eine gewisse Intensität oder Extremität aufweisen, um den Assimilationseffekt überdecken zu können. Der Vergleich mit Albert Einstein hätte also zur Folge, dass man sich selbst als weit weniger intelligent einschätzt und somit auch konträr, weniger intelligent, handelt.

In Dijksterhuis und van Knippenbergs Experiment 1 und 2 (1998) wurden die Versuchspersonen mit dem Stereotyp von Professoren geprimt, die die Konzepte Intelligenz und Wissen repräsentieren sollten. Die Versuchsteilnehmer wurden instruiert, sich fünf Minuten lang einen Professor (Versuchsgruppe) oder einen Sekretär (Kontrollgruppe 1) vorzustellen und dabei Verhaltensweisen, Lifestyle und

Erscheinungsbild aufzulisten, die ihnen dabei in den Sinn kamen. Zusätzlich wurde eine zweite Kontrollgruppe eingeführt, in der gar keine Prime-Manipulation stattfand (Kontrollgruppe 2). Die Annahme war, dass die Versuchsgruppe in einem nachfolgenden Allgemeinwissenstest besser abschneidet als die Kontrollgruppen, die zuvor nicht auf das Konzept „Intelligenz“ geprimt wurden. In Experiment 3 wurden die Versuchspersonen mit dem Stereotyp von Fußballhooligans geprimt, die in einer Vorstudie als dumm eingeschätzt worden waren. Die geprimten Versuchspersonen sollten im anschließenden Wissenstest schlechter abschneiden als die nicht geprimte Kontrollgruppe. In Experiment 4 wurde erforscht, ob das Priming mit Eigenschaften (intelligent vs. dumm) den gleichen Effekt auf die Leistung der Versuchspersonen im Wissenstest hat, wie das Priming mit Stereotypen, die mit diesen Eigenschaften verbunden sind (Professor vs. Hooligans). In allen vier Experimenten berichteten die Autoren von signifikanten Haupteffekten in Richtung der erwarteten Hypothesen.

Shanks et al. versuchten die Experimente 1 und 2 mit kleinen Abänderungen in den Versuchsbedingungen und die Experimente 3 und 4 nach völlig identem Ablauf zu replizieren. Shanks et al. nahmen in ihrem ersten Experiment eine Änderung in der Priming Manipulation vor, indem sie ein ausgedehnteres Verfahren anwandten, das aus einem 8-minütigem Video Clip mit diskutierenden Professoren oder aus einer Dokumentation über Fußball Hooligans bestand. Nach Dijksterhuis und van Knippenberg würde eine längere Primingdauer den Effekt verstärken. Anstatt nach der Priming Aufgabe einen Allgemeinwissenstest mit Aufgaben des Spiels *Trivial Pursuit* (bestehend aus 42 Multiple Choice Fragen) durchzuführen, sollten die Versuchspersonen in der Replikation von Shanks et al. (2013) sowohl vor als auch nach dem Ansehen des Video Clips, den *Raven's Advanced Progressive Matrices Test* lösen. Wenn sich die Priming Effekte, wie beschrieben, tatsächlich auf die Motivation und die Lösungsstrategien auswirken, dann wäre ein allgemeiner Wissenstest eine sehr schlechte Wahl, angesichts der binären Natur von Wissen (wissen/nicht wissen einer Frage). Ein Test, der analytisches Denken und Problemlösen erfasst, sollte dagegen eine bessere Alternative darstellen, um Priming Effekte zu erfassen.

Um die Veränderung in der Testleistung vor und nach dem Priming in Experiment 1 und 2 zu analysieren, wurde der Baseline-Score von dem Post-Priming Score abgezogen und Effektgrößen für die Differenzen berechnet. In beiden Priming Gruppen (Professor und Hooligan) zeigte sich eine Verbesserung in der Testleistung, die jedoch auf den Übungseffekt zurückzuführen ist. Entgegen der Hypothese von Dijksterhuis und van Knippenberg war die Leistungssteigerung in der Professor-Gruppe nicht größer als in der

Hooligan-Gruppe. Es war sogar ein Trend in die umgekehrte Richtung zu beobachten, allerdings unterschieden sich die Mittelwerte nicht signifikant voneinander.

In Experiment 3 gab es keinen Pretest und als Daten dienten die richtigen Antworten (in Prozent) im Allgemeinwissenstest. Die Analyse ergab keinen signifikanten Unterschied zwischen den beiden Versuchsbedingungen. Auch in Experiment 4 stellte sich kein signifikanter Priming Effekt heraus.

In der psychologischen Forschung wird das Scheitern, signifikante Effekte zu erzielen (=Null Results), oft als ergebnislose Studie bezeichnet. Aus der Perspektive der bayesschen-Statistik ist es dagegen sinnvoll, sowohl für die Nullhypothese als auch für die Alternativhypothese, die Wahrscheinlichkeit zu berechnen, mit der diese eintreten könnte. In der Bayesschen Statistik werden immer bedingte Wahrscheinlichkeiten betrachtet, das heißt, wie wahrscheinlich es ist, dass Aussage A eintritt, wenn B eingetreten ist (Shanks et al., 2013).

Diese Wahrscheinlichkeiten werden dann in einem Likelihood Ratio Test miteinander verglichen und daraus entsteht der sogenannte „Bayes Faktor“ (=Likelihood Ratio), der das Verhältnis der Auftretenswahrscheinlichkeit für die beiden Hypothesen angibt. Wenn dieses Verhältnis (=Odds Ratio) sehr klein ist, beispielsweise 1:10, dann kann man davon ausgehen, dass die Alternativhypothese mit einem Faktor von 10 bekräftigt wird. Wenn es allerdings sehr groß ist (z.B. 10:1), dann wird die Nullhypothese mit einem Faktor von 10 bestärkt. Die vier replizierten Experimente von Shanks et al. (2013) erzielten Bayes Faktoren zwischen 2.54:1 und 5.58:1 und bekräftigten somit die Nullhypothese, dass bei diesen Studien nicht von einem Intelligenz-Priming Effekt ausgegangen werden kann.

Die Autoren Dijksterhuis, Spears et al. (1998) postulierten in ihrer Studie, dass die Aktivierung durch ein konkretes Bild (z.B. Albert Einstein) zu einem unterschiedlichen Verhalten führt, als die Aktivierung durch eine Eigenschaft („intelligent“) oder eines Stereotyps („Professor“). In ihrem ersten Experiment demonstrierten sie Assimilations- bzw. Kontrasteffekte durch die Verbesserung oder Verschlechterung in einem Intelligenztest. Die Versuchsteilnehmer wurden mit Stereotypen geprimt, die mit ihrem Ausmaß an Intelligenz assoziiert werden sollten (Professor = intelligent vs. Supermodel = dumm). Weitere Versuchspersonen wurden mit den konkreten Beispielen dieser Kategorien, mit Albert Einstein oder Claudia Schiffer geprimt. Danach beantworteten sie einen allgemeinen Wissenstest, wobei die Anzahl korrekter Antworten als Index der intellektuellen Leistung herangezogen wurde. Für die stereotype Priming-Bedingung (Professor vs. Supermodel) ergab sich ein signifikanter Assimilationseffekt, das heißt, die Versuchsteilnehmer der Professor Gruppe erzielten im Wissenstest höhere Leistungen als die Teilnehmer der Supermodel Gruppe. Für das Priming mit konkreten Vorbildern

ergaben sich wie erwartet Kontrasteffekte, so dass die Versuchspersonen der Claudia Schiffer Gruppe im Wissenstest besser abschnitten als die Versuchspersonen der Albert Einstein Gruppe.

Dass der Unterschied zwischen den Prime-Stimuli Kategorie vs. konkretes Bild jedoch nicht die Schlüsselvariable für den Kontrasteffekt in Dijksterhuis et al.s Experiment war, behaupteten LeBoeuf und Estes (2004) in ihrer späteren Replikation. Sie stellten die Alternativhypothese auf, dass ein Prime, unabhängig davon, ob es eine Kategorie oder ein konkretes Bild ist, sowohl einen Assimilationseffekt als auch einen Kontrasteffekt hervorrufen kann, je nachdem wie relevant ein Vergleichsstandard für einen selbst eingeschätzt wird. LeBoeuf und Estes erhoben die Vergleichsrelevanz, in dem sie die Versuchspersonen baten, entweder Ähnlichkeiten (hohe Relevanz) oder Unterschiede (niedrige Relevanz) zwischen ihnen selbst und dem präsentierten Prime aufzulisten. Die Idee dahinter war, dass das Auflisten von Ähnlichkeiten zwischen Prime und einem selbst zur Überlegung führt, den Prime als relevanten Vergleichsstandard in Betracht zu ziehen. Die teilnehmenden Personen wurden gebeten, aufzuschreiben, wie ähnlich (hohe Relevanz) oder verschieden (niedrige Relevanz) sie sich entweder zu einem Professor (stereotyper Prime) oder Albert Einstein (konkreter Prime) beurteilen würden. Unmittelbar nach der Priming Manipulation mussten die Teilnehmer einen allgemeinen Wissenstest in Multiple Choice Format beantworten. Die Ergebnisse zeigten, dass die Versuchsteilnehmer in der niedrigen Relevanz-Bedingung (in der sie Unterschiede auflisteten), bessere Testleistungen erzielten als in der hohen Relevanz-Bedingung (in der sie Ähnlichkeiten aufzählten). Versuchsteilnehmer in der niedrigen Relevanz-Bedingung, die Unterschiede zwischen ihnen und einem Professor oder Albert Einstein auflisteten, aktivierten dadurch die allgemeine Eigenschaft „intelligent“ und zeigten mit einer besseren Leistung einen Assimilationseffekt. Die Aufgabe in der hohen Relevanz-Bedingung, Ähnlichkeiten zwischen einem selbst und einem Professor oder Einstein aufzulisten, führte zu schlechteren Leistungen im Wissenstest, da die Primes als Vergleichsstandard herangezogen wurden und somit zu einem Kontrasteffekt führten. Der größte Unterschied ergab sich zwischen den Gruppen Professor-Selbst Ähnlichkeiten und Einstein-Selbst Unterschieden, wobei letztere Gruppe wesentlich besser abschnitt als die erste.

Als abhängige Variable wurde in der Replikation von Shanks et al. (2013) die Prozentanzahl richtig beantworteter Fragen im Allgemeinwissenstest herangezogen. Der Unterschied zwischen den Scores in der Professor-Ähnlichkeitsgruppe und der Einstein-Unterschiedsgruppe war nicht signifikant und ging sogar in die falsche Richtung. Das Experiment von LeBouef und Estes konnte also nicht repliziert werden, jedoch zeigte sich ein signifikanter Effekt für den Prime-Typ (Stereotyp vs. konkretes Beispiel) im anschließenden Wissenstest.

In einer weiteren Studie zur Untersuchung der Moderatorvariablen von Assimilations- und Kontrasteffekten, demonstrierten Gordijn und Stapel (2006), dass soziale Vergleiche automatisch dann zu Kontrast- oder Assimilationseffekten führen, wenn das zu vergleichende Ziel entweder als „in-group“ oder „out-group“ Mitglied kategorisiert wird. Die Autoren gingen davon aus, dass ein Gruppenkontext dazu führen kann, dass die soziale Identität für einen zugänglicher ist als die persönliche Identität und demzufolge Assimilation dann auftritt, wenn das zu vergleichende Objekt als Gruppenmitglied wahrgenommen wird. Wenn das Objekt jedoch nicht als Gruppenmitglied wahrgenommen wird, sollte ein Kontrasteffekt resultieren.

In ihrer Studie wurde ein Gruppenkontext geschaffen, indem Studenten der Groningen Universität gesagt wurde, dass ihre Testleistungen mit Studenten der Amsterdam Universität verglichen werden. Als Priming Aufgabe sollten die Studenten anschließend einige Persönlichkeitseigenschaften einer fiktiven Person einschätzen, die entweder als intelligent oder unintelligent und als Student der Universität Groningen (*in-group*) oder Amsterdam (*out-group*) beschrieben wurde. Zuletzt erhielten sie wieder einen Allgemeinwissenstest. Den Hypothesen entsprechend führte das Priming mit einem „in-group“ Personenvergleich zu einer Assimilation (das heißt, besserer Leistung bei Vergleich mit einem intelligenten Studenten aus Groningen und schlechterer Leistung bei einem unintelligenten) und der „out-group“ Personenvergleich zu einem Verhaltenskontrast (schlechtere Leistung bei intelligentem Studenten aus Amsterdam und bessere Leistung bei einem unintelligenten).

Shanks et al. (2013) versuchten Gordijn und Stapels Ergebnisse zu replizieren. Deren Artikel wurde kurz nach der Durchführung des Replikationsversuches von der Zeitschrift *British Journal of Social Psychology* aufgrund Stapels Geständnisses, dass er manipulierte Ergebnisse enthält, zurückgezogen.

Shanks et al.s replizierter Versuchsablauf folgte dem Originalexperiment sehr genau, bis auf eine kleine Änderung in der Intelligenz Manipulation. Die Zielperson wurde entweder als Professor oder als Fußballhooligan, anstatt als „intelligent“ oder „unintelligent“, beschrieben. Das Experiment bestand also aus einem 2 (Professor vs. Hooligan) x 2 (in-group vs. out-group) Design. Es wurde eine 2x2 *Between Subjects ANOVA* mit den Prozentzahlen korrekter Antworten im Wissenstest berechnet. Die Varianzanalyse ergab keine signifikanten Haupteffekte, weder für die Professor/Hooligan-Variation noch für die soziale Identität (in-group vs. out-group) und auch keine Interaktionseffekte. Es stellten sich ebenfalls keine einfachen Effekte für die „in-group“ Bedingung heraus, in der das Priming mit dem Professor zu einer besseren Leistung im Wissenstest als in der Hooligan-Prime Bedingung führen hätte sollen und auch nicht in der „out-group“

Bedingung, in der der Eindruck eines Professors die Testleistung verschlechtern hätte sollen.

Gordijn und Stapels Studie wurde laut *Web of Science* 3 Mal zitiert. Die zitierenden Artikel greifen die Erkenntnisse der niederländischen Autoren in ihrem Theorieteil auf, allerdings unterstützen auch andere Autoren deren „ingroup-outgroup“ Theorie, so dass nicht davon ausgegangen werden kann, dass allein durch den betrugshaften Artikel von Gordijn und Stapel (2006) fälschliche Theorien weiter ausgebaut und somit zeitliche Forschungsressourcen verschwendet wurden.

Ein weiterer Replikationsfehlschlag: John Bargh et al. (1996) im Kreuzfeuer der Kritik

Die Original Studie von Bargh, Chen und Burrows wurde 1996 im *Journal of Personality and Social Psychology* veröffentlicht und galt seitdem als einer der ersten Belege für die Existenz von sozialen Priming Effekten. In ihrem zweiten Experiment verfolgten die Autoren das Ziel, durch die Aktivierung des Stereotyps alter Leute („the elderly“) das menschliche Verhalten in Bezug auf die Gehgeschwindigkeit zu beeinflussen. Sie stellten die Hypothese auf, dass Personen, die zuvor mit dem Konzept alter Leute geprimt werden, beim Verlassen des Testraums langsamer gehen, als Personen, die zuvor nicht mit diesem Konzept geprimt werden.

Für ihre Studie wurden für die erste Durchführung 30 Psychologiestudierende der *New York University* und für ihre eigene Replikation erneut 30 Studenten ausgewählt. Den Versuchspersonen wurde mitgeteilt, dass sie an einem sprachlichen Leistungstest teilnehmen. Dieser Test bestand aus einer „Scrambling-Sentences“ Aufgabe, bei der die Versuchsteilnehmer 30 Sets mit je 5 Wörtern präsentiert bekamen, aus denen sie mit nur 4 Wörtern einen grammatikalisch richtigen Satz bilden und aufschreiben sollten. Die Versuchspersonen wurden dabei zufällig einer von zwei Versuchsbedingungen zugeordnet. Der Kontrollgruppe wurden zur Lösung der Aufgabe neutrale Wörter vorgegeben, die nicht in Verbindung mit dem Konzept alter Leute gebracht werden konnten (z.B. „thirsty“, „clean“, „private“), während die Versuchsgruppe zur Bildung grammatikalisch richtiger Sätze Wörter präsentiert bekam, die mit dem Stereotyp alter Leute assoziiert werden konnten (z.B. „worried“, „lonely“, „grey“, „sentimental“, „wise“, „bingo“, „stubborn“, „retired“, „wrinkle“, „traditional“, „dependent“ etc.). Für die Versuchsgruppe wurden dabei bewusst keine Wörter gewählt, die sich auf geschwindigkeitsbezogene Assoziationen beziehen (z.B. „schnell“, „langsam“). Der Versuchsleiter ließ jede Versuchsperson ein Kuvert ziehen, in dem sich bereits die Aufgabe befand, so dass er bezüglich der Zuteilung jeder Versuchsperson zu einer der beiden Bedingungen nicht Bescheid wusste. Für die Bearbeitungszeit wurden die

Teilnehmer allein gelassen und nach Beendigung der Aufgabe wurden sie über den scheinbaren Inhalt der Studie aufgeklärt und vom Experimentator verabschiedet. Ab Verlassen des Testraums stoppte ein weiterer Mitarbeiter des Experiments mit einer Stoppuhr die Zeit (in Sekunden), die die Versuchspersonen für den insgesamt 9.75 Meter langen Weg bis zum Erreichen des Fahrstuhls brauchten. Danach wurden sie erneut von einem eingeweihten Mitarbeiter aufgehalten, über den wahren Inhalt der Studie aufgeklärt und gefragt, ob sie die Relevanz der ausgewählten Wörter in der Versuchsbedingung in irgendeinen Zusammenhang mit der Ursache des Experiments gebracht hatten. Diese Frage verneinten alle Teilnehmer. Die Ergebnisse bestätigten die Annahme von Bargh et al. (1996): Ein t-Test mit der Priming-Bedingung als unabhängige Variable und der Gehgeschwindigkeit als abhängige Variable ergab, dass die Versuchspersonen, die zuvor mit dem Stereotyp alter Leute geprimt worden waren (Versuchsgruppe), nach Verlassen des Testraums langsamer den Flur entlang gingen ($M=8.23$ sec.) als die Teilnehmer der Kontrollgruppe ($M=7.30$ sec.). Und auch bei der Wiederholung des Experiments wurde die Hypothese abermals bestätigt.

Trotz der Berühmtheit Bargh et al.s Studie, wurde sie bis vor kurzem kaum im Rahmen eines externen Replikationsversuchs in Frage gestellt. Die Autoren Cesario, Plaks und Higgins wagten 2006 zwar ein Experiment mit ähnlichem Versuchsaufbau, bei dem gezeigt werden konnte, dass Versuchspersonen mit positiven impliziten Einstellungen gegenüber alten Menschen langsamer gehen, nachdem sie zuvor mit dem Stereotyp alter Leute geprimt worden waren und vice versa, Versuchspersonen, die negative implizite Einstellungen gegenüber alten Menschen hegten, schneller gingen, nachdem sie mit dem Konzept „the elderly“ geprimt worden waren, doch diese Replikationsversuche kamen dem originalen Versuchsdesign nie nahe genug, um von dessen Gültigkeit ausgehen zu können.

Diese Tatsache forderte Stephane Doyen und seine Kollegen 2012 heraus, Barghs Studie einer exakten Replikation zu unterziehen. In ihrem ersten Versuch scheiterten die Autoren trotz einer vergrößerten Anzahl an Versuchspersonen und der Verwendung einer automatischen Zeitmessmethode via elektronischen Sender, der mit der Aufzeichnung ab der ersten Bewegung eines Teilnehmers startete. Auch in ihrem zweiten Replikationsversuch folgten die Autoren dem ursprünglichen Versuchsaufbau sehr genau, bis auf eine weitere Änderung bezüglich des Versuchsleiters, der nun gegen zehn voneinander unabhängigen Experimentatoren eingetauscht wurde, die über das Ziel der Studie nicht Bescheid wussten. Doyen et al. vermuteten nämlich, dass sich Barghs Forschungsteam den Versuchspersonen gegenüber unbewusst in einer Art und Weise verhalten hatte, welche sie dazu veranlasst hat, sich entsprechend der zugewiesenen Versuchsbedingung zu verhalten. Indem sie sich beispielsweise während der

Versuchsdurchführung selbst langsamer fortbewegten, langsamer sprachen oder ihren Händedruck lockerten, bewirkten sie ein ähnliches Verhalten bei den Versuchspersonen in Bezug auf ihre Gehgeschwindigkeit. Um diese Annahme zu testen, wiederholten Doyen und seine Kollegen das Experiment mit 50 neuen Versuchspersonen und 10 Experimentatoren, wobei er letzteren jeweils mitteilte, dass die Versuchsteilnehmer gemäß der Macht des vorangehenden Primings entweder schneller oder langsamer aus dem Testraum gehen werden. Die Autoren fanden tatsächlich, dass sich die Versuchspersonen nur dann langsamer bewegten, wenn den Experimentatoren zuvor gesagt wurde, dass sie dies zu erwarten haben.

Die Autoren des Replikationsversuchs stellten ebenfalls die Vermutung auf, dass Barghs Forschungsmitarbeiter, nachdem die Versuchspersonen sich ein Kuvert mit einer „Scrambling-Sentences“ Aufgabe aussuchen durften und demnach einer Versuchsbedingung zugeteilt wurden, einen kurzen Blick auf das Versuchsmaterial werfen konnten und somit erkannten, ob es sich um eine Person der Versuchs- oder Kontrollgruppe handelte, und infolge ihr Verhalten sofort unbewusst im Sinne einer selbsterfüllenden Prophezeiung an den Versuchsteilnehmer anpassten. Bargh wehrte sich in seinem eigenen Blog *Psychology Today* (2012) vehement gegen diese Vorwürfe und betonte, dass der Versuchsleiter in seiner Studie vollkommen „blind“ gegenüber dem Zweck des Experiments war und niemals Einblick in die zugewiesene Versuchsbedingung eines Teilnehmers bekam. Ein Psychologe, der den Eintrag Barghs auf seinem Blog kommentierte, Matt Craddock, meinte, dass es jedoch einen Studienmitarbeiter gegeben haben muss, der die Aufgaben in die Kuverts verpackte und somit ausreichend Zeit hatte, sich mit den Versuchsbedingungen vertraut zu machen. Diese Tatsache wird in Barghs Beschreibung des Versuchsablaufs allerdings nicht explizit erwähnt und stellt die erste einer Reihe von Inkonsistenzen seiner Argumentation dar (Yong, 2012d).

Barghs zweite Kritik an Doyens Replikation ist, dass seine Studie sein Original-Experiment nicht exakt nachgestellt habe, wodurch der Priming Effekt verschwunden sei. Beispielsweise habe Doyen den Versuchsteilnehmern kurz vor dem Verlassen des Testraums gesagt, dass sie den Gang einfach gerade aus hinunter gehen sollen, wodurch sie laut Bargh an einen automatisch ablaufenden Prozess (das Gehen) erinnert wurden, welcher den Priming Effekt überschattet hätte. Doyen antwortete darauf wiederum, dass er diese Worte in seiner Studie nie benutzt habe und er außerdem den Unterschied zu Barghs Instruktion der Versuchsteilnehmer („The experimenter told the participant that the elevator was down the hall.“) nicht feststellen könne. Des Weiteren kreierte Bargh Doyen et al. an, dass sie für die Aufgabe der Versuchsgruppe zu viele altersbezogene Wörter benutzt hätten, wodurch die Versuchspersonen den Zweck des Experiments erraten konnten und sich somit umso stärker in Richtung der erwarteten Hypothesen verhielten.

Bargh meint: „Doyen et al. apparently did not check to make sure their participants possessed the same stereotype of the elderly as our participants did“ (zitiert nach Yong, 2012d). Dem kann jedoch entgegen gehalten werden, dass sich Doyen et al. bei ihrer Auswahl der Priming-Wörter auf eine Online-Umfrage beriefen, in der die Teilnehmer Adjektive nannten, die sie am ehesten mit dem Konzept alter Leute in Verbindung brachten. Bargh stützte sich im Vergleich dazu, nicht auf einen solch empirischen Pre-Test. Ein weiterer Einwand, den Bargh gegen seinen Kontrahenten vorbrachte, war die Behauptung, dass in Doyens Studie die gleiche Person, welche die sprachliche Aufgabe durchführte, die Zeit bis zur Ankunft der Versuchspersonen am Ende des Flurs maß, was jedoch erneut eine Falschaussage darstellte. Doyen et al. verwendeten in ihrem Experiment nämlich einen elektronischen Infrarotsender zur Aufzeichnung der Geschwindigkeit und keine Stoppuhr wie Bargh, welche im Übrigen für die Verzerrung von Messergebnissen die wesentlich anfälligere Methode darstellt.

Ein Co-Autor Doyens, Axel Cleeremans, antwortete auf die hervorgebrachte Kritik Barghs folgendermaßen:

The fact is that we failed to replicate this experiment, despite having twice as many participants and using objective timing methods. Regardless of the arguments one may come up with that explain why his study worked and ours did not, this suggests that unconscious behavioural priming is not as strong as it is cast to be. If the effect were truly robust, it shouldn't depend on minute differences. The fact that we did manage to replicate the original results when both experimenters and participants were appropriately primed suggests interesting avenues for further research and should be taken as an opportunity to better delineate the conditions under which the effect is observed. (Yong, 2012b)

Auf diese zutreffende Stellungnahme reagierte Bargh letztendlich sogar mit direkten persönlichen Angriffen auf die Autoren Doyen et al., indem er sie als „incompetent and ill-informed“ und die Zeitschrift, in der ihr gescheiterter Replikationsversuch publiziert wurde (*PloS One*), als ein Journal „who does not receive the usual scientific journal standards of peer-review scrutiny“ bezeichnete (Yong, 2012b). Auf seinem Online-Blog *Psychology Today*, nennt er neben der erfolgreichen Replikation seines Experiments von den Autoren Cesario et al. (2006) eine weitere von Hull, Slone, Metayer und Matthews (2002), welche den Priming Effekt bezüglich der Gehgeschwindigkeit für Personen mit einem hohen Selbstbewusstsein nachweisen konnten. Außerdem hätten mindestens zwei Fernsehsender den „elderly-walking-slow-effect“ erfolgreich repliziert (*South Korean national television* und *Great Britain's BBC1*). Die Feldstudie des Senders *BBC1* könne sogar auf YouTube verfolgt werden (http://www.youtube.com/watch?v=5g4_v4JStOU).

Angesichts Barghs Zitate erfolgreicher Replikationen seines Priming Effekts, stellt sich natürlich die Frage, inwieweit diese konzeptuelle Replikationen darstellen, welche weitere Variablen, wie das Selbstbewusstsein oder implizite Einstellungen gegenüber alten Menschen, hinzuziehen, anstatt sich auf das Versuchsdesign des Originalartikels zu konzentrieren. Der einzige Replikationsversuch, der diese Bedingung erfüllte und Barghs Experiment sogar mit verbesserten Methoden erneut durchführte, war und ist der von Doyen und Kollegen, welcher die Gültigkeit des Priming Effekts nicht bestätigen konnte.

Konsequenzen der gescheiterten Replikationsversuche

Aufgrund der Anzahl nicht signifikanter Ergebnisse, forderte Shanks Ap Dijksterhuis dazu auf, ein detailliertes Versuchsprotokoll zur Verfügung zu stellen, um den einst publizierten Priming Effekt exakt nachvollziehen zu können. Nach anfänglicher Weigerung Dijksterhuis, ein solches Protokoll zu veröffentlichen, gab die Zeitschrift *Nature* in einem späteren Kommentar den Hinweis, dass er dies bald doch tun wolle (Abbott, 2013).

Die gescheiterten Replikationsversuche rund um Ap Dijksterhuis und seine Kollegen, die Kontroverse zwischen John Bargh und Doyen et al. und noch einige weitere nicht replizierbare Ergebnisse anderer Psychologen (zum Beispiel die Diskussion um Dr. Daryl Bem, der in seinem Artikel, der im *Journal for Personality and Social Psychology* publiziert wurde, behauptete, dass Menschen in die Zukunft sehen können) lösten unlängst eine scharfe E-Mail Debatte unter den Fachexperten aus, die von dem Nobelpreisträger Daniel Kahneman, einem höchst angesehenen Psychologen der *Princeton University* in New Jersey, durch ein offenes Schreiben an die betroffene Wissenschaftsgemeinde ausgelöst wurde. Die am 26. September 2012 ausgesendete E-Mail mit dem Betreff „A proposal to deal with questions about priming effects“ kritisiert zwei methodologische Forschungsansätze jener Sozialpsychologen, die sich im Feld des „Social Primings“ bewegen, nämlich den häufigen Gebrauch von Meta-Analysen und der Vorzug von groben konzeptuellen gegenüber detailliert protokollierten Replikationen. Das multiple Scheitern, fehlgeschlagene Priming-Experimente zu berichten, auch bekannt als „File-Drawer Problem“, sei ebenfalls ein Grund der wachsenden Zweifel für Kritiker der psychologischen Forschungsmethoden.

For all these reasons, right or wrong, your field is now the poster child for doubts about the integrity of psychological research. Your problem is not with the few people who have actively challenged the validity of some priming results. It is with the much larger population of colleagues who in the past accepted your surprising results as facts when they were published. These people have now attached a question mark to the field, and it is your responsibility to remove it. (Kahneman, 2012)

Dijksterhuis Unterstützer, etwa der Sozialpsychologe John Bargh von der *Yale University* in Connecticut, der in seinem weltbekannten Experiment untersuchte, dass Personen langsamer gehen, wenn sie zuvor mit dem Stereotyp alter Leute geprimt werden (Bargh, Chen & Burrows, 1996), verteidigen sich damit, dass die Replikation sozialer Priming Experimente extrem schwierig sei, da die kleinste Veränderung in der Versuchsanordnung, das Ergebnis beeinflussen könne. Hal Pashler, ein Kognitionspsychologe der *University of California*, San Diego, und langjähriger Kritiker des sozialen Primings, stellte dem entgegen, dass die Effekte, die in den Originalartikeln postuliert wurden, dermaßen groß waren, dass es sehr unwahrscheinlich ist, dass diese bei einem Replikationsversuch mit kleinsten Änderungen an den Versuchsbedingungen einfach verschwinden (Abbott, 2013).

Daniel Kahneman betont ebenfalls sein Verständnis dafür, dass es sehr schwierig ist, Priming Experimente zu replizieren, da es dazu jahrelange Erfahrung und hohes Geschick braucht, allerdings gehe es auch nach Shanks et al. (2013) um wesentlich mehr, als um einige gescheiterte Replikationsversuche (Abbott, 2013).

Kahneman sieht hauptsächlich den psychologischen Nachwuchs in Gefahr:

My reason for writing this letter is that I see a train wreck looming. I expect the first victims to be young people on the job market. Being associated with a controversial and suspicious field will put them at a severe disadvantage in the competition for positions. (Kahnemann, 2012)

Und auch die Weigerung Dijksterhuis, die genaue Versuchsanordnung seines Priming Experiments bekannt zu geben, schärfte die Vermutung bei Kritikern, dass es sich nach Stapel, Smeesters und Sanna, die alle drei in der Priming-Forschung tätig waren, erneut um einen niederländischen Betrugsfall in der Sozialpsychologie handeln könnte. Dijksterhuis hatte zunächst auf die wiederholten Anfragen, ein Protokoll zu veröffentlichen, mit den Worten: „Focusing on a single phenomenon is not that helpful and won't solve the problem“, geantwortet. Dem fügte er hinzu, dass die Sozialpsychologie zwar genauer arbeiten müsse, dass sich diese Gründlichkeit jedoch auf die Zukunft und nicht auf historische Experimente konzentrieren sollte. Er meinte, dass die Debatte um das *Social Priming* sowieso noch länger anhalten wird, da es da draußen eine Ideologie gäbe, die vehement nicht davon ausgehen wolle, dass unser Verhalten durch die Umwelt beeinflusst werden kann (Abbott, 2013).

Kahneman entgegnete der langen Abwehrhaltung Dijksterhuis mit Worten, die vermutlich die Gedanken vieler mitverfolgenden Leser widerspiegeln:

This refusal to engage in a legitimate scientific conversation invites the interpretation that the believers are afraid of the outcome. (Abbott, 2013).

Es muss hier allerdings explizit festgehalten werden, dass Ap Dijksterhuis bisher von keiner offiziellen Instanz des wissenschaftlichen Betruges beschuldigt und auch seine Artikel zum „Intelligence Priming“ nicht von betroffenen Fachzeitschriften zurückgewiesen wurden.

Dennoch besteht Kahneman in seinem Schreiben darauf, dass etwas gegen die wachsende Rufschädigung der Psychologie getan werden muss.

I believe that you should collectively do something about this mess. To deal effectively with the doubts you should acknowledge their existence and confront them straight on, because a posture of defiant denial is self-defeating. (Kahneman, 2012)

Kahnemans Lösungsvorschlag

Als ersten Lösungsansatz schlägt er vor, dass ein gemeinsamer Versuch gestartet werden sollte, die berühmtesten Priming Studien zu replizieren und dabei nach einem genauen Protokoll vorzugehen, welches die Glaubwürdigkeit für Kollegen außerhalb des Feldes widerherstellen würde.

Dazu sollte ein betreffender Forscher fünf Labors auswählen, dessen führende Wissenschaftler einen guten Ruf genießen und denen genügend Studenten als geeignete Versuchsteilnehmer zur Verfügung stehen. Jedes dieser fünf Labors solle sich einen bekannten Priming Effekt auswählen, von dem sie glauben, dass er robust und der Replikation würdig ist. Danach sollte von den obersten Verantwortlichen der Forschungseinrichtungen eine öffentliche Meldung zu den fünf spezifischen Effekten und deren Replikationsversuchen gemacht werden. Zur Umsetzung der Replikationen soll eine Labor-Kette nach dem Beispielmuster A-B-C-D-E-A gebildet werden, wobei jedes Labor die Studie replizieren soll, welche vom Nachbarinstitut ausgesucht wurde (z.B. B repliziert A, C repliziert B etc.). Um zu prüfen, ob das Experiment auch ordnungsgemäß durchgeführt wird, soll ein Mitarbeiter der benachbarten Forschungseinrichtung einen Kontrollbesuch abstatten. Damit die Replikation über ausreichend statistische Power verfügt, sollen genügend Versuchspersonen rekrutiert und zur sicheren Dokumentation der Durchführung, eine Videokamera benutzt werden. Zuletzt sollen alle Ergebnisse veröffentlicht und die Daten für weitere Analysen zur Verfügung gestellt werden.

Kahneman glaubt, dass ein Erfolg (diesen definiert er als 4-5 gelungene Replikationen von Priming Experimenten) das Feld umgehend rehabilitieren würde. Noch mehr widersprüchliche Ergebnisse wären natürlich schmerzhaft, allerdings könnte niemand den Ruf dieser Forscher in Frage stellen, die aufrichtig an ihrer Arbeit festhalten, auch wenn sie manchmal scheitern. Abschließend appelliert Kahneman (2012) an den Gemeinschaftssinn der Wissenschaftswelt:

The protocol I outlined is just an example of something you might do. The main point of my letter is that you should do something, and that you must do it collectively. No single individual will be able to overcome the doubts, but if you act as a group and avoid defensiveness you will be credible.

Dass Kahnemans E-Mail für großen Wirbel unter den angesprochenen Sozialpsychologen führte, bestätigte Brian Nosek, ein Sozialpsychologe der *University of Virginia* in Charlottesville: „Kahnemann is a hard man to ignore. I suspect that everybody who got a message from him read it immediately“ (Yong, 2012a).

Ob seine Forderung einer Forschungskette jedoch großen Anklang findet, bleibt fraglich, wie auch David Funder, ein Psychologe der *University Carolina* und gewählter Präsident der *Society for Personality and Social Psychology*, meint. Er glaubt zwar, dass Kahnemans E-Mail genau den richtigen Ton getroffen habe, um Änderungen zu bewirken, allerdings würden sich viele Kollegen dadurch angegriffen fühlen und die Debatte über soziales Priming würde allmählich in eine wütende Abwehrhaltung abdriften (Yong, 2012a).

Der Sozialpsychologe Norbert Schwarz von der *University of Michigan* in Ann Arbor erhielt ebenfalls Kahnemans E-Mail und wäre dazu bereit, an einer Forschungskette teilzunehmen, allerdings nur dann, wenn sich diese nicht nur auf einen einzigen Forschungsbereich konzentriert. Dieser Meinung stimmt Brian Nosek zu und hofft, dass ein solcher Forschungsansatz eine breitere Bewegung in der Psychologie auslösen wird, der irgendwann zur alltäglichen Vorgehensweise in der Wissenschaft wachsen wird (Yong, 2012a).

b. Zeitliche Ressourcenverschwendung durch den Studienaufbau auf manipulierten Erkenntnissen

Wesentlich schwieriger als die Suche nach direkten Replikationen Diederik Stapels Arbeiten, ist die Frage, welche Autoren ihre Studien auf den betrügerischen Inhalten des Niederländers aufgebaut und damit wertvolle zeitliche Ressourcen verschwendet haben. Zur Beantwortung dieser Frage wurde wiederum eine Literaturrecherche im *Web of Science* durchgeführt.

Dazu wurden alle 123 Einträge (diese Zahl entspricht der Gesamtanzahl aller Suchergebnisse minus der Anzahl der *Retraction Statements*) des Namens „Stapel, DA“, die eine „retracted“ Markierung (die offizielle Zurückweisung des Artikels von der jeweiligen Fachzeitschrift) aufwiesen, nach ihren Folgezitierungen durch andere Autoren untersucht. Insgesamt wurden 47 Artikel Stapels im *Web of Science* als zurückgewiesen

markiert, weitere 7 finden sich auf einer online Webseite, die sich „Retraction Watch“ nennt (<http://retractionwatch.com/>) und auf der die aktuellsten Zurückweisungen Stapels Artikel durch die betreffenden Fachzeitschriften angegeben werden. Diese 7 zurückgewiesenen Artikel sind derzeit im *Web of Science* noch nicht offiziell als solche markiert, daher wurden sie in die folgende Recherche nicht miteinbezogen. Die 47 Artikel Stapels wurden insgesamt 444 Mal zitiert, davon wurden die Selbstzitate, Zitierungen in Büchern und das jeweilige *Retraction Statement* abgezogen und die verbliebenen 295 Artikel auf den Aufbau auf Stapels fälschlichen Ergebnissen untersucht. Abbildung 3 stellt diesen Analyse Prozess graphisch dar.

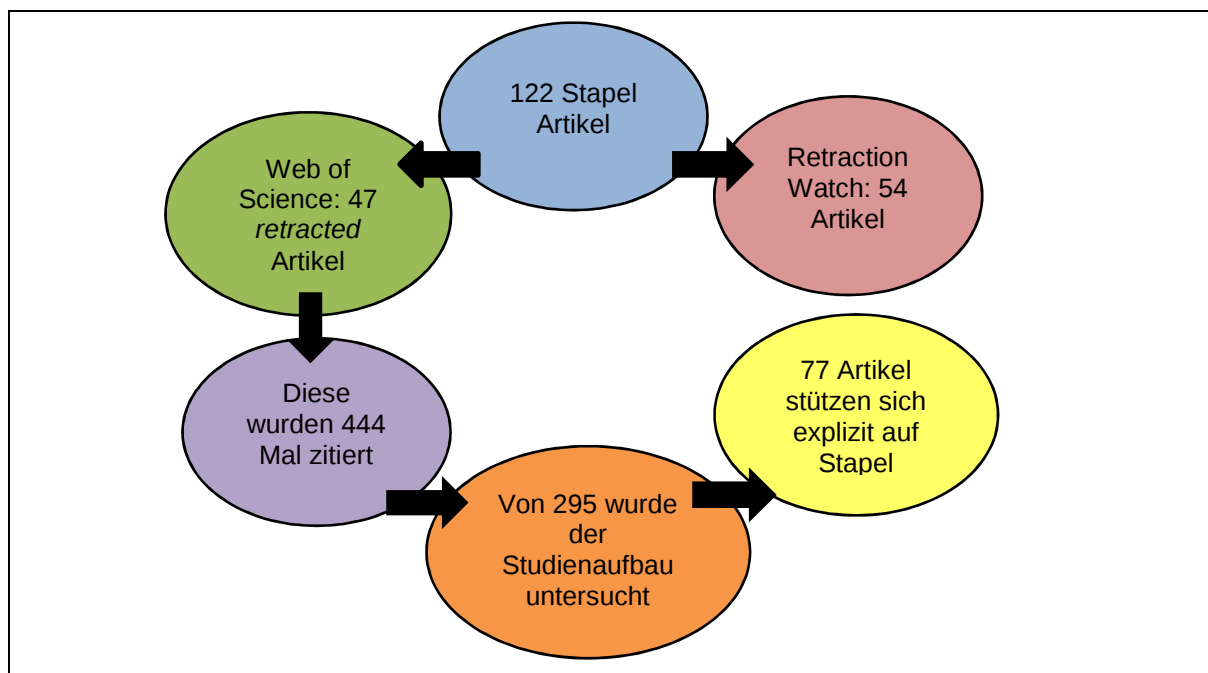


Abbildung 3: Literaturrecherche

Für insgesamt 77 Artikel wurde befunden, dass sie explizit auf den Erkenntnissen eines oder mehrerer zurückgewiesener und damit betrugshaften Studien Stapels aufbauen. Als Kriterien für den expliziten Aufbau auf Stapels Erkenntnissen wurden folgende definiert:

1. Formulierte Hypothesen wurden explizit von Stapels Studienergebnissen abgeleitet beziehungsweise trugen seine Erkenntnisse in entscheidender Weise zum Versuchsaufbau bei. Die bloße Erwähnung seines Namens im Theorieteil, da er beispielsweise gemeinsam mit vielen anderen Wissenschaftlern zur Erforschung eines Themas beigetragen hat, reichte nicht aus, um in die Liste der aufbauenden Artikel aufgenommen zu werden.

2. Im zitierenden Artikel wurde eine von Stapel und Kollegen entwickelte Priming Methode oder der Aufbau eines Versuchsdesigns übernommen.
3. Theoretische Metaanalysen, die ohne den Erkenntnissen Stapels manipulierter Studien nicht ausgekommen wären.

In Tabelle 4 sind die Ergebnisse dieser Literaturrecherche dargestellt. Die 47 zurückgewiesenen Artikel im *Web of Science* wurden nach ihrer Zitathäufigkeit gerangreicht.

Spalte 6 stellt in dieser Tabelle die Anzahl untersuchter Artikel (alle angezeigten Zitate minus Stapels Selbstzitate, *Retraction Statements* und Buchbeiträge) dar. Darin sind einige Einträge mit einem oder zwei Sternen (*) markiert. Dies bedeutet, dass ein (*) oder zwei (**) Artikel davon Diederik Stapel zitieren, diese Artikel im Fachzeitschriftenkatalog der Universität Wien jedoch nicht frei zugänglich waren. Daher wurden die Autoren per E-Mail kontaktiert und die gewünschten Artikel angefordert. Alle fehlenden Artikel wurden mir netterweise zugeschickt und konnten auf deren Studienaufbau untersucht werden, bis auf die 6 markierten, die leider nicht in englischer oder deutscher Sprache existieren. Dabei handelt es sich um folgende Artikel, die entweder in einer französischen oder niederländischen Fachzeitschrift veröffentlicht wurden. Dazu angegeben ist jeweils die Artikelnummer ihres zugehörigen zitierenden Stapel Artikels (siehe Tabelle 4).

Artikel Nr. 4:

Berjot, S., Girault-Lidvan, N., Gillet, N. et al. (2010). How french students from arabic origin appraise and cope with stereotype threat. *Année Psychologique*, 110, 427-451.

Artikel Nr. 7:

Brunot, S. & Juhel, J. (2012). Social and temporal comparisons, self-esteem and job-search activity among long-term unemployed people. *Année Psychologique*, 112, 197-226.

Dijkstra, P., Barelds, D. P. H. & Kuyper, H. (2011). Social comparison in the Dutch classroom: Discussion and conclusion. *Pedagogische Studien*, 88, 204-213.

Artikel Nr. 12:

Chateignier, C., Chekroun, P., Nugier, A. et al. (2011). „Death awaiting women drivers“: Effects of stereotype threat and anger on women performance to a driving rules task. *Année Psychologique*, 111, 673-700.

Artikel Nr. 15:

De Kort, M. J. M. & Velthuijsen, A. S. (2011). Do you wash your hands after using the bathroom? Two studies about communicating the injunctive and descriptive norm and practicing informational and normative social influence in order to promote handwashing behavior. *Tijdschrift voor Communicatiewetenschap*, 39, 41+.

Artikel Nr. 28:

Zwier, S. & Bolink, W. (2011). „Betaserc (R) is better“. The third-person effect in reactions to direct-to-consumer advertising. *Psychologie & Gezondheid*, 39, 224-231.

Tabelle 4: Ergebnisse der Literaturrecherche

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
1	Stapel, D. A., Koomen, W. & van der Pligt, J. (1996)	54	19	2	33	13
2	Trampe, D., Stapel, D. A. & Siero, F. W. (2007)	51	3	2	46	5
3	Stapel, D. A. & Semin, G. R. (2007)	37	3	7	27	9
4	Marx, D. M. & Stapel, D. A. (2006)	24	2	0	22*	7
5	Johnson, C. S. & Stapel, D. A. (2007)	21	3	1	17	7
6	Stapel, D. A. & Spears, R. (1996)	20	7	1	12	4
7	Blanton, H. & Stapel, D. A. (2008)	18	3	0	15**	3
8	Stapel, D. A. & Koomen, W. (2006)	15	7	0	8	0

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
9	Avramova, Y. R. & Stapel, D. A. (2008)	14	6	0	8	3
10	Renkema, L. J., Stapel, D. A., Maringer, M. & van Yperen, N. W. (2008)	13	4	1	8	2
11	Stapel, D. A. & Schwarz, N. (1998)	13	3	1	9	4
12	Marx, D. M. & Stapel, D. A. (2006)	12	1	0	11*	2
13	Ruys, K. I. & Stapel, D. A. (2008)	12	4	0	8	1
14	Johnson, C. S. & Stapel, D. A. (2007)	9	6	0	3	1
15	Joly, J. F., Stapel, D. A. & Lindenberg, S. M. (2008)	9	3	0	6*	1
16	Noordewier, M. K. & Stapel, D. A. (2010)	8	6	0	2	0

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
17	Stapel, D. A. & Koomen, W. (2001)	8	5	0	3	0
18	Maringer, M. & Stapel, D. A. (2007)	7	3	1	3	0
19	Stapel, D. A. & Johnson, C. S. (2007)	7	3	0	4	2
20	Stapel, D. A. & Marx, D. M. (2007)	7	2	1	4	1
21	Avramova, Y. R., Stapel, D. A. & Lerouge, D. (2010)	7	1	0	6	1
22	Stapel, D. A. & Marx, D. M. (2006)	7	4	0	3	1
23	Stapel, D. A. & Lindenberg, S. (2011)	6	1	0	5	2
24	Häfner, M. & Stapel, D. A. (2010)	5	2	0	3	2

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
25	Gordijn, E. H. & Stapel, D. A. (2006)	5	1	0	4	1
26	Van den Bos, A. & Stapel, D. A. (2009)	5	3	0	2	0
27	Johnson, C. S., Norton, M. I., Nelson, L. D., Stapel, D. A. & Chartrand, T. L. (2008)	5	4	0	1	0
28	Trampe, D. & Stapel, D. A. (2010)	4	1	0	3*	0
29	Stapel, D. A. & Noordewier, M. K. (2009)	4	3	0	1	0
30	Stapel, D. A. & van der Linde (2011)	4	1	0	3	1
31	Ruys, K. I. & Stapel, D. A. (2009)	4	2	0	2	1
32	Noordewier, M. K. & Stapel, D. A. (2009)	3	1	0	2	0

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
33	Meijers, M. H. C. & Stapel, D. A. (2011)	3	1	0	2	1
34	Joly, J. F. & Stapel, D. A. (2009)	2	1	0	1	1
35	Johnson, C. S. & Stapel, D. A. (2011)	2	1	0	1	1
36	Trampe, D., Stapel, D. A. & Siero, F. W. (2011)	2	1	0	1	0
37	Johnson, C. S. & Stapel, D. A. (2010)	2	1	0	1	0
38	Stapel, D. A. & Marx, D. M. (2007)	2	1	0	1	0
39	Lammers, J. & Stapel, D. A. (2011)	2	1	0	1	0
40	Schwinghammer, S. A. & Stapel, D. A. (2011)	2	1	0	1	0

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
41	Noordewier, M. K., van Horen, F., Ruys, K. I. & Stapel, D. A. (2010)	2	1	0	1	0
42	Lindenberg, S., Joly, J. F. & Stapel, D. A. (2011)	2	1	0	1	0
43	Noordewier, M. K. & Stapel, D. A. (2011)	1	1	0	0	0
44	Stapel, D. A. & Noordewier, M. K. (2011)	1	1	0	0	0
45	Noordewier, M. K. & Stapel, D. A. (2011)	1	1	0	0	0
46	Avramova, Y. R., Stapel, D. A. & Lerouge, D. (2010)	1	1	0	0	0
47	Nelissen, R. M. A., van de Ven, N. & Stapel, D. A. (2011)	1	1	0	0	0

Artikel Nr.	Autoren	Gesamt zitiert	Selbstzitationen + Retraction (S+R)	Beiträge in Büchern (B)	Gesamt zitiert minus (S+R+B)	Auf Stapel aufbauende Artikel
gesamt		444	132	17	295 (289)	77

Inhaltliche Darstellung der auf Stapel aufbauenden Artikel

Im Folgenden wird dargestellt, welche Erkenntnisse die Autoren der 77 Artikel aus den Studien von Stapel und Kollegen übernommen haben. Die Artikel Stapels sind nach ihrer Zitathäufigkeit von 1 bis 47 nummeriert und in den mit Pfeilen markierten Aufzählungen finden sich die jeweils zitierenden Artikel weiterer Forscher. Alle Zitate entstammen den genannten Originalartikeln von Stapel und Kollegen beziehungsweise den angegebenen zitierenden Artikeln und können dort nachgelesen werden. Zu jedem der 47 Stapel Artikel wurde das Original *Retraction Statement* hinzugefügt.

1. Stapel, D. A., Koomen, W. & van der Pligt, J. (1996). The referents of trait inferences: The impact of trait concepts versus actor-trait links on subsequent judgments. *Journal of Personality and Social Psychology*, 70, 437-450. (Retracted article. See vol. 103, pg. 565, 2012)

Stapel et al. untersuchten in ihrem Artikel, welchen Unterschied die Aktivierung von Eigenschaftskonzepten (durch Priming) im Vergleich zur Aktivierung von Personen-Eigenschaftsverbindungen auf ein nachfolgendes soziales Urteil einer Person ausmachen kann. Wenn sich geprimte Eigenschaftsbeschreibungen auf ein abstraktes Verhalten beziehen, dann dienen sie als grundlegender Interpretationsrahmen und führen im nachfolgenden Urteil zu Assimilationseffekten bezüglich des zu bewertenden Objekts, während Beschreibungen, die sich auf eine spezifische Personen-Eigenschaftsverbindung beziehen, als Vergleichsanker dienen und Kontrasteffekte in der Beurteilung auslösen. Versuchspersonen, die sich im 1. Experiment Eigenschafts-implizierende Sätze einprägen sollten („memorization“ Bedingung), zeigten einen Assimilationseffekt, wohingegen Versuchspersonen, die sich einen konkreten Eindruck von der beschriebenen Zielperson in den Sätzen verschaffen sollten („impression“ Bedingung), einen Kontrasteffekt zeigten. In Experiment 1 konnte demonstriert werden, dass Eigenschafts-implizierende Sätze, in denen Person mit echtem Namen und einem Foto vorgestellt wurden, sowohl in der „memorization“ als auch in der „impression“ Bedingung in einem Kontrasteffekt bezüglich der Personenbewertung resultierten.

Stapel et al.s Artikel wurde 54 Mal zitiert, dabei handelt es sich bei einer Zitation um das *Retraction Statement* der publizierenden Fachzeitschrift und bei 18 Einträgen um Selbstzitate. Folgende Artikel bauen auf den Erkenntnissen von Stapel, D. A., Koomen, W. & van der Pligt, J. (1996) auf:

→ Graffin, S. D., Boivie, S. & Carpenter, M. A. (2013). Examining CEO succession and the role of heuristics in early-stage CEO evaluation. *Strategic Management Journal*, 34, 383-403.

Stapel et al.s Annahme folgend - je spezifischer ein Vergleichsobjekt, etwa eine Person oder eine konkrete Eigenschaft, desto eher entsteht ein Kontrasteffekt – stützen die Autoren ihre Hypothese 1a und 1b auf der Tatsache, dass ein neuer CEO (*Chief executive officer* = US-amerikanische Bezeichnung für Vorsitzender, Präsident, Geschäftsführer) im Vergleich zum geprimten ehemaligen Star CEO von einem Expertengremium eines Unternehmens schlechter bewertet wird.

Hypothesis 1a: Succeeding a star CEO is negatively related to changes in the newly appointed CEO's compensation in the year following the CEO succession.

Hypothesis 1b: Succeeding a star CEO is positively related to the likelihood of a CEO being dismissed early in his or her tenure (within two years).

→ Smeesters, D., Wheeler, S. C. & Aaron, C. (2010). Indirekt prime-to-behavior effects: The role of perceptions of the self, others and situations in connecting primed constructs to social behavior. *Advances in Experimental Social Psychology*, 42, 259-317.

Smeesters et al. geben in ihrem Review-Artikel einen Überblick über die Forschung der letzten Jahre im Bereich des indirekten Primings. Indirektes Priming beeinflusst das soziale Verhalten von Individuen (im Vergleich zu einer direkten Prime-Verhaltens-Verbindung) durch die Änderung unterschiedlicher Wahrnehmungsprozesse. Die Autoren beschreiben in ihrem Rückblick drei indirekte Priming Mechanismen: die Selbstwahrnehmung, die Fremdwahrnehmung und situative Wahrnehmungsprozesse und zitieren dabei die Forschungsergebnisse ganzer neun Artikel von Stapel und Kollegen, von denen drei von den jeweiligen Fachzeitschriften zurückgewiesen wurden (mit * gekennzeichnet) und vier von der Stapel Untersuchungskommission (Stapel Investigation, 2012) als „fraudulent“ (mit # gekennzeichnet) betitelt wurden (Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D. A., Koomen, W., van Knippenberg, A. et al., 1998; Marx, D. M. & Stapel, D. A., 2006; Spears, R., Gordijn, E., Dijksterhuis, A. & Stapel, D. A., 2004; Stapel, D. A., 2007#; Stapel, D. A. & Koomen, W., 2001; Stapel, D. A., Koomen, W. & van der Pligt, J., 1996#*; Stapel, D. A., Koomen, W. & van der Pligt, J., 1997; Stapel, D. A. & Suls#*, J., 2004; Stapel, D. A. & van der Zee, K. I., 2006#*).

→ Bless, H. & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. *Advances in Experimental Social Psychology*, 42, 319-373.

Bless und Schwarz geben in ihrem Artikel eine Zusammenfassung über die Forschungsergebnisse der vergangenen Jahre zum „Inclusion-Exclusion“ Modell (IEM; Schwarz & Bless, 1992a), in dem erklärt wird, wann die Aktivierung eines mentalen Konzepts zu einem Assimilationseffekt und wann zu einem Kontrasteffekt im menschlichen Verhalten oder in sozialen Urteilen führt. Da Stapel und Kollegen einen wesentlichen Beitrag zu den Grundlagen und zur Erweiterung des IEMs geleistet haben, werden sie in diesem Artikel auch 8 Mal zitiert, wovon zwei Artikel offiziell zurückgewiesen und vier von der Untersuchungskommission (Stapel Investigation, 2012) als „fraudulent“ gekennzeichnet wurden (Stapel, D. A., 2007#; Stapel, D. A. & Blanton, H., 2004#*; Stapel, D. A. & Koomen, W., 1998a#; Stapel, D. A. & Koomen, W., 1998b; Stapel, D. A. & Koomen, W., 2001; Stapel, D. A., Koomen, W. & van der Pligt, J., 1996#*; Stapel, D. A., Koomen, W. & van der Pligt, J., 1997; Stapel, D. A. & Schwarz, N., 1998; Stapel, D. A. & Winkielman, 1998).

→ Ledgerwood, A. & Chaiken, S. (2007). Priming us and them: Automatic assimilation and contrast in group attitudes. *Journal of Personality and Social Psychology*, 93, 940-956.

Ledgerwood und Chaiken primten in ihrer ersten Studie die Versuchspersonen subliminal mit ihrer politischen „ingroup“ oder „outgroup“, wodurch sie im Vergleich zur Kontrollgruppe (ohne Priming) extremere politische Einstellungen angaben. In ihrer zweiten Studie demonstrierten die Autoren, dass die Art des Prime-Stimulus den Polarisierungseffekt in den *ingroup* und *outgroup* Bedingungen (durch Assimilation oder Kontrast) zusätzlich verstärkt. Die Ergebnisse lassen vermuten, dass ein Wechsel von einem Assimilations- zu einem Kontrasteffekt dann stattfindet, wenn eine Gruppe nicht mehr innerhalb eines Kontexts, sondern als Vergleichsanker wahrgenommen wird.

Die Autoren stützen sich in ihrer Studie auf Stapel et al.s Theorie, dass bestimmte Charakteristika eines Prime-Stimulus bestimmen, ob Assimilations- oder Kontrasteffekte auftreten. Wenn ein Stimulus von einschließender, abstrakter und moderater Natur ist (z.B. die allgemeine Eigenschaft „fröhlich“), dann führt er eher zu einem Assimilationseffekt, wohingegen ein konkreter, eng gefasster und extremer Prime (z.B. eine spezifische Person wie „Santa Claus“) eher einen Kontrasteffekt verursacht. Stapel et al. (1996) ließen die Versuchspersonen in ihrer Studie Schlüsse aus entweder weiter gefassten Verhaltensmustern oder einer spezifischen Personen-Eigenschaftsverbinding

ziehen, indem sie ihnen Sätze mit unterschiedlicher Informations-Spezifität präsentierten (z.B. eine Eigenschafts-implizierende Handlung in einem Satz wurde durch ein Foto der handelnden Person ergänzt). Danach wurde den Versuchspersonen ein Absatz über einen gewissen „Donald“ zu lesen gegeben, welcher mehrdeutige Verhaltensbeschreibungen enthielt. Anschließend sollten die Versuchspersonen Donald nach bestimmten Kriterien beurteilen. Versuchspersonen, die die Sätze in der „memorization“ Bedingung lasen, näherten sich in ihrer Bewertung den aktivierten Eigenschaften entsprechend an, wenn sie zuvor durch weit gefasste Eigenschaftsbeschreibungen geprimt wurden, zeigten jedoch einen Kontrasteffekt, wenn sie durch eine spezifische Personen-Eigenschaftsverbindung (Foto) geprimt wurden. Weit gefasste Verhaltensbeschreibungen schufen nach Stapel et al. einen einschließenden Bezugsrahmen, innerhalb dessen Donalds Verhalten interpretiert wurde, während eine enger gefasste Personen-Eigenschaftsverbindung als Ankerpunkt fungierte, der Donalds Verhalten durch einen Vergleich kontrastierte.

In ihrer Studie schlussfolgerten Ledgerwood et al., dass Stapels Erkenntnisse zu Assimilations- und Kontrasteffekten auch auf den Bereich der Einstellungsänderung angewandt werden können, so dass die Aktivierung einer sozialen Gruppe, sofern diese als impliziter Kontext (*ingroup*) für die eigene Einstellung fungiert, zu einem Assimilationseffekt führt, wohingegen das Priming mit einer Gruppe als Vergleichsstandard (*outgroup*) einen Kontrasteffekt verursacht. Die Autoren stellten daher in ihrer ersten Studie die Hypothesen auf, dass a) Versuchspersonen, die mit ihrer politischen *ingroup* geprimt werden, eine vergrößerte Einstellungspolarisation zeigen (Zustimmung mit der *ingroup* und Ablehnung der *outgroup*) und b), dass Versuchspersonen, die mit ihrer politischen *outgroup* geprimt werden, ebenfalls eine vermehrte Einstellungspolarisation zeigen (Ablehnung der *outgroup* und Zustimmung mit der *ingroup*).

→ DeCoster, J. & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational bias. *Personality and Social Psychology Review*, 8, 2-27.

In ihrem Review-Artikel fassen die Autoren drei mögliche Effekte zusammen, die infolge einer Informationspräsentation in einem bestimmten Kontext, auf die Einstellungen von Personen bezüglich eines sozialen Ziels in einem anderen Kontext auftreten können. Wenn der Prime-Stimulus in den Eindruck des zu bewertenden Ziels integriert werden kann, entsteht ein Assimilationseffekt, wenn er als Vergleichsstandard herangezogen wird, resultiert ein Kontrasteffekt und manchmal kann ein Prime-Stimulus auch dazu

führen, dass eine Person ihr Urteil bezüglich eines Objekts bewusst verändert (= Korrektur).

Die Autoren zitieren in ihrem Review acht Artikel von Stapel und seinen Kollegen (Stapel & Koomen, 1996#; Stapel & Koomen, 1997#; Stapel & Koomen, 1998; Stapel & Koomen, 2001; Stapel, Koomen & van der Pligt, 1996*#; Stapel, Koomen & van der Pligt, 1997; Stapel, Koomen & Zeelenberg, 1998; Stapel & Schwarz, 1998*#), wovon zwei Artikel von der jeweiligen Fachzeitschrift zurückgewiesen und vier von der Untersuchungskommission als betrugshaft identifiziert wurden. Die Hälfte der zitierten Stapel Literatur beruht also auf betrügerischer Forschung.

→ Markman, K. D. & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review*, 7, 244-267.

Der Artikel von Marman & McMullen gibt einen Überblick über die bestehende Literatur zu sozialen, temporalen und widersprüchlichen Vergleichen und legt das „Reflection and Evaluation Model“ (REM) als organisierende Instanz fest. Im Mittelpunkt des REMs stehen zwei unterschiedliche psychologische Mechanismen der mentalen Simulation, die während des vergleichenden Denkens aktiv sind: die Reflexion („als wenn“-Denken), welche simuliert, dass die Information über das zu vergleichende Ziel der Wahrheit oder einem Teil des Selbst entspricht, und der Evaluation, bei der die Information über das zu vergleichende Ziel als Referenzpunkt herangezogen wird, anhand dessen der eigene gegenwärtige Standpunkt evaluiert wird. Reflexion tritt auf, wenn die Information über das Ziel in das eigene Selbstkonzept integriert werden kann und führt anschließend zur affektiven Assimilation. Evaluation entsteht, wenn die Zielinformation vom Selbstkonzept ausgeschlossen wird, wodurch ein affektiver Kontrasteffekt resultiert.

In ihrer Spezifikation der zugrundeliegenden Mechanismen, die im vergleichenden Denken zu Assimilations- oder Kontrasteffekten führen können, haben die Autoren ihre Annahmen auf zwei theoretische Modelle gestützt, die sie zu ihrem „Reflektion and Evaluation Model“ integriert haben: Das „Inclusion-Exclusion Model“ von Schwarz und Bless (1992) und Stapel und Koomen (2000) und das „Selective-Accessibility Model“ von Mussweiler (2003).

Die Autoren zitieren in ihrer Ableitung die wesentlichsten Erkenntnisse aus Stapels sozialer Vergleichsforschung und stützen sich auf fünf Artikel des Niederländers und seiner Kollegen (Dijksterhuis, Spears, Postmes et al., 1998; Stapel & Koomen, 2000; Stapel & Koomen, 2001#; Stapel, Koomen & van der Pligt, 1996*#; Stapel, Koomen & van der Pligt, 1997), von denen einer von der betreffenden Fachzeitschrift zurückgewiesen

wurde und in zweien Hinweise auf betrügerische Inhalte von der Stapel Untersuchungskommission (2012) festgestellt werden konnten.

→ Innes-Ker, A. & Niedenthal, P. M. (2002). Emotion concepts and emotional states in social judgment and categorization. *Journal of Personality and Social Psychology*, 83, 804-816.

Die Aufgabe „Unscrambling Sentences“ wurde entwickelt, um die Konzepte Glück, Traurigkeit oder neutrale Ideen zu primen. In Pilot Studien konnte gezeigt werden, dass das Entwirren von emotionalen Sätzen nicht direkt die emotionale Stimmung beeinflusst, sondern semantisch verwandte Wörter aktiviert. In Experiment 1 demonstrierten die Autoren, dass die Aktivierung einer emotionalen Stimmung (durch Musik oder Filme), und nicht die „Unscrambling Sentences“ Aufgabe, emotions-kongruente Urteile verursacht. Ihr Versuchsdesign in Experiment 1 stützen die Autoren unter anderem auf ähnliche Experimente von Stapel und Kollegen, in denen die Erfassung von Priming-Effekten, durch die Vorgabe von Prime-Stimuli in einer Aufgabe, und die resultierende emotionale Aktivierung in einer zweiten Aufgabe (z.B. durch die Vorgabe einer Geschichte, die mehrdeutig interpretierbare Informationen über die emotionale Stimmung der handelnden Person enthält), gemessen wurde.

→ Dijksterhuis, A. & Smith, P. K. (2002). Affective habituation: Subliminal exposure to extreme stimuli decreases their extremity. *Emotion*, 2, 203-214.

Die Autoren gehen in ihrer Studie davon aus, dass die subliminale Präsentation von extrem positiven oder negativen Stimuli dazu führt, dass diese danach weniger extrem wahrgenommen werden. In Experiment 1 wurden die Versuchspersonen zuvor mit extrem positiven oder negativen Wörtern geprimt und schätzten ihre Extremität in einer nachfolgenden Beurteilungsaufgabe geringer ein, als Wörter, mit denen sie zuvor nicht geprimt worden waren. In Experiment 2 wurden diese Ergebnisse mit einem impliziten Evaluationsmaß als Extremitätseinschätzung repliziert. In den Experimenten 3 und 4 führte das subliminale Priming mit extrem positiven oder negativen Wörtern dazu, dass diese danach nur mehr als moderat positiv oder negativ eingeschätzt wurden, im Vergleich zu extremen Wörtern, die ohne einer Prime-Bedingung präsentiert wurden.

Wenn ein Ziel-Stimulus zuvor mit einem positiven oder negativen Stimulus geprimt wird, dann kann dieser Prime-Stimulus die Reaktion einer Person auf den Zielstimulus beeinflussen. Zum Beispiel wenn Versuchsteilnehmer so schnell wie möglich entscheiden sollen, ob das Wort „Liebe“ ein positives oder negatives ist, dann reagieren die

Versuchsteilnehmer schneller, wenn vor dem Wort „Liebe“ ein anderes positives Wort präsentiert wurde. In diesem Fall würde also ein Assimilationseffekt entstehen. Stapel et al. (und auch andere Autoren z.B. Glaser & Banaji, 1999) demonstrierten in ihren Studien, dass ein Kontrasteffekt dagegen dann resultiert, wenn der vorangehende Prime ein extremer ist. Versuchsteilnehmer würden also länger brauchen, um das Wort „Liebe“ als positives zu identifizieren, wenn sie zuvor ein extrem positives Wort präsentiert bekommen hätten. In ihren Experimenten 1-4 replizierten die Autoren diese Erkenntnisse und gingen daher von der Hypothese aus, dass moderate Primes zu Assimilation und extreme Primes zu Kontrasteffekten führen.

→ Dijksterhuis, A., Spears, R. & Lepinasse, V. (2001). Reflecting and deflecting stereotypes: Assimilation and contrast in impression formation and automatic behavior. *Journal of Experimental Social Psychology*, 37, 286-299.

Die Autoren stützen ihre Grundannahmen über die zugrunde liegenden Mechanismen der Assimilation und des Kontrastes im automatischen Verhalten unter anderem auf die Arbeiten von Stapel und seinen Kollegen. Diese fanden heraus, dass Assimilation in sozialen Urteilsprozessen dann auftritt, wenn abstrakte Konzepte (z.B. Eigenschaften oder Stereotype) als Interpretationsrahmen für die geprimte Information dienen. Die Aktivierung der Eigenschaft „feindselig“, bevor Versuchsteilnehmer einen Mann namens „Donald“ beurteilen mussten, führte dazu, dass sein Verhalten dann als feindselig interpretiert wurde, wenn dieses Hand in Hand mit dem aktivierten Konzept ging. Konkrete Beispiele führten dagegen zu Kontrasteffekten, da sie als Vergleichsstandard herangezogen wurden. Die zu beurteilende Person (Donald) wurde mit dem präsentierten Beispiel (Adolf Hitler) verglichen und erschien dadurch weniger feindselig.

→ Moskowitz, G. B. & Skurnik, I. W. (1999). Contrast effects as determined by the type of prime: Trait versus exemplar primes initiate processing strategies that differ in how accessible constructs are used. *Journal of Personality and Social Psychology*, 76, 911-927.

Die Autoren demonstrierten in vier Experimenten, dass Kontrasteffekte in der Personenwahrnehmung vom Prime-Typ und der Extremität der geprimten Information abhängen. Zwei Priming-Effekt-Modelle, das „Standard Comparison Model“ (von Herr, 1986; Stapel et al., 1996) und das „Set-Reset Model“ (Martin, 1986; Martin, Seta & Crelia, 1990) gehen von unterschiedlichen Annahmen bezüglich des Einflusses der Prime-Extremität auf Kontrasteffekte aus. Konkrete Beispiele (z.B. „Dracula“) bewirkten einen größeren Kontrast, wenn sie extrem anstatt moderat waren und bestätigten damit die

Annahme des „Standard-of-comparison“ Modells, bei dessen Entwicklung Stapel et al. (1996) maßgeblich beteiligt waren. Geprimte moderate Eigenschaftsbegriffe (z.B. „übelwollend“) führten zu größeren Kontrasteffekten als extreme Eigenschaftsbegriffe und bestätigten die Annahme des „Set-Reset-Modells“. In den Experimenten 3 und 4 konnte demonstriert werden, dass die kontrastauslösenden Mechanismen für beide Prime-Typen unterschiedlicher Natur sind. Bei ersterem Modell werden Primes als Vergleichsanker genutzt, mit Hilfe dessen Urteile gefällt werden. Je extremer ein Prime in Bezug auf das zu bewertende Ziel ist, desto größer fällt der Kontrasteffekt aus. Im „Set-Reset“-Modell hängen Assimilation und Kontrast davon ab, ob Personen das geprimte Konstrukt als fehlerhafte Information wahrnehmen. In ihrem Versuch, alle möglichen Fehlerquellen aus ihren Urteilen auszuschließen, korrigieren sie ihr Urteil, indem sie sich von der störenden Informationsquelle distanzieren. Die Schwierigkeit zu differenzieren, ob ihr Urteil nun vom zu beurteilenden Objekt oder vom geprimten Konstrukt beeinflusst wurde, führt dazu, dass entweder ein Assimilations- oder Kontrasteffekt auftritt. Kontrast entsteht dann, wenn einer Person zunächst die geprimte Information als Reaktion auf das zu bewertende Ziel zur Verfügung steht, sie ihr Urteil danach jedoch korrigiert, da sie darin den Prime-Einfluss erkennt.

→ Lepore, L. & Brown, R. (2002). The role of awareness: Divergent automatic stereotype activation and implicit judgment correction. *Social Cognition*, 20, 321-351.

Lepore und Brown untersuchten in zwei Experimenten die automatische Aktivierung von Stereotypen und die Korrektur dieser Stereotype. In ihrem ersten Experiment primten sie ihre Versuchspersonen mittels einer Lexikon-Entscheidungsaufgabe mit der Kategorie „Blacks“. Personen mit großen Vorurteilen gegenüber schwarzen Personen beurteilten diese nach dem Priming negativer als Personen mit geringeren Vorurteilen. In Experiment 2 konnten die Autoren demonstrieren, dass das Bewusstsein um die Verbindung zwischen dem vorhergehenden Priming und der anschließenden Bewertung von Zielpersonen zu einem umgekehrten Effekt in der Beurteilung der Versuchspersonen führte. Personen, die sich der externen Manipulation bewusst waren, korrigierten ihre Urteile über die Zielperson in die entgegengesetzte Richtung (z.B. Personen mit großen Vorurteilen bewerteten schwarze Zielpersonen besser). Ihre Hypothesen stützten Lepore und Brown unter anderem auf die Erkenntnisse von Stapel und Kollegen (Stapel, Koomen & van der Pligt, 1996; Stapel, Koomen & Zeelenberg, 1998), welche Kontrasteffekte bereits in zwei Kategorien einteilten, nämlich in Vergleichs- und Korrektoreffekte. Ein vergleichender Kontrasteffekt entsteht dann, wenn ein konkretes Beispiel oder Sätze, die eine konkrete Verbindung zu einer Person herstellen, als Vergleichsstandard dienen, mit

denen das Zielobjekt verglichen wird. Der Vergleich zwischen dem Prime-Stimulus und dem Zielobjekt verlangt nicht unbedingt das Bewusstsein um diesen Prozess, wohingegen der korrigierende Kontrast dies sehr wohl tut. Nach dem „Set-Reset model of contrast effects“ (Martin & Stapel, 1998) kommt es zur Korrektur des eigenen Urteils, wenn man sich dessen bewusst wird, dass die geprimte Information die eigene Bewertung beeinflussen könnte.

→ Zarate, M. A., Uleman, J. S. & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, 19, 295-323.

In zwei Experimenten untersuchten die Autoren Zarate et al. (2001) kulturelle Unterschiede in der spontanen Anwendung von Eigenschaften. In einer Diskussion über unterschiedliche Rückschlüsse auf Personeneigenschaften in einer ihrer Lehrveranstaltungen antworteten 95% der anwesenden Studenten auf die erste assoziierte Eigenschaft zu dem Satz „He wondered where the stars come from“ mit „curious“, was übersetzt so viel wie *neugierig* oder *eigenartig* bedeutet. Ein Student indischer Abstammung antwortete dahingegen sofort mit „child“ und unterstützte damit die bereits erforschte Theorie, dass kollektivistische Kulturen eher Bezug auf soziale Rollen nehmen, wohingegen individualistische europäisch-amerikanische Kulturen spontane Rückschlüsse auf die Eigenschaften der handelnden Person ziehen. Warum es zu diesen unterschiedlichen Schlussfolgerungen bezüglich assoziierter Eigenschaften einer aktivierten Information kommt, versuchten die Autoren in ihren beiden Studien herauszufinden. Grundsätzlich sind bei spontanen Rückschlüssen auf Personeneigenschaften (STIs = *Spontaneous trait inferences*: „Trait inferences that occur unintentionally, almost effortlessly and usually without awareness“) zwei Prozesse involviert: Die Aktivierung eines Eigenschaftskonzepts und die Verbindung dieses Konzepts an eine aktuell verfügbare Person oder eine handelnde Person aus dem Langzeitgedächtnis. Unterschiede in den STIs können beispielsweise durch die Reaktionszeiten in einer Lexikon-Entscheidungsaufgabe gemessen werden.

In ihrem theoretischen Aufbau berufen sich Zarate et al. auf die Erkenntnisse von Stapel und Kollegen (1996), die bereits zeigen konnten, dass unterschiedliche Prime-Stimuli (das Lesen eigenschaftsimplizierender Sätze vs. Sätze, die eine Personen-Eigenschaftsverbindung herstellen) zu unterschiedlichen Effekten (Assimilation vs. Kontrast) in der Personenbeurteilung führen können.

→ Corneille, O., Vescio, T. K. & Judd, C. M. (2000). Incidentally activated knowledge and stereotype based judgments: A consideration of primed construct-target attribute match. *Social Cognition*, 18, 377-399.

Aus langjähriger empirischer Forschung steht mittlerweile fest, dass Priming Effekte durch unterschiedliche Faktoren, wie die Extremität des Prime-Stimulus, dessen Anwendbarkeit, dessen Deutlichkeit und die Ähnlichkeit zwischen Prime-Stimulus und Bewertungsobjekt beeinflusst werden können. Vor allem die beiden Faktoren der Zugänglichkeit und der interpretativen Anwendbarkeit von Prime-Stimuli dienten als Grundlage für verschiedenste Assimilations- und Kontrastmodelle in der Personenwahrnehmung (e.g. Schwarz & Bless, 1992; Stapel et al., 1997). In der Erforschung mehrerer Aspekte der Anwendbarkeit geprimter Eigenschaften (sozial anwendbare Eigenschaften als Stimuli vs. die Aktivierung einer sozial anwendbaren Kategorie oder einer Kombination aus beidem) beziehen sich die Autoren auf die Priming-Erkenntnisse aus vier Studien von Stapel und Kollegen (Stapel & Koomen, 1997#; Stapel & Koomen, 1998; Stapel, Koomen & van der Pligt, 1996#*; Stapel, Koomen & van der Pligt, 1997), wovon zwei von der Untersuchungskommission als *fraudulent* und einer von der betreffenden Fachzeitschrift zurückgewiesen wurde.

2. Trampe, D., Stapel, D. A. & Siero, F. W. (2007). On models and vases: Body dissatisfaction and proneness to social comparison effects. *Journal of Personality and Social Psychology*, 92, 106-118. (Retracted article. See vol. 103, pg. 398, 2012)

In sechs Studien gingen die Autoren Trampe, Stapel und Siero (2007) der Frage nach, wann und warum attraktive Frauen, die in den Medien präsentiert werden, die Selbstbewertung zusehender Frauen beeinflussen. Ob solche Bilder die Selbstbewertung von Frauen beeinflusst, hängt von verschiedenen Charakteristiken des präsentierten Objekts und der Zuseherin ab. In Studie 1 führte das Bild einer physisch attraktiven Frau im Vergleich zu einem attraktiven Model zu einer geringeren Selbstbewertung bei den Versuchsteilnehmerinnen. In Studie 2 berichteten Frauen, die mit ihrem Körper unzufrieden waren, im Vergleich zu körperzufriedenen Frauen, dass sie sich in höherem Ausmaß mit dem Körper anderer Frauen vergleichen. Körperunzufriedene Frauen ließen sich im Vergleich zu körperzufriedenen Frauen leichter von attraktiven Models und attraktiven normalen Frauen beeinflussen (Studie 3) und körperunzufriedene Frauen bewerteten sich im Vergleich zu den Zufriedenen negativer, nachdem sie ein Bild einer schmalen (vs. breiten) Vase gesehen hatten. In Studie 5 replizierten die Autoren ihre bisherigen Ergebnisse, indem sie die Körperunzufriedenheit experimentell manipulierten

und zeigten in Studie 6, dass die Unzufriedenheit mit dem eigenen Körper mit einer erhöhten Neigung zu sozialen Vergleichen einhergeht, da die Körperunzufriedenheit eine erhöhte Selbstaktivierung bewirkt. Trampe et al.s Artikel wurde 51 Mal zitiert, darin enthalten sind ein *Retraction Statement* und zwei selbstzitierende Artikel.

Folgende Artikel bauen auf den Erkenntnissen von Trampe, Stapel und Siero (2007) auf:

→ McFerran, B., Dahl, D. W., Fitzsimons, G. J. & Morales, A. C. (2010). Might an overweight waitress make you eat more? How the body type of others is sufficient to alter our food consumption. *Journal of Consumer Psychology*, 20, 146-151.

Ausgehend von Trampe et al.s Resultat, dass Frauen für soziale Vergleiche bezüglich ihres Körpers empfänglicher sind als Männer, wurden als Stichprobe für die vorliegende Studie nur weibliche Studenten ausgewählt.

→ McFerran, B., Dahl, D. W., Fitzsimons, G. J. & Morales, A. C. (2010). I'll have what she's having: Effects of social influence and body type on the food choices of others. *Journal of Consumer Research*, 36, 915-929.

In der Studie von McFerran et al. geht es darum, wie der Körperbau eines Konsumenten die Essenswahl anderer anwesender Konsumenten beeinflussen kann. Die Autoren fanden, dass sich Konsumenten an der Essensmenge anderer Personen orientieren und diese dem Körpertyp der beobachteten Person anpassen. Die Versuchspersonen wählten eine größere Portionsgröße, wenn sich eine andere Person zuvor auch eine große Essensmenge genommen hatte, jedoch nahmen sie sich weniger, wenn die beobachtete Person übergewichtig (vs. schlank) war. Die Mengengleichung an andere war bei Konsumenten mit einem niedrigen Selbstwert bezüglich ihres körperlichen Erscheinungsbilds größer ausgeprägt und ließ sich durch eine „Cognitive-Load“ Aufgabe abschwächen.

Da nach Trampe, Stapel und Siero (2007) Frauen empfänglicher für soziale Vergleiche bezüglich ihres Körperbaus sind, wählten die Autoren auch für diese Studie nur Frauen als Versuchspersonen aus. Dies trifft vor allem auf Frauen zu, die mit ihrem Körper unzufrieden sind, daher schlussfolgerten McFerran et al. (2010) aus den Erkenntnissen von Trampe et al. (2007), dass Frauen, die unzufrieden mit ihrer körperlichen Erscheinung sind, in höherem Maße dazu neigen, ihre Portionsgröße anderen Personen, die sie als soziales Vergleichsobjekt angesehen, anzupassen.

→ Svaldi, J., Caffier, D. Blechert, J. & Tuschen-Caffier, B. (2009). Body-related film clip triggers desire to binge in women with binge eating disorder. *Behaviour research and therapy*, 47, 790-796.

Die bisherige Literatur zu Essstörungen geht davon aus, dass die Themen Figur und Gewicht bezüglich der Selbstbewertung eine große Rolle bei Frauen mit einer „Binge Eating Disorder“ (BED) spielen. Bisher ist allerdings wenig über direkte Einflussfaktoren auf die Binge-Episode bekannt. In der vorliegenden Studie zeigten die Autoren 27 Frauen mit BED und 25 übergewichtigen gesunden Frauen einen körperbezogenen Film-Clip. Dieser dauerte 90 Sekunden lang und zeigte zwei Frauen, die gemeinsam in einem Modegeschäft herumschlenderten und einkauften. Vor der Präsentation des Film-Clips und danach sollten die Versuchsteilnehmerinnen ihren Drang nach einer Binge-Episode und ihre Emotionen auf einer Rating-Skala einschätzen. Zusätzlich wurden bei jeder Teilnehmerin die Herzrate, der Fingerpuls und die elektrodermale Aktivität gemessen. Die Ergebnisse zeigten für die BED Gruppe einen signifikanten Anstieg im Drang nach einer Binge-Episode, erhöhte Traurigkeit und Angst und einen Anstieg der psychophysiologischen Messungen nach dem Ansehen des körperbezogenen Video-Clips. Bei der Suche nach Erklärungsansätzen, warum der Video-Clip bei BED Teilnehmerinnen einen verstärkten Binge-Drang ausgelöst hat, zitieren sie die Studie von Trampe, Stapel und Siero (2007), in der gezeigt werden konnte, dass körperunzufriedene Frauen ihren Körper in höherem Maße mit den Körpern anderer Frauen vergleichen als körperzufriedene Frauen. Außerdem bewerten sich körperunzufriedene Frauen nach der Präsentation eines schlanken Objekts (z.B. einer schmalen Vase) negativer als nach einem runderen Objekt. Der körperbezogene Video-Clip könnte laut Svaldi et al. (2009) also negative Bewertungsprozesse bei den BED Teilnehmerinnen ausgelöst haben, die signifikant unzufriedener mit ihrem Körper waren (die Körperzufriedenheit wurde zusätzlich mit dem „Body Shape Questionnaire“ erhoben) als die gesunden Teilnehmerinnen. Trampe et al. fanden in ihrer Studie außerdem, dass Körperunzufriedenheit die Neigung zu sozialen Vergleichen steigert, da die Unzufriedenheit die kognitive Aktivierung des Selbst erhöht. Somit könnte die höhere körperliche Unzufriedenheit der BED Gruppe zu einer höheren Verfügbarkeit von selbstbezogenen Kognitionen geführt haben, die einen sozialen Vergleich auslöste und in Folge einen verstärkten Binge-Drang bewirkte.

→ Gordijn, E. H. (2010). When thinking that you are fat makes you feel worthless: Activation and application of meta-stereotypes when appearance matters. *Social Cognition*, 28, 20-39.

In Gordijns Studie wurde in vier Experimenten der Frage nachgegangen, ob normalgewichtige Personen, die selbst davon überzeugt sind, übergewichtig zu sein, glauben, dass andere Personen sie negativ beurteilen, wenn ihr Erscheinungsbild in den Fokus gerückt wird. Es wurde außerdem erhoben, ob diese negativen „Meta-Stereotype“ den betroffenen Personen ein Gefühl der Wertlosigkeit vermitteln. In Studie 1 konnte gezeigt werden, dass bei Frauen, die mit ihrem Körpergewicht unzufrieden sind, automatisch negative Meta-Stereotype aktiviert werden, wenn sie glaubten, dass ihr Erscheinungsbild bewertet wird. Da aus den Ergebnissen von Studie 1 nicht hervor ging, ob die mit ihrem Gewicht unzufriedenen Personen negative Urteile über ihr Gewicht nur im Zuge der Bewertung ihres Erscheinungsbilds trafen, oder ob sie dies aufgrund einer generellen Unzufriedenheit mit sich selbst taten, wurden sie in Studie 2 nach grundsätzlichen positiven und negativen Eigenschaften gefragt, die sie mit sich selbst in Verbindung bringen. Die weiblichen Versuchspersonen sollten sich dazu das fiktive Szenario vorstellen, dass ein attraktiver Mann vor ihnen steht und sich überlegt, sie zu einem Rendezvous auszuführen. Anschließend wurden die Selbstevaluation, die Meta-Stereotype bezüglich des Körpergewichts und die generelle Körperunzufriedenheit erhoben. Zur Messung der Körperunzufriedenheit griff die Autorin auf eine niederländische Übersetzung von Trampe et al. (2007) des Originals „Body Dissatisfaction subscale of the Eating Disorders Inventory“ von Garner et al. (1983) zurück. Dieser Fragebogen beinhaltete 15 Items mit einer jeweils 9-stufigen Ratingskala (1=absolutely not bis 9=absolutely) zur Bewertung einzelner Körperteile (z.B.: „I think that my thighs are too fat“; „I think that my hips are too wide“ etc.).

In Studie 3 sollten die Ergebnisse von Studie 2 repliziert werden, nur dass diesmal die Körperunzufriedenheit nach einer Methode von Trampe et al. (2007) experimentell manipuliert wurde. Dazu sollten die Versuchspersonen einen Fragebogen namens „Body Experience“ ausfüllen, bei dem die Teilnehmer in der zufriedenen Körpergewichts-Bedingung angeleitet wurden, eine kurze Geschichte darüber zu schreiben, warum sie mit ihrem Körper zufrieden seien und dabei die vier Wörter „me“, „beautiful“, „thin“ und „attractive“ zu verwenden. Versuchspersonen in der unzufriedenen Körpergewichts-Bedingung sollten eine Geschichte darüber schreiben, warum sie mit ihrem Körper unzufrieden sind und dabei die vier Wörter „me“, „ugly“, „fat“ und „unattractive“ benutzen.

Da beide Versuchsmaterialien aus der als betrügerisch identifizierten Studie von Trampe, Stapel und Siero (2007) stammen, kann für deren Qualität und praktische Anwendbarkeit nicht garantiert werden, so dass davon ausgegangen werden kann, dass Gordijn

erhebliche zeitliche Ressourcen damit verschwendet hat, auf den fälschlichen Erkenntnissen der drei Autoren aufzubauen.

→ Häfner, M. (2009). Knowing you, knowing me: Familiarity moderates comparison outcomes to idealized media images. *Social Cognition*, 27, 496-508.

In Häfners Studie wurde die kontraintuitive Annahme untersucht, dass besonders berühmte und damit vertraute Bilder in den Medien zu Assimilationseffekten in sozialen Vergleichssituationen führen. In Experiment 1a wurden Versuchspersonen entweder berühmte idealisierte oder unbekannte Medienbilder präsentiert und es zeigte sich, dass Personen, die berühmte Bilder gezeigt bekamen, sich danach schöner fühlten als Personen, die unbekannte Medienbilder betrachteten. In Experiment 1b konnten diese überraschenden Ergebnisse repliziert werden mit der Erweiterung, dass diesmal der Faktor der Vertrautheit experimentell manipuliert wurde. Versuchspersonen, die wiederholt mit einem Vergleichsstandard subliminal geprimt wurden, zeigten einen Assimilationseffekt (fühlten sich danach schöner), während Versuchsteilnehmer, die der Vergleichsstandard nur einmal präsentiert wurde, einen Kontrasteffekt zeigten. Die Erhebung der abhängigen Variable, die Einschätzung der eigenen Schönheit nach der Priming-Session, erfolgte in Anlehnung an Trampe et al.s (2007) Vorgabe, bei der den Versuchspersonen für 30 Sekunden ein Bild von einer jungen Frau präsentiert wurde, deren Aussehen sie danach beschreiben sollten. Im Anschluss sollten die Teilnehmer einige persönliche demographische Informationen angeben, die angeblich dazu dienten, um zu erheben, ob ihre Persönlichkeitsfaktoren und ihre Selbstwahrnehmung einen Einfluss auf die nachfolgende Beschreibung der jungen Frau haben würden. In der Erhebung dieser demographischen Daten war auch die Frage versteckt, wie schön sich die Versuchsteilnehmer in diesem Moment fühlten. Darauf konnten sie auf einer 100-stufigen Rangskala von *1=not at all* bis *100=very much* antworten. Um die verkaufte Geschichte aufrecht zu erhalten, sollten die Versuchspersonen im Anschluss das Aussehen der Frau auf dem Foto beschreiben.

Häfner stützt seine experimentelle Manipulation also auf eine manipulierte Studie Stapels, für deren Validität nicht garantiert werden kann.

3. Stapel, D. A. & Semin, G. R. (2007). The magic spell of language: Linguistic categories and their perceptual consequences. *Journal of Personality and Social Psychology*, 93, 23-33. (Retracted article. See vol. 104, pg. 197, 2013)

Folgende Artikel bauen auf den Erkenntnissen von Stapel und Semin (2007) auf, welcher 37 Mal zitiert wurde, mit einem *Retraction Statement* und zwei Selbstzitaten:

→ Hagvedt, H. & Patrick, V. M. (2011). Turning art into mere illusion: Concretizing art renders its influence context dependent. *Personality and Social Psychology Bulletin*, 37, 16-24-1632.

Ein großer Teil der Kunst-Forschung hat sich bisher auf die psychologische Wahrnehmung von Kunst konzentriert, jedoch nicht auf die Prozesse, die diesen Erfahrungen zugrundeliegen. Genau dies wollten die Autoren Hagvedt et al. (2011) in ihrer Studie untersuchen und stellten die Frage, wie sich die Wahrnehmung eines Betrachters ändert, wenn a) der Inhalt des Kunstobjekts als „Kunst“ betont wird oder b) das Kunstobjekt auf einer wenig konkreten konstruktbasieren Stufe präsentiert wird, so dass das Objekt die Sicht des Betrachters nicht länger nur mit dem Stempel „das ist Kunst“ beeinflusst. Stattdessen erscheint das Kunstobjekt im Falle b) als Illustration, deren Einfluss auf den Betrachter vom umliegenden Kontext abhängig ist.

Ein Beispiel: Person X und Person Y sehen sich ein Werbeplakat für ein neues Restaurant an, das ein Bild des Malers Rubens zeigt, auf dem zwei üppige Frauenkörper dargestellt sind. Person X erkennt dieses Bild als Kunstobjekt und zieht aus der guten Qualität des Bildes oder dessen Kunststatus Schlüsse auf das beworbene Restaurant. Person Y nimmt das Bild als gemalte Illustration wahr und schließt daraus, dass das Essen des neuen Restaurants bestimmt fettig und ungesund sein wird. Diese Schlüsse sind grundlegend verschieden, können jedoch von ein und demselben Bild ausgelöst werden. In Studie 3 demonstrierten die Autoren, dass die Ebene, auf der ein Betrachter ein Bild wahrnimmt (abstrakt vs. konkret), die differenzierte Verarbeitung des Objekts als „Illustration“ oder als „Kunstobjekt“ bedingt. Ausgehend von Stapel und Semins Erkenntnissen (2007), dass abstrakte vs. konkrete Sprache einen globalen vs. konkreten Informationsverarbeitungsstil begünstigen kann, schlussfolgerten die Autoren der vorliegenden Studie, dass die Lenkung der Aufmerksamkeit auf den Inhalt vs. die Art und Weise der Gestaltung in unterschiedlichen Verarbeitungseffekten resultiert.

→ Hamilton, R. V., Kathleen, D., Sellier, A. L. & Meyvis, T. (2011). Being of two minds: Switching mindsets exhausts self-regulatory resources. *Organizational Behavior and Human Decision Processes*, 115, 13-24.

Die menschliche Psyche ist mit der Fähigkeit ausgestattet, Probleme mittels unterschiedlichen mentalen Denkweisen (sogenannte „Mindsets“) zu lösen. Ein „Mindset“ wird als grundsätzliche Bereitschaft, auf ein Problem in einer bestimmten Art und Weise zu reagieren, definiert. Unterschiedliche „Mindsets“ können zu unterschiedlichen Entscheidungsprozessen und Urteilen führen. Um als Individuum eine andere Perspektive auf ein Problem einnehmen zu können, ist es auch möglich, seine Denkweise zu ändern. Die Autoren Hamilton et al. (2011) gehen davon aus, dass das Hin- und Herspringen von „Mindsets“ psychologische Ressourcen verbraucht und sich somit negativ auf nachfolgende Entscheidungen auswirken kann, welche von denselben exekutiven Funktionen gesteuert werden, wie beispielsweise die Selbstregulation. In ihrem dritten Experiment untersuchten Hamilton et al., welchen Effekt das Ändern der Denkweise in einem sprachlichen Kontext auf das Ausmaß der Selbstregulation hat. Dafür nahmen zweisprachige Versuchspersonen am Experiment teil, die einen Persönlichkeitstest entweder auf Englisch, in ihrer Zweitsprache oder in einer Kombination aus beiden Sprachen ausfüllten. Jeweils vor und nach der Vorgabe des Persönlichkeitstests, wurde mit ihnen der Handgriff-Test („Handgrip-Test“) durchgeführt, bei dem die Versuchspersonen ein eigens dafür angefertigtes Gerät in einer Hand halten und dabei den Haltegriff mit vier Fingern so weit zurückziehen, dass damit ein Blatt Papier gehalten werden kann. Aufbauend auf Stapel und Semins Erkenntnissen (2007), dass das Sprechen einer Sprache das dazugehörige „Mindset“ aktiviert, stellten die Autoren die Hypothese auf, dass das Ändern der Sprache innerhalb eines Tests psychologische Ressourcen beansprucht, so dass im nachfolgenden Handgriff-Test das Blatt aufgrund verminderter Selbstregulation nicht so lange gehalten werden kann wie bei Versuchspersonen, die den Fragebogen nur in einer Sprache bearbeiten müssen. Die Ergebnisse bestätigten Hamilton et al.s (2007) Annahme.

→ Walther, E., Langer, T., Weil, R. & Komischke, M. (2011). Preferences surf on the currents of words: Implicit verb causality influences evaluative conditioning. *European Journal of Social Psychology*, 41, 17-22.

Es ist eine alte psychologische Binsenweisheit, dass Gedanken unsere Sprache formen. Die umgekehrte Idee, dass die Sprache unser Denken beeinflusst, wurde dahingegen in der Philosophie, Linguistik und Psychologie weit weniger erforscht. Daher war das Ziel der vorliegenden Studie, zu untersuchen, welchen Einfluss verschiedene Verb-Klassen auf

das Gefallen oder die Abneigung von Personen gegenüber anderen Personen in einem interpersonellen Beurteilungskontext haben. Eine experimentelle Antwort auf diese Frage bietet die Forschung zur evaluativen Konditionierung („Evaluative Conditioning“ = EC), bei der ein Urteil eines Individuums aufgrund der Paarung zweier Stimuli geändert wird. Die positive Bewertung einer Person aufgrund eines konditionierten Stimulus („Conditioned Stimulus“ = CS) kann durch die Hinzunahme eines zweiten unkonditionierten Stimulus („Unconditioned Stimulus“ = US) geändert werden. Entgegen der traditionellen EC Annahme, dass durch die Kombination von CS und US die Valenz der Beurteilung beeinflusst wird, fanden die Autoren in ihren Experimenten, dass die Sprache das EC moderieren kann, indem die linguistische Wortart des US geändert wird. Stapel und Semin (2007) konnten bereits demonstrieren, dass linguistische Kategorien grundlegende Wahrnehmungsprozesse beeinflussen können. Sie fanden, dass eine abstrakte Sprache (z.B. Adjektive) eher zu einem globalen Wahrnehmungsfokus führt, während eine konkrete Sprache (z.B. Verben) eine lokale Wahrnehmung zur Folge hat. Ausgehend von Stapel et al.s Annahme (2007), dass die Sprache die Aufmerksamkeit von Individuen zu einem bestimmten Wahrnehmungsfokus führt, schlussfolgern Walther et al. (2011), dass die evaluative Konditionierung ebenfalls durch linguistische Stimuli manipuliert werden kann.

→ Price, T. F. & Harmon-Jones, E. (2010). The effect of embodied emotive states on cognitive categorization. *Emotion*, 10, 934-938.

Bisherige Forschung hat ergeben, dass ein positiver Affekt die Fähigkeit zur kognitiven Kategorisierung einschränkt. Das Motivations-Dimensions-Modell geht jedoch davon aus, dass ein positiver Affekt kein eindimensionales Konstrukt ist, sondern dass es mehrere Arten positiven Affekts gibt, die in ihrer Intensität der motivationalen Herangehensweise („Motivational Intensity Approach“) variieren können. Diesem Modell zufolge, führt nur ein positiver Affekt mit einer geringen motivationalen Intensität zu einer Ausweitung kognitiver Denkprozesse, während ein positiver Affekt mit hoher motivationaler Intensität kognitive Denkprozesse einengt. Ein positiver „high approach“ Affekt fokussiert die Aufmerksamkeit und sollte daher auch zu einer Einschränkung in einer Kategorisierungs-Aufgabe führen und vice versa für positive „low approach“ Affekte. Die Autoren überprüften diese Annahme in zwei Experimenten, in denen sie Versuchspersonen in drei verschiedenen Sitzhaltungen eine Kategorisierungs-Aufgabe durchführen ließen. Die Manipulation der Sitzhaltung diente der Manipulation des Ausmaßes der „Approach Motivation“ (zurückgelehntes Sitzen: niedrige Motivation, aufrechtes Sitzen: moderate Motivation, vorgelehntes Sitzen: hohe Motivation). Der positive Affekt wurde durch die Anweisung des Experimentators erreicht, die Mundwinkel während der gesamten Aufgabe nach oben zu

ziehen (angeblich um die zur Verschleierung aufgeklebten Elektroden zur Messung der Hirnaktivität durch die Gesichtsmuskelkontraktion zu unterstützen). Als Kategorisierungsaufgabe wurde den Versuchspersonen ein nach Stapel und Semin (2007) adaptierter Test vorgegeben, bei dem die Versuchspersonen 10 Items (z.B. „Karotte“) einer übergeordneten Kategorie (z.B. „Gemüse“) auf einer 7-stufigen Ratingskala (1=*definitely does belong to the category* bis 7=*definitely does not belong to the category*) zuordnen mussten. Die Items variierten von schwach, moderat bis stark repräsentativ für eine Kategorie. Wie erwartet, wurde die Enge der Kategorisierung von der Sitzposition beeinflusst. Während eine zurückgelehnte Sitzhaltung und ein lachendes Gesicht dazu führten, dass Versuchspersonen auch schwach repräsentative Items der vorgegebenen Kategorie zuordneten, taten dies lachende und vorgelehnte Versuchspersonen nicht.

→ Malkoc, S. A., Zauberan, G. & Bettman, J. R. (2010). Unstuck from the concrete: Carryover effects of abstract mindsets in intertemporal preferences. *Organizational Behavior and Human Decision Processes*, 113, 112-126.

Der sogenannte „Present-Bias“ beschreibt, dass Individuen umso weniger ungeduldig sind, je länger sie auf ein Objekt der Begierde warten müssen. Bisherige Forschung konnte eindrucksvoll demonstrieren, dass Personen, die sich beispielsweise online eine Kamera bestellt haben, mehr Geld pro Tag ausgeben würden, um die Lieferzeit zu verkürzen, wenn die Ankunftszeit 3 Tage betragen würde, im Vergleich zu einer angeschriebenen Lieferzeit von 10 Tagen. In der vorliegenden Studie erhoben die Autoren die dem „Present-Bias“ zugrunde liegenden psychologischen Variablen und betrachteten dabei vor allem das Timing einer Entscheidung als wichtigen Faktor. Malkoc et al. (2010) gingen von der Annahme aus, dass die Abstraktheit der Informationsverarbeitung (das Fokussieren auf Details vs. auf ganzheitlichere Aspekte) durch frühere (themenbezogene vs. nicht themenbezogene) Kaufentscheidungen beeinflusst wird und einen systematischen Einfluss auf das Ausmaß des „Present-Bias“ in nachfolgenden Entscheidungen hat. In ihrem dritten Experiment manipulierten die Autoren das abstrakte Denken durch eine supraliminale Priming Aufgabe, in der entweder konkrete oder abstrakte Wörter vorgegeben wurden. Die bisherige Forschung zur linguistischen Abstraktion hat ergeben, dass manche Wörter (Verben) konkreter verarbeitet werden als andere Wörter (Adjektive). Stapel und Semin (2007) fanden beispielsweise, dass das Priming mit Adjektiven (im Vergleich zu Verben) die Versuchspersonen dazu führte, eine weniger kontextabhängige, globalere Sichtweise zu haben und in einer Kategorisierungsaufgabe mehr inkludierende Entscheidungen zu treffen. Diesem linguistischen Effekt zufolge, verwendeten die Autoren ebenfalls eine Priming Aufgabe, bei der die Versuchspersonen in einer Wort-Such-Aufgabe nach Nomen oder Adjektiven

suchen sollten. Die Annahme war, dass bei Versuchspersonen, die mit abstrakten Wörtern geprimt werden, ein abstraktes Denkmuster („Mindset“) aktiviert wird (im Vergleich zu Versuchspersonen, die mit konkreten Wörtern geprimt werden), welches zu einem geringeren „Present-Bias“ führen sollte.

→ IJzerman, H. & Semin, G. R. (2009). The thermometer of social relations: Mapping social proximity on temperature. *Psychological Science*, 20, 1214-1220.

Die gängigen Sprichwörter „Warme Gefühle gegenüber jemandem hegen“ oder „Jemandem die kalte Schulter zeigen“ verdeutlichen bereits, dass die Temperatur etwas mit dem Ausmaß sozialer Nähe zu tun haben könnte. Die Autoren wollten diesen Metaphern experimentell nachgehen und fanden in drei Experimenten, dass in einer warmen Versuchsbedingung (die Versuchspersonen sollten kurz ein warmes vs. kaltes Getränk in der Hand halten) die soziale Nähe zu einer frei ausgewählten Person, die gemeinsam mit sich selbst auf der „Inclusion of other in self“ Skala (Aron, Aron & Smollan, 1992) beschrieben wurde, größer eingeschätzt wurde als in der kalten Versuchsbedingung (Experiment 1). In ihrem zweiten Experiment untersuchten die Autoren, welchen Effekt die Manipulation der Raumtemperatur (warm vs. kalt) nicht nur auf die soziale Nähe, sondern auch auf den Sprachgebrauch hat. Die Annahme der Autoren war, dass wenn eine höhere Raumtemperatur (22-24 Grad) zu einer höheren Einschätzung der sozialen Nähe (dieses Mal sollte die Nähe zum Experimentator bewertet werden) führt, dann sollte die Beschreibung sozialer Ereignisse in der warmen Bedingung konkreter ausfallen als in der kalten Bedingung (15-18 Grad Raumtemperatur). Nachdem die Versuchspersonen in dem experimentell temperierten Raum Platz genommen hatten, wurde ihnen ein kurzer Video-Clip vorgespielt, der animierte Schachfiguren zeigte, die verschiedene Bewegungen machten, die nichts mit klassischen Schachzügen zu tun hatten. Danach sollten die Versuchspersonen in eigenen Worten beschreiben, was sie im Video gesehen hatten. Die Auswahl des Videos basiert auf Stapel und Semins Studie (2007), die dieses Versuchsmaterial für ihre Forschung ebenfalls eingesetzt haben. Nach der Videobeschreibung teilte ein vom Experiment unabhängiger Experte die Beschreibungen in die beiden Kategorien „abstrakt“ und „konkret“ nach dem „Linguistic Category Model“ von Semin und Fiedler (1988) ein. Die Ergebnisse bestätigen die Annahme, dass sich Versuchspersonen in der warmen Bedingung dem Experimentator sozial näher fühlten als Versuchspersonen in der kalten Bedingung.

→ Liberman, N. & Forster, J. (2009). Distancing from experienced self: How global-versus-local perception affects estimation of psychological distance. *Journal of Personality and Social Psychology*, 97, 203-216.

Liberman und Forster untersuchten in ihrem Artikel den Effekt der globalen vs. lokalen Wahrnehmung auf die Einschätzung der egozentrischen psychologischen Distanz. Die Autoren trafen die Annahme, dass die Aktivierung einer globaleren Sichtweise dazu führt, dass Versuchspersonen größere egozentrische Distanzen schätzen als aus einer detaillierteren Perspektive. Eine egozentrische Distanz ist der mental repräsentierte Abstand zwischen einem selbst im Hier und Jetzt und anderen Objekten, die nicht teil der direkten Erfahrung sind, beispielsweise der psychologische Abstand zu Ereignissen aus der Vergangenheit oder der Zukunft (zeitliche Distanz), der Abstand zu entfernten Lokalitäten (räumliche Distanz), der Abstand zu Personen, deren Eigenschaften oder Erfahrungen unseren unähnlich sind (soziale Distanz) oder der Abstand zu Ereignissen, deren Auftretenswahrscheinlichkeit als gering eingeschätzt wird (hypothetische Distanz). Der „Construal level theory of psychological distance“ (CLT; Liberman & Trope, 2008) zufolge, werden Objekte auf einer niedrigen Abstraktionsstufe („low-level construal“) konkret und kontextspezifisch repräsentiert, während Objekte auf einer hohen Abstraktionsstufe („high-level construal“) zu abstrakten und globalen mentalen Repräsentationen führen. Laut CLT beeinflussen sich die psychologische Distanz und die mentale Abstraktion wechselseitig, das heißt Individuen denken über entferntere Objekte abstrakter nach und abstrakte mentale Repräsentationen begünstigen die Einschätzung einer weiteren psychologischen Distanz. Stapel und Semin (2007) demonstrierten in ihrer Studie einen Effekt der Abstraktionsstufe auf die Wahrnehmung: Sie primten Versuchspersonen mit Adjektiven („high-level constructs“) oder mit konkreten Verben („low-level constructs“) und zeigten, dass Versuchspersonen in der abstrakten Priming-Bedingung (high-level) ein Objekt globaler wahrnahmen als Versuchspersonen in der niedrigen Abstraktions-Bedingung (low-level), die das Objekt lokal detaillierter beurteilten. Stapel et al.s (2007) Arbeit und die weiterer Forscher zur Verknüpfung von Wahrnehmungs- und Abstraktionsprozessen dienten den Autoren als theoretische Grundlage für die Durchführung ihrer vier Experimente.

→ Semin, G. R. (2008). Language puzzles – A prospective retrospective on the linguistic category model. *Journal of Language and Social Psychology*, 27, 197-209.

Semin schreibt in seinem Literatur-Review über die bisherigen Erkenntnisse rund um das „Linguistic Category Model“ und diskutiert diese aus verschiedenen Perspektiven. In der

Beschreibung der Anwendbarkeit des LCMs geht er auf die Studie von Stapel und Semin (2007) ein, in der gezeigt werden konnte, dass die Sprache einen metasemantischen Effekt auf die Kognition haben kann. Stapel et al. (2007) fanden, dass der Gebrauch von aktiven Verben (z.B. „Mary schlägt Jill“) die Aufmerksamkeit des Zuhörers auf lokale, situationsbezogene Eigenschaften des Ereignisses lenkt, während abstraktere Adjektive (z.B. „Mary ist aggressiv“) zu einer globalen Wahrnehmung der Situation führen und lieferten damit einen wesentlichen Beitrag für die zukünftige psycho-linguistische Forschung.

→ Monga, A. B. & John, D. R. (2010). What makes brands elastic? The influence of brand concept and styles of thinking on brand extension evaluation. *Journal of Marketing*, 74, 80-92.

Monga und John gingen in ihrem Artikel der Frage nach, warum sich das Sortiment einiger Marken leichter erweitern lässt als bei anderen. Bisherige Forschung zu diesem Thema ergab, dass Konsumenten die Erweiterung entfernterer Produktkategorien eher bei Marken mit Prestige-Konzepten (z.B. Rolex) akzeptieren als bei Marken mit rein funktionellen Konzepten (z.B. das einfache Uhrendesign von Timex). In dem vorliegenden Artikel analysieren die Autoren den Einfluss der beiden Denkstile des analytischen vs. des holistischen Denkens auf die Ausweitungsakzeptanz bei Konsumenten von prestigeträchtigen vs. funktionalen Marken.

In Experiment 5 manipulierten die Autoren den Denkstil der Versuchspersonen experimentell, wobei sie sich auf die Studie von Stapel und Semin (2007) beriefen, die fanden, dass unterschiedliche linguistische Kategorien (Adjektive vs. Verben) unterschiedliche Effekte in der Aufmerksamkeit verursachen können. Adjektive schaffen einen holistischen Interpretationsrahmen, indem der Aufmerksamkeitsfokus auf globale, abstrakte Beziehungen gelegt wird, während Verben einen analytischen Rahmen kreieren, bei dem das Augenmerk auf spezifische Eigenschaften und Details gelegt wird. Auf dieser Unterscheidung aufbauend, untersuchten Monga und John, wie analytisch oder holistisch denkende Konsumenten die Erweiterung eines Markensortiments beurteilen, wenn ihnen die Produkte durch Beschreibungen angepriesen werden, welche analytische (Verben) oder holistische (Adjektive) Informationen enthalten.

4. Marx, D. M. & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, 91, 243-254. (Retracted article. See vol. 104, pg. 196, 2013)

Viele Forscher haben sich bereits mit dem Thema auseinandergesetzt, dass das Primen negativer stereotyper Eigenschaften zu einer wahrgenommenen stereotypen Bedrohung führt.

Nach Steele und Aronson (1995) versteht man unter „Stereotype threat“:

[...] one of the most heavily studied topics in social psychology over the past decade -- refers to the concern that is experienced when one feels at risk of confirming, as selfcharacteristic, a negative stereotype about one's group.

Wenn Frauen beispielsweise ein Mathematiktest vorgegeben wird, von dem sie glauben, dass er als Diagnostik für ihre mathematischen Fähigkeiten herangezogen wird, erzielen sie eine schlechtere Testleistung als Männer, da durch die Betonung des diagnostischen Werts der stereotype Charakter einer möglichen schlechteren Testleistung betont wird. Wenn zu Beginn der Testdurchführung jedoch angegeben wird, dass Frauen und Männer in dem nachfolgenden Test für gewöhnlich gleiche Werte erzielen, wird die stereotype Bedrohung für Frauen irrelevant und sie schneiden nicht schlechter als Männer ab (Steele & Aronson, 1995).

Aufgrund der Allgegenwärtigkeit stereotyper Bedrohungseffekte, erschien es Marx und Stapel (2006) wichtig, zwischen Prozessen zu differenzieren, die alle zu einer Bestätigung stereotyper Verhaltensmuster (z.B. schlechte Mathematik-Testleistung) führen, nämlich stereotype Priming Effekte und die Angst, negative Gruppen-Stereotype zu erfüllen. In drei Experimenten demonstrierten die Autoren, dass die stereotype Bedrohung die Testleistung und die bedrohungsbezogenen Überlegungen von Zielpersonen (z.B. Frauen), aber nicht von Nicht-Zielpersonen (z.B. Männern), beeinflusst, da nur das soziale Selbstkonzept der Zielpersonen mit dem negativen Gruppen-Stereotyp verbunden ist. Das Primen von Eigenschaften hatte dagegen einen negativen Effekt auf die Testleistung beider Versuchsgruppen (z.B. Frauen und Männer), allerdings nicht auf ihre bedrohungsbezogenen Gedanken, da das Primen von Eigenschaften keine Verbindung zwischen dem sozialen Selbstkonzept und einem Gruppen-Stereotyp voraussetzt.

Marx und Stapels Artikel (2006) wurde 24 Mal zitiert, davon handelt es sich bei einem Zitat und das *Retraction Statement* und bei einem weiteren um ein Selbstzitat. Folgende Artikel bauen auf Marx und Stapels Erkenntnissen auf:

→ Cromley, J. G., Perez, T., Wills, T. W., Tanaka, J. C., Horvat, E. M. & Agbenyega, E. T. B. (2013). Changes in race and sex stereotype threat among diverse STEM students: Relation to grades and retention in the majors. *Contemporary Educational Psychology*, 38, 247-258.

Die Autoren Cromley et al. konnten in ihren Laborstudien zeigen, dass eine induzierte stereotype Bedrohung zu negativen Effekten auf die akademische Leistung und das Lernen führt. Um zu überprüfen, ob diese Effekte auch auf ein naturalistisches Setting umgelegt werden können, wurden 1358 Studenten der STEM Wissenschaften („Science, Technology, Engineering, Mathematics“) 4 Mal im Semester während eines einführenden Chemie oder Biologie Kurses neu entwickelte Items zu einem Geschlechter- und Rassen-Stereotyp vorgelegt. Die Entwicklung dieser Items beruht auf einer „Threat-Measure“ Methode von Marx und Stapel (2006), deren Items zur Messung der Anfälligkeit für Rassen- und Geschlechter-Stereotype umformuliert wurden. Ein hoher Wert in diesem Test, der aus jeweils fünf Items besteht, würde für eine hohe Vulnerabilität für eine stereotype Bedrohung bezüglich der Rasse oder des Geschlechts sprechen. Ein Beispiel für ein Item des „Sex Stereotype Bias“ wäre: „I believe that my ability to perform well on chem/bio tests is affected by my gender“ und ein Beispiel für ein Item des „Race Stereotype Bias“ wäre: „I believe that if I perform poorly on a test, the professor will attribute my poor performance to my race“. Die Ergebnisse der Studie können bei Cromley et al. (2013) nachgelesen werden.

→ Swift, H. J., Abrams, D. & Marques, S. (2013). Threat or boost? Social comparison affects older people's performance differently depending on task domain. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 68, 23-30.

Swift et al. (2013) untersuchten in ihren Studien, ob der soziale Vergleich mit jüngeren Personen eine stereotype Bedrohung oder einen Leistungsanstieg bei älteren Personen verursacht. Dazu wurden in Studie 1 national repräsentative Daten herangezogen, um zu erforschen, in welchen Domänen ältere Personen im Vergleich zu jüngeren stereotyp begünstigt oder benachteiligt sind. In Studie 2 wurde getestet, welchen Einfluss ein potentiell bedrohlicher vs. Kontroll- vs. bestärkender Vergleich mit jüngeren Personen auf die Leistung älterer Personen in negativen und positiven stereotypen Domänen hat. Nach der Erhebung der abhängigen Variablen, wurden den Versuchsteilnehmern in Studie 2 zwei Items nach Marx und Stapel (2006) vorgegeben, die die Wirkung der manipulierten stereotypen Alters-Bedrohung erfassen sollte. Die beiden Items lauteten: „Were you worried that your ability to perform well on the test was affected by your age? “ und „Were

you worried that if you performed poorly on the test, the researcher would attribute your poor performance to your age? ". Die Versuchsteilnehmer konnten darauf mit 1=*not at all* bis 7=*very much* antworten. Daraus wurden Mittelwerte berechnet, wobei höhere Werte für einen höheren „Threat Concern“ stehen.

→ Jamieson, J. P. & Harkins, S. G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes & Intergroup Relations*, 15, 291-304.

Aus bisheriger Literatur ist bekannt, dass sowohl die stereotype Bedrohung als auch stereotypes Priming einen Einfluss auf die Leistung in einem nachfolgenden Test haben können. Die Autoren zitieren dabei Marx und Stapel (2006), die davon ausgingen, dass diese beiden psychologischen Phänomene zwar zu denselben Effekten in der Testleistung führen, jedoch auf unterschiedlichen kognitiven Prozessen beruhen. Sie meinen, dass der Effekt der stereotypen Bedrohung darauf basiert, dass die betreffende Person über das Stereotyp Bescheid weiß (z.B. „Frauen sind schlecht in Mathematik“) und sich gleichzeitig auch dessen bewusst ist, dass sie selbst Mitglied dieser stereotypisierten Gruppe ist („Ich bin eine Frau“). Beim Priming hingegen wird das Verhalten (meist unbewusst) durch die Aktivierung von Informationen beeinflusst, die mit dem geprimten Konstrukt in Verbindung stehen. Das Ziel der vorliegenden Studie war es also, Marx und Stapels theoretische Annahmen (2006) zu erweitern und die Hypothese zu testen, dass sich die stereotype Bedrohung und das stereotype Priming nicht nur durch die „threat-based concerns“ (= bedrohungsbasierten Gedanken), sondern auch durch verschiedene motivationsbezogene Muster in der Testbearbeitung unterscheiden. Die Ergebnisse zeigten, dass Frauen, die durch die Möglichkeit, negative Gender-Mathematik Stereotype bestätigen zu werden, bedroht wurden, motiviert waren, diesem Stereotyp nicht zu entsprechen, während Frauen, die vor der Testdurchführung durch ein weibliches Geschlechter-Konstrukt geprimt wurden, keine erhöhte Motivation zeigten, eine gute Leistung zu erzielen, sondern ihre Bemühungen im Mathematiktest sogar noch reduzierten. Die stereotype Bedrohung führte also zu einer erhöhten Motivation, dem aktivierten Stereotyp nicht nachzukommen, aber dennoch zu einer schlechteren Testleistung im Vergleich zur nicht bedrohten Kontrollgruppe. Frauen, die mit einem Gender-Mathematik Stereotyp geprimt wurden, zeigten hingegen keine Motivation, durch ihre Leistung das Gegenteil zu beweisen.

→ Gaillard, M., Desmette, D. & Keller, J. (2011). Regulatory focus moderates the influence of age-related stereotypic expectancies on older adults' test performance and threat-based concerns. *European Review of Applied Psychology*, 61, 23-29.

Gaillard et al. (2011) verwendeten, genauso wie Swift et al. (2013), die von Marx und Stapel (2006) entwickelten zwei Items, um das Ausmaß der „threat-based concerns“ zu erheben.

→ Rydell, R. J., Rydell, M. T. & Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology*, 99, 883-896.

Die bisherige „Stereotype threat“ Literatur ist hauptsächlich darauf eingegangen, welche direkten Effekte eine stereotype Bedrohung auf eine Testleistung haben kann. In der vorliegenden Studie behandeln die Autoren die Frage, ob negative Gender-Mathematik Stereotype die Leistung nicht auch indirekt über die Hemmung des Lernens mathematischer Regeln und Rechenwege, die für die Lösung der Aufgaben essentiell sind, beeinflussen können. Ausgehend von Marx und Stapels Annahme (2006), dass das negative Stereotyp „Frauen sind schlecht in Mathematik“, zu Sorgen bei den Versuchsteilnehmerinnen führt, dass sie diesem Stereotyp entsprechen könnten („stereotype threat concern“), stellten sie die Hypothese auf, dass diese Sorge die Kapazität ihres Arbeitsgedächtnisses reduziert, ihr körperliches Erregungsniveau erhöht und ihre Motivation zu lernen reduziert. Diese Faktoren führen anschließend dazu, dass die Enkodierung mathematischer Informationen (z.B. mathematische Regeln und Rechenwege) gehemmt und somit eine schlechtere Leistung im Mathematiktest erzielt wird. Zur Messung der „stereotype threat concerns“ wurden auch in dieser Studie die von Marx und Stapel (2006) entwickelten Items vorgegeben.

→ Jamieson, J. P. & Harkins, S. G. (2010). Evaluation is necessary to produce stereotype threat performance effects. *Social Influence*, 5, 75-86.

Jamieson und Harkins stützten sich in der Ableitung ihrer Forschungshypothese auf die Arbeit von Marx und Stapel (2006), die bereits zeigten, dass sowohl die stereotype Bedrohung als auch das stereotype Priming die Leistung in einem Mathematiktest schwächen können, dass sich die zugrunde liegenden kognitiven Prozesse jedoch voneinander unterscheiden (siehe oben). Jamieson et al. (2006) erweiterten diese Annahme und meinten, dass der Faktor der Evaluation eine entscheidende Rolle in der Entstehung von stereotypen Bedrohungseffekten auf die Leistung spielt. Die

wahrgenommene Bedrohung wird nämlich durch die Angst ausgelöst, den negativen Stereotypen entsprechend zu versagen und diese Stereotype können nur dann bestätigt werden, wenn im Anschluss eine Evaluation der Testleistung von anderen oder einem selbst durchgeführt wird.

→ Nguyen, H. H. D. & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and woman? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93, 1314-1334.

Die Autoren führten eine Meta-Analyse zu „Stereotype threat“ Effekten mit einer gesamt-durchschnittlichen Effektgröße von 1.261 durch. In einer Serie von Analysen zu möglichen Moderator-Variablen ergaben sich unterschiedliche Effekte für rassen- vs. geschlechterbasierte Stereotype. In einer Tabelle über „Stereotype activating cues“ zitieren die Autoren die Ergebnisse von Marx und Stapels Studien (2006).

5. Johnson, C. S. & Stapel, D. A. (2007). No pain, no gain: The conditions under which upward comparisons lead to better performance. *Journal of Personality and Social Psychology*, 92, 1051-1067. (Retracted article. See vol. 103, pg. 415, 2012)

Johnson und Stapel (2007) untersuchten in drei Studien den Zusammenhang zwischen bedrohlichen aufsteigenden sozialen Vergleichen und der Leistung in einem „Remote Association Task“ (RAT, Mednick, 1962; McFarlin & Blascovich, 1984), bei dem die Versuchspersonen drei Wörter präsentiert bekommen (z.B. Kaffee, Kuchen, Butter) und anschließend das dazu passende vierte Wort angeben müssen (z.B. Tasse). In Studie 1 wurden die Versuchspersonen einem Vergleich mit Zielpersonen ausgesetzt, der ihre Selbstbewertung entweder bedrohte oder förderte (durch das Lesen einer Geschichte, in der ein moderat beziehungsweise extrem erfolgreicher Student beschrieben wird). Versuchspersonen in der extremen Bedrohungsbedingung zeigten in der RAT-Aufgabe bessere Leistungen als die Versuchspersonen in der moderaten Bedrohungsbedingung und die Kontrollgruppe, die gar keiner Selbstbewertungsbedrohung ausgesetzt wurde, erzielte die schlechtesten Leistungen. In Studie 2 konnten die Autoren demonstrieren, dass eine der Bedrohungsbedingung vorausgehende Selbst-Affirmation, den Effekt der Leistungsverbesserung ausschaltete. In Studie 3 konnten die Versuchspersonen ihre Leistung nur dann verbessern, wenn die durchgeführte Aufgabe von der Aufgabe der Vergleichsperson in der Geschichte verschieden war. Diese Ergebnisse ließen vermuten, dass eine Leistungssteigerung infolge eines sozialen Vergleichs durch die Motivation einer Person entsteht, ihren Selbstwert zu erhalten bzw. wieder herzustellen.

Johnson und Stapel (2011) wurden 21 Mal zitiert, davon stammt ein Zitat von Stapel selbst und ein Zitat enthält das *Retraction Statement* der zuständigen Fachzeitschrift. Folgende Artikel bauen auf den Ergebnissen der Autoren auf:

→ Melwani, S. & Barsade, S. G. (2011). Held in contempt: The psychological, interpersonal and performance consequences of contempt in a work context. *Journal of Personality and Social Psychology*, 101, 503-520.

In drei Studien untersuchten die Autoren, welche psychologischen, interpersonellen und leistungsbezogenen Auswirkungen ein missbilligendes Feedback auf Personen in einer Arbeitssituation hat. Dazu nahmen alle Versuchspersonen an einer computerisierten Arbeitssimulation teil, bei der sie nach Beendigung ihrer Aufgabe ein (missbilligendes, wütendes, auf Fehler hinweisendes oder neutrales) Feedback von einem angeblichen Arbeitspartner bekamen, der, wie den Versuchspersonen gesagt wurde, über einen anderen Computer mit ihnen gemeinsam an der Lösung verschiedener Aufgaben arbeitete. In Wirklichkeit wurde ihnen vom Computer, je nach Versuchsbedingung, das jeweilige Feedback zugeschaltet.

Johnson und Stapels (2007) Erkenntnis folgend, dass Individuen eine Selbstwertbedrohung erfahren, wenn sie in eine niedrige Status-Position gedrängt werden, und in Folge Handlungen unternehmen, um ihren Selbstwert wiederherzustellen, indem sie sich mehr Mühe geben bzw. härter arbeiten, nahmen die Autoren an, dass ein missbilligendes Feedback zu einer gesteigerten Qualität der Arbeitsleistung führen wird. In der ersten Studie bestätigten die Ergebnisse die Hypothese der Autoren. Versuchspersonen, die vom Computer ein missbilligendes Feedback erhielten, zeigten im Vergleich zu den anderen Feedback-Bedingungen eine signifikant bessere Arbeitsleistung und eine ebenfalls signifikant höhere Aggressivität gegenüber ihrem virtuellen Arbeitspartner.

→ Henderson, C. M., Beck, J. T. & Palmatier, R. W. (2011). Review of the theoretical underpinnings of loyalty programs. *Journal of Consumer Psychology*, 21, 256-276.

In ihrem Review der psychologischen Faktoren, die dem Konstrukt der Loyalität zugrunde liegen, verweisen die Autoren in einer Graphik auf die Forschungsergebnisse zu sozialen Vergleichseffekten nach Johnson und Stapel (2007), in der sie beschreiben, dass absteigende, aufsteigende und gleichwertige Vergleiche, die auf den eigenen

Statuscharakteristiken basieren, dazu führen, dass Individuen sich anderen Status-Gruppen anpassen, von ihnen differenzieren oder sie überragen wollen.

→ Van de Ven, N., Zeelenberg, M. & Pieters, R. (2011). Why envy outperforms admiration. *Personality and Social Psychology Bulletin*, 37, 784-795.

Die Autoren testeten in vier Studien die Hypothese, dass gutartiger Neid Individuen motiviert, sich selbst zu verbessern, dass dies die Emotionen des bösartigen Neids und der Bewunderung jedoch nicht tun. Der wesentliche Unterschied zwischen den beiden Formen des Neids ist, dass der gutartige Neid ein Gefühl der Handlungstendenz bei Individuen auslöst, sich selbst zu verbessern, um dem Neidobjekt in der bewunderten Dimension ähnlicher zu werden, während der bösartige Neid in einer Handlungstendenz resultiert, die beneidete Person abzuwerten. Die bloße Bewunderung ist ein positives Gefühl, das jedoch nicht dazu führt, sich selbst verbessern zu wollen, während der gutartige Neid ein Gefühl der Frustration verursacht, das wiederum in einer erhöhten Selbstverbesserungsmotivation resultiert. Ihre Hypothesen stützen die Autoren auf Johnson und Stapels „no pain, no gain“ Prinzip (2007), das besagt, je schlechter sich Personen im Zuge eines ansteigenden Vergleichs mit anderen fühlen, desto härter arbeiten sie und erzielen somit eine bessere Leistung. Stapel et al.s Annahme, dass eine gewisse Frustration und Selbstwertbedrohung („pain“) bei einem ansteigenden Vergleich notwendig ist, um eine bessere Leistung zu erzielen („gain“), ergänzten die Autoren damit, dass der Schmerz (z.B. Neid) an sich nicht ausreichend ist, um seine Leistung zu steigern. Beim bösartigen Neid besteht zwar die schmerzvolle Emotion, jedoch kein motivationaler Anstieg, etwas an der Situation zu ändern.

Um die Gefühle der Bewunderung, des gutartigen und bösartigen Neids experimentell zu manipulieren, griffen die Autoren auf die von Stapel et al. angepasste Priming-Methode einer Geschichte über „Hans de Groot“, einem niederländischen Studenten der Tilburg Universität, zurück. Hans wurde als exzellenter Student beschrieben, der für seinen Erfolg und seine Fähigkeiten einen angesehenen Preis gewonnen hat. Nach dem Lesen der Geschichte sollten die Versuchspersonen auf einer 9-stufigen Skala einschätzen, wie stark ihre Bewunderung/ihr gutartiger/ bösartiger Neid Hans gegenüber sei. Anschließend wurde ihnen der „Remote Association Task“ vorgegeben. Die Ergebnisse bestätigten die von Johnson und Stapel (2007) erweiterte Hypothese, dass Individuen nach einer Selbstwertbedrohung im Zuge eines ansteigenden sozialen Vergleichs eine gesteigerte Testleistung zeigen, jedoch nur dann, wenn die schmerzvolle Emotion mit einer erhöhten Motivation einher geht, die eigene Leistung verbessern zu wollen, so wie es die Erfahrung

gutartigen Neids, im Vergleich zum böartigen Neid oder bloßer Bewunderung, bewirken kann.

→ Moreau, C. P. & Herd, K. B. (2010). To each his own? How comparisons with others influence consumers' evaluations of their self-designed products. *Journal of Consumer Research*, 36, 806-819.

Morau und Herd untersuchten in ihrem Artikel, welchen Einfluss verschiedene Variablen auf die Konsumentenbewertung selbst designerter Produkte haben können. In ihrer dritten Studie verwendeten die Autoren das Setting eines realen online Design Wettbewerbs, um der Frage nachzugehen, welchen Einfluss das Timing eines sozialen Vergleichs auf die Richtung des Vergleichs (auf- oder absteigend) und die anschließende Selbstbewertung hat. Bisherige unterschiedliche Ergebnisse in der sozialen Vergleichsforschung, dass einem überlegene Personen als Vorbilder dienen und zu einer Leistungssteigerung motivieren oder aber, dass aufsteigende Vergleiche den Selbstwert reduzieren und somit in schlechteren Leistungen resultieren, führen Stapel und Johnson (2007) beispielsweise darauf zurück, dass die Versuchsmanipulation einer Möglichkeit, den bedrohten Selbstwert wiederherzustellen, bei einigen Experimenten angewandt wurde und bei anderen nicht. Bei der Möglichkeit zur Wiederherstellung des Selbstwerts wird den Versuchspersonen die Option geboten, eine weitere Aufgabe in einem Bereich zu lösen, in dem zuvor eine Selbstwertbedrohung stattgefunden hat. In ihrer Studie fanden Stapel und Johnson (2007), dass eine hohe Selbstwertbedrohung zu einer hohen Leistungssteigerung in einer anschließenden Aufgabe führte. Dazu manipulierten sie das Vergleichsobjekt (extremer ansteigender Vergleich vs. moderat ansteigender Vergleich) und maßen den Effekt auf die Selbstbewertung und die Leistungssteigerung. In ihrem Experiment erhoben sie jedoch nicht den Einfluss der gesteigerten Leistung auf eine nachfolgende Selbstbewertung. Sie meinten lediglich:

[...] future studies examining the efficacy of performance as a means of self-regard repair would be useful. (S. 1055)

Morau und Herd (2010) setzten sich genau diese Aufgabe zum Ziel ihrer dritten Studie. Die Richtung des sozialen Vergleichs wurde im Setting des realen Online-Wettbewerbs durch die Beschreibung der Teilnahme weiterer Mitbewerber (andere Konsumenten vs. professionelle Produktdesigner) manipuliert. Die Variablen des Timings des sozialen Vergleichs und die Möglichkeit zur Wiederherstellung des Selbstwerts wurden so gestaltet, dass den Versuchsteilnehmern nach der Information über ihre Mitbewerber (Profis=ansteigender Vergleich vs. Konsumenten=gleichwertiger Vergleich) entweder explizit vor Beginn ihres Design-Prozesses oder erst danach, gesagt wurde, dass sie um

den Gewinn eines Preisgeldes und die öffentliche Aufmerksamkeit, dass ihr Produkt nach ihnen benannt und sechs Monate lang auf der Homepage des Wettbewerb Betreibers zu sehen sein wird, mitkämpfen. Die Möglichkeit, die eigenen Design-Fähigkeiten der Öffentlichkeit zu präsentieren, sollte eine starke Wiederherstellung des Selbstwerts im Zuge einer vorangehenden Bedrohung (Profi-Designer als Mitbewerber), darstellen. Die Autoren schlussfolgerten daher, dass wenn nach dem sozialen Vergleich mit Mitbewerbern eine Möglichkeit zur Wiederherstellung des Selbstwerts geboten wird (durch die Produktgestaltung vor der Nennung des Preises), ein aufsteigender Vergleich (mit Profi-Designern) zu einer positiveren Bewertung der selbst designten Produkte führt:

H3a: When a social comparison is followed by a repair opportunity, upward comparison targets (i.e., professional designers) will yield higher evaluations of self-designed products than will equivalent comparison targets (i.e., consumer designers).

H3b: When a social comparison is not followed by a repair opportunity, evaluations of self-designed products will be unaffected by the type of comparison target.

→ Pettit, N. C. & Lount, R. B. (2010). Looking down and ramping up: The impact of status differences on effort in intergroup contexts. *Journal of Experimental Social Psychology*, 46, 9-20.

Johnson und Stapel (2007) fanden in ihrer Studie, dass die Erfahrung einer sozialen Selbstwertbedrohung zu einer Motivations- und Leistungssteigerung führen kann, vor allem dann, wenn die Bedrohung durch einen Vergleich mit einer dem Individuum auf einer bestimmten Ebene überlegenen Person stattfindet. Von dieser Erkenntnis ausgehend, schlussfolgerten Pettit und Lount, dass für einen Gruppenkontext gegenteilige Regeln gelten. Demnach führt der Vergleich eines Gruppenmitglieds mit einem Mitglied einer anderen Gruppe („outgroup“), welche einen geringeren sozialen Status hat oder weniger angesehen ist (=absteigender Vergleich), zu einer erhöhten Motivation und einer Leistungssteigerung. Dieser umgekehrte Effekt lässt sich darauf zurückführen, dass durch den Vergleich mit einer minderwertigeren Gruppe, die Selbstkategorisierung dahingehend bedroht wird, dass ein schlechteres Abschneiden gegen das outgroup-Mitglied für sich selbst und andere bedeuten würde, dass man der Mitgliedschaft in der angeseheneren Gruppe nicht würdig ist. In Studie 4 untersuchten die Autoren ebenfalls eine auf Johnson und Stapel (2007) aufbauende Hypothese, nämlich dass eine der Selbstbedrohung durch den Vergleich mit einer im Status niedrigeren/höheren sozialen Gruppe vorausgehende Selbstaffirmation dazu führt, dass die Versuchspersonen sich weniger bedroht fühlen und

sich in einer nachfolgenden Aufgabe somit weniger anstrengen als Versuchspersonen, die zuvor keine Selbstaffirmation durchführen.

→ Pliner, P., Rizvi, S. & Remick, A. K. (2009). Competition affects food choice in women. *International Journal of Eating Disorders*, 42, 557-564.

Pliner et al. untersuchten in ihrer Studie, welchen Einfluss ein sozialer Vergleich durch eine aufsteigende Bedrohungssituation auf die anschließende Essenswahl von Frauen hat. Als Versuchspersonen wurden weibliche Psychologiestudentinnen des ersten Semesters ausgewählt, die gemeinsam mit einer zweiten angeblichen Psychologiestudentin (in Wirklichkeit war diese eine eingeweihte Mitarbeiterin des Experiments) an einem Tisch in einem Testraum drei verschiedene Aufgaben zur Erhebung der Kreativität, des sozialen Einfühlungsvermögens und des kognitiven Denkens lösen sollten. Für die Teilnehmerinnen in der Versuchsgruppe stellte die zweite Studentin eine bedrohende Wettbewerbssituation dar, da ihnen gesagt wurde, dass ihre Leistungen in den drei Aufgaben im Anschluss mit den Leistungen ihrer Konkurrentin verglichen werden und die Siegerin einen Gutschein von 100 Dollar erhält. Um die angebliche Überlegenheit der zweiten Studentin hervorzuheben, wurde diese instruiert, während der Bearbeitung der Aufgaben Sätze wie „Ich hab's“ oder „Ich werde bestimmt gewinnen“, lautstark von sich zu geben. Nach Beendigung der drei Aufgaben sollten sich die Teilnehmerinnen aus der Reihe verschiedener Nahrungsmittel eines auswählen (sie wurden instruiert, dass es sich dabei um ein unabhängiges Experiment zur Konsumenteneinschätzung verschiedener Produkte handle). Die Autoren Pliner et al. (2009) stellten auf Basis der Erkenntnisse von Johnson und Stapel (2007) die Hypothese auf, dass ein abgeschwächter Selbstwert, durch eine schlechtere Leistung im Vergleich zu einer überlegenen Person, wiederhergestellt werden kann, indem die Leistung in einer anderen Domäne verbessert wird. Daher sollten Studienteilnehmerinnen, die zuvor eine Selbstwertbedrohung erfahren haben, höher motiviert sein, ihren Selbstwert durch die Wahl eines kalorienarmen bzw. nährstoffreichen Nahrungsmittels wieder herzustellen. Die Ergebnisse bestätigten diese Hypothese und zeigten zusätzlich, dass sich besonders Diäthalterinnen in ihrer Essenswahl von sozialen Vergleichssituationen, die eine Selbstwertbedrohung darstellen, beeinflussen ließen.

→ Zarankin, T. G. & Wall, J. A. Jr. (2007). Negotiator's information sharing: The effects of opponent behavior and information about previous negotiator's performance. *Negotiation and Conflict Management Research*, 5, 162-181.

Zarankin et al. untersuchten innerhalb eines wirtschaftlichen Unternehmenskontexts welchen Einfluss verschiedene Faktoren auf die Bereitschaft, in einer Verhandlung mit dem Verhandlungspartner Informationen zu teilen (z.B.: BATNA = „best alternative to negotiated agreement“ oder Informationen über die eigenen Interessen, Ressourcen oder Ausgaben), haben. Dabei wurde unter anderem erforscht, ob die Information über das frühere Verhalten des Verhandlungspartners das Ausmaß der eigenen geteilten Information beeinflusst. In sozialen Vergleichssituationen, also auch bei geschäftlichen Verhandlungen, strebt der Mensch nach einer Erhaltung oder Erhöhung des Selbstwerts und nach einer Selbstevaluation. Im Aufbau ihrer Hypothesen stützen sich Zarankin et al. auf die Autoren Johnson und Stapel (2007), die bereits fanden, dass Individuen zu aufsteigenden sozialen Vergleichen tendieren, um die eigene Leistung voran zu treiben und somit eine hohe Selbstevaluation zu erzielen. Im Kontext einer Verhandlung sollte dieses Selbsterhaltungsziel das Erfolgsstreben der Verhandlungspartner erhöhen, indem mehr Informationen geteilt werden. Die Autoren erstellten also die Hypothese, dass Verhandlungspartner dann zu aufsteigenden sozialen Vergleichen tendieren, wenn Informationen über das frühere Verhalten verschiedener Verhandlungspartner existieren, wobei einige sehr gut und andere eher schlecht bei der Verhandlung abschnitten. Im Folgenden sind die Hypothesen im originalen Wortlaut dargestellt:

Hypothesis 4a: Negotiators presented with information indicating that some previous negotiators achieved high outcomes and some netted low outcomes will develop higher goals/aspirations than will those given information of moderate previous outcomes.

Hypothesis 4b: The aspirations of negotiators receiving information about previous negotiators' performance are positively related to their disclosure of their BATNA.

Hypothesis 4c: The aspirations of negotiators receiving information about previous negotiators' performance are positively related to their disclosure of their payoffs.

6. Stapel, D. A. & Spears, R. (1996). Event accessibility and context effects in causal inference: Judgment of a different order. *Personality and Social Psychology Bulletin*, 22, 979-992. (Retraction, vol 22, pg 982, 1996)

Stapel und Spears (1996) untersuchten in zwei Studien, auf welche Weise ein verfügbares Kontext-Ereignis kausale Urteile bezüglich eines mehrdeutigen Zielobjekts beeinflusst, je nachdem in welcher Reihenfolge die zwei Ereignisse interpretiert werden. Ob die Beurteilung eines Zielereignisses zu einem Assimilations- oder Kontrasteffekt führt, hängt nach den Autoren von der Deutlichkeit der Kontextinformation zum Zeitpunkt der Beurteilung ab. Die Ergebnisse unterstützen die Annahme, dass Versuchspersonen,

die ein Kontext-Ereignis vor einem Ziel-Ereignis bewerten sollten, das Kontext-Ereignis als markanter wahrnahmen, was zu einem Vergleich und somit zu einem Kontrasteffekt führte, während die umgekehrte Beurteilungsreihenfolge dazu führte, dass das Kontext-Ereignis als weniger markant wahrgenommen wurde und in einem Assimilationseffekt resultierte.

Stapel und Spears wurden 20 Mal zitiert, 6 Mal von Stapel selbst einmal von der zuständigen Fachzeitschrift, als ihr Artikel zurückgezogen wurde. Folgende Artikel bauen auf den Erkenntnissen der beiden Autoren auf:

→ Seta, J. J., Seta, C. E. & McElroy, T. (2003). Attributional biases in the service of stereotype maintenance: A schema-maintenance through compensation analysis. *Personality and Social Psychology Bulletin*, 29, 151-163.

Seta et al. (2003) gingen in sechs Studien der Frage nach, ob sich der Attributionsstil einer Person durch ihre Neigung, stereotyp inkonsistente Verhaltensweisen eines Gruppenmitglieds zu kompensieren, ändern kann. Individuen nutzen die Kompensation, um ein Schema aufrecht zu erhalten, da sie ansonsten bestehende Stereotype angesichts einer widerlegenden Information aufgeben müssten. Kompensation wird also dann angewandt, wenn ein Ereignis oder eine Erfahrung mit einem Schema oder Stereotyp inkonsistent ist und die beobachtende Person implizit oder explizit motiviert ist, diese Stereotype durch die Fähigkeit, kompensatorische Informationen zu generieren, aufrecht zu erhalten. Wenn die beobachtende Person die inkonsistenten Handlungen einer Zielperson nicht kompensieren kann, kann es vorkommen, dass sie sich für die Aufrechterhaltung ihres Schemas ein anderes Gruppenmitglied für den Kompensationsprozess sucht. In den von Seta et al. (2003) durchgeführten Studien wurden die Versuchspersonen in der Experimentalbedingung mit einer stereotyp inkonsistenten Information über ein Gruppenmitglied (Krankenschwester oder Minister) konfrontiert und danach gebeten, ihr Urteil über ein stereotyp konsistentes Verhalten dieser Person („same target condition“) oder eines weiteren Gruppenmitglieds („fellow group member condition“) abzugeben. In der „same target“ Bedingung lasen die Versuchspersonen einen Text über einen Minister, der einen 7-jährigen Jungen sexuell belästigt und einem Jugendlichen Heroin verkauft hat. Anschließend bekamen sie die Information, dass dieser Minister am Wochenende freiwillig bei einer gemeinnützigen Organisation mitarbeitet. Die Versuchspersonen sollten nun auf einer 7-stufigen Ratingskala einschätzen, warum der Minister bei der gemeinnützigen Organisation mitarbeitet (1=um einen guten Eindruck zu machen bis 7=weil er von der Organisation total überzeugt ist). Die Versuchspersonen der „fellow group member“ Bedingung lasen

denselben Text über das normabweichende Verhalten des Ministers, sollten anschließend jedoch über einen anderen unabhängigen Minister urteilen, warum er am Wochenende einer gemeinnützigen Organisation hilft. Die Ergebnisse zeigten, dass Versuchspersonen, die zuerst über das normabweichende Verhalten des Ministers informiert wurden, sein soziales Engagement, das mit der vorherigen Information inkonsistent erschien, darauf zurückführten, dass er damit andere beeindrucken möchte. Die Versuchspersonen in der „fellow group member“ Bedingung attribuierten das soziale Engagement des Ministers auf die Dimension, die mit der grundsätzlichen Disposition eines Ministers konsistenter ist, nämlich dass ihm die gemeinnützige Organisation persönlich viel bedeutet.

In Experiment 3 erweiterten die Autoren ihre Annahmen auf Basis der Erkenntnisse von Stapel und Spears (1996), dass die Verhaltensweisen einer weiteren Zielperson im Vergleich zum normabweichenden Verhalten einer handelnden Person (z.B. Krankenschwester) extremer wahrgenommen werden, unabhängig davon, ob die weitere Zielperson der gleichen Gruppe angehört oder nicht. Daher bekamen die Versuchspersonen in Experiment 3, in der inkonsistenten Bedingung, einen Text zu lesen, der eine Krankenschwester beschrieb, die einem Kind mitten ins Gesicht schlug. Anschließend fragte eine Person eine andere Krankenschwester („same group identity condition“) oder eine Verkäuferin („different group identity condition“), ob sie einen freien Sonntag opfern würde, um bei einer gemeinnützigen Organisation mitzuhelfen. Nach dem Lesen des Textes sollten die Versuchspersonen wieder einschätzen, auf welche Dimension (Selbstdarstellung vs. Großzügigkeit) sie die Wahrscheinlichkeit zurückführen, dass die Krankenschwester oder die Verkäuferin an einem Sonntag aushelfen. Die Ergebnisse replizierten die Annahmen der ersten beiden Experimente, folgten allerdings nicht der Hypothese von Stapel und Spears (1996), dass die Versuchspersonen die Attribution der Verhaltensweisen einer Person ändern, die nicht mit der Zielgruppe assoziiert wird. Versuchspersonen führten extreme interne Attributionen nur für verwandte Gruppenmitglieder durch und bestätigten somit die anfängliche Hypothese, dass die Attribution aus dem Streben resultiert, stereotype Schemata aufrecht zu erhalten.

→ Dijksterhuis, A., Spears, R. & Lepinasse, V. (2001). Reflecting and deflecting stereotypes: Assimilation and contrast in impression formation and automatic behavior. *Journal of Experimental Social Psychology*, 37, 286-299.

In ihrer Studie zu moderierenden Variablen von Assimilations- und Kontrasteffekten im automatischen Verhalten, untersuchten die Autoren die unterschiedliche Wirkung von stereotypen Begriffen vs. konkreten Beispielen als Priming-Stimuli, die Präsentation einer einzelnen Person vs. einer einzelnen Person während einer „Cognitive Load“ Aufgabe

und den Effekt der Wahrnehmung einer Gruppe, der zusätzlich durch den Faktor der Motivation beeinflusst werden kann. In ihrem theoretischen Aufbau zitieren sie die Erkenntnisse aus fünf Studien von Stapel und Kollegen (Dijksterhuis et al., 1998; Stapel et al., 1996*#; Stapel et al., 1997; Stapel et al., 1998; Stapel & Spears, 1996*#), wovon zwei Artikel wegen betrügerischer Inhalte von den Fachzeitschriften zurückgewiesen wurden.

→ Seta, J. J., Seta, C. E. & Hundt, G. M. (2001). Exaggerating the differences between relatively successful groups: Identity orientation as a perceptual lens. *Group Dynamics-Theory Research and Practice*, 5, 19-32.

Aufbauend auf Stapel und Spears Erkenntnissen (1996) und anderer Autoren (e.g., Martin & Seta, 1983) stellten Seta et al. (2001) in ihrem ersten Experiment die Hypothese auf, dass wenn sich für ein Individuum der Wert als Mitglied einer Gruppe erhöht, gleichzeitig auch seine Wahrnehmung für Unterschiede in verschiedenen Fähigkeitsdimensionen zwischen erfolgreichen und nicht erfolgreichen Gruppen verstärkt wird. Individuen mit einer hohen Bewertung für eine Gruppenmitgliedschaft werden auch erfolgs-bezogene Aktivitäten einer erfolgreichen Gruppe mehr schätzen als die einer nicht erfolgreichen Gruppe, auch wenn sie selbst nicht Mitglied dieser Gruppe sind. Zur Erhebung der individuellen Wichtigkeit einer Gruppenmitgliedschaft wurde der „Aspects of Identity Questionnaire“ (Cheek, 1989) vorgegeben. Versuchspersonen mit einer hohen kollektiven Identitäts-Orientierung sollten nach Seta et al.s Hypothese die Aspekte ihres Selbstkonzepts höher bewerten, die auf einer Gruppenmitgliedschaft beruhen. Den Versuchspersonen wurde ein Szenario vorgegeben, in dem eine Gruppe X (Gewinnergruppe) und eine Gruppe Z (Verlierergruppe) gemeinsam mit 20 anderen Firmen an einem monatelangen nationalen Wettkampf mit Aufgaben zu kognitiven Fähigkeiten teilnahmen. In diesen Aufgaben traten immer zwei Gruppen gegeneinander an und der Sieger bekam pro Gewinn einen Punkt. In der Geschichte wurde erzählt, dass Gruppe X und Z eine intensive Rivalität pflegten und X schlussendlich mit sieben Siegen und fünf Niederlagen und Z mit fünf Siegen und fünf Niederlagen den Wettkampf beendeten. Anschließend sollten die Versuchspersonen das Ausmaß der Fertigkeiten von Gruppe X und Z und den Wunsch, an einer Rede der beiden Gruppen teilzunehmen, auf einer 9-stufigen Rating-Skala einschätzen. Die Ergebnisse zeigten, dass eine hohe Ausprägung der kollektiven Identität mit einer hohen Differenz des Fähigkeits-Ratings der beiden Gruppen (Verlierer vs. Gewinner) in Zusammenhang stand. Damit wurde die Hypothese der Autoren bestätigt, dass eine hohe kollektive Identität zu einer verstärkten Wahrnehmung der Fähigkeitsunterschiede zwischen einer erfolgreichen und einer nicht erfolgreichen Gruppe führt.

→ Weary, G. T., Tobin, S. J. & Reich, D. A. (2001). Chronic and temporary distinct expectancies as comparison standards: Automatic contrast in dispositional judgments. *Journal of Personality and Social Psychology*, 80, 365-380.

In Weary et al.s Studie (2001) wurde in vier Experimenten erhoben, welchen Einfluss Kontext-Informationen im Rahmen sozialer Urteile auf Assimilations- und Kontrasteffekte haben können. Die Autoren stützen ihre Hypothesen auf die theoretischen Modelle Stapels und seinen Kollegen (Stapel & Spears, 1996; Stapel & Winkielman, 1998), die besagen, dass Assimilation dann entsteht, wenn das zu bewertende Zielobjekt und der vorhandene Kontext derselben Kategorie zugeordnet werden können und Kontrast, wenn Zielobjekt und Kontext verschiedenen Kategorien zugeteilt werden. Wenn eine verfügbare und anwendbare Kontext-Information subliminal geprimt wird oder auf andere Art und Weise in undeutlicher Form („indistinct“) präsentiert wird, kommt es eher zu einer einschließenden Kategorisierung von Kontext und Zielobjekt und in Folge zur Assimilation. Wenn die Kontext-Information für die Bewertung des Zielobjekts jedoch klar erkennbar („distinct“) ist und deutliche Unterschiede zwischen Objekt und Kontext wahrgenommen werden können, wird die Kontext-Information als Vergleichsanker herangezogen und hat einen Kontrasteffekt zur Folge.

In Studie 1 demonstrierten die Autoren, dass das Deutlich-Machen von zukunftsbezogenen Leistungserwartungen eines Individuums dazu führt, dass es diese Erwartungen in einer nachfolgenden Leistungsbeurteilung einer Zielperson als Vergleichsstandard heranzieht. Die Versuchspersonen wurden instruiert, die Leistung eines Kindes in verschiedenen Tests zum räumlichen Vorstellungsvermögen zu beobachten und anschließend das Ausmaß der kindlichen Intelligenz zu schätzen. Danach bekamen sie ein Video sehen, das ein Kind beim Lösen dieser Aufgaben zeigte. Je nachdem, ob die Versuchspersonen die Instruktion bekamen, sich auf ihre Erwartungen bezüglich der Testleistung des Kindes zu konzentrieren oder sich während der Videopräsentation eine 8-stellige Nummer merken mussten („Cognitive Load“ Aufgabe), schätzten sie die Testleistung des Kindes, das allgemeine Intelligenzniveau und die Schwierigkeit der Aufgabe ein. Die Ergebnisse entsprachen der Annahme, dass Versuchsteilnehmer, auf deren unmittelbare Leistungserwartungen klar aufmerksam gemacht wurde, diese als Vergleichsstandard heranzogen und das Kind entgegen ihrer Erwartungen bewerteten (Kontrasteffekt). Das Merken einer 8-stelligen Nummer führte bei den Versuchsteilnehmern, im Vergleich zur Kontrollgruppe, die sich keine Zahl merken musste, ebenfalls dazu, dass die relativ automatisch gefassten Schlüsse zu den dispositionalen Eigenschaften des Kindes in einem Kontrasteffekt zu den Leistungserwartungen resultierten.

In Studie 4 wandten die Autoren eine Methode von Stapel und seinen Kollegen an (Stapel & Spears, 1996; Stapel & Winkielman, 1998), um die Deutlichkeit der Leistungserwartungen durch die explizite Evaluation der Kontextinformation unmittelbar vor der Bewertung der Zielperson (des Kindes) zu manipulieren. Vor der Präsentation des Videos wurde allen Versuchspersonen ein ausgefülltes Auswertungsblatt gegeben, das die erfolgreichen vs. nicht erfolgreichen Testergebnisse des Kindes vor drei Jahren veranschaulichen sollte. Die Hälfte der Versuchspersonen sollte auf Basis dieser Kontextinformation bereits vor dem Ansehen des Videos das Fähigkeits- und Intelligenzniveau des Kindes einschätzen und der vorangehenden Theorie zufolge einen Kontrasteffekt zeigen. Alle weiteren Ergebnisse können in Weary et al.s Studie (2001) nachgelesen werden.

7. Blanton, H. & Stapel, D. A. (2008). Unconscious and spontaneous and ... Complex: The three selves model of social comparison assimilation and contrast. *Journal of Personality and Social Psychology*, 94, 1018-1032. (Retracted article. See vol. 104, pg. 198, 2013)

Einige theoretische Perspektiven gehen davon aus, dass soziale Vergleiche zu einfachen automatischen Effekten führen, wenn sie außerhalb unseres Bewusstseins ablaufen. Manche Theorien gehen davon aus, dass ein Vergleich zwischen einem aktivierten Konstrukt und uns selbst zu einem Assimilationseffekt führen, andere postulieren unter bestimmten Umständen einen Kontrasteffekt. Die Autoren untersuchten in dieser Hinsicht die beiden Einflussfaktoren des sozialen Kontextes und des Typs der Selbstkonzept Aktivierung. Sie fanden dabei, dass die Aktivierung des persönlichen Selbst zu einem Kontrast und die Fokussierung auf ein kollektives oder mögliches Selbst, zu einem Assimilationseffekt führen. Diese Ergebnisse lassen vermuten, dass soziale Vergleichseffekte spontan und außerhalb unseres Bewusstseins auftreten können.

Blanton und Stapel (2008) wurden 18 Mal zitiert. Zwei Zitierungen sind Selbstzitate Stapels und eine betrifft das *Retraction Statement* der Fachzeitschrift. Folgende Artikel bauen auf den Erkenntnissen der Autoren auf:

- Dyczewski, E. A. & Markman, K. D. (2012). General attainability beliefs moderate the motivational effects of counterfactual thinking. *Journal of Experimental Social Psychology*, 48, 1217-1220.

Wenn Personen ein Feedback über ihre erbrachte Leistung in einem Test erhalten, dann werden meistens Vergleichsprozesse der eigenen Leistung mit gewissen Standards

aktiviert. Ein solcher Vergleichsprozess wird in der Fachliteratur als „counterfactual thinking“, also kontrafaktisches Denken bezeichnet und involviert immer die mentale Simulation einer Standard-Information. Die kontrafaktischen Standards reflektieren manchmal eine vorgestellte bessere Realität („upward counterfactuals“) und manchmal eine schlechtere („downward counterfactuals“). Bisherige Literatur konnte zeigen, dass „upward counterfactuals“ als Folge einer unterdurchschnittlichen Leistung in wiederholten Aufgaben motivierender auf die Versuchsteilnehmer wirkten als ein „downward“ Vergleich. In der vorliegenden Studie gingen die Autoren also davon aus, dass „upward counterfactuals“ nur dann einen motivierenden Charakter haben, wenn die Versuchsteilnehmer daran glauben, dass eine Verbesserung grundsätzlich erreichbar ist. Dagegen sollten „upward counterfactuals“ die Motivation verringern und „downward counterfactuals“ die Motivation steigern, wenn die Versuchsteilnehmer nicht daran glauben, dass eine Verbesserung möglich ist. Die Ergebnisse zweier Experimente zeigten, dass die inkrementellen Theoretiker (Versuchsteilnehmer, die glaubten, dass Intelligenz-bezogene Fähigkeiten verbessert werden können) mehr Motivation und eine bessere Testleistung in Folge eines „upward“ Vergleichs im Zuge ihres Feedbacks zeigten, während die Einheitstheoretiker (Versuchsteilnehmer, die nicht daran glaubten, dass Intelligenz-bezogene Fähigkeiten verbessert werden können), größere Motivation und eine bessere Testleistung erzielten, wenn sie „downward“ Feedback erhielten. In Studie 2 erhielten die Versuchsteilnehmer ein negatives Feedback auf ihre Leistung in einem verbalen Intelligenztest und kurz darauf noch ein negatives Feedback zu einer Aufgabe, die mit verbaler Intelligenz in Verbindung stand. Danach wurden die Versuchsteilnehmer entweder mit einem „counterfactual-upward“ Vergleich („imagine how your performance on the first task could have been better“) oder einem „counterfactual-downward“ Vergleich („imagine how your performance on the first task could have been worse“) konfrontiert. Anschließend sollten sie das Ausmaß ihrer Motivation auf einer 9-stufigen Ratingskala angeben, noch eine verbale Intelligenzaufgabe lösen zu wollen. Um die Versuchsteilnehmer in die beiden Gruppen inkrementelle Theoretiker vs. Einheitstheoretiker einzuteilen, füllten sie zuvor den „Theories of Intelligence Scale“ (Dweck, 2000) aus und wurden danach nach dem von Stapel und Blanton (2008) entwickelten Verfahren in zwei Gruppen eingeteilt (nur Versuchsteilnehmer, die sich im oberen oder unteren Drittel der Scores bewegten, durften am nachfolgenden Experiment teilnehmen).

→ Corcoran, K. & Mussweiler, T. (2010). The cognitive miser's perspective: Social comparisons as a heuristic in self-judgements. *European Review of Social Psychology*, 21, 78-113.

Die Autoren geben aus der Sichtweise des metaphorischen „kognitiven Geizhalses“ („cognitive miser“) einen Überblick über die bisherige soziale Kognitionsforschung. Der kognitive Geizhals ist ein Bild für einen Menschen, der nur über limitierte kognitive Ressourcen im sozialen Denken verfügt und daher auf schnelle und effiziente Informationsverarbeitungsstrategien zurückgreifen muss. Eine solche Strategie stellt beispielsweise der soziale Vergleich dar, der als Heuristik angewandt wird, um sich selbst zu bewerten. Da Stapel einen erheblichen Beitrag zur sozialen Vergleichsforschung geleistet hat, gehen die Autoren auf mehrere Ergebnisse seiner Studien ein und zitieren insgesamt acht seiner Artikel als Erst- oder Co-Autor (Blanton & Stapel, 2008*#; Marx, Stapel & Muller, 2005; Stapel & Suls, 2004#; 2007#; Stapel & Blanton, 2004#; Stapel & Koomen, 1998; 2000; 2001), wovon einer von der betreffenden Fachzeitschrift zurückgewiesen und vier von der Untersuchungskommission als betrugschaft identifiziert wurden.

→ Want, S. C. (2009). Meta-analytic moderators of experimental exposure to media portrayals of women on female appearance satisfaction: Social comparisons as automatic processes. *Behaviour Research and Therapy*, 47, 790-796.

Der Autor gibt einen Überblick über die bisherige Forschung zu dem Effekt, dass idealisierte Frauenbilder in den Medien zu sozialen Vergleichseffekten bei Frauen führen und in Folge die Zufriedenheit mit dem eigenen Körper beeinflussen können.

In der Forschung zu Medieneffekten ist ein Pre-Post-Design zur Erhebung der Zufriedenheit mit dem eigenen körperlichen Erscheinungsbild üblich, um überhaupt feststellen zu können, welchen Effekt präsentierte Werbungen oder Filme auf den Konsumenten haben können. Jedoch fanden Forscher (Hamilton, Mintz & Kashubeck-West, 2007), dass ein Pre-Test zur Erhebung der körperlichen Erscheinung, die Sensibilität von Versuchsteilnehmern bezüglich körperbezogener Themen primen und nach Blanton und Stapel (2008) zu negativen Selbstbewertungen im Zuge eines aufsteigenden Vergleichs mit den präsentierten Bildern führen kann. Daher schlussfolgerte Want (2009) für seine 7. Hypothese:

H7: Effect sizes will be larger when participants are asked questions regarding appearance before media exposure.

8. Stapel, D. A. & Koomen, W. (2006). The flexible unconscious: Investigating the judgmental impact of varieties of unaware perception. *Journal of Experimental Social Psychology*, 42, 112-119. (Retracted article. See vol. 49, pg. 782, 2013)

Stapel und Koomen (2006) wurden 15 Mal zitiert, darunter befinden sich das *Retraction Statement* und sechs Selbstzitate. Keiner der untersuchten Artikel baut direkt auf den Erkenntnissen der beiden Autoren auf.

9. Avramova, Y. R. & Stapel, D. A. (2008). Moods as spotlights: The influence of mood on accessibility effects. *Journal of Personality and Social Psychology*, 95, 542-554. (Retracted article. See vol. 103, pg. 471, 2012)

In drei Studien explorierten Avramova und Stapel (2008), auf welche Weise die Stimmung den Gebrauch verfügbarer Informationen auf soziale Urteile beeinflussen kann. Je nachdem, ob die Versuchspersonen mit abstrakten Eigenschaften oder konkreten Personen-Eigenschafts-Verbindungen geprimt worden waren, führte eine positive und negative Stimmung zu unterschiedlichen Effekten in der Informationsverarbeitung. In der ersten Studie wurde der Einfluss positiver vs. negativer Stimmung auf die Beurteilung eigenschafts-implizierender Verhaltensweisen gemessen, wobei sich herausstellte, dass eine positive Stimmung zu einer Assimilation in der Bewertung führte und eine negative Stimmung zu einem Kontrasteffekt. In Studie 2 wurde dieser Effekt in einem subliminalen Priming Design repliziert. In der dritten Studie konnte demonstriert werden, dass der durch die eigenschafts-implizierenden Verhaltensweisen aktivierte Informationstyp ebenfalls von der Stimmung abhängig ist, so dass in einer positiven Stimmung eher abstrakte Eigenschaftsinformationen aktiviert werden und in einer negativen Stimmung spezifische Personen-Eigenschafts-Verbindungen hergestellt werden.

Avramova und Stapel (2008) wurden 14 Mal zitiert, 5 Mal von Stapel selbst und einmal durch das *Retraction Statement* der Fachzeitschrift. Folgende Artikel bauen auf deren Erkenntnissen auf:

→ Chang, H. Y., Teng, C. I., Chu, T. L., Chang, H. T. & Hsu, W. H. (2012). Impact of nurse agreeableness and negative mood of nursing supervisors on intention to help colleagues. *Journal of Advanced Nursing*, 68, 636-646.

Chang et al. (2012) erforschten in ihrer Studie den Zusammenhang zwischen der Verträglichkeit von Krankenschwestern und der negativen Stimmung ihrer Vorgesetzten auf die Bereitschaft von Krankenschwestern, ihren Kolleginnen zu helfen. Der „Mood influence theory“ von Avramova und Stapel zufolge (2008), führt eine negative Stimmung,

die von Gefühlen der Feindseligkeit, Angst und Sorgen gekennzeichnet ist, zu einem verminderten gehorsamen und kooperativen Verhalten und kann in interpersonellen Konflikten resultieren. Damit kann eine negative Stimmung auch zum potentiellen Störfaktor am Arbeitsplatz werden.

→ Huntsinger, J. R. (2011). Mood and trust in intuition interactively orchestrate correspondence between implicit and explicit attitudes. *Personality and Social Psychology Bulletin*, 37, 1245-1258.

Huntsinger geht in seiner Studie davon aus, dass die affektive Stimmung die Tendenz einer Person vorhersagt, ihrer Intuition zu vertrauen oder zu misstrauen. Wenn das Vertrauen in die eigene Intuition geprimt wird, dann sollte eine induzierte positive Stimmung zu einer größeren Übereinstimmung zwischen der expliziten (bewusste, kognitiv verarbeitete, verbalisierte) Einstellung und der impliziten (schnelle, automatische, unbewusste) Einstellung führen. Ein geprimtes Misstrauen in die eigene Intuition sollte bei einer positiven Stimmung dagegen zu einer geringeren Einstellungsübereinstimmung führen. Für die experimentelle Manipulation der Stimmung wandte Huntsinger (2011) für sein viertes Experiment ein von Avramova und Stapel (2008) entwickeltes Verfahren an. Dabei lasen die Versuchspersonen eine von zwei Geschichten über eine junge Künstlerin, in deren Leben entweder eine Reihe positiver Ereignisse (z.B. ein Stipendium für ihr Kunststudium) oder negativer Schicksalsschläge (z.B. das Erkranken an rheumatoider Arthritis, welche ihre körperliche Leistungsfähigkeit als Künstlerin lähmte), eingetreten sind.

→ Kim, M. G. & Mattila, A. S. (2010). The impact of mood states and surprise cues on satisfaction. *International Journal of Hospitality Management*, 29, 432-436.

Kim & Mattila erhoben in ihrer Studie (2010) den Zusammenhang zwischen den Variablen der Stimmung und Überraschungseffekten auf die Konsumentenzufriedenheit. Dazu wurde zu Beginn des Experiments die Stimmung von Konsumenten geprimt und danach sollten sie sich ein hypothetisches Szenario eines Abendessens im Restaurant vorstellen, wobei die Versuchspersonen in einer Version der Geschichte ein gratis Dessert als positive Überraschung erhielten und in der zweiten Version einen verschimmelten Kuchen als negative Überraschung. Im Anschluss sollten sie auf einer 7-stufigen Ratingskala angeben, wie zufrieden sie mit diesem Abend gewesen wären. Um die Stimmung der Versuchspersonen zu Beginn des Experiments zu primen, wurde ihnen eine fröhliche oder traurige Musik vorgespielt, zu der sie gleichzeitig negative/positive Ereignisse aus ihrem Leben aufschreiben sollten. Um sicher zu gehen, dass diese Technik auch

tatsächlich die Stimmung der Versuchspersonen manipulierte, griffen die Autoren auf die Musikauswahl von Stapel und Avramova (2008) zurück, die mit den Stücken „Eine kleine Nacht Musik“ von Mozart (fröhliche Stimmung) und „Adagio for Strings“ von Samuel Barber (traurige Stimmung) ihre Versuchspersonen bereits erfolgreich primen konnten.

10. Renkema, L. J., Stapel, D. A., Maringer, M. & van Yperen, N. W. (2008). Terror management and stereotyping: Why do people stereotype when mortality is salient? *Personality and Social Psychology Bulletin*, 34, 553-564. (Retracted article. See vol. 39, pg. 264, 2013)

Renkema et al. untersuchten in drei Studien zwei Möglichkeiten, wie eine Bedrohung der eigenen Sterblichkeit zu stereotypen Urteilen führen kann. Individuen fällen stereotype Urteile, da sie immer eines von zwei Zielen verfolgen: Entweder versuchen sie dadurch ihre soziale Umwelt besser zu verstehen („comprehension goal“), indem sie im Falle limitierter kognitiver Ressourcen oder einer mehrdeutigen Information auf Stereotype zurückgreifen, um diese „Lücken“ zu füllen, oder sie versuchen dadurch den eigenen Selbstwert zu erhalten beziehungsweise zu stärken („enhancement goal“), indem negative Stereotype über eine Person oder eine Gruppe durch den Vergleich mit sich selbst zur Aufwertung des Selbstwerts herangezogen werden. Das Bewusstsein um die eigene Sterblichkeit („mortality salience“) aktiviert nach Renkema et al. (2008) sowohl das „comprehension goal“ als auch das „enhancement goal“. In einer Wettbewerbssituation zwischen Gruppen oder bei Vorhandensein eines Konflikts wird eher das Selbstwert-Erhaltungsziel verfolgt, während das Verständnisziel angestrebt wird, wenn keine Konfliktsituation vorhanden ist. Wenn das gepriimte Sterblichkeitsbewusstsein ein „comprehension goal“ aktiviert, treten beim Individuum sowohl positive als auch negative Stereotype auf, während bei der Aktivierung eines „enhancement goals“ lediglich negative stereotype Urteile gefällt werden.

Renkema et al.s Arikel (2008) wurde 13 Mal zitiert, darin enthalten sind drei Selbstzitate und ein *Retraction Statement* der Fachzeitschrift. Folgende Artikel stützen ihre Annahmen auf die Erkenntnisse der Autoren:

→ Henry, E. A., Bartholow, B. D. & Arndt, J. (2010). Death on the brain: Effects of mortality salience on the neural correlates of ingroup and outgroup categorization. *Social Cognitive and Affective Neuroscience*, 5, 77-87.

Aus bisheriger Forschung ist bekannt, dass Gedanken an den eigenen Tod („mortality salience“; MS) soziale Prozesse innerhalb einer Gruppe beeinflussen können, jedoch

wurde bisher kaum erforscht, in welchem Ausmaß die MS auf die neuronale Hirnaktivität wirkt, welche maßgeblich für die Personenwahrnehmung verantwortlich ist. In der vorliegenden Studie wurde gemessen, welchen Einfluss die Aktivierung der MS, im Vergleich zu einer Kontrollgruppe, auf die evaluativen Prozesse von weißen Versuchspersonen haben, die in einer computerisierten Aufgabe, weißen („ingroup“) oder schwarzen („outgroup“) Personen den richtigen Gesichtsausdruck („happy vs. angry“) zuordnen mussten. Die Studie von Renkema et al. (2008) konnte bereits zeigen, dass die Erinnerung an die eigene Sterblichkeit zu stereotypen Denkmustern und einer negativeren Einstellung gegenüber „outgroup“ Mitgliedern führt. Demnach schlussfolgerten auch die Autoren Henry et al. (2010), dass wenn die Aktivierung der MS die Sensitivität für „ingroup“ bezogene Ziele erhöht, die N2 Amplitude (ein EEG Maß, das die elektrische Aktivität des Gehirns infolge eines präsentierten Stimulus misst) bei weißen Gesichtern in der MS Bedingung größer ist als bei schwarzen Gesichtern. Und da das aktivierte Bewusstsein um die eigene Sterblichkeit die Präferenz für konsistente Zusammenhänge erhöht (Renkema et al., 2008), sollte die MS zu einem kognitiven Konflikt und einem Interaktionseffekt (der Variablen Rasse und Gesichtsausdruck in der N2 Amplitude) bei der Präsentation von bösen weißen Gesichtern und/oder fröhlichen schwarzen Gesichtern führen.

→ Landau, M. J., Greenberg, J. & Rothschild, Z. K. (2009). Motivated cultural worldview adherence and culturally loaded test performance. *Personality and Social Psychology Bulletin*, 35, 442-453.

Die Autoren untersuchten in ihrem Artikel, warum ein kognitiver Test, der ein sogenanntes „Cultural loading“ aufweist, die Testleistung einer Person beeinflussen kann. Ein Test ist dann kulturspezifisch gestaltet, wenn die Inhalte der Testitems kulturelles Wissen, Werte und Erfahrungen reflektieren, die im Grunde nicht ausschlaggebend für das Wissen oder die Fähigkeit sind, die mit diesem Test erfasst werden sollen. Aus früherer Literatur ist bekannt, dass Individuen eine bessere Testleistung erzielen, wenn der Testinhalt kulturell vertraute Erfahrungen widerspiegelt. Landau et al. (2009) testeten, welche Auswirkungen die Aktivierung des Bewusstseins um die eigene Sterblichkeit („mortality salience“) im Sinne einer aufrechterhaltenden Weltsicht auf die Leistung in einem kognitiven Test hat. In zwei Studien konnte bereits demonstriert werden, dass die Aktivierung der MS bei amerikanischen Versuchspersonen dazu führte, dass sie „outgroup“ Mitglieder (Afroamerikaner) bevorzugten, welche sich kulturellen stereotypen entsprechend verhielten, im Vergleich zu solchen, die diese Stereotype nicht bestätigten (Schimmel et al., 1999). Renkema, Stapel, Maringer und van Yperen (2008) fanden, dass die MS die Tendenz von Versuchspersonen erhöht, bei der Beschreibung von zweideutigen

Verhaltensweisen auf Stereotype zurückzugreifen. Ausgehend von diesen Erkenntnissen, stellten die Autoren die Hypothese auf, dass die Aktivierung der MS die Motivation von Versuchspersonen erhöht, die Testitems konform der vorherrschenden kulturellen Stereotype zu beantworten, um ihre bestehende Weltsicht nicht zu zerstören. Die Ergebnisse bestätigten diese Annahme und zeigten, dass Versuchspersonen, die zuvor durch die MS geprimt wurden, eine stärkere Leistung im kognitiven Test erzielten als die Kontrollgruppe (ohne MS Primes), wenn die richtigen Antworten die vorherrschenden kulturellen Stereotype bestätigten und eine schlechtere Leistung, wenn die richtigen Antworten den stereotypen Erwartungen der eigenen Kultur widersprachen.

11. Stapel, D. A. & Schwarz, N. (1998). Similarities and differences between the impact of traits and expectancies: What matters is whether the target stimulus is ambiguous or mixed. *Journal of Experimental Social Psychology*, 34, 227-245. (Retraction of vol 34, pg 227, 1998)

In ihrer Studie zu den Effekten von Erwartungen auf die Urteilsbildung, primten Stapel und Schwarz (1998) bei Versuchspersonen die Erwartung über eine Zielperson („Michael ist freundlich“) und gaben anschließend Personenbeschreibungen vor, die entweder zweideutiger Natur (das gleichzeitige Implizieren von Freundlichkeit und Unehrlichkeit) oder gemischter Natur (einige Verhaltensweisen wurden eindeutig als freundlich, andere als unehrlich dargestellt) waren. Die Autoren fanden, dass die Personenbewertung Hand in Hand mit der geprimten Erwartung ging, wenn die Beschreibung eindeutige aber gemischte Verhaltensweise enthielt. Das Primen der Erwartung „Michael ist freundlich“ führte also zu positiveren Urteilen über Michael, wenn er zuvor durch ein zweideutiges Verhalten (freundlich oder unehrlich) beschrieben wurde und zu negativeren Urteilen, wenn er eindeutig als „unehrlich“ beschrieben wurde.

Stapel und Schwarz (1998) wurden 13 Mal zitiert, darin enthalten sind 2 Selbstzitate und ein *Retraction Statement*. Folgende Artikel basieren auf den Erkenntnissen der Autoren:

- Lerouge, D. & Smeesters, D. (2008). Knowledge activation after information encoding: Implications of trait priming on person judgment. *Journal of Experimental Social Psychology*, 44, 429-436.

Bisher wurde empirisch bestätigt, dass das Urteil über eine Person von einer Kombination des Typs der erhaltenen Information über diese Person und den zugänglichen kognitiven Konzepten, zum Beispiel Eigenschaften, mit denen die Person verbunden wird, abhängig ist (Stapel & Schwarz, 1998). Mehrere Studien konnten bereits demonstrieren, dass die

Aktivierung von Eigenschaftskonzepten (durch Priming), vor der Enkodierung der präsentierten Information über eine Person, zu Assimilationseffekten im nachfolgenden Urteil führte. Demzufolge passt sich das Personenurteil an die gegebene Information an, indem die Information konsistent zu den geprimten Eigenschaftskonzepten enkodiert und interpretiert wird. In vorangehender Literatur fand man jedoch keine Hinweise darauf, dass die Effekte der aktivierten Eigenschaftskonzepte auch dann auftreten können, wenn das Priming der Eigenschaften erst nach der Enkodierung der Personeninformation erfolgt. Nach Stapel, Koomen und van der Pligt (1997) führt das „Post-Enkodierungs-Priming“ ziemlich unwahrscheinlich zu Assimilationseffekten in der Personenbewertung, da die Information während der Enkodierphase nicht als Bezugsrahmen herangezogen werden kann. Stapel und Schwarz (1998) primten Versuchspersonen vor der Konfrontation mit positiven oder negativen Informationen mit positiven Wörtern und konnten zeigen, dass die Versuchspersonen in der positiven Priming-Bedingung positivere Personenurteile abgaben und somit einen Assimilationseffekt zeigten. Die Autoren Lerouge et al. (2008) versuchten in ihrer Studie die Annahme von Stapel et al. (1997, 1998, 2001) zu widerlegen, dass geprimte Eigenschaftskonzepte ein Personenurteil nicht mehr beeinflussen können, sobald die Information enkodiert wurde. Sie stimmten in dem Punkt mit Stapel et al. überein, dass aktivierte Eigenschaften nicht als Interpretationsrahmen für die Personenbeurteilung dienen können, wenn die gegebene Information vage und mehrdeutig ist, jedoch wäre dies möglich, wenn es sich um eine Information mit gemischtem Inhalt (positiv und negativ) handeln würde. In diesem Sinne wurde den Versuchspersonen zuerst eine Information gegeben (z.B. „The librarian carries the old woman's groceries accros the street“) und erst danach wurden sie mit einem Wort geprimt (z.B. „kindness“), welches ihnen bei der Bewertung einer Zielperson als „retrieval cue“ diente, auf Basis dessen sie ihr Urteil bei einer gemischten Information (wenn die Information positive und negative Verhaltenseigenschaften einer Person enthält) an die danach aktivierte Eigenschaft anpassten und einen Assimilationseffekt in ihrer Bewertung zeigten.

→ Geers, A. L. & Lassiter, G. D. (2005). Affective assimilation and contrast: Effects of expectations and prior stimulus exposure. *Basic and Applied Social Psychology*, 27, 143-154.

Die Autoren Wilson und Klaaren (1992) beschrieben in ihrem „Affective Expectation Model“ (AEM), dass bei der Bildung unserer Affekte und Emotionen Erwartungen eine wichtige Rolle spielen. Sie definierten affektive Erwartungen als: „people's predictions about how they will feel in a particular situation or toward a specific stimulus“. Nach dem AEM können Erwartungen zu Assimilations- oder Kontrasteffekten in unserem affektiven

Erleben führen, je nachdem ob das Erlebnis mit unserer Erwartung konsistent vs. inkonsistent und uns der Grad dieser (Nicht-) Übereinstimmung bewusst oder unbewusst ist. Die Autoren Geers und Lassiter (2005) trafen in der vorliegenden Studie die Annahme, dass eine Erfahrung, die mit der affektiven Erwartung diskrepant ist, beim Bewusstsein um diese Diskrepanz zu einer affektiven Reaktion führt, die im Kontrast zur Erwartung steht. Wenn die Diskrepanz zwischen Erwartung und Erfahrung jedoch unentdeckt bleibt, wird die affektive Reaktion der Erwartung angepasst (=Assimilationseffekt). In zwei Experimenten untersuchten die Autoren die Rolle der affektiven Erwartung auf nachfolgende Reaktionen, indem sie den Versuchspersonen entweder einen erfahrungs-inkonsistenten Stimulus oder gar keinen Stimulus zuvor präsentierten. Die Ergebnisse zeigten, dass Versuchspersonen, die einem erfahrungs-inkonsistenten Stimulus ausgesetzt wurden, im Vergleich zur Kontrollgruppe ohne Stimulus Präsentation, ihre affektive Reaktion in Kontrast zu ihrer affektiven Erwartung setzten. Die theoretische Annahme der Autoren, dass ein vorangehender Stimulus den Einfluss der affektiven Reaktion moderieren kann, baut auf den Forschungsergebnissen der Studien über soziale Urteile von Stapel und seinen Kollegen auf (Stapel, Koomen & Van der Pligt, 1997; Stapel & Schwarz, 1998; Stapel & Winkielman, 1998). Diese fanden unter anderem, dass die Beurteilung neuer und mehrdeutiger Stimuli den eigenen Erwartungen und den Kontextinformationen angepasst werden (Assimilation), während ein bereits bekannter oder eindeutig interpretierbarer Stimulus als Vergleichsstandard zu den eigenen Erwartungen herangezogen wird, so dass oftmals ein Kontrasteffekt resultiert.

→ DeCoster, J. & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational biases. *Personality and Social Psychology Review*, 8, 2-27.

In ihrem Review Artikel geben die Autoren DeCoster et al. einen Überblick über die Primingforschung bis zum Jahre 2004 und tragen Informationen zur Frage zusammen, welchen Einfluss aktivierte (geprimte) Eigenschaftskonzepte auf soziale Urteile von Personen haben können. Die untersuchte Primingliteratur geht dabei von drei unterschiedlichen Effekten aus: Zum einen können geprimte Stimuli in die Bewertung des Zielobjekts integriert werden (Assimilation), zum anderen können sie als Vergleichsanker herangezogen werden (Kontrast) und manchmal können sie dazu führen, dass Individuen bewusst ihr Urteil ändern (Korrektur). In der Meta-Analyse dieser drei Effekte zitieren die Autoren die Forschungsergebnisse ganzer acht Artikel von Stapel und seinen Kollegen, wovon zwei aufgrund betrügerischer Inhalte oder Forschungsmethoden von der jeweiligen Fachzeitschrift zurückgewiesen und weitere zwei von der Stapel Untersuchungskommission als „fraudulent“ entlarvt wurden (Stapel & Koomen, 1996#;

Stapel & Koomen, 1997#; Stapel & Koomen, 1998; Stapel & Koomen, 2001; Stapel, Koomen & van der Pligt, 1996#*; Stapel, Koomen & van der Pligt, 1997; Stapel, Koomen & Zeelenberg, 1998; Stapel & Schwarz, 1998#*).

→ Bohner, G., Ruder, M. & Erb, H. P. (2002). When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology*, 41, 495-519.

In ihrem Artikel über Assimilations- und Kontrasteffekte in der Überredungskunst untersuchten die Autoren die weit verbreitete Tatsache, dass sich Personen mehr von Experten überzeugen lassen als von Nicht-Experten, da sie die Informationsquelle eines Experten als schnell verfügbare und anwendbare Heuristik ansehen, mithilfe derer Entscheidungen mit minimalem kognitiven Aufwand getroffen werden können. Nach dem „Heuristics-Systemic-Model“ (HSM; Bohner, Moskowitz & Chaiken, 1995; Chaiken, Liberman & Eagly, 1989; Chen & Chaiken, 1999) dienen Heuristiken als alleinige Entscheidungsgrundlage wenn die Informationsverarbeitungsmotivation gering ist, die kognitive Beanspruchung hoch wäre und eine detailliertere Information nicht verfügbar ist. Unter diesen Umständen übernehmen Individuen die Information eines Experten, ohne dessen Inhalt tiefer zu verarbeiten. Wenn ein höheres Motivationsausmaß und ausreichend kognitive Kapazität vorhanden sind, wenden Individuen ebenfalls Heuristiken an, allerdings verarbeiten sie dann Argumente und weitere relevante Informationen auf eine sorgfältigere und systematischere Art und Weise. Die Heuristik (der Informationsquelle) und die systematische Informationsverarbeitung tragen in diesem Falle in gleicher Weise zur Urteilsbildung bei. Im vorliegenden Artikel prüften die Autoren neue Hypothesen im Rahmen des HSMs zum Zusammenspiel aus niedrigem und hohem Arbeitsaufwand in der kognitiven Verarbeitung. In zwei Experimenten fanden sie, dass Studenten negativere Gedanken und Einstellungen berichteten, nachdem ihnen von einer hohen (vs. niedrigen) Expertenquelle schwache Argumente präsentiert wurden und positivere Gedanken und Einstellungen nach der Präsentation von starken Argumenten einer niedrigen (vs. hohen) Expertenquelle. Die Gedanken und Einstellungen von Studenten waren außerdem dann positiver, wenn eine hohe (vs. niedrige) Expertenquelle mehrdeutige Argumente präsentierte. Im Zuge ihrer Experimente gingen die Autoren auch auf die Integration von Assimilations- und Kontrasteffekten in das HSM ein, da die Begriffe der Verfügbarkeit und Anwendbarkeit von Heuristiken und der „construct accessibility“ in der sozialen Kognitionsforschung thematisch nah beieinander liegen und miteinander verglichen werden können. Bohner et al. (2002) zitieren dabei wesentliche Ergebnisse von Stapel und Kollegen (Stapel, Koomen & van der Pligt, 1996; Stapel, Koomen & van der Pligt, 1997; Stapel & Schwarz, 1998), die bereits fanden, dass der Typ der

verfügbaren Information (Eigenschaften, Situationen oder Erwartungen) und das zu beurteilende Zielobjekt (eindeutiges vs. mehrdeutiges Verhalten) dafür ausschlaggebend sind, ob ein Urteil dem aktivierten Konstrukt angepasst (Assimilation) oder entgegengesetzt (Kontrast) wird. Während bestimmte Konstrukte (z.B. Eigenschaften) zu Assimilationseffekten führen, lösen andere Konstrukte (z.B. Erwartungen) dann Assimilation aus, wenn es sich um eine mehrdeutige Information handelt und Kontrast, wenn eine gemischte Zielinformation (wenn die Information mit der Erwartung teilweise inkonsistent ist) präsentiert wird. Wenn die vorgegebene Zielinformation außerdem eindeutig interpretierbar und von extremer Qualität ist, dann wird sie als Vergleichsstandard herangezogen, anhand dessen ein Kontrasteffekt im nachfolgenden Urteil entsteht.

12. Marx, D. M. & Stapel, D. A. (2006). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology*, 36, 687-698. (Retracted article. See vol. 42, pg. 933, 2012)

In der „Stereotype threat“ Literatur wird meistens davon ausgegangen, dass die wahrgenommene Bedrohung durch eine erhöhte Ängstlichkeit entsteht, ein negatives stereotypes Klischee über die eigene Gruppe zu bestätigen (Steele, 1997). Die Variation von verschiedenen Typen emotionaler Reaktionen, die vor und nach einer Testdurchführung bei Individuen infolge einer stereotypen Bedrohung entstehen kann, wurde jedoch noch nie systematisch erhoben. Marx und Stapel (2006) setzten sich diese Aufgabe zum Ziel ihrer Studie und untersuchten die emotionale Spezifität zu unterschiedlichen Zeitpunkten infolge einer stereotypen Bedrohungssituation. Die Ergebnisse zeigten, dass die Zielpersonen eines negativen Stereotyps (im Vergleich zu den nicht durch den Stereotyp betroffenen Personen) eine höhere Ängstlichkeit vor der Testdurchführung und eine erhöhte Frustration nach Beendigung des Tests angaben. In der Kontrollbedingung (ohne stereotype Bedrohung) wurden keine Effekte gefunden.

Marx und Stapels Artikel (2006) wurde 12 Mal zitiert, kein einziges Mal von Stapel selbst und einmal von der zuständigen Fachzeitschrift, die diesen Artikel aufgrund betrügerischer Inhalte zurückwies. Folgende Artikel bauen auf den Erkenntnissen der beiden Autoren auf:

- Brodisha, A. B. & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology*, 45, 180-185.

Die Autoren untersuchten in ihrer Studie, dass eine stereotype Bedrohung die Leistung in einem Mathematik-Test durch niedrig gesetzte Zielerwartungen und die Sorge über die eigene Leistung verringern kann. Dazu sollten weibliche Versuchsteilnehmer in einer stereotypen Bedrohungs-Bedingung ihre testbezogenen Leistungsziele vor und nach der Durchführung eines schweren mathematischen Tests angeben. Zusätzlich schätzten die Versuchsteilnehmerinnen das Ausmaß ihrer momentanen Test-Angst („State-Test-Anxiety“) und ihrer Sorgen unmittelbar vor und nach der Testdurchführung auf einer Rating-Skala ein. Die Manipulation des Timings der Erhebung der emotionalen Verfassung vor und nach der Durchführung eines Tests in einer stereotypen Bedrohungssituation, haben außer den Autoren Brodish et al. (2009) bisher nur Marx und Stapel (2006) erforscht. In deren Studie berichteten die Versuchsteilnehmer von einer unterschiedlichen Affektqualität in Abhängigkeit davon, wann ihre emotionale Verfassung erhoben wurde. Sie berichteten eine höhere Ängstlichkeit vor dem Test und eine größere Frustration nach dem Test. Jedoch fanden Marx und Stapel (2006) keine Leistungsunterschiede in Abhängigkeit von dem Timing der Emotionserhebung. In der vorliegenden Studie von Brodish et al. (2009) erreichten die Versuchsteilnehmerinnen, die vor dem Test ihr Angstaussmaß einschätzten, eine signifikant schlechtere Leistung als die Teilnehmerinnen, die ihre Ängstlichkeit erst nach dem Test angaben. Brodish et al. (2009) glauben, dass die Frage vor dem Mathematik-Test, ob die Versuchsteilnehmerinnen nervös oder besorgt seien, genau diese Emotionen in ihnen ausgelöst haben und sie deswegen eine schlechtere Leistung erzielten als die Frauen, die erst nach Beendigung des Tests zu ihrer Gefühlslage befragt wurden. Die Autoren erweiterten somit Marx und Stapels Erkenntnis, dass bei der Erhebung von psychologischen Variablen auf jeden Fall immer das Timing bewusst gewählt werden sollte.

→ Kit, K. A., Tuokko, H. A. & Mateer, C. A. (2008). A review of the stereotype threat literature and its application in a neurological population. *Neuropsychology Review*, 18, 132-148.

In ihrem Review-Artikel gehen die Autoren auf die wissenschaftlichen Grundlagen des Phänomens der stereotypen Bedrohung ein und zitieren dabei zwei Artikel von Marx und Stapel (2006a; 2006b), die dazu einen wesentlichen Beitrag geleistet haben und beide aufgrund betrügerischer Forschung von der jeweiligen Fachzeitschrift zurück gewiesen wurden.

13. Ruys, K. I. & Stapel, D. A. (2008). How to heat up from the cold: Examining the preconditions for (unconscious) mood effects. *Journal of Personality and Social Psychology*, 94, 777-791. (Retracted article. See vol. 103, pg. 451, 2012)

Ruys und Stapel untersuchten in ihrem Artikel die grundlegenden Techniken der Stimmungsmanipulation und fanden, dass das wichtigste Element dabei die evaluative Komponente („evaluative tone“, siehe Stapel & Koomen, 2005, für eine nähere Definition) und nicht so sehr die deskriptive Bedeutung der geprüften Information ist.

Der Artikel von Stapel und Ruys (2008) wurde 12 Mal zitiert, davon 3 Mal von Stapel selbst und einmal von der Fachzeitschrift als *Retraction Statement*. Folgender Artikel baut auf den Ergebnissen der beiden Autoren auf:

→ Ziegler, R. & Diel, M. (2011). Mood and multiple source characteristics: Mood congruency of source consensus status and source trustworthiness as determinants of message scrutiny. *Personality and Social Psychology Bulletin*, 37, 1016-1030.

In einem 4-faktoriellen Design untersuchten die Autoren, welchen Einfluss die Stimmung (positiv vs. negativ), die Übereinstimmungsquelle der Information (Mehrheit vs. Minderheit), die Vertrauenswürdigkeit der Information (hoch vs. niedrig) und die Informationsstärke (stark vs. schwach) auf die Prozesse der Überzeugung und auf die Einstellungen einer Person haben können. Ziegler und Diel (2011) testen in ihren Experimenten zwei verschiedene Informationsverarbeitungsansätze. Nach der „Mood as information“ Hypothese dient eine negative Stimmung als Warnsignal, dass es sich um eine problematische Situation handeln könnte und dass es folglich wichtig ist, verfügbare Informationen gründlich zu verarbeiten. Eine positive Stimmung signalisiert dagegen eine gutartige Situation, in der es nicht notwendig erscheint, viel Energie in die Informationsverarbeitung zu stecken. Bei vorhandener positiver Stimmung wird daher viel eher auf die Verwendung von Heuristiken zurückgegriffen, zum Beispiel auf die Meinung einer Mehrheit, ohne dass deren Einstellungen kritisch hinterfragt werden, vor allem wenn diese Mehrheit als vertrauenswürdig eingeschätzt wird. Nach der „Mood as information“ Hypothese wenden Individuen die gleichen Verarbeitungsstrategien an, wie jene, die von Personen mit einer niedrigen „Need for Cognition“ (NC) Ausprägung genutzt werden. „Need for Cognition“ beschreibt die dispositionelle Ausprägung bei Individuen, gerne über Problematiken nachzudenken und die Bereitschaft, ein bestimmtes Ausmaß an kognitiven Ressourcen in deren Lösung zu investieren. Eine positive Stimmung und eine niedrige Ausprägung in NC führen also zu heuristischen Informationsverarbeitungsprozessen, wohingegen eine negative Stimmung und eine hohe Ausprägung in NC eine

gewissenhafte Informationsverarbeitung bedingen. Ruys und Stapel (2008) fanden in ihrer Studie außerdem, dass eine induzierte Stimmung das NC Ausmaß von Individuen beeinflussen kann (eine positive Stimmung führte zu niedrigem NC, eine negative Stimmung zu hohem NC). Ausgehend von dieser Erkenntnis, stellten Ziegler et al. (2011) die Hypothese auf, dass Individuen in einer positiven Stimmung mehr Aufwand für die Verarbeitung einer überzeugenden Information betreiben, wenn es sich bei der Informationsquelle um eine inkongruente Kombination handelt (vertrauenswürdige Minderheit vs. nicht-vertrauenswürdige Mehrheit). Personen in einer negativen Stimmung, deren Informationsverarbeitungsstrategie nach Ruys und Stapel (2008) denen von Individuen mit hohem NC gleichen, sollten die Information in jedem Falle gewissenhaft verarbeiten, unabhängig von den Charakteristiken der Informationsquellen. Diesem Ansatz stellten die Autoren die Theorie der „Mood-congruent expectancies“ entgegen, wonach Personen, sowohl in positiver als auch negativer Stimmung, Informationen dann gründlich und gewissenhaft verarbeiten, wenn diese nicht ihren stimmungsbasierten Erwartungen entsprechen.

14. Johnson, C. S. & Stapel, D. A. (2007). When different is better: Performance following upward comparison. *European Journal of Social Psychology*, 37, 258-275. (Retracted article. See vol. 42, pg. 799, 2012)

Johnson und Stapel erforschten 2007 die Effekte eines aufsteigenden sozialen Vergleichs auf die Leistung in einem anschließenden Test. Sie fanden, dass der Vergleich mit einer Zielperson, welche beispielsweise in einem analytischen Fähigkeitsbereich erfolgreich war, zu einer Leistungssteigerung in einem Test zur Erfassung einer anderen Fähigkeitsdimension (beispielsweise der verbalen Fähigkeiten) führte. Die Verbesserung der eigenen Leistung in einem komplementären Bereich, führten die Autoren auf die Möglichkeit einer Wiederherstellung des Selbstwerts zurück, welche ihnen durch die Vorgabe eines Tests in einer von der vergleichenden Zielperson verschiedenen Domäne geboten wurde.

Johnson und Stapel wurden 9 Mal zitiert, davon entstammen fünf Selbstzitate und ein *Retraction Statement* der Fachzeitschrift. Folgender Artikel baut auf den Erkenntnissen der beiden auf:

→ Van de Ven, N., Zeelenberg, M. & Pieters, R. (2011). Why envy outperforms admiration. *Personality and Social Psychology Bulletin*, 37, 784-795.

Die Autoren verwendeten für die experimentelle Manipulation des böartigen/gutartigen Neids und der Bewunderung eine nach Johnson und Stapel (2011) angepasste Geschichte über einen Studenten namens „Hans de Groot“ (siehe Artikel: „No pain, no gain: The conditions under which upward comparisons lead to better performance“).

15. Joly, J. F., Stapel, D. A. & Lindenberg, S. M. (2008). Silence and table manners: When environments activate norms. *Personality and Social Psychology Bulletin*, 34, 1047-1056. (Retracted article. See vol. 38, pg. 1378, 2012)

Joly et al. testeten in zwei Studien, unter welchen Bedingungen die Umgebung (z.B. eine Bibliothek oder ein Restaurant) Versuchspersonen dazu bringen kann, sich den umgebungsspezifischen sozialen Normen (z.B. leise zu sein oder Tischmanieren zu befolgen) entsprechend zu verhalten. Deren Artikel wurde 9 Mal zitiert, 2 Mal von Stapel selbst und einmal durch die Zurückweisung der zuständigen Fachzeitschrift. Folgender Artikel baut auf den Ergebnissen der Autoren auf:

→ Neal, D. T., Wood, W., Labrecque, J. S. & Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental and Social Psychology*, 48, 492-498.

Neal et al. gehen in ihrer Studie der Frage nach, welche psychologischen Mechanismen hinter unseren täglichen Gewohnheiten stecken. In zwei Experimenten fanden sie Hinweise darauf, dass unsere Gewohnheiten sehr stark durch Kontexthinweise, die mit vergangenen Aktivitäten verbunden sind (z.B. bestimmte Orte und Lokalitäten), beeinflusst werden. In ihrer zweiten Studie untersuchten die Autoren daher, ob Personen, die zuvor mit Bildern eines Fußballstadions geprimt wurden (im Vergleich zur Kontrollgruppe, die mit Küchenbildern geprimt wurde), danach die Lautstärke ihrer Stimme erheben. Da bereits Joly, Stapel und Lindenberg (2008) und auch Aarts und Dijksterhuis (2000; 2003) herausfanden, dass unsere Gewohnheiten auch zielabhängig sind und sich Personen, die zuvor mit Bildern einer Bibliothek geprimt wurden, nur dann leise verhielten, wenn sie vor hatten, eine Bibliothek in geraumer Zeit zu besuchen, wurde den Versuchspersonen in Neal et al.s Experiment (2012) gesagt, dass sie für den finalen Teil der Studie ein Sportstadion aufsuchen werden. Die Ergebnisse zeigten, dass Personen, denen versprochen wurde, anschließend ein Stadion zu besuchen, tatsächlich die Lautstärke ihrer Stimme erhoben.

16. Noordewier, M. K. & Stapel, D. A. (2010). Affects of the unexpected: When inconsistency feels good (or bad). *Personality and Social Psychology Bulletin*, 36, 642-654. (Retracted article. See vol. 38, pg. 1378, 2012)

Noordewier und Stapel untersuchten in ihrem Artikel welche affektiven Reaktionen eine unbestätigte Erwartung bei Individuen auslösen kann und fanden, dass diese davon abhängig sind, wie Personen grundsätzlich den Wert der Konsistenz bewerten und sich infolge eher auf die Erwartungskongruenz oder den Inhalt einer Information (positiv vs. negativ) konzentrieren.

Noordewier und Stapels Artikel wurde 8 Mal zitiert, darin enthalten sind fünf Selbstzitate und das *Retraction Statement* der Fachzeitschrift. Keine der Zitierungen baut jedoch explizit auf den Erkenntnissen der Autoren auf.

17. Stapel, D. A. & Koomen, W. (2001). sil all means: Interpretation goals facilitate accessibility and stereotyping effects. *Personality and Social Psychology Bulletin*, 27, 915-929. (Retracted article. See vol. 39, pg. 264, 2013)

Stapel und Koomen (2001) wurden 8 Mal zitiert, darin enthalten sind eine Zurückweisung der Fachzeitschrift und vier Selbstzitate. Keiner dieser Artikel baut konkret auf Stapels Ergebnissen auf.

18. Maringer, M. & Stapel, D. A. (2007). Unfinished business: How completeness affects the impact of emotional states and emotion concepts on social judgment. *Journal of Experimental Social Psychology*, 43, 712-718. (Retracted article. See vol. 49, pg. 780, 2013)

Maringer und Stapel untersuchten 2007 welchen Einfluss die volle Darbietung einer Prime-Episode im Vergleich zu einer kürzeren abgebrochenen Prime-Episode auf kongruente bzw. inkongruente Urteile bezüglich einer zu bewertenden Information hat. Die Autoren fanden dabei, dass der Effekt der Vollständigkeit der Prime-Episode nicht davon abhängt, ob die emotionale Stimmung (durch die Beschreibung eines glücklichen/unglücklichen Ereignisses in seinem Leben und das gleichzeitige Hören fröhlicher/trauriger Musik) oder Emotionskonzepte (durch die grammatikalisch richtige Anordnung von durchgemischten Wörtern, die zusammengesetzt traurige/fröhliche Sätze ergeben) geprimt werden. Sowohl eine positive als auch negative geprimte Information führte bei einer unvollständigen Prime-Episode (vorzeitiges Abschalten der Musik oder

Beenden der Satzaufgabe nach 20 Sätzen) zu Assimilationseffekten, während die durch eine vollständige Prime-Episode aktivierte Information zu Kontrasteffekten führte.

Maringer und Stapel wurden 7 Mal zitiert, einmal bezüglich der Zurückweisung der Fachzeitschrift und 2 Mal von Stapel selbst. Keiner dieser Artikel baut jedoch direkt auf den Erkenntnissen zur Vollständigkeit einer Prime-Episode auf.

19. Stapel, D. A. & Johnson, C. S. (2007). When nothing compares to me: How defensive motivations and similarity shape social comparison effects. *European Journal of Social Psychology*, 37, 824-838. (Retracted article. See vol. 42, pg. 664, 2012)

Ein wesentliches menschliches Motiv ist es, ein positives Selbstbild zu bewahren. Um dies zu bewerkstelligen, suchen wir nach selbstrelevanten Informationen. Eine Quelle selbstrelevanter Information ist der soziale Vergleich. Wenn wir mit einem Zielobjekt konfrontiert werden, mit dem ein Vergleich für unsere Selbstevaluation schädigend wäre, wenden wir zur Interpretation der Information eine verteidigende Strategie des sozialen Vergleichs („defensive social comparison“) an. Defensive soziale Vergleiche sind durch einen asymmetrischen Effekt auf die Selbstbewertung gekennzeichnet: Aufsteigende Vergleiche mit uns überlegenen Personen schädigen die Selbstbewertung nicht unbedingt, während absteigende Vergleiche mit uns unterlegenen Personen zu einem Anstieg der positiven Selbstbewertung führen. Der Vergleich mit einem unterlegenen Zielobjekt führt zu einem Kontrasteffekt, weil er der Selbstbewertung gut tut, jedoch führt ein Vergleich mit einem überlegenen Zielobjekt nicht zum Kontrast, da ansonsten die Selbstbewertung geschädigt werden würde. Die Information in einer sozialen Vergleichssituation wird also nur dann in das Selbst integriert, wenn die Information für die Selbstbewertung förderlich ist. Allerdings fanden Stapel und Kollegen auch, dass die defensive Reaktion auf eine soziale Vergleichsinformation nicht auftritt, wenn die Information eindeutig ähnlich und für das Selbst hoch relevant ist (z.B. der Vergleich meines Studienfortschritts mit einem anderen Psychologiestudenten). In diesem Fall würde ein aufsteigender Vergleich zu einem Kontrasteffekt und ein absteigender Vergleich zur Assimilation führen.

Stapel und Johnsons Artikel über verteidigende Strategien des sozialen Vergleichs wurde 7 Mal zitiert, mit eingeschlossen zweier Selbstzitate und einem *Retraction Statement*. Folgende zwei Artikel bauen auf den Erkenntnissen der beiden Autoren auf:

→ McIntyre, K. P. & Eisenstadt, D. (2011). Social comparison as a self-regulatory measuring stick. *Self and Identity*, 10, 137-151.

McIntyre und Eisenstadt untersuchten in ihrer zweiten Studie, auf welche Weise der Zusammenhang zwischen chronischen auf- und absteigenden Vergleichen und bestehenden Selbstdiskrepanzen zu Assimilations- oder Kontrasteffekten führt. Ihre Ergebnisse stimmten mit Stapel et al.s Ansatz überein.

→ Lalwani, A. K. & Shavitt, S. (2009). The „Me“ I Claim to Be: Cultural Self-Construal Elicits Self-Presentational Goal Pursuit. *Journal of Personality and Social Psychology*, 97, 88-102.

In ihrem Artikel gehen die Autoren in 12 Experimenten der Annahme nach, dass das aktivierte kulturelle Selbstkonstrukt beeinflusst, nach welchen Selbstpräsentationszielen wir effektiv streben. Ist es uns auf einer Dinnerparty, bei der wir Gastgeber sind, beispielsweise wichtiger, als selbstsicher, kompetent und geschickt eingeschätzt zu werden, oder würden wir uns lieber als sozial-einfühlsam und der Gruppe angepasst präsentieren? Lalwani und Shavitt gehen aus den Erkenntnissen früherer Literatur von zwei Arten von Selbstpräsentationszielen aus: Der Selbsttäuschenden Aufwertung („self-deceptive-enhancement“=SDE), bei der Individuen eine aufgeblasene Bewertung ihrer eigenen Fähigkeiten darstellen, und das „Impression Management“ (=IM), bei dem Individuen ihre Schwächen und Unzulänglichkeiten hinunter spielen, um sich der Norm anzupassen.

Darüber hinaus verfügen Individuen zur selben Zeit über mehrere Selbstkonzepte. Im unabhängigen Selbstkonzept („independent self-construal“) sieht man sich selbst als autonom, selbstständig und verschieden von anderen an, während man sich beim abhängigen Selbstkonzept („interdependent self-construal“) als Teil einer sich selbst einschließenden Gemeinschaft ansieht, welche das eigene Verhalten durch ihre Gedanken, Gefühle und Handlungen beeinflusst. Personen mit abhängigem Selbstkonzept tendieren eher dazu, mit sozialen Normen und den Entscheidungen anderer konform zu gehen.

In ihrer zweiten Studie untersuchten die Autoren, ob Versuchspersonen, bei denen zuvor durch Priming das interdependente Selbstkonstrukt aktiviert wurde, in einem fiktiven „Geschenks-Szenario“ einer anderen Person gegenüber sozial einfühlsamer sind und ihr Verhalten von Beobachtern auch genauso eingeschätzt wird. Zur Überprüfung ihrer Hypothese wandten die Autoren die Methode der „Impression Formation“ nach Stapel und Johnson (2007) an, bei der sich die Versuchspersonen vorstellen sollten, dass sie in einem kleinen Betrieb arbeiten und einer ihrer Kollegen und guter Freund, namens Chris,

den Betrieb verlassen. Sie wurden anschließend gefragt, welches Geschenk sie Chris zum Abschied schenken würden, wie viel Zeit und welchen Weg sie für die Besorgung auf sich nehmen würden und wie viel Geld sie bereit wären, auszugeben. Die Ergebnisse bestätigten die Annahme, dass das zuvor aktivierte interdependente Selbstkonzept die Versuchsteilnehmer in einem sozialen Szenario dazu führte, sich einfühlsamer zu verhalten und mehr Geld, Zeit und Wegdistanz für den Einkauf des Abschiedsgeschenks zu investieren.

Die Zusammenhänge zwischen der Aktivierung eines Selbstkonzepts und den Selbstpräsentationszielen verschwanden jedoch, wenn sich die Versuchspersonen nicht auf ihre, sondern auf die Selbstpräsentationsziele anderer Personen konzentrierten (Studie 3) oder wenn ihnen zuvor die Möglichkeit geboten wurde, ihre Ziele durch eine Selbstaffirmation zu befriedigen (Studien 4-9). Die Selbstaffirmationsaufgabe wurde nach Johnson und Stapel (2007) gestaltet und lautete, dass die Versuchspersonen vier Situationen aufschreiben sollen, in denen sie die Erwartungen ihrer Familie oder Freunde erfüllt oder übertroffen haben. Die Kontrollgruppe sollte dagegen vier Eigenschaften eines Baums nennen. Diese Methodik wurde in allen weiteren Experimenten mit einer Selbstaffirmations-Manipulation verwendet und beruht auf Johnson und Stapels betrügerischem Artikel „No pain, no gain: The conditions under which upward comparisons lead to better performance“, der von der Zeitschrift *Journal of Personality and Social Psychology* zurückgewiesen wurde.

20. Stapel, D. A. & Marx, D. M. (2007). Distinctiveness is key: How different types of self-other similarity moderate social comparison effects. *Personality and Social Psychology Bulletin*, 33, 439-448. (Retracted article. See vol. 39, pg. 264, 2013)

Stapel und Marx (2007) erforschen in ihrem Artikel den Einfluss der Ähnlichkeit zwischen einem selbst und anderen („self-other similarity“) auf soziale Vergleichseffekte. Sie gingen davon aus, dass die Ausgeprägtheit der Ähnlichkeit ein entscheidender Faktor in dieser Beziehung ist. In zwei Experimenten demonstrieren sie, dass eine ausgeprägte „self-other“ Ähnlichkeit zu Assimilationseffekten führt, während eine unklare Ähnlichkeit in einem Kontrasteffekt resultiert. Die Ergebnisse lassen vermuten, dass es also mehr auf den Ähnlichkeitstyp als auf das Ähnlichkeitsausmaß ankommt. Die Ähnlichkeit auf einer markanten Dimension führt eher zur Assimilation als die Ähnlichkeit auf mehreren unklaren Dimensionen. Assimilationseffekte treten vor allem dann auf, wenn die vergleichende Dimension für einen persönlich unwichtig ist und Kontrast, wenn einem die Dimension persönlich wichtig ist. Stapels Erkenntnisse replizierten und erweiterten Tessers „Self-Evaluation Maintenance Model“.

Stapel und Marx wurden 7 Mal zitiert, einmal von Stapel selbst und einmal bezüglich der Zurückweisung der betreffenden Fachzeitschrift. Folgender Artikel baut auf deren Erkenntnissen auf:

→ Gebauer, J. E., Broemer, P., Haddock, G. & von Hecker, U. (2008). Inclusion-exclusion of positive and negative past selves: Mood congruence as information. *Journal of Personality and Social Psychology*, 95, 470-487.

In ihrem Artikel untersuchten Gebauer et al. die weit verbreitete Annahme, dass das gedankliche Zurückholen eines vergangenen positiven Selbst den Selbstwert erhöht, während die Erinnerung an ein negatives Selbst den Selbstwert verringert. In vier Experimenten konnten die Autoren zeigen, dass bei chronisch fröhlichen Versuchsteilnehmern bei der Erinnerung sowohl an ein positives als auch an ein negatives Selbst, eine Selbstwertsteigerung auftrat, wohingegen chronisch traurige Versuchsteilnehmer eine relative Abnahme des Selbstwerts zeigten, wenn sie zuvor ihr positives oder negatives vergangenes Selbst zurückriefen. Diese Effekte lassen sich auf unterschiedliche Wahrnehmungen bezüglich der Stimmungskongruenz zwischen dem vergangenen und dem gegenwärtigen Selbst erklären. Fröhliche Versuchsteilnehmer nahmen eine hohe Stimmungskongruenz zwischen ihrem positiven vergangen Selbst und ihrem gegenwärtigen Selbst und eine geringe Stimmungskongruenz zwischen ihrem negativen vergangenen Selbst und ihrem gegenwärtigen Selbst wahr. Für chronisch traurige Versuchsteilnehmer traf der gegenteilige Effekt zu. Unabhängig von der chronischen Stimmung, führte die Wahrnehmung einer Stimmungskongruenz zu einem Gefühl der zeitlichen Aktualität, während eine Stimmungsinkongruenz zu einem Gefühl der zeitlichen Distanz führte. Dem „inclusion-exclusion model“ sozialer Urteile folgend, führte eine wahrgenommene zeitliche Aktualität zu einer Assimilation des Selbstwerts und eine zeitliche Distanz zu einem Kontrasteffekt. Die vorangehende theoretische Annahme, dass die wahrgenommene Stimmungskongruenz den entscheidenden Faktor („feature overlap“) für die Auswirkung auf das Gefühl der zeitlichen Distanz darstellt, stammt von Stapel und Marx (2007), die bereits darauf aufmerksam machten, dass einige Faktoren beim Auftreten von Assimilations- und Kontrasteffekten eine wichtigere Rolle spielen als andere.

21. Avramova, Y. R., Stapel, D. A. & Lerouge, D. (2010). Mood and context-dependence: Positive mood increases and negative mood decreases the effects of context on perception. *Journal of Personality and Social Psychology*, 99, 203-214. (Retracted article. See vol. 103, pg. 488, 2012)

Avramova et al. (2010) konnten in ihren Studien demonstrieren, dass eine induzierte negative Stimmung zu einer Aufmerksamkeitslenkung zu einem Zielobjekt führt, während eine positive Stimmung sowohl das Zielobjekt als auch die Kontextinformationen für eine Bewertung einschließt. Die Autoren zeigten, dass Personenurteile über die Temperatur, das Gewicht oder die Größe dann kontextbezogener ausfielen, wenn die Versuchspersonen zuvor in eine positive Stimmung versetzt wurden. Diese Effekte konnten auch auf soziale Urteile ausgeweitet werden, so dass sich glückliche Versuchspersonen in ihren Bewertungen eher von Kontextvariablen beeinflussen ließen als unglückliche Versuchspersonen.

Avramova et al. wurden 7 Mal zitiert, davon stammt eine Zitierung von der Zurückweisung des *Journal of Personality and Social Psychology*. Folgender Artikel baut auf den Ergebnissen der Autoren auf:

→ Lynn, S. K., Zhang, X. & Barrett, L. F. (2012). Affective state influences perception by affecting decision parameters underlying bias and sensitivity. *Emotion*, 12, 726-736.

Lynn et al. (2012) beschäftigen sich in ihrer Studie mit den direkten Effekten, die ein (geprimter) affektiver Gemütszustand bei einer Person auf ihre Wahrnehmung haben kann. Avramova et al. (2010) konnten in ihrer beispielsweise Studie demonstrieren, dass der Affekt einer Person die Sensitivität für die Schätzung der Temperatur und des Gewichts beeinflussen kann. Lynn et al. (2012) wollten in ihrer Studie zeigen, dass der affektive Gemütszustand als internaler Kontextrahmen dient, der die Wahrnehmung einer sozialen Bedrohung durch drei verschiedene externe Umweltmanipulationen beeinflussen kann. Dafür kreierten sie eine bedrohliche Wahrnehmungsaufgabe („threat perception task“), bei der den Versuchspersonen Bilder von Personen mit bösem, finsterem Gesichtsausdruck gezeigt wurden und die in ihrer Intensität von mehr bis weniger bedrohlich variierten. Pro richtiger Entscheidung der Versuchspersonen, ob es sich um ein mehr oder weniger bedrohliches Gesicht handelt, bekamen sie einen Punkt bzw. bei falscher Entscheidung einen Punkt abgezogen. Die Valenz des Gemütszustands („feeling of pleasantness“ vs. „unpleasantness“) und das Erregungsausmaß („feelings of low activation“ vs. „high activation“) der Versuchspersonen wurden zuvor durch die Vorgabe von Bildern und Musik geprimt. Avramova et al. (2010) konnten bereits

demonstrieren, dass Personen, die sich in einer unangenehmen Stimmungslage befinden, nicht sehr feinfühlig in ihrer Bewertung von Stimuli in einem diskriminativen Wahrnehmungskontext sind. Die experimentelle Manipulation des Gemütszustands (Valenz und Erregungsausmaß) sollte nach Lynn et al. (2012) also die Leistung der Versuchspersonen in der „Threat perception task“ in drei unterschiedlichen Wahrnehmungsbedingungen beeinflussen.

22. Stapel, D. A. & Marx, D. M. (2006). Hardly thinking about close and distant others: On cognitive busyness and target similarity in social comparison effects. *Journal of Experimental Social Psychology*, 42, 397-405. (Retracted article. See vol. 49, pg. 781, 2013)

In zwei Studien untersuchten Marx und Stapel die Auswirkungen der „other-self“ Ähnlichkeit auf soziale Vergleichseffekte, wenn die urteilende Person währenddessen kognitiv beschäftigt vs. nicht beschäftigt ist.

Ihr Artikel wurde 7 Mal zitiert, drei Zitate stammen von Stapel selbst und eines von der betreffenden Fachzeitschrift via *Retraction Statement*. Folgender Artikel beruht auf den Erkenntnissen der Autoren:

→ Graffin, S. D., Boivie, S. & Carpenter, M. A. (2013). Examining CEO succession and the role of heuristics in early-stage CEO evaluation. *Strategic Management Journal*, 34, 383-403.

In Graffin et al.s Studie geht es um die Erhebung von Faktoren, die bei der frühen Evaluation eines neuen CEOs (Chief Executive Officer) eine Rolle spielen (siehe auch zitierende Artikel von „The referents of trait inferences: The impact of trait concepts versus actor-trait links on subsequent judgments“). Nach Stapel und Marx (2006) kommt es eher dann zu Assimilationseffekten in der Personenbewertung, wenn sich der Beurteiler und die zu beurteilende Person ähnlich sind und der Beurteiler kognitiv beschäftigt ist. Da viele der neuen CEO beurteilenden Personen selbst einmal Walter dieses Amtes waren, sollte diese wahrgenommene Ähnlichkeit zur Zielperson zu einem Assimilationseffekt in der Bewertung führen und darüber hinaus auch die Tatsache, dass es sich bei den beurteilenden Personen um Vorsitzende des Unternehmens handelt, die mit ihrem Beruf voll ausgelastet und kognitiv beschäftigt sind.

23. Stapel, D. A. & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, 332, 251-253. (Retracted article. See vol. 334, pg. 1202, 2011)

Einer von Stapels zuletzt publizierten Artikeln handelt davon, dass Menschen mehr Vorurteile gegenüber farbigen Personen haben, wenn sie sich in einer unsauberen Umgebung befinden. Diese gänzlich erfundene Studie, welche in Wirklichkeit niemals in der beschriebenen U-Bahn Station in den Niederlanden, durchgeführt wurde, sorgte medial für großes Aufsehen und wurde wenig später bereits von *Science* aufgrund Stapels Betrugsgeständnisses zurückgezogen. „Coping with chaos“ wurde 6 Mal zitiert, davon einmal durch das *Retraction Statement*. Folgende Artikel, die sich bereits mit den Ausmaßen der Stapel Affäre befassen, erwähnen seine gefälschten Ergebnisse:

→ Stroebe, W., Postmes, T. & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives of Psychological Science*, 7, 670-688.

Der Artikel von Stroebe et al. (2012) war Ausgangspunkt für die Entwicklung der bearbeiteten Forschungshypothesen und wurde in der vorliegenden Diplomarbeit bereits mehrmals zitiert.

→ Yong, E. (2012). Bad Copy. *Nature*, 485, 298-300.

Auch der Artikel von Ed Yong wurde für die Beantwortung der Forschungsfrage zur Replikationsproblematik bereits mehrfach in dieser Arbeit erwähnt.

24. Häfner, M. & Stapel, D. A. (2010). Information to go: Fluency enhances the usability of primed information. *Journal of Experimental Social Psychology*, 46, 73-84. (Retracted article. See vol. 49, pg. 318, 2013)

Häfner und Stapel konnten in ihrer Studie demonstrieren, dass die flüssige Verarbeitung von geprimten Eigenschaften als Hinweisreiz für Personen dient, dass es sich um eine anwendbare Information handelt, auch wenn diese eigentlich nicht zur Zielbeschreibung passt. Sie manipulierten die Verarbeitungsflüssigkeit in Experiment 1 durch reimende/nicht reimende Wörter und in Experiment 2 und 3 durch die Schriftart und -größe. Flüssig präsentierte Eigenschaften hatten einen Einfluss auf die wahrgenommene Anwendbarkeit der Information.

Die Autoren wurden 5 Mal zitiert, einmal von Stapel selbst und einmal durch die Zurückweisung des Artikels durch die betreffende Fachzeitschrift. Folgende zwei Artikel bauen auf Häfner und Stapels Ergebnissen auf:

→ Rennekamp, K. (2012). Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research*, 50, 1319-1354.

Die „Securities and Exchange Commision“ betont die klare und lesbare Formulierung von finanziellen Offenlegungen. Nach psychologischer Literatur führt eine Prozessflüssigkeit in einer lesbareren Offenlegung dazu, dass Investoren glauben, dass sie sich auf diese Information verlassen können. Die Prozessflüssigkeit ist ein subjektives Gefühl eines Individuums, wie leicht eine Information verarbeitet werden kann. Das Gefühl der Prozessflüssigkeit kann in Beurteilungen/Entscheidungen unbewusst als Urteilsheuristik herangezogen werden, dass es sich um eine zuverlässige Informationsquelle handelt (Hafner & Stapel, 2010). Daraus schlussfolgert die Autorin ihre zweite und dritte Hypothese:

H2: More readable disclosures increase investors' reliance on the information in the disclosure, and this effect operates through increased feelings of processing fluency.

H3: Investors' judgments of management credibility will be higher when a disclosure is more readable, regardless of whether the disclosure conveys good or bad news about the firm.

→ Loersch, C. & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior and motivation. *Perspectives on Psychological Science*, 6, 234-252.

Auf der Basis bisheriger Priming-Literatur erstellten die Autoren Loersch und Payne (2011) das „Situated inference model“, das davon ausgeht, dass viele unterschiedliche Einflüsse, die im Rahmen eines Priming-Experiments entstehen können, auf drei grundlegende Prozesse zurückgeführt werden können und dass die Anwendung geprimter Information von der unmittelbaren Situation der urteilenden/handelnden Person abhängig ist. Auf der ersten Stufe des Modells („Prime Exposure“) wird Personen mittels unterschiedlichster Priming Verfahren eine bestimmte Information höchst zugänglich gemacht. Ein Prime-Stimulus aktiviert die grundlegende Verfügbarkeit eines mentalen Konstrukts und wird in der zweiten Stufe des Modells vom Individuum als eigene natürliche Wahrnehmung eines Zielobjekts interpretiert („Misattribution“). Die

fehlattribuierte Information wird in der dritten Stufe herangezogen, um die am meisten hervorstechendste Frage bezüglich eines Objekts, einer Person oder der Situation zu beantworten („Afforded Questions“).

In einer Tabelle (*Table 1*) listen Loersch und Payne Studien auf, die sie als Beleg für die Entwicklung ihres 3-Stufen Modells herangezogen haben. Für die erste Stufe der Prime-Exposition dienten ihnen u.a. die Ergebnisse von Stapel und Koomen (2001), nämlich dass die Aktivierung unterschiedlicher „mindsets“ beeinflusst, auf welche Weise Personen Prime-Stimuli anwenden. Für die zweite Stufe der Missattribution stützen sie sich auf die Erkenntnisse von Dijksterhuis et al. (1998), dass das Primen sozialer Kategorien zu Assimilation und das Primen mit konkreten (Vor-)bildern bzw. Personen zu Kontrasteffekten führt.

Eine weitere Tabelle (*Table 2*) zeigt verschiedene Priming-Verfahren auf, welche die Autoren in ihrem Artikel miteinander vergleichen. Darunter befinden sich auch die Studien von Stapel und Koomen (1996; 2001), Dijksterhuis et al. (1998) und Häfner und Stapel (2010).

25. Gordijn, E. H. & Stapel, D. A. (2006). Behavioural effects of automatic interpersonal versus intergroup social comparison. *British Journal of Social Psychology*, 45, 717-729. (Retracted article. See vol. 51, pg. 498, 2012)

Gordijn und Stapels Studie über das Primen intelligenten Verhaltens in Abhängigkeit des aktivierten Gruppenkontexts konnte unlängst nicht von Shanks et al. (2013) repliziert werden. Ihr Artikel wurde 5 Mal zitiert, darin enthalten ist das *Retraction Statement* der Fachzeitschrift.

→ Shanks, D. R., Newell, B. R., Lee, E. H. Balakrishnan, D., Ekelund, L., Cenac, Z. Kavvadia, F. & Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *Plos One*, 8 (4).

Die Autoren replizieren in ihrem Artikel neun Experimente bekannter Intelligenz-Priming Studien, darunter auch die von Gordijn und Stapel (2006), die aufgrund betrügerischer Inhalte von der Fachzeitschrift *Social Psychology* zurückgewiesen wurde. Zu genaueren Informationen über die Replikationsstudien und deren Ergebnisse siehe oben.

26. Van den Bos, A. & Stapel, D. A. (2009). Why people stereotype affects how they stereotype: The differential influence of comprehension goals and self-enhancement goals on stereotyping. *Personality and Social Psychology Bulletin*, 35, 101-113. (Retracted article. See vol. 38, pg. 1378, 2012)

Van den Bos und Stapels Artikel wurde 5 Mal zitiert, davon einmal von der zuständigen Fachzeitschrift bezüglich der *Retraction* und 2 Mal von Stapel selbst. Die übrigen zwei Artikel bauen nicht auf den Erkenntnissen der Autoren auf.

27. Johnson, C. S., Norton, M. I., Nelson, L. D., Stapel, D. A. & Chartrand, T. L. (2008). The downside of feeling better: Self-regard repair harms performance. *Self and Identity*, 7, 262-277. (Retracted article. See vol. 12, pg. 237, 2013)

Johnson et al. erforschten in ihrem Artikel, dass eine Bedrohung des Selbstwerts in einer sozialen Vergleichssituation unter bestimmten Umständen nicht immer als unangenehm empfunden werden muss, sondern dass durch eine gebotene Möglichkeit, seine Kompetenz zu demonstrieren, auch die Motivation sich selbst zu ändern bzw. zu verbessern, erhöht werden kann.

Die Autoren wurden 5 Mal zitiert, davon stammt ein *Retraction Statement* und drei Selbstzitate Stapels. Der verbliebene zitierende Artikel baut nicht explizit auf Johnson et al.s Ergebnissen auf.

28. Trampe, D. & Stapel, D. A. (2010). Beauty as a tool: The effect of model attractiveness, product relevance, and elaboration likelihood on advertising effectiveness. *Psychology & Marketing*, 27, 1101-1121. (Retracted article. See vol. 29, pg. 805, 2012)

Trampe und Stapels Artikel wurde 4 Mal zitiert, davon stammt eine Zitierung von der Zurückweisung der Fachzeitschrift. Die drei verbliebenen zitierenden Artikel bauen nicht explizit auf den Erkenntnissen der Autoren auf.

29. Stapel, D. A. & Noordewier, M. K. (2009). Stop making sense: The ultimate fear. *Psychological Inquiry*, 20, 245-248. (Retracted article. See vol. 23, pg. 316, 2012)

Stapel und Noordewier wurden 4 Mal zitiert, darin enthalten sind ein *Retraction Statement* und zwei Selbstzitate. Der verbliebene vierte Artikel baut nicht explizit auf den Ergebnissen der Autoren auf.

30. Stapel, D. A. & van der Linde (2011). What drives self-affirmation effects? On the importance of differentiating value affirmation and attribute affirmation. *Journal of Personality and Social Psychology*, 101, 34-45. (Retracted article. See vol. 103, pg. 505, 2012)

Stapel und van der Linde demonstrierten in ihren Studien, dass unterschiedliche Typen der Selbstaffirmation zu unterschiedlichen Effekten im Selbstkonzept führen können. Die Affirmation mit persönlich wichtigen Werten („value affirmation“) erhöhte die Selbstklarheit, welche nach Campell et al. (1996) als „the extent to which the contents of an individual's self-concept are clearly and confidently defined, internally consistent, and temporally stable“ definiert ist, während die Affirmation positiver Eigenschaften eines selbst („attribute affirmation“) zu einem erhöhten Selbstwert, allerdings nicht zu einer erhöhten Selbstklarheit führte. Die attributive Affirmation scheint im Vergleich zur Werte-Affirmation also effektiver, um eine den Selbstwert bedrohende Information im Zuge eines sozialen Vergleichs abzuwehren.

Stapel und van der Linde wurden 4 Mal zitiert, davon einmal durch die Zurückweisung des Artikels von der Fachzeitschrift. Folgender Artikel baut auf den Ergebnissen der beiden Autoren auf:

→ Pelham, B. W. & Hardin, C. (2011). Opportunities for sharpening the stereotype inoculation model. *Psychological Inquiry*, 22, 285-290.

In ihrem Review Artikel über die dem „Stereotype Inoculation Model“ zugrunde liegenden Faktoren, gehen die Autoren auf die Fragen ein, welche Rolle Mitglieder der eigenen sozialen Gruppe in der Entwicklung und Aufrechterhaltung der Selbstkonzepte von stigmatisierten Gruppenmitgliedern spielen und auf welche Weise „ingroup“ Vorbilder den Einfluss von stereotypen Informationen auf die stereotyp-bezogenen Fähigkeiten, die Selbstwirksamkeitserwartung und die Motivation, stereotypen Bildungswegen entgegen zu handeln, minimieren bzw. maximieren können. Von Stapel und van der Linde (2011) stammt die Idee, dass mithilfe der Selbstaffirmation der Einfluss einer wahrgenommenen stereotypen Bedrohung oder eines Minderheitsstatus auf eine Testleistung in einem intellektuellen Fähigkeitstest drastisch reduziert werden kann. Für eine Diskussion idealer Bedingungen, unter denen die Selbstaffirmation in verschiedenen Settings eine positive Wirkung zeigen kann, wird auf Stapel und van der Lindes Studie (2011) verwiesen.

31. Ruys, K. I. & Stapel, D. A. (2009). The unconscious unfolding of emotions. *European Review of Social Psychology*, 20, 232-271. (Retracted article. See vol. 23, pg. 386, 2012)

Ruys und Stapel gingen in ihrem Artikel der Frage nach, wie emotionale Reaktionen entstehen und fanden, dass die Verarbeitung von Informationen und deren emotionale Reaktion von einem globalen Prozess ausgeht und sich erst später zu einen spezifischeren Prozess hin entwickelt. Die anfängliche emotionale Reaktion basiert auf einer grundsätzlichen Evaluation der Wertigkeit einer Information (positiv vs. negativ) während spätere Reaktionen auf einer spezifischeren und abgestuften Informationsverarbeitung beruhen. Globale positive oder negative emotionale Reaktionen beeinflussen die Stimmung von Personen, wohingegen fein-stufige emotionale Reaktionen spezifischere Emotionen, wie beispielsweise Angst, Ekel oder Freude auslösen können.

Ruys und Stapel wurden 4 Mal zitiert, einmal von Stapel selbst und einmal von der zuständigen Fachzeitschrift via *Retraction Statement*. Folgender Artikel baut auf den Erkenntnissen der Autoren auf:

→ Rohr, M. Degner, J. & Wentura, D. (2012). Masked emotional priming beyond global valence activations. *Cognition and Emotion*, 26, 224-244.

In Rohr et al.s Studie geht es um das viel erforschte Thema, das emotionale Gesichtsausdrücke von Individuen unbewusst verarbeitet werden können. Das sofortige und adäquate Erkennen emotionaler Reaktionen ist für die soziale Interaktion mit unseren Mitmenschen essentiell, um das eigene Verhalten entsprechend anpassen zu können. Bisherige Forschungsergebnisse gehen davon aus, dass die Verarbeitung von Gesichtsausdrücken im Wesentlichen auf der Unterscheidung zwischen einer positiven und einer negativen Emotion basiert (globale Verarbeitung), dass zwischen spezifischen Emotionen jedoch nicht differenziert wird. Eine weitere Forschungsfrage in diesem Zusammenhang war, ob der emotionale Gesichtsausdruck auch genauso schnell unter begrenzten Wahrnehmungsbedingungen verarbeitet werden kann. Darum untersuchten Rohr et al. in drei Experimenten die Spezifität der Emotionsverarbeitung unter einer begrenzten Wahrnehmung, in dem den Versuchspersonen für 33ms (Studie 1) oder 14ms (Studie 2 und 3) Gesichter mit glücklichem, bösem, traurigem, angstvollem oder neutralem Ausdruck als Prime-Stimuli präsentiert wurden, bevor sie die Emotion eines Zielgesichtsausdrucks (Studie 1 und 2) oder emotionale Adjektive (Studie 3) richtig kategorisieren sollten. Ruys und Stapel (2008) waren das erste Forscherteam, das empirisch bestätigte, dass die emotionale Spezifität von Gesichtsausdrücken auch unter

limitiertem Bewusstsein verarbeitet werden kann. In ihrer Studie wurde den Versuchspersonen einer von vier Gesichtsausdrücken (Ekel, Angst, Ärger und ein neutraler Ausdruck) mehrfach hintereinander in einem parafovealen Priming-Design (für 120ms) präsentiert. Die unbewusste Wahrnehmung der Gesichtsausdrücke führte in einer nachfolgenden Wortergänzungsaufgabe dazu, dass Versuchspersonen Wörter, deren ersten 2-3 Anfangsbuchstaben vorgegeben wurden, konsistent zu ihrer geprimten Emotion beendeten (z.B. das Primen der Emotion „disgust“ führte dazu, dass Versuchspersonen die Buchstaben „si“ am ehesten mit „sick“ anstelle von „sing“ beendeten). Versuchspersonen, denen die Gesichter in der Priming Bedingung jedoch nur für 40ms vorgegeben wurden, unterschieden in der nachfolgenden Aufgabe im Wesentlichen zwischen der Wertigkeit (positive oder negative Bedeutung) eines Wortes und ergänzten die Anfangsbuchstaben konsistent zur Emotionsvalenz, allerdings nicht zur Spezifität. Sie schlussfolgerten daraus, dass sich die Wahrnehmung von Emotionen von einer globalen hin zu einer spezifischeren Ebene entfaltet, je nachdem wie lange ein Stimulus präsentiert wird. Rohr et al. (2012) übernahmen Ruy und Stapels Priming-Design und adaptierten es für die Untersuchung ihrer Hypothesen.

32. Noordewier, M. K. & Stapel, D. A. (2009). Judging the unexpected: Disconfirmation of situation-specific expectancies. *European Journal of Social Psychology*, 39, 944-956. (Retracted article. See vol. 42, pg. 665, 2012)

Noordewier und Stapel gingen in ihrem Artikel der Frage nach, wann und warum die situationsbezogene Spezifität einer Erwartung (z.B. „Michael ist in der Arbeit nett“) nachfolgende soziale Urteile beeinflussen kann. Ihr Artikel wurde 3 Mal zitiert, einmal durch die Zurückweisung der Fachzeitschrift, die übrigen zwei Artikel bauen nicht explizit auf den Erkenntnissen der Autoren auf.

33. Meijers, M. H. C. & Stapel, D. A. (2011). Me tomorrow, the others later: How perspective fit increases sustainable behavior. *Journal of Environmental Psychology*, 31, 14-20. (Retracted article. See vol. 33, pg. 96, 2013)

Die Autoren demonstrierten in zwei Studien, dass Individuen nachhaltigere Entscheidungen treffen, wenn es zu einer Passform der zeitlichen Perspektive (kurzfristig vs. langfristig) und der Personenperspektive (ich vs. andere) kommt. Wenn die Versuchspersonen eine globale Zeitperspektive (langfristig) und eine globale Personenperspektive (andere) oder eine spezifische Zeitperspektive (kurzfristig) zusammen mit einer spezifischen Personenperspektive (ich) einnahmen, sprachen sie der Nachhaltigkeit eine wichtigere Bedeutung zu.

Meijers und Stapel wurden 3 Mal zitiert, einmal durch das *Retraction Statement* der Fachzeitschrift und 2 Mal von anderen Forschergruppen, wobei eine Gruppe ihren Artikel auf den Erkenntnissen der beiden Autoren stützt:

→ Lee, T. H., Jan, F. H. & Yang, C. C. (2013). Conceptualizing and measuring environmentally responsible behaviors from the perspective of community-based tourists. *Tourism Management*, 36, 454-468.

Die beste Möglichkeit für viele Wissenschaftler, einen nachhaltigeren Tourismus zu schaffen, bei dem die Umwelt nicht zerstört wird, ist über die Erforschung des umweltverantwortlichen Verhaltens des Menschen („environmentally responsible behavior“= ERB). Lee et al. (2012) listen in einer Tabelle verschiedene Verhaltensweisen und psychologische Variablen auf, die dazu beitragen können, die Umwelt zu schützen. In dieser Tabelle tauchen auch die Forschungsergebnisse von Meijers und Stapel (2011) auf, die besagen, dass Individuen, deren Überlegungen zu zukünftigen Konsequenzen hoch ausgeprägt sind, sich nachhaltiger verhalten und nachhaltigere Entscheidungen treffen.

34. Joly, J. F. & Stapel, D. A. (2009). Staff, miter, book, share: How attributes of Saint Nicholas induce normative behavior. *European Journal of Social Psychology*, 39, 145-153. (Retracted article. See vol. 42, pg. 800, 2012)

Joly und Stapel demonstrierten in ihrer Studie, dass Kinder mehr Süßigkeiten mit anderen Kindern teilten, wenn sie zuvor mit Eigenschaften von St. Nikolaus geprimt wurden (im Vergleich zur Kontrollgruppe, die mit einem Zwerg geprimt wurde) und den Aspekt des Teilens positiver bewerteten als die Kontrollgruppe. Dieses Ergebnis stellte den ersten empirischen Beweis dafür dar, dass die Eigenschaften normativer Symbole normatives Verhalten auslösen und Norm-Beurteilungen ändern können.

Joly und Stapel wurden 2 Mal zitiert, davon stammt ein Zitat von der Zurückweisung durch die Fachzeitschrift. Der verbliebene Artikel baut auf den Erkenntnissen der beiden Autoren auf:

→ Broeders, R., van den Bos, K., Muller, P. A. et al. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology*, 47, 923-934.

Broeders et al. untersuchten in ihrem Artikel, wie Menschen moralische Dilemmata lösen, in denen unterschiedliche moralische Regeln miteinander in Konflikt stehen. Eines der bekanntesten Beispiele stellt das sogenannte „Footbridge Dilemma“ dar. Darin geht es um die Frage, ob man das Leben einiger Menschen opfern soll, um mehrere retten zu können. Die Autoren fanden, dass die moralische Regel, die während des Entscheidungsprozesses kognitiv am ehesten zugänglich ist („Save lives“ oder „Do not kill“) darüber entscheidet, wie Personen dieses Dilemma lösen. Ausgehend von Joly & Stapels Erkenntnissen (2009), dass Symbole eine wichtige Rolle im menschlichen Verhalten spielen und im Stande sind, soziale Normen zu primen, benutzten die Autoren in ihrer zweiten Studie Symbole, um die beiden Botschaften „Save lives“ (rotes Kreuz, Symbol eines Krankenhauses, Bild eines Rettungsschwimmers) und „Do not kill“ („Peace“ Logo, Bild der 10 Gebote, Logo einer niederländischen Organisation gegen sinnlose Gewalt) durch ein 9-teiliges Puzzle zu aktivieren, das zusammen gebaut jeweils eines der Symbole abbildete. Die „Save lives“ Symbole führten dazu, dass sich die Versuchspersonen bereit fühlten, im Dilemma einzugreifen im Vergleich zur „Do not kill“ Bedingung und zur Kontrollgruppe, die mit den Symbolen eines Fahrrads, Tisches und eines Sessels geprimt wurde.

35. Johnson, C. S. & Stapel, D. A. (2011). Happiness as alchemy: Positive mood leads to self-serving responses to social comparisons. *Motivation and Emotion*, 35, 165-180. (Retracted article. See vol. 36, pg. 180, 2012)

Die Autoren Johnson und Stapel fanden in ihrer Studie 2011, dass Personen in einer positiven Stimmung Informationen auf die Art und Weise verarbeiten, dass ihre positive Stimmung erhalten bleibt. Positiv gestimmte Personen interpretieren in einer sozialen Vergleichssituation zweideutige Informationen über eine Zielperson in einer selbstförderlichen Weise, so dass sie sich aufsteigenden sozialen Vergleichen in ihrer Bewertung annähern (Assimilation) und sich von absteigenden Vergleichen distanzieren (Kontrast). Die Stimmung der Versuchspersonen wurde in vier Experimenten durch eine autobiographische Methode geprimt, bei der sie sich an ein positives vs. negatives Ereignis erinnern und anschließend niederschreiben sollten, welche Erfahrung sie damals und noch heute glücklich/unglücklich macht(e).

Johnson und Stapels Artikel wurde 2 Mal zitiert, einmal von der Fachzeitschrift *Motivation and Emotion* bezüglich dessen Zurückweisung und einmal von Pelham und Hardin (2011), die ihren Artikel auf die Ergebnisse der Autoren stützten:

→ Pelham, B. W. & Hardin, C. (2011). Opportunities for sharpening the stereotype inoculation model. *Psychology Inquiry*, 22, 285-290.

Pehlham und Hardin erstellten in ihrem Artikel (2011) das sogenannte „Stereotype inoculation model“, also ein Modell, welches Faktoren beschreibt, die Vorurteilen gegenüber stereotypisierten Gruppenmitgliedern entgegenwirken können. Die wesentlichste Frage dabei ist, welche Rolle „ingroup“ Mitglieder in der Entwicklung und Aufrechterhaltung von Selbstkonzepten der stigmatisierten Gruppenmitglieder spielen. Erkenntnisse aus der sozialen Vergleichsforschung liefern hierzu wichtige Antworten, da bereits demonstriert werden konnte, dass aufsteigende soziale Vergleiche auch zu einer Inspiration, Motivation und somit Assimilation von Personen gegenüber der vergleichenden überlegenen Person führen können, sofern sich die vergleichende Person nicht in einer Wettbewerbssituation mit der Zielperson wahrnimmt (Stapel & Koomen, 2005) oder motiviert ist, ihre positive Stimmung aufrecht zu erhalten (Johnson & Stapel, 2011). Außerdem ist bekannt, dass Personen, die Informationen über die Eigenschaften anderer Personen im Vergleich zu Informationen über ein einzelnes Individuum erhalten, dazu tendieren, sich in ihrer Selbstsicht dem ausgesetzten Standard anzupassen (Stapel, Koomen & van der Pligt, 1997).

Die soziale Vergleichstheorie spielt also eine wesentliche Rolle im „Stereotype inoculation model“, was durch das Beispiel veranschaulicht werden kann, dass sich Frauen, die in einem vier-Personen Team, bestehend aus drei Frauen und einem Mann, arbeiten, wohler fühlen, als Frauen in einem vierer-Team, bestehend aus zwei Frauen und zwei Männern, da der Vergleich mit nur einer zweiten Frau einen Vergleichsstandard darstellt, zu dem man sich selbst in Kontrast setzt, während sich eine Frau mit zwei anderen Frauen als „ingroup“ Mitglied fühlt und sich in ihrer Bewertung oder ihrem Verhalten der Gruppe anpasst.

Aus bisheriger Forschung ist ebenfalls bekannt, dass die Selbstaffirmation (selbstbegründende Sätze) den Einfluss einer stereotypen Bedrohung oder einen Minderwertigkeitsstatus auf die intellektuelle Leistung in einem Test wesentlich reduzieren kann. Für eine Diskussion idealer Bedingungen, unter denen verschiedene Variationen der Selbstaffirmation in unterschiedlichen Settings zu besseren Leistungen führen können, verweisen die Autoren auf die Literatur von Stapel und van der Linde (2011).

Die folgenden sieben Artikel von Stapel als Erst- oder Co-Autor wurden alle 2 Mal zitiert, davon jeweils einmal durch die offizielle Zurückweisung der zuständigen Fachzeitschrift und einmal durch verschiedene Autoren, die ihre Artikel jedoch alle nicht explizit auf den Ergebnissen von Stapel und Kollegen aufbauen:

36. Trampe, D., Stapel, D. A. & Siero, F. W. (2011). The self-activation effect of advertisements: Ads can affect whether and how consumers think about the self. *Journal of Consumer Research*, 37, 1030-1045. (Retracted article. See vol. 39, pg. 444, 2012)
37. Johnson, C. S. & Stapel, D. A. (2010). It depends on how you look at it: Being versus becoming mindsets determine responses to social comparisons. *British Journal of Social Psychology*, 49, 703-723. (Retracted article. See vol. 51, pg. 497, 2012)
38. Stapel, D. A. & Marx, D. M. (2007). Making sense of war: Using the interpretation comparison model to understand the Iraq conflict. *European Journal of Social Psychology*, 37, 401-420. (Retracted article. See vol. 42, pg. 934, 2012)
39. Lammers, J. & Stapel, D. A. (2011). Racist biases in legal decisions are reduced by a justice focus. *European Journal of Social Psychology*, 41, 375-387. (Retracted article. See vol. 42, pg. 666, 2012)
40. Schwinghammer, S. A. & Stapel, D. A. (2011). Measure by measure: When implicit and explicit social comparison effects differ. *Self and Identity*, 10, 166-184. (Retracted article. See vol. 12, pg. 467, 2013)
41. Noordewier, M. K., van Horen, F., Ruys, K. I. & Stapel, D. A. (2010). What's in a name? 361.708 euros: The effects of marital name change. *Basic and Applied Social Psychology*, 32, 17-25. (Retracted article. See vol. 34, pg. 192, 2012)
42. Lindenberg, S., Joly, J. F. & Stapel, D. A. (2011). The norm-activating power of celebrity: The dynamics of success and influence. *Social Psychology Quarterly*, 74, 98-120. (Retracted article. See vol. 75, pg. 289, 2012)

Zuletzt bleiben noch fünf Artikel von Stapel und seinen Kollegen übrig, die jeweils einmal zitiert wurden und zwar aufgrund der Zurückweisung von den jeweiligen Fachzeitschriften. Von weiteren Autoren wurden die Artikel nicht zitiert, daher wurden aufgrund dieser Arbeiten keine nachfolgenden Forschungsressourcen verschwendet:

43. Noordewier, M. K. & Stapel, D. A. (2011). When failure feels better than success: Self-salience, self-consistency and affect. *British Journal of Social Psychology*, 50, 321-330. (Retracted article. See vol. 51, pg. 404, 2012)
44. Stapel, D. A. & Noordewier, M. K. (2011). The mental roots of system justification: System threat, need for structure and stereotyping. *Social Cognition*, 29, 238-254. (Retracted article. See vol. 30, 2012)
45. Noordewier, M. K. & Stapel, D. A. (2011). Stereotype disconfirmation affect: When sweet hooligans make you happy and honest salesmen make you sad. *Basic and Applied Social Psychology*, 33, 1-6. (Retracted article. See vol. 34, pg. 192, 2012)
46. Avramova, Y. R., Stapel, D. A. & Lerouge, D. (2010). The influence of mood on attribution. *Personality and Social Psychology Bulletin*, 36, 1360-1371. (Retracted article. See vol. 38, pg. 1378, 2012)
47. Nelissen, R. M. A., van de Ven, N. & Stapel, D. A. (2011). Status concerns and financial debts in adolescents. *Social Influence*, 6, 39-56. (Retracted article. See vol. 8, pg. 216, 2013)

Laut der „Retraction Watch“ (<http://retractionwatch.com/category/by-author/diederik-stapel/>) wurden offiziell bereits 54 Artikel Diederik Stapels von den jeweiligen Fachzeitschriften zurückgewiesen. Die verbliebenen sieben Artikel scheinen im *Web of Science* offiziell noch nicht als „retracted articles“ auf und wurden somit nicht in meine Literaturrecherche miteinbezogen:

Ruys, K. I. & Stapel, D. A. (2008). Emotion elicitor or emotion messenger? Subliminal priming reveals two faces of facial expressions. *Psychological Science*, 19, 593-600. (Wurde 26 Mal zitiert)

Stapel, D. A., Koomen, W. & Ruys, K. I. (2002). The effects of diffuse and distinct affect. *Journal of Personality and Social Psychology*, 83, 60-74. (Wurde 70 Mal zitiert)

Stapel, D. A. & Van der Zee, K. I. (2006). The self salience model of other-to-self effects: Integrating principles of self-enhancement, complementarity, and imitation. *Journal of Personality and Social Psychology*, 90, 258-271. (Wurde 32 Mal zitiert)

Stapel, D. A. & Suls, J. (2004). Method matters: Effects of explicit versus implicit social comparisons on activation, behavior, and self-views. *Journal of Personality and Social Psychology*, 87, 860-875. (Wurde 47 Mal zitiert)

Stapel, D. A. & Blanton, H. (2004). From seeing to being: Subliminal social comparisons affect implicit and explicit self-evaluations. *Journal of Personality and Social Psychology*, 87, 468-481. (Wurde 97 Mal zitiert)

Van den Bos, A. & Stapel, D. A. (2008). The Impact of comprehension versus self-enhancement goals on group perception. *Social Psychology*, 39, 222-230. (Wurde 1 Mal zitiert)

Ruys, K. I. & Stapel, D. A. (2008). The secret life of emotions. *Psychological Science*, 19, 385-391. (Wurde 31 Mal zitiert)

6.1.4 Forschungsfrage 1.3.: Welche von Stapel verfassten Buchkapitel basieren auf den Erkenntnissen manipulierter Studien?

Da Diederik Stapel nicht nur Artikel in Fachzeitschriften publizierte, sondern auch Autor von Beiträgen in Büchern war, deren Inhalte sich zum Teil auf dessen betrügerische Literatur stützen, untersuchte die Kommission (Stapel Investigation, 2012) um wie viele Bücher es sich dabei handelt. Die Untersuchung hat ergeben, dass bei folgenden 9 Büchern mit Diederik Stapel als Erst- oder Co-Autor davon ausgegangen werden kann, dass manipulierte Studien als Literaturgrundlage verwendet wurden:

Häfner, M. & Stapel, D. A. (2007). Attraktive Models in der Werbung: Auswirkungen sozialer Vergleiche auf den Betrachter und die beworbenen Produkte. In S. Trepte & E. H. Witte (Hrsg.), *Sozialpsychologie und Medien* (S. 159-170). Lengerich: Pabst.

Martin, L. L. & Stapel, D. A. (1998). Metacognition and correction processes in social judgment: Are people naive dogmatists or naive empiricists? In V. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and Social Dimensions* (pp. 228-247). London: Sage.

Martin, L. L., Strack, F. & Stapel, D. A. (2001). What's on our mind and how did it get there? The role of knowledge accessibility in social psychological phenomena. In A. Tesser & N. Schwarz (Eds.), *Intraindividual processes* (pp. 236-256). Malden, MA: Blackwell Publishers.

Rosema, M., Jost, J. T. & Stapel, D. A. (2008). Social psychology and the study of politics. In L. Steg, A. P. Buunk & T. Rothengatter (Eds.), *Applied Social Psychology: Understanding and Managing Problems* (pp. 291-315). New York, NY, USA: Cambridge University Press.

Ruys, K. I., Stapel, D. A. & Aarts, H. (2011). From (unconscious) perception to emotion: A global-to-specific unfolding view of emotional responding. In I. Nyklicek, A. J. J. M. Vingerhoets, & M. Zeelenberg (Eds.), *Emotion Regulation and Well-Being* (pp. 49-66). New York: Springer.

Stapel, D. A. & Koomen, W. (2001). Let's not forget the past when we go to the future: On our knowledge of knowledge accessibility effects. In G.B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 229-246). Mahwah NJ: Lawrence Erlbaum.

Stapel, D. A. (2003). Making sense of hot cognition: Why and when description influences our feelings and judgments. In J. P. Forgas, K. D. Williams & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 227-250). Cambridge: Cambridge University Press.

Stapel, D. A. (2007). In the mind of the beholder: The interpretation comparison model of accessibility effects. In D. A. Stapel & J. Suls (Eds.), *Assimilation and Contrast in Social Psychology* (pp. 313-327). New York: Psychology Press.

Stapel, D. A. (2011). Priming as proxy: Understanding the subjectivity of social life. In K. C. Klauer, A. Voss & C. Stahl (Eds.), *Cognitive Methods in Social Psychology* (pp. 148-183). New York: Guilford Publications.

6.1.5 Forschungsfrage 1.4.: Welche (Lehr-) Bücher stützen ihre Erkenntnisse auf Stapels manipulierte Studien?

Aufgrund der Popularität Stapels Forschungsschwerpunkte, vor allem im Bereich der Assimilations- und Kontrasteffekte im Zuge der sozialen Urteilsforschung, tauchen Zitierungen seiner manipulierten Studien auch immer wieder in (Lehr-) Büchern anderer Autoren der Sozialpsychologie auf. Um einen Überblick darüber zu bekommen, in welchem Ausmaß seine Erkenntnisse zitiert wurden, wurde eine stichprobenartige Literatursuche durchgeführt.

Dazu wurde zunächst im online Universitätskatalog der Universität Wien via der Suchfunktion „U-Search“ nach Büchern mit dem Titel „Sozialpsychologie“ beziehungsweise „Social Psychology“ gesucht. Die dadurch angezeigten Bücher wurden anschließend aus den Bibliotheken der Universität Wien ausgeliehen, sofern sie verfügbar waren, und auf Stapel Zitierungen untersucht.

Zusätzlich wurde eine Lehrbuchliste von der Webseite *Social Psychology Network* (<http://www.socialpsychology.org/texts.htm>) zu Rate gezogen, auf der die populärsten Lehr- und Sachbücher der Sozialpsychologie nach Kategorien schwerpunktmäßig geordnet sind. Von dieser Liste wurden hauptsächlich aus den Kategorien *Introductory Social Psychology*, *Applied Social Psychology*, *Handbooks*, *Social Cognition* und *Prejudice and Discrimination* mehrere Bücher stichprobenartig entnommen.

Eine Liste der insgesamt 48 untersuchten Bücher findet sich im Anhang wider. Von diesen 48 Büchern wurden in 10 einer oder mehrere Artikel Diederik Stapels zitiert. Tabelle 5 bietet eine Zusammenfassung der Anzahl an Artikeln, die in den Büchern jeweils gefunden wurden und ob diese von der Stapel Untersuchungskommission (Stapel Investigation, 2012) als betrugshaft (=fraudulent) beziehungsweise auch von der betreffenden Fachzeitschrift offiziell zurückgewiesen wurden. In 10 Büchern wurden insgesamt 41 Artikel Stapels gefunden, davon wurden 13 für betrugshaft befunden und 8 von den Fachzeitschriften zurückgezogen.

Tabelle 5: Stapel Literatur in (Lehr-) Büchern

Buch Nr.	Autoren	Enthaltene Stapel Literatur	Davon „fraudulent“	Davon von den Fachzeitschriften zurückgewiesen
1	Aronson, E., Wilson, T. D. & Akert, R. M. (2008)	1	0	0
2	Baumeister, R. F. & Vohs, K. D. (2012)	1	1	0
3	Bless, H., Fiedler, K. & Strack, F. (2004)	4	1	1
4	Dovidio, J. F. (2010)	2	0	0
5	Frey, D. (2002)	1	0	0
6	Kassin, S., Fein, S. & Markus, H. R. (2010)	1	0	0
7	Kruglanski, A. W. & Higgins, E. T. (2007)	18	7	3
8	Myers, D. G. (2013)	2	2	2
9	Taylor, S. E., Peplau, A. L. & Sears, D. O. (2006)	7	2	2
10	Werth, L. & Mayer, J. (2007)	4	0	0
Gesamt		41	13	8

6.1.5.1 Inhaltliche Darstellung der Stapel Literatur in den (Lehr-) Büchern

Im Folgenden sind die untersuchten Bücher alphabetisch von 1-10 geordnet und dazu ist angegeben, welche Artikel Stapels im jeweiligen Buch zitiert werden, auf welcher Seite und in welchem Kapitel sich diese Zitate befinden und worum es sich inhaltlich dabei handelt. Alle angeführten Zitate entstammen den angegebenen Werken und können in diesen nachgelesen werden.

1. Aronson, E., Wilson, T. D. & Akert, R. M. (2008). *Sozialpsychologie*. München: Pearson Studium.

Darin enthaltene Stapel Literatur:

Stapel, D. A. & Koomen, W. (2000). How far do we go beyond the information given? The impact of knowledge activation on interpretation and inference. *Journal of Personality and Social Psychology*, 79, 19-37.

Kapitel 3.1. Im Autopilot – Denken ohne Mühe (S. 63)

Bezüglich der Frage, welche Schemata bei einem Priming Vorgang zur Anwendung kommen, gehen die Buchautoren auf den Faktor der Zugänglichkeit eines kognitiven Schemas ein. Zur Veranschaulichung nennen sie ein Beispiel:

Ein Mann setzt sich in U-Bahn neben Sie, starrt alle Fahrgäste an, fährt sich wiederholt mit den Händen übers Gesicht und murmelt Zusammenhangsloses vor sich hin. Wie würden Sie sein Verhalten interpretieren? Dabei können verschiedene Schemata zur Anwendung kommen, zum Beispiel das Schema eines Alkoholikers, eines Geisteskranken etc. Die Zugänglichkeit des Schemas „Alkoholiker“ ist wahrscheinlich dann erhöht, wenn die beobachtende Person zuvor zum Beispiel mit einer guten Freundin über das Thema Alkoholismus gesprochen hat.

Grundsätzlich unterscheiden die Autoren zwischen drei Situationen, in denen ein Konzept zugänglicher ist als ein anderes und zitieren dabei die Studienergebnisse von Stapel et al.:

Chronische Zugänglichkeit kann aufgrund vorausgegangener Erfahrung entstehen, zweitens kann die Zugänglichkeit aufgrund des Bezugs zu einem aktuellen Ziel beeinflusst werden und drittens können Schemata auch aufgrund des gerade Erlebten zeitweilig zugänglich gemacht werden (Stapel & Koomen, 2000).

2. Baumeister, R. F. & Vohs, K. D. (2012). *New Directions in Social Psychology*. London: Sage Publications.

Darin enthaltene Stapel Literatur:

Stapel, D. A. (2007). In the mind of the beholder: The interpretation comparison model of accessibility affects. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (pp. 143-164). New York: Psychology Press.

Kapitel 7: Construal-related Extensions (S. 406)

Um die Erkenntnisse zur Entstehung von Assimilations- und Kontrasteffekten zu veranschaulichen, nennen Baumeister und Vohs (2012) folgendes Beispiel:

Wenn man die Arbeit eines Studenten (A) mit der besten Klassenarbeit, die eines anderen Studenten (B), vergleicht, dann ähnelt die Arbeit von A womöglich der exzellenten Arbeit von B und erscheint dadurch besser. Oder die Arbeit von A wird durch den Vergleich mit B in Kontrast zu diesem gesetzt und erscheint dadurch wesentlich schlechter. Die Frage, welche der beiden gegensätzlichen Prinzipien wann zum Einsatz kommt, ist also entscheidend (Mussweiler, 2001; Schwarz & Bless, 2007; Stapel, 2007).

3. Bless, H., Fiedler, K. & Strack, F. (2004). *Social cognition: How individuals construct social reality*. New York: Psychology Press.

Darin enthaltene Stapel Literatur:

Stapel, D. A., Koomen, W. & van der Pligt, J. (1996). The referents on trait inferences: The impact of trait concepts versus actor-trait links on subsequent judgments. *Journal of Personality and Social Psychology*, 33, 47-76.

Stapel, D. A., Reicher, S. D. & Spears, R. (1995). Contextual determinants of strategic choice: Some moderators of the availability heuristic. *European Journal of Social Psychology*, 25, 141-158.

Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D. A., Koomen, W., van Knippenberg, A. & Scheepers, D. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology*, 75, 862-871.

Martin, L. L., Strack, F. & Stapel, D. A. (2001). How the mind moves: Knowledge accessibility and the fine-tuning of the cognitive. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (Vol. 1) (pp. 236-256). London: Blackwell.

Differential use of activated information (S. 75)

Im Kapitel über die unterschiedliche Anwendung (durch Priming) aktivierter Informationen, greifen die Autoren auf das „Inclusion/Exclusion Model“ zurück (Bless & Schwarz, 1998; Schwarz & Bless, 1992a). In diesem Modell wird die Anwendung aktivierter Informationen erklärt, welche folgend auch von Stapel, Koomen und van der Pligt (1996) in einem Experiment demonstriert wurde:

Aus deren Forschung ist bekannt, dass soziale Urteile Assimilationseffekte reflektieren, wenn der Prime-Stimulus aus einer abstrakten Eigenschaftskategorie stammt und Kontrasteffekte, wenn der Prime-Stimulus die Eigenschaften mit einer konkreten Person verbindet, die mit Namen und Foto präsentiert wird. In Stapel et al.s Experiment (1996) wurden Versuchspersonen in zwei unterschiedlichen Versuchsbedingungen geprimt. Der ersten Versuchsgruppe wurden Wortgruppen zum Thema Abhängigkeit vorgegeben und die zweite Versuchsgruppe bekam ein Foto mit einer Frau namens Linda präsentiert, die als besonders abhängig beschrieben wurde. Danach bekamen beide Gruppen die Aufgabe, eine Person, die mit mehrdeutigen Verhaltensweisen beschrieben wurde, auf mehreren Dimensionen zu beurteilen. In der Foto-Gruppe schätzten die Versuchspersonen Linda als wesentlich abhängiger als die Zielperson ein, wodurch in ihrer Bewertung ein Kontrasteffekt resultierte, der die Zielperson als weit weniger abhängig erscheinen ließ, im Gegensatz zur Bewertung der Versuchsgruppe, die zuvor eine abstrakte Eigenschaftsbeschreibung erhalten hatte. Eine geprimte Eigenschaft kann also viel eher in das Schema einer zu beurteilenden Person inkludiert werden, wohingegen eine konkrete Personenbeschreibung einen starken Kontrast zum zu bewertenden Zielobjekt darstellt, welche nicht in ein Schema integriert werden kann und somit zu einem direkten Vergleich führt.

Die Buchautoren ergänzen diese Informationen damit, dass ähnliche Ergebnisse auch beim Priming von manifesten Handlungen festgestellt werden konnten, siehe Bargh et al. (1996), der demonstrierte, dass das Primen des Konzepts „alt“ zu einer langsameren Gehgeschwindigkeit bei Studenten führte, oder Dijksterhuis et al. (1998), die zeigten, dass das Priming mit einem Professors (Einstein) oder einem Model (Claudia Schiffer) die Intelligenz-Leistung von Versuchspersonen in einem nachfolgenden Test beeinflussen konnte.

Possibilities and conditions of the flexible use of judgmental heuristics (S. 115)

Eine Studie von Stapel, Reicher und Spears (1995, Experiment 5) demonstriert ein Beispiel für die flexible Anwendung von Urteilsheuristiken (hier: Verfügbarkeitsheuristik) als eine Funktion ihrer Anwendbarkeit für eine Aufgabe. Den Versuchspersonen wurden Listen mit 19 Frauen und 20 Männernamen präsentiert, wobei sie einschätzen sollten, welche Namen (Frauen oder Männer) in der Liste häufiger vorkommen (siehe auch Punkt 5. Frey, D. (2002). Theorien der Sozialpsychologie - 3. Motivations-, Selbst- und Informationsverarbeitungstheorien. Bern: Huber.).

Anwendbarkeit aktivierter Informationen (S. 127)

Auch an einer anderen Stelle in ihrem Buch, gehen die Autoren nochmals auf die Anwendung aktivierter Informationen und deren Konsequenz für Assimilations- oder Kontrasteffekte ein. Sie schreiben, dass eine aktivierte Information nicht direkt in unser Urteil einfließen muss. Manchmal ist die durch Priming aktivierte Information einfach nicht auf das gefragte Urteil anwendbar. Zur Veranschaulichung nennen Bless et al. (2004) ein Beispiel:

Wenn man die Vertrauenswürdigkeit eines Politikers namens Smith einschätzen soll und einem durch eine zuvor geprimte Aufgabe der skandalträchtige Politiker Miller in den Sinn kommt, dann ist diese Information zwar nützlich, jedoch kann das Beispiel Miller nicht einfach in das Beispiel Smith integriert werden. Also fungiert Miller als ein Standard, mit dem der Politiker Smith verglichen wird. Infolgedessen erscheint Smith im Gegensatz zu Miller wesentlich vertrauenswürdiger als er dies ohne die Aktivierung durch den Skandalpolitiker eingeschätzt worden wäre. Wenn man allerdings nun die Vertrauenswürdigkeit von Politikern allgemein einschätzen soll (und zuvor mit dem Skandalpolitiker Miller geprimt wurde), dann lässt sich das Beispiel Miller gut in die gefragte Kategorie integrieren und man schätzt die allgemeine Vertrauenswürdigkeit geringer ein als ohne vorangegangenes Priming.

Dieses praktische Beispiel unterstützt die Hypothese von Stapel et al. (1996), dass es bei der Aktivierung von abstrakten Eigenschaftskategorien zu Assimilationseffekten kommt und beim Primen einer spezifischen Person eher ein Kontrasteffekt entsteht. Eine geprimte Eigenschaft kann leichter in die mentale Repräsentation einer Person integriert werden, was für das Primen mit einer konkreten Person nicht der Fall ist.

Das „Inclusion-Exclusion“ Modell (S. 133)

Die Buchautoren gehen im Zuge der Informationsverarbeitung auf das „Inclusion-Exclusion“ Modell ein, welches beschreibt, dass die gleiche Art von Information

entgegengesetzte Effekte haben kann, je nachdem wie sie verwendet wird. Es gibt drei unterschiedliche Filter, die bestimmen, wie Informationen genutzt werden:

1. Individuen schließen Informationen aus, wenn sie glauben, dass diese Information von einer irrelevanten Quelle stammt (zum Beispiel, wenn man sich dessen bewusst ist, dass man extern manipuliert wird)
2. Information wird ausgeschlossen, wenn sie für das Ziel nicht repräsentativ erscheint. Die Kategorisierung von Informationen wird von einer Vielzahl von Variablen beeinflusst, zum Beispiel durch die Extremität und Typizität oder das Präsentationsformat (abstrakt/Bild) und darauf bezogene Kontextvariablen (Martin, Strack & Stapel, 2001).
3. Ein dritter Filter ist das Kommunikationsverhalten, das steuert, welche Informationen in Gesprächen verwendet werden. Informationen werden ausgeschlossen, wenn ihr wiederholter Gebrauch die Gesprächsnorm der Nonredundanz (Wortwiederholung) verletzen würde.

Informationen, die durch alle drei Filter gelangen, werden in die Repräsentation inkludiert und resultieren in Assimilationseffekten. Informationen, die an einem oder mehreren Filtern hängen bleiben, werden aus der Repräsentation ausgeschlossen und können als Standard fungieren, mit dem das Zielobjekt verglichen und kontrastiert wird.

4. Dovidio, J. F. (2010). *The Sage handbook of prejudice, stereotyping and discrimination*. Los Angeles: Sage.

Darin enthaltene Stapel Literatur:

Marx, D. M. & Stapel, D. A. (2006). Understanding stereotype lift: On the role of the social self. *Social Cognition*, 24, 776-791.

Marx, D. M., Stapel, D. A. & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology*, 88, 432-446.

Part IV: Social Impact of Prejudice, Stereotyping and Discrimination (S. 388 ff.)

Im vierten Teil ihres Buches versuchen die Autoren die Frage zu beantworten, in welcher Weise auf den Effekt einer stereotypen Bedrohung eingegriffen und wie dieser abgeschwächt werden kann. Eine Strategie besteht darin, die Bedeutung der stereotypen Identität dahingehend zu verändern, dass positive Aspekte mehr hervorstechen und negative weniger sichtbar werden. Die Autoren Inzlicht und Ben-Zeev (2000) konnten beispielsweise feststellen, dass Frauen in Gruppen, nur mit Frauen als weiteren

Gruppenmitgliedern, bei einem Mathematiktest besser abschnitten als in Gruppen, in denen sich auch Männer befanden (siehe auch Marx, Stapel & Muller, 2005). Dieses Ergebnis zeigt, dass trotz einer bestehenden stereotypen Bedrohung, die Leistung verbessert werden kann, wenn negative Stereotype vermieden werden und die betreffende Person die Wahrnehmung einer Bedrohung reduzieren kann.

Stereotype Lift (S. 389)

Obwohl sich die Forschung zur Bedrohlichkeit von Stereotypen hauptsächlich auf deren negative Auswirkungen konzentriert, von welchen Mitglieder stigmatisierter Gruppen betroffen sind, konnte festgestellt werden, dass eine stereotype Bedrohung auf die nicht stereotypisierte Gruppe sogar leistungsfördernd wirken kann. Männer, schnitten in einem Mathematiktest sogar noch besser ab, wenn dieser für die mathematischen Fähigkeiten als hoch diagnostisch und schwierig angepriesen wurde, wohingegen sie eine schlechtere Leistung erzielten, wenn bei der Beschreibung des Tests die stereotype Bedrohung für Frauen durch Wörter wie „genderfair“ reduziert wurde.

Die Buchautoren erklären, dass die Ursache für diesen „Lift“ (Leistungsanstieg) in einem sozialen Abwärtsvergleich liegt. Eine wahrgenommene Überlegenheit in der stereotypen Bedrohungsbedingung gegenüber der stereotypisierten Gruppe, führt zu einem höheren Selbstvertrauen und einer höheren Leistungsfähigkeit im Mathematiktest.

Eine weitere Möglichkeit, um einen „Lift-Effekt“ zu erzielen, besteht darin, dass die Gruppenunterschiede stark betont werden und die „in-group“ (z.B. Männer) sich dadurch überlegen fühlt. Interessant ist lediglich, warum manche Studien einen „Lift-Effekt“ zeigen, obwohl die Gruppenunterschiede nicht explizit betont wurden. Marx und Stapel (2006) meinen hierzu, dass die Gruppenidentität für die Personen der positiv stereotypisierten Gruppe nicht so stark hervor stechen, wenn sie nicht explizit daran erinnert werden. Die Autoren zeigten in ihrer Studie, dass sich der „Lift-Effekt“ in dem Ausmaß steigerte, in dem Männer auf ihre eigene Geschlechteridentität hingewiesen wurden.

5. Frey, D. (2002). *Theorien der Sozialpsychologie - 3. Motivations-, Selbst- und Informationsverarbeitungstheorien*. Bern: Huber.

Darin enthaltende Stapel Literatur:

Stapel, D. A., Reicher, S. D. & Spears, R. (1995). Contextual determinants of strategic choice: Some moderators of the availability heuristic. *European Journal of Social Psychology*, 25, 141-158.

8.1. Möglichkeiten und Bedingungen der flexiblen Verwendung von Urteilsheuristiken (S. 377)

Ein Beispiel für die flexible Verwendung einer Urteilsheuristik (hier der Verfügbarkeitsheuristik) in Abhängigkeit von deren Angemessenheit für die Aufgabe liefert eine Untersuchung von Stapel, Reicher und Spears (1995). Die Autoren präsentierten den Versuchsteilnehmern Listen mit 19 weiblichen und 20 männlichen Namen (oder umgekehrt), wobei die selteneren Namen berühmter waren als die häufigeren. Die Versuchspersonen sollten anschließend beurteilen, ob mehr Frauen- oder Männernamen in der Liste enthalten waren. Wie bei der Anwendung einer Verfügbarkeitsheuristik erwartet, schätzte die Mehrzahl der Versuchspersonen diejenige Kategorie als größer ein, die aus berühmten Namen bestand, obwohl sie in jedem Fall kleiner war als die mit nicht berühmten Namen. Den Versuchspersonen wurde zusätzlich die Möglichkeit geboten, die Aufgabe drei weitere Male zu bearbeiten – in diesem Fall natürlich in Kenntnis der eigentlichen Aufgabe, also der Schätzung des Anteils weiblicher und männlicher Namen. Diesmal verringerte sich der „Verfügbarkeits-Bias“ so weit, dass im vierten Durchgang nur noch 7 von 22 Personen falsche Urteile abgaben. Offensichtlich hatten die Versuchspersonen von ihrer anfänglichen Urteilsstrategie Abstand genommen und beim zweiten Mal die erfolgreichere gewählt.

6. Kassin, S., Fein, S. & Markus, H. R. (2010). *Social Psychology*. Boston: Houghton Mifflin.

Darin enthaltene Stapel Literatur:

Stapel, D. A. & Blanton, H. (2006). *Social comparison theories*. New York: Psychology Press.

Chapter 3: The Social Self - Influences of other people (S. 65)

Kapitel drei handelt unter anderem davon, welchen Einfluss andere Personen auf das soziale Selbst haben. Die Buchautoren zitieren dabei den „Großvater“ der sozialen Vergleichstheorie, Leon Festinger, der 1954 behauptete, dass Menschen, die sich ihrer Fähigkeiten und Meinungen unsicher sind und wenn objektive Informationen nicht zur Verfügung stehen, dazu tendieren, sich durch den Vergleich mit anderen selbst zu bewerten. Über die Jahre haben sich Sozialpsychologen mit der Überprüfung dieser Theorie beschäftigt und sich dabei auf zwei Fragen fokussiert (nach Stapel & Blanton, 2006):

1. Wann wenden wir uns an andere, um vergleichende Informationen einzuholen?
2. Welche Personen suchen wir uns für unseren Vergleich mit uns selbst aus?

Chapter 4: Perceiving Persons - Implicit Personality Theories (S. 130)

In der Erforschung impliziter Persönlichkeitstheorien war Solomon Asch (1946) der erste Wissenschaftler, der entdeckte, dass das Vorhandensein einer einzelnen Eigenschaft oftmals das Vorhandensein weiterer Eigenschaften impliziert. Asch erzählte einer Gruppe von Versuchsteilnehmern, dass eine Person intelligent, talentiert, vorsichtig etc. und kalt sei und einer anderen Gruppe nannte er die gleichen Eigenschaften, nur dass er das Wort „kalt“ durch „warm“ ersetzte. Die „warme Wort-Gruppe“ sprach der beschriebenen Person wesentlich bessere Eigenschaften zu, wie beispielsweise glücklicher, großzügiger etc. als der kalten Person. Wenn Asch jedoch zwei andere Eigenschaften variierte (höflich vs. klump), dann wurden weitere Unterschiede wesentlich weniger betont. Asch führt dies darauf zurück, dass „warm“ und „kalt“ zwei zentrale Eigenschaften sind, die das Vorhandensein weiterer Eigenschaften implizieren und einen großen Einfluss auf den finalen Eindruck einer Person haben. Andere Wissenschaftler haben ähnliche Effekte beobachtet (Stapel & Koomen, 2000).

7. Kruglanski, A. W. & Higgins, E. T. (2007). *Social Psychology: Handbook of Basic Principles*. New York: Guilford Press.

Darin enthaltende Stapel Literatur:

Stapel, D. A. & Koomen, W. (2001a). The impact of interpretation versus comparison mindsets on knowledge accessibility effects. *Journal of Experimental Social Psychology*, 37, 134-149.

Stapel, D. A. & Koomen, W. (2001b). Let's not forget the past when we go to the future: On our knowledge of knowledge accessibility. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the legacy and future of social cognition* (pp. 229-246). Mahwah: Erlbaum.

Stapel, D. A., Koomen, W. & van der Pligt, J. (1996). The referents on trait inferences: The impact of trait concepts versus actor-trait links on subsequent judgments. *Journal of Personality and Social Psychology*, 33, 47-76.

Stapel, D. A. (2007). In the mind of the beholder: The interpretation comparison model of accessibility effects. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology*. New York: Psychology Press.

- Stapel, D. A. & Koomen, W. (2000). Distinctness of others, mutability of selves: Their impact on self evaluations. *Journal of Personality and Social Psychology*, 79, 1068-1087.
- Stapel, D. A. & Koomen, W. (2001). I, we and the effects of others on me: How self-construal moderates social comparison effects. *Journal of Personality and Social Psychology*, 80, 766-781.
- Stapel, D. A., Koomen, W. & van der Pligt, J. (1997). Categories of category accessibility: The impact of trait concept versus exemplar priming on person judgments. *Journal of Experimental Social Psychology*, 33, 47-76.
- Stapel, D. A., Koomen, W. & Velthuisen, A. S. (1998). Assimilation or contrast? Comparison relevance, distinctness and the impact of accessible information on consumer judgments. *Journal of Consumer Psychology*, 7, 1-24.
- Stapel, D. A. & Suls, J. (2004). Method matters: Effects of explicit versus implicit social comparisons on activation, behavior and self-views. *Journal of Personality and Social Psychology*, 87, 860-875.
- Stapel, D. A., Martin, L. L. & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin*, 24, 797-806.
- Stapel, D. A. & Blanton, H. (2004). From seeing to being: Subliminal social comparisons affect implicit and explicit self-evaluations. *Journal of Personality and Social Psychology*, 87, 468-481.
- Stapel, D. A. & Koomen, W. (1996). Differential consequences of trait inferences: A direct test of the trait-referent hypothesis. *European Journal of Social Psychology*, 26, 827-837.
- Stapel, D. A. & Koomen, W. (1997). Using primed exemplars during impression formation: Interpretation or comparison? *European Journal of Social Psychology*, 27, 357-367.
- Stapel, D. A. & Koomen, W. (1998). When stereotype activation results in (counter) stereotypical judgments: Priming stereotype-relevant traits and exemplars. *Journal of Experimental Social Psychology*, 34, 136-163.
- Stapel, D. A. & Koomen, W. (2000). How far do we go beyond the information given? The impact of knowledge activation on interpretation and inference. *Journal of Personality and Social Psychology*, 78, 19-37.
- Stapel, D. A. & Schwarz, N. (1998). The Republican who did not want to become president: Colin Powell's impact on evaluations of the Republican party and Bob Dole. *Personality and Social Psychology Bulletin*, 24, 690-698.

Stapel, D. A. & Winkielman, P. (1998). Assimilation and contrast as a function of context-target similarity, distinctness and dimensional relevance. *Personality and Social Psychology Bulletin*, 24, 634-646.

Spears, R., Gordijn, E., Dijksterhuis, A. & Stapel, D. A. (2004). Reaction in action: Intergroup contrast in automatic behavior. *Personality and Social Psychology Bulletin*, 30, 605-616.

Chapter 7: Automatic Thought (S. 142, 152, 154, 156, 159, 160)

5 grundlegende Prinzipien der Automatisität:

Kruglanski et al. (2007) diskutieren in ihrem 7. Buchkapitel, ob Informationsverarbeitungsprozesse nicht gleichzeitig automatisch und kontrolliert ablaufen können und zitieren dabei Studienergebnisse von Stapel und Kollegen (Stapel & Koomen, 2001b; Stapel, Koomen & van der Pligt, 1997), die meinen, dass Vergleichsprozesse im Sinne von Kontrasteffekten ausschließlich im Zuge einer automatischen Informationsverarbeitung auftreten.

Zu den fünf grundlegenden Prinzipien der Automatisität der Informationsverarbeitung zählt die Anwendung von aktivierten Eigenschaftskonzepten auf eine zu beurteilende Zielperson. Stapel und Koomen (2001b) fanden diesbezüglich, je verfügbarer ein Konzept ist, desto eher wird es auf ein potentiell relevantes und mehrdeutiges Verhalten angewandt.

Ein weiteres Prinzip stellt die Anwendungshemmung aufgrund widersprüchlicher Informationen dar. Ältere Literatur ging davon aus, dass eine Informationsaktivierung immer zu einer Anwendung dieser Information führt und dass man bei einer Anwendung automatisch von einer vorausgegangenen Aktivierung ausgehen kann. Die neue Literatur zieht jedoch eine Grenze zwischen den beiden Begriffen und meint, dass eine Aktivierung nicht unbedingt zu einer Anwendung führen muss. Ein geprimtes Eigenschaftskonzept muss nicht auf das Verhalten übertragen werden, wenn zwar die Wertigkeit übereinstimmt, das Konzept jedoch nicht anwendbar ist. Wenn man zum Beispiel auf einen feindseligen Tiernamen (Tiger oder Hai) geprimt wird, dieses Konzept aber nicht auf die menschliche Zielperson übertragbar ist, wird die Zielperson auch nicht als feindselig beurteilt werden (Stapel et al., 1997).

Eine weitere wichtige Funktion automatischer Informationsanwendung ist das „Goal of Comprehension“. Ein fundamentales menschliches Motiv im psychologischen Funktionieren ist es, von anderen Menschen verstanden zu werden. Das Ziel, Personen zu verstehen, kann Stereotype verhindern, wenn die Person das Gefühl hat, dass ein

Stereotyp das Verstehen verhindern würde. Wenn diese Ziele geprimt werden, kann das dazu führen, dass Personen aufmerksamer und sorgsamer in ihrer Beurteilung sind. Andererseits können diese Verständnis-Ziele eine Person auch dazu bringen, Stereotype anzuwenden, wenn sie dazu dienen, eine andere Person besser zu verstehen (Kunda & Spencer, 2003; Stapel & Koomen, 2001a).

In ihrer Darstellung zur Entstehung von Assimilations- und Kontrasteffekten in der Informationsanwendung, gehen die Buchautoren auf das „Standard of Comparison Model“ nach Stapel und Koomen (2001a) ein.

In der Priming-Literatur gilt die Basisregel: Wenn eine geprimte soziale Information nicht zu extrem ist, dann wird ein Urteil gefällt, dass sich dem Prime-Stimulus annähert, wohingegen extreme Kategorien einen Kontrasteffekt auslösen. Nach dem „Standard of comparison model“ dient das geprimte Konstrukt als Vergleichsstandard, womit nachfolgende Informationen implizit verglichen werden. Der Vergleichsprozess wird dabei ohne bewusste Überwachung und Kontrolle ausgeführt (Stapel & Koomen, 2001b).

Nun gibt es zwei gegensätzliche Modelle, die sich dahingehend widersprechen, wann Kontraste auftreten. Das „Standard of comparison“ Modell besagt, dass Kontraste durch extreme Prime-Stimuli ausgelöst werden, wohingegen das „Correction-model“ besagt, dass moderate Reize ausschlaggebend sind. Bei beiden Modellen spielt die Aufmerksamkeit eine wesentliche Rolle. Beim „Standard of comparison“ Modell werden aktivierte Muster auf automatische, unbewusste Weise mit einem Standard verglichen, beim „Correction-Model“ führen geprimte Eigenschaften zu einem Vergleich und einer bewussten Korrektur in der Beurteilung, wodurch Kontrasteffekte entstehen können. Die Zielaktivierung („comprehension“ vs. „interpretation“) spielt also eine entscheidende Rolle dabei, auf welche Weise aktivierte Konstrukte angewandt werden (Stapel & Koomen, 2001a).

Chapter 9: Knowledge Activation (S. 208, 209)

In ihrem neunten Buchkapitel gehen die Autoren auf die Gestaltung von Prime-Stimuli ein und demonstrieren durch Verweise auf mehrere Studien, dass die Aktivierung von Informationen unterschiedliche Effekte in sozialen Bewertungsprozessen haben kann, je nachdem ob es sich bei den präsentierten Stimulus Materialien um Eigenschaften, Kategorien oder konkrete Personenbeispiele handelt, wie lange die Pause zwischen Priming und Zielbeurteilung andauert oder wie ähnlich und relevant diese Reize für einen persönlich sind. Auch unterschiedliche Ziele einer Person („interpretation“ vs. „comprehension“) können dafür ausschlaggebend sein, ob im nachfolgenden Urteil ein Assimilations- oder Kontrasteffekt auftritt.

Traits, Categories and Exemplars

In einer Serie von Experimenten fanden Stapel, Koomen & Velthuijsen (1998) den klassischen Assimilationseffekt: Sie primten Versuchspersonen mit Sätzen zu Abhängigkeit und ließen sie anschließend das abhängige Verhalten einer beschriebenen Person beurteilen. Die Versuchsgruppe beurteilte die Zielperson als abhängiger als die Kontrollgruppe, welche zuvor nicht geprimt wurde. In einem weiteren Experiment wurde die Versuchsgruppe auf eine fiktive Person namens Linda geprimt, die als besonders abhängig beschrieben wurde. Plötzlich beurteilte die Versuchsgruppe eine weitere Zielperson als weit weniger abhängig als Linda, im Gegensatz zur Kontrollgruppe, welche nur durch abhängigkeitsbezogene Wörter und nicht durch die konkrete Personenbeschreibung Linda geprimt worden war. (Stapel, Koomen & van der Pligt, 1996).

Closeness and Similarity

Stapel und Koomen (2000; 2001; Stapel & Suls, 2004) untersuchten Variablen in Zusammenhang mit der zeitlichen Distanz zwischen Prime-Stimulus und Bewertung. In einer Serie von Experimenten konnten sie zeigen, dass eng verbundene, hoch relevante, ähnliche und unklare Standards zu Assimilation und nicht zu Kontrast führen. Außerdem führte eine affektive Assimilation (z.B. sich durch einen aufsteigenden sozialen Vergleich glücklich fühlen) dazu, dass Vergleichsstandards als erreichbar und kontrollierbar eingeschätzt wurden (siehe auch Markman & McMullen, 2003).

Extremity of Primes

Dijksterhuis et al. (1998) stellten fest, dass extreme und vom zu bewertenden Zielobjekt deutlich verschiedene Stimuli automatisch zu einem Kontrasteffekt führen (siehe auch Stapel et al., 1996, 1997).

Interpretation versus Comparison

Assimilation und Kontrast entstehen durch unterschiedliche Ziele. In ihrem „Interpretation comparison model of social comparison“ konnten Stapel und seine Kollegen (Stapel, 2007; Stapel & Koomen, 2000, 2001; Stapel & Suls, 2004) die Unterscheidung zwischen einer interpretierenden Denkweise, bei der Personen versuchen, den Sinn einer Zielinformation zu erfassen und einer vergleichenden Denkweise, bei der Personen die Zielinformation mit einem Standard vergleichen, treffen. Die interpretierende Denkweise führt im sozialen Urteil zur Assimilation, wohingegen die vergleichende Denkweise (sofern ein halbwegs extremer Stimulus präsentiert wird) eher zum Kontrast führt.

Chapter 9: Standards (S. 309, 312, 317, 323)

Stapel und Kollegen führten eine Reihe von Studien zu Faktoren durch, welche zu Assimilations- und Kontrasteffekten beitragen und fassten diese in ihrem „Interpretation-comparison“ Modell zusammen (Stapel & Koomen, 1998, 2000; Stapel & Winkielman, 1998). Nach ihrem Modell werden zugängliche und anwendbare Hinweisreize als Interpretationshilfe herangezogen und führen zu Assimilationseffekten, wohingegen entferntere Hinweisreize als Vergleichstandards dienen und zu Kontrasteffekten führen (Higgins, 1996; Stapel & Koomen, 1998). Nicht nur die Ähnlichkeit sondern auch die Eindeutigkeit von Standards und Kontexten entscheidet darüber, ob es zu Kontrast- oder Assimilationseffekten kommt (Stapel, Koomen & Van der Pligt, 1996, 1997).

Weitere Einflussfaktoren sind die Klarheit (distinctness) vs. die Breite (breadth) von Prime-Stimuli. Weitgefasstere, abstrakte Verhaltensmuster als Primes führen eher zur Assimilation, spezifische personenbezogene Verbindungen oder anschauliche, konkrete Beispiele (z.B. Albert Einstein) führen zu Kontrasteffekten (Stapel & Koomen, 1996; Stapel & Schwarz, 1998). Gemäß dem ICM Modell werden enger gefasste Primes als vergleichende Standards für die Beurteilung herangezogen, sie schließen das Ziel aus und schaffen Kontrast. Abstrakte Primes schaffen einen interpretierenden Rahmen und schließen das Ziel nicht aus (Stapel & Koomen, 1996, 1997).

False Consensus and False Uniqueness (S. 317)

Individuen gehen nicht von einer Ähnlichkeit zwischen dem Selbst und anderen aus. Das Selbst wird als verschieden von anderen gesehen und schafft somit einen Kontrast (Stapel & Koomen, 1998).

Others as Standards for Judging the Self: Social Comparison (S. 323)

Locke (2003) geht davon aus, dass zu anderen Menschen in bestimmten Situationen eine gewisse Verbindung oder Solidarität durch einen vergleichenden Stimulus hergestellt werden kann, der dazu führt, dass sich die betreffende Person einer Kategorie zugehörig fühlt und somit Assimilation geschaffen wird. Zum Beispiel durch gemeinsame Gruppenmitgliedschaften, gemeinsame Geburtstage oder ein geprimtes Gefühl, dass das Selbst sozial integriert ist (Stapel & Koomen, 2001).

Chapter 11: The Role of Metacognition in Social Judgment (S. 272+274)

In der kognitiven Überwachung sozialer Urteilsprozesse spielt die Korrektur bereits gefällter Entscheidungen eine wichtige Rolle. Die Forschung zur „theory-based correction“ zeigt auf, wie breit gefächert die Bereiche sind, in denen Bias-Korrekturen auftreten können. Stapel, Martin und Schwarz (1998) lieferten mit ihrer Studie einen wichtigen Beitrag zum Forschungsbereich der Eindrucksbildung. Sie fanden, dass es eine Reihe

von Settings gibt, in denen eine erinnerungsbasierte Neubewertung durchführbar ist. In ihrem Experiment gaben sie den Versuchspersonen die Möglichkeit, Denkfehler zu korrigieren, sofern sie meinten, dass welche aufgetreten seien.

Chapter 18: Social Identity and Self-Regulation

Identity-Based Motivation: Integrating a Social Self-Schema Approach with Social Cognition Frameworks (S. 441)

In mehreren Experimenten zur Identitätsbezogenen Motivation konnten Stapel und Koomen (2001) zeigen, dass das Primen des persönlichen Fürworts „ich“ zu einem starken Kontrast und das Fürwort „wir“ zur Assimilation mit anderen führten. Dabei stellten sie ein selbstdienliches asymmetrisches Muster in diesen Kontrasten fest. Die Kontrasteffekte waren größer, wenn sich die Versuchspersonen selbst mit niedrigeren Standards verglichen oder wenn sich die positiveren Erfolge eines anderen auf einer für die Versuchsperson persönlich unwichtigen Domäne abspielten. Wenn die Versuchspersonen jedoch zuvor mit den Wörtern „compare“, „distinguish“, „differ“, oder „opposition“ geprimt wurden, verschwand diese Asymmetrie und die Versuchsteilnehmer verglichen sich mit Standards, selbst wenn diese in negativen Selbst-Definitionen resultierten.

Chapter 24: Attitude Change (S. 572)

Das „Social Judgment Model“ geht davon aus, dass durch einen automatischen Wahrnehmungsmechanismus der Assimilation oder des Kontrasts, die Art und Weise beeinflusst wird, wie Menschen auf ein Überredungs- bzw. ein Überzeugungsargument reagieren.

Es konnte vielzählig demonstriert werden, dass Assimilation und Kontrast eine wichtige Rolle in der Prüfung von Prozessen spielen, bei denen Personen andere beurteilen (Stapel & Winkelman, 1998).

Chapter 30: The Social Psychology of Intergroup Relations

Separating Intergroup Relations from Group Cognition (S. 697)

Ein Beweis für den Unterschied zwischen stereotyp basierten Prozessen und „ingroup-outgroup“ Prozessen stammt aus der Priming-Forschung. Das Primen des Labels oder des Inhalts einer sozialen Kategorie aktiviert automatisch stereotypbezogene Konzepte, welche zu Verhaltensweisen führen, die mit dem aktivierten Konzept verbunden werden (Assimilation). Beispielsweise schneiden Studenten, die mit dem Konzept Hooligan geprimt wurden in einem nachfolgenden kognitiven Test schlechter ab als Studenten, die mit dem Konzept Professor geprimt wurden. Das Primen der „out-group“ Kategorie führte

dahingegen automatisch zu einem Kontrasteffekt (Spears, Gordijn, Dijksterhuis & Stapel, 2004).

Chapter 34: Cultural Processes

Self and Contrast Effect (S. 796)

Personen, die mit einer sozialen Kategorie geprimt werden, passen sich in ihrem Verhalten dieser Kategorie an. Studenten, die mit der Kategorie „alt“ geprimt wurden, verlangsamten danach ihre Gehgeschwindigkeit. Die Aktivierung von sozialen Stereotypen kann jedoch auch zu einem automatischen Verhaltenskontrast zu der geprimten Kategorie führen, indem ein Vergleich eines selbst mit der geprimten Gruppe provoziert wird. Wenn Studenten beispielsweise mit den Wörtern „selbst“ und „alt“ geprimt werden, dann werden sie zu einem Vergleich gezwungen, der zu Verhaltensweisen führt, die für die verglichene Kategorie untypisch ist (z.B. in normaler oder schnellerer Geschwindigkeit gehen; Spears, Gordijn, Dijksterhuis & Stapel, 2004).

8. Myers, D. G. (2013). *Psychology*. New York: Worth Palgrave Macmillan.

Darin enthaltene Stapel Literatur:

Joly, J. F. & Stapel, D. A. (2009). Staff, miter, book, share: How attributes of Saint Nicholas induce normative behavior. *European Journal of Social Psychology*, 39, 145-153.

Ruys, K. I. & Stapel, D. A. (2008). The secret life of emotions. *Psychological Science*, 19, 385-391.

Chapter 12: Emotions, Stress and Health

Cognition may not precede emotion: Zajonc, LeDoux and Lazarus (S. 62)

Ist das Herz dem Verstand immer untergeordnet? Müssen wir eine Erregung/Aktivierung immer zuerst interpretieren bevor wir eine Emotion spüren können? Robert Zajonc (1980, 1984a) meint, dass wir viele emotionale Reaktionen haben, bevor wir die Situation überhaupt interpretieren. Man kann also etwas oder jemanden mögen, bevor man weiß warum. Wir haben ein so sensibles automatisches Radar für emotional signifikante Informationen, dass sogar subliminal dargebotene Reize uns dahingehend primen können, dass wir uns nach einem folgenden Stimulus besser oder schlechter fühlen (Ruys & Stapel, 2008).

Chapter 8: Memory – Priming (S. 316)

Die Autoren Joly und Stapel (2009) demonstrierten in ihrer Studie einen Primingeffekt, der auf der Erinnerung bekannter Verhaltensnormen basiert. Sie aktivierten bei niederländischen Kindern Verhaltenskonzepte, die mit der Figur „Santa Claus“ in Verbindung stehen. Die geprimten jungen Versuchsteilnehmer teilten mehr Süßigkeiten mit anderen Kindern als Kinder, die mit einer Figur geprimt wurden, die nicht mit Freundlichkeit und Großzügigkeit in Verbindung stand.

9. Taylor, S. E., Peplau, A. L. & Sears, D. O. (2006). *Social Psychology*. New York: Prentice Hall.

Darin enthaltene Stapel Literatur:

Stapel, D. A. & Koomen, W. (2000). Distinctiveness of others, mutability of selves: Their impact on self evaluations. *Journal of Personality and Social Psychology*, 79, 1068-1087.

Stapel, D. A. & Koomen, W. (2001). When we wonder what it all means: Interpretation goals facilitate accessibility and stereotyping effects. *Personality and Social Psychology Bulletin*, 27, 915-929.

Stapel, D. A., Koomen, W. & Ruys, K. I. (2002). The effects of diffuse and distinct affect. *Journal of Personality and Social Psychology*, 83, 60-74.

Stapel, D. A., Koomen, W. & Zeelenberg, M. (1998). The impact of accuracy motivation on interpretation, comparison and correction processes: Accuracy X knowledge accessibility effects. *Journal of Personality and Social Psychology*, 74, 878-893.

Stapel, D. A., Martin, L. L. & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgements? *Personality and Social Psychology Bulletin*, 24, 797-806.

Stapel, D. A. & Tesser, A. (2001). Self-activation increases social comparison. *Journal of Personality and Social Psychology*, 81, 742-750.

Gordijn, E. H., Koomen, W. & Stapel, D. A. (2001). Level of prejudice in relation to knowledge of cultural stereotypes. *Journal of Experimental Social Psychology*, 37, 150-157.

Chapter 2: Person Perception: Forming Impressions of Others (S. 39, 45)

Die Wahrnehmung anderer Personen und unsere sozialen Urteile basieren weitgehend auch auf Kontexten, in denen sie gefällt werden. Es gibt zwei Kontext-Faktoren, die sich hauptsächlich auf soziale Urteile auswirken: Die Assimilation und der Kontrast.

Wodurch wird bestimmt, ob Kontrast- oder Assimilationseffekte auftreten? Assimilation tritt eher auf, wenn Personen Informationen über andere Personen auf einer kategorischen oder stereotypen Ebene einholen. Assimilation tritt hingegen seltener auf, wenn die Informationssuche über eine andere Person gründlich und systematisch (Thompson, Roman, Moskowitz, Chaiken & Bargh, 1994) und akkurat (Stapel, Koomen & Zeelenberg, 1998) erfolgt.

Neben den Kontextvariablen spielen auch noch der kognitive und emotionale Gemütszustand der beurteilenden Person eine wichtige Rolle in der Entstehung sozialer Urteile. Manchmal haben die Eigenschaften, die wir anderen Personen zuschreiben und die uns selbst nicht unbedingt bewusst sind, nichts mit der objektiven Beurteilung dieser Person zu tun. Eine Frau, die beispielsweise ihren Partner verlässt, weil ihr dieser zu kontrollsüchtig war, wird vielleicht alle weiteren Verabredungen nach dem Ausmaß ihrer ausübenden Kontrolle beurteilen. Diese Tendenz, andere Personen durch eine aktivierte Eigenschaft zu beurteilen, kann nur schwer wieder korrigiert werden (Stapel, Koomen & Zeelenberg, 1998).

Chapter 3: Social Cognition: Understanding the Social World (S. 78)

Um soziale Urteile schnell und effizient treffen zu können, verlassen sich Menschen auf bereits existierende Konzepte und „inferential shortcuts“ (= Heuristiken). Die Urteilsfehler, die dabei oftmals entstehen, können problematische Konsequenzen haben. Personen sind weder sehr gut darin, zwischen ihren fehlerhaften und guten Beurteilungen zu unterscheiden, noch sind sie sehr wachsam, selbst wenn man sie darauf hinweist, dass ihre Schlussfolgerungen Denkfehler enthalten könnten (Stapel, Martin & Schwarz, 1998). Menschen in geeigneten Schlussfolgerungsstrategien zu trainieren, wäre eine essentielle Grundlage, um bessere Entscheidungen treffen zu können (Bushman & Wells, 2001).

Chapter 4: The Self: Learning About the Self (S. 122, 123)

Seit der Publikation der sozialen Vergleichstheorie nach Leon Festinger (1954) wissen wir, dass Menschen stets dazu neigen, sich mit anderen zu vergleichen. Seitdem hat sich diese Theorie umfassend weiterentwickelt. Mittlerweile ist bekannt, dass Menschen soziale Vergleiche nicht nur bezüglich ihrer Fähigkeiten und Einstellungen ziehen, sondern auch bezüglich ihrer Emotionen, ihrer Persönlichkeit, ihrem Einkommen, Ansehen und Erfolg. Soziale Vergleiche bestimmen also große Bereiche unseres

Soziallebens und können immer dann ausgelöst werden, sobald das Selbst in den Vordergrund tritt (Stapel & Tesser, 2001).

Vieles der sozialen Vergleichsforschung konzentriert sich darauf, wie Personen sich mit anderen anhand objektiver Charakteristika, die zur Bewertung einladen, vergleichen. Solche Vergleiche resultieren häufig in Kontrasteffekten, wobei sich die Person selbst als verschieden von anderen wahrnimmt (Stapel & Koomen, 2001). Allerdings können Personen auch ihre Gefühle oder Reaktionen mit anderen in Situationen vergleichen, durch die eine gewisse emotionale Verbindung empfunden werden kann (Stapel & Kooman, 2001). Zum Beispiel beobachten Personen, die in einem durch Turbulenzen gebeutelten Flugzeug sitzen, die Reaktionen anderer nicht nur deswegen, um das Ausmaß ihrer eigenen Angstreaktion einzuschätzen, sondern um mit den Mitfliegenden durch ein Gefühl der Solidarität verbunden zu sein, solch ein außergewöhnliches Erlebnis zu teilen.

Chapter 6: Prejudice (S. 173)

Als Voreingenommenheit oder Vorurteil (= „Prejudice“) versteht man die Bewertung einer Gruppe oder eines einzelnen Individuums aufgrund der Gruppenzugehörigkeit dieser Person.

„Most studies show that highly prejudiced people are the most likely to recognize conventional stereotypes“ (Gordijn, Koomen & Stapel, 2001). Es ist offensichtlich, dass die am weitest verbreitetsten Stereotype jedem bekannt sind, egal ob man diesen zustimmt oder nicht. Jedoch sind weniger häufige und tiefer gehende Stereotype gerade den Personen am geläufigsten, die selbst mit Vorurteilen konfrontiert sind.

10. Werth, L. & Mayer, J. (2007). *Sozialpsychologie*. Heidelberg: Springer.

Darin enthaltene Stapel Literatur:

Stapel, D. A. & Koomen, W. (2000). How far do we go beyond the information given? The impact of knowledge activation on interpretation and inference. *Journal of Personality and Social Psychology*, 79, 19-37.

Stapel, D. A. & Koomen, W. (2000). Distinctness of others, mutability of selves: Their impact on self evaluations. *Journal of Personality and Social Psychology*, 79, 1068-1087.

Stapel, D. A., Koomen, W. & van der Pligt, J. (1997). Categories of category accessibility: The impact of trait concept versus exemplar priming on person judgments. *Journal of Experimental Social Psychology*, 33, 47-76.

Stapel, D. A., Reicher, S. D. & Spears, R. (1995). Contextual determinants of strategic choice: some moderators of the availability bias. *European Journal of Social Psychology*, 25, 141-158.

Kapitel 2: Grundlagen sozialer Informationsverarbeitung und sozialen Verhaltens (S. 32)

Nachdem wir einen Reiz wahrgenommen haben, ist der nächste Schritt, ihn zu enkodieren, das heißt ihm eine Bedeutung zu verleihen. Dies beinhaltet, dass ein externer Stimulus in eine interne Repräsentation umgewandelt wird. Eine Bedeutung erhält er dabei dadurch, dass er zu dem, was wir bereits an Wissenstruktur haben, in Beziehung gesetzt wird.

Entscheidend ist, wie zugänglich eine solche bereits bestehende Wissenskategorie ist, denn dies entscheidet darüber, mit welcher Kategorie der neue Reiz am wahrscheinlichsten in Verbindung gebracht wird.

Priming beschreibt eine Voraktivierung durch Reize, die assoziativ mit dem Zielreiz verknüpft sind oder werden. Die Wirkung von Priming unterliegt gewissen Einschränkungen, etwa ob das verfügbar gemachte Konzept semantisch oder evaluativ für den Informationsverarbeitungsprozess anwendbar ist.

Geprierte Inhalte, die eine starke negative oder positive Konnotation beinhalten, wirken sich auf die Anwendbarkeit aus, selbst wenn sie keine eindeutige Schlussfolgerung auf eine bestimmte Charakterisierung ermöglichen (also inhaltlich oder semantisch nicht eindeutig anwendbar sind). Dies ist darauf zurückzuführen, dass eine allgemeine evaluative Valenz (gut vs. schlecht) oder eine starke evaluative Komponente (z.B. süß als etwas sehr Positives oder aggressiv als etwas sehr Negatives), eine valente Reaktion bahnen, so dass die nachfolgende Bewertung einer Zielperson beispielsweise als sympathisch (positive Valenz) bzw. unsympathisch (negative Valenz) ausfällt (Stapel & Koomen, 1999).

Kapitel 3: Heuristiken (S. 64)

Was sind die Faktoren, die unabhängig von der tatsächlichen Auftretenswahrscheinlichkeit die Verfügbarkeit bzw. die Vorstellbarkeit von Beispielen für ein Ereignis beeinflussen und uns damit zu fehlerhaften Urteilen verleiten?

Es ist die häufige („frequency“) oder erst kürzliche („recency“) Aktivierung von Inhalten, die Auffälligkeit („salience“) beziehungsweise die Lebhaftigkeit („vividness“) eines Ereignisses und die Ereignisverknüpfungen im Sinne illusorischer Korrelationen.

Frequency und recency

Informationen sind leichter abrufbar, wenn diese in der Vergangenheit häufig („frequency“) oder erst vor kurzem („recency“) aktiviert wurden. Wird eine bestimmte Information häufig aktiviert, kann das daran liegen, dass das Ereignis tatsächlich häufiger vorkommt. In diesem Fall führt die Anwendung der Verfügbarkeitsheuristik zu korrekten Urteilen. Andererseits kann die häufige Aktivierung auch daher rühren, dass über ein Ereignis beispielsweise überproportional viel in den Medien berichtet wird. Die Leichtigkeit des Abrufs als Urteilsgrundlage kann dann zu einer Überschätzung der tatsächlichen Auftretenshäufigkeit führen (McKelvie, 1995, 2000; Stapel et al., 1995; Tversky & Kahnemann, 1973). Dies demonstriert beispielsweise ein bekanntes Experiment von Tversky & Kahnemann (1973), bei dem berühmte Frauen- und Männernamen in ihrer Häufigkeit überschätzt wurden.

Kapitel 6: Das Selbst - Selbsterkenntnis durch die Beobachtung anderer Personen (S. 184)

Wie unsere (Selbst-) Einschätzung ausfällt und welche emotionalen und motivationalen Konsequenzen dies hat, hängt zum einen in bedeutendem Maße davon ab, mit wem wir uns vergleichen, dem sogenannten Vergleichsstandard. Zum anderen ist entscheidend, welchen Fokus wir während des Vergleichens einnehmen. Letzterer beeinflusst, ob sich die Selbsteinschätzung an den Vergleichsstandard annähert (Assimilation) oder davon entfernt (Kontrast) (z.B. Mussweiler, 2001a, b; Mussweiler et al., 2004a, b; für weitere Faktoren, die Assimilation und Kontrast auslösen können, siehe auch Stapel & Koomen, 2000, Stapel et al., 1997).

6.1.6 Rufschädigung der (Sozial-) Psychologie

Der Schaden für den Ruf der Sozialpsychologie ist ebenso immens hoch. Diederik Stapel wurde jahrelang als „Rising Star“ in den Niederlanden und in der internationalen Wissenschaftswelt gefeiert. Die Universität Amsterdam betitelte ihn als „Golden Boy“ und auch unter seinen Kollegen und Studenten genoss er jahrelang den Ruf eines höchst charismatischen, freundlichen und ehrgeizigen Wissenschaftlers. Seine unglaublich hohe Machtposition, die er als Dekan einer Universität, durch die vielen Publikationen in den renommiertesten Fachzeitschriften, seinen internationalen Kontakten, den gewonnenen Wissenschaftspreisen und erhaltenen Forschungsgeldern inne hatte, war schlussendlich ein ausschlaggebender Grund dafür, warum seine Missachtung der wissenschaftlichen Integrität über Jahre bestehen konnte und warum die Aufdeckung seines Fehlverhaltens

dermaßen hohe Wellen der Kritik und Fassungslosigkeit in der Wissenschaftsgemeinde geschlagen hat.

Vor allem das Feld der Sozialpsychologie wurde seitdem heftig attackiert und in Frage gestellt. Benedict Carey publizierte im November 2011 beispielsweise einen Artikel in der *New York Times* mit dem Titel „Fraud Case Seen as a Red Flag for Psychology Research“, in dem er kritisierte, dass der Fall Stapel nur einer von vielen peinlichen Vorfällen innerhalb einer Wissenschaft sei, welche ihre Forschungsmethoden unbedingt überarbeiten sollte (Carey, 2011). Ähnlich schrieb Hank Campell (2011) in einer Kolumne in *Science* mit dem Titel „Another World-Class Psychology Fraud“, dass Stapel wohl nur einer von vielen Betrugsfällen in der Psychologie sei. Campell stellte darin die Frage, warum Stapel nicht schon viel früher entlarvt wurde und nannte selbst den Grund, dass die Sozialpsychologie, im Gegensatz zu anderen Disziplinen wie die Biologie oder die Physik, wissenschaftlich viel zu „unscharf und schwammig“ ist, um Betrüger schnell überführen zu können. Han Oud, ein niederländischer Statistiker, der für die Universität von Nijmegen arbeitet, schrieb 2011 in einem Artikel der nationalen Zeitung *NRC Handelsblad* mit dem Titel „Fraud is Too Easy in the Social Sciences“, dass die Sozialwissenschaften aufgrund fehlender Replizierbarkeit anfälliger für Betrug sind als die Naturwissenschaften (zitiert nach Strobe et al., 2012).

Diese Reaktionen und noch viele mehr führen zu der Frage, ob wissenschaftlicher Betrug in der (Sozial-) Psychologie tatsächlich häufiger vorkommt als in anderen Wissenschaften. Um diese Forschungsfrage beantworten zu können, muss zunächst geklärt werden, wie oft Betrug in der gesamten Welt der Wissenschaft vorkommt, um Vergleiche zur Psychologie anstellen zu können.

6.2 Forschungsfrage 2: Wie weit verbreitet ist wissenschaftlicher Betrug?

Es gibt keine genauen Informationen zur Prävalenz von wissenschaftlichem Betrug und auch keine Zahl an jährlich aufgedeckten Betrugsfällen. Die Autoren Stroebe et al. (2012) beziehen sich jedoch auf eine Metaanalyse von Fanelli (2009), der Forscher aus 21 Studien anonym befragte, ob sie jemals Forschungsbetrug begangen hätten. Zwei Prozent aller antwortenden Wissenschaftler gab zu, bereits einmal Daten manipuliert oder fabriziert zu haben. Dennoch ist die Information, die man aus dieser Metaanalyse von Fanelli (2009) gewinnen kann, gering, da es sich nicht um eine repräsentative Stichprobe handelt und die Befragung nicht mit einschließt, wie oft die betroffenen Personen in ihren Laufbahnen bereits betrogen haben.

Eine genauere Schätzung der Prävalenz wissenschaftlichen Fehlverhaltens stammt von einer Studie der *Gallup Organization* (2008), welche von der ORI (*U.S. Office of Research Integrity*) gegründet wurde. Anstatt Wissenschaftler zu fragen, ob sie selbst schon einmal Forschungsbetrug begangen hätten, wollten sie wissen, wie oft sie in den letzten drei Jahren Zeuge eines wissenschaftlichen Fehlverhaltens in ihrer eigenen Forschungseinrichtung geworden sind. Dabei befragten die Verantwortlichen jeweils nur einen Wissenschaftler pro Forschungsinstitut, um zu vermeiden, dass ein und derselbe Fall von mehreren Personen genannt werden kann. Die Analyse ergab, dass 1,5% aller Forschungsarbeiten jährlich durch betrügerisches Verhalten zustande kommen.

Die Schätzung der *National Institutes of Health* (NIH) ergab 1993, dass es jährlich zu 2335 Vorfällen möglichen Fehlverhaltens kommt, wobei 60% dieser Vorfälle die Manipulation oder Fabrikation von Daten betrifft.

Grauzonen wissenschaftlichen Fehlverhaltens

Daten bewusst zu manipulieren oder gänzlich zu fabrizieren ist ein klarer Fall von wissenschaftlichem Fehlverhalten, allerdings gibt es auch eine viel größere Grauzone an Verhaltensweisen, die nicht so eindeutig als Betrug zu identifizieren sind. Die Autoren Simmons, Nelson und Simonsohn publizierten 2011 einen Artikel in der Fachzeitschrift *Psychological Science*, in dem demonstriert wird, dass unter Forschern weit verbreitete Methoden, Daten zu analysieren und zu berichten, in einer inakzeptabel hohen Rate dazu führen, dass negative Ergebnisse als positiv valide dargestellt werden. Einfache Methoden, wie das Weglassen von abhängigen Variablen, das Nichtberichten von Untersuchungsbedingungen, das Abändern der Untersuchungsteilnehmer während des Experiments, um signifikante Ergebnisse zu erhalten oder das Ignorieren von Ergebnissen, die seltsam zufällig aussehen, können schnell zum Anstieg sogenannter „False-Positives“ führen.

Eine weitere Studie von John, Loewenstein und Prelec (2012) zeigt eindrucksvoll, dass diese fragwürdigen Forschungsmethoden (= *Questionable Research Practices*, QRP) in den meisten Forschungslabors weit verbreitet sind. Die Autoren befragten 2000 Psychologen in einem anonymen Antwortformat und boten den Studienteilnehmern eine Belohnung für das Erzählen der Wahrheit. Das Bieten eines Anreizes hatte einen positiven Effekt auf die Häufigkeit, in der von fragwürdigen Forschungsmethoden berichtet wurde, insbesondere dann, wenn es sich um Methoden handelte, die die Versuchspersonen selbst als gering vertretbar erachteten. Die Tatsache, dass ein Drittel aller Befragten zugab, QRPs anzuwenden, lässt die Alarmglocken läuten, dass solche weit verbreiteten Forschungsmethoden der Wissenschaft vielleicht sogar mehr Schaden zufügen, als es ein in den Medien auftauchender Betrugsskandal tut.

6.3 Forschungsfrage 3: Ist die Psychologie tatsächlich anfälliger für Betrug als andere Wissenschaften?

Stroebe et al. (2012) untersuchten in ihrer Analyse von bekannt gewordenen Betrugsfällen, ob Diederik Stapel, als Vertreter der Sozialpsychologie, eine Ausnahme für die Wissenschaft der Psychologie darstellt bzw. wie häufig wissenschaftlicher Betrug in anderen Disziplinen vorkommt. Außerdem interessierte sie, durch welche Methoden der jeweilige Betrug aufgedeckt wurde. Für ihre Recherche stützten sich die Autoren auf Publikationen zu Fällen wissenschaftlichen Fehlverhaltens, auf Informationen aus dem Internet und auf den jährlichen ORI Report (Office of Research Integrity, 2001), einer Zusammenfassung öffentlich gewordener Betrugsfälle. Dabei schlossen sie nur solche Fälle in ihre Analyse mit ein, bei denen Forschungsbetrug vorlag (Manipulation oder Fabrikation von Daten), betroffene Publikationen aus den Fachzeitschriften zurückgezogen wurden und es eine Informationsquelle zur Methode der Betrugsentdeckung (zumeist aus Zeitungsartikeln) gab.

Durch die Selektion nach diesen Kriterien resultierten 40 Vergleichsfälle von Forschungsbetrug, wovon zwei Drittel aller Personen aus dem Bereich der biomedizinischen Wissenschaften stammen. Dies lässt auf den ersten Blick vermuten, dass die biomedizinischen Wissenschaften anfälliger für Forschungsbetrug sind als andere Disziplinen. Dennoch kann man das Verhältnis von Betrugsfällen in der Psychologie versus in der Medizin nicht generalisieren, da man nicht die genauen Populationsgrößen der Wissenschaftler in den jeweiligen Disziplinen kennt. Außerdem sind in den von den Autoren aufgelisteten Fällen nur diejenigen enthalten, welche sowohl erfolgreich aufgedeckt wurden als auch mediales Interesse nach sich zogen. Viele Betrugsfälle kommen entweder gar nicht ans Tageslicht, oder falls sie aufgedeckt werden, dann wird diese Information von den beteiligten Universitäten gerne unter Verschluss gehalten, um den guten Ruf nicht zu gefährden. Ein weiterer Grund für die Überrepräsentation der Biomedizin in den Betrugsfällen ist, dass für die Recherche der offizielle jährliche Report der ORI (2001) herangezogen wurde, in dem alle untersuchten Fälle aufgelistet werden, die durch eine staatliche Förderung unterstützt wurden. Da die Medizin für ihre Forschung wesentlich mehr Forschungsgelder erhält als die Psychologie, ist es auch nicht verwunderlich, dass am Ende mehr biomedizinische Projekte von Betrug betroffen sind als psychologische (Stroebe et al., 2012).

Es ist jedoch auch möglich, dass es zwischen verschiedenen Disziplinen einen Unterschied dahingehend gibt, mit welcher Leichtigkeit Betrugsfälle aufgedeckt und bewiesen werden können. In der medizinischen Forschung müssen Projekte von den zuständigen Instituten bewilligt und Patientenaufzeichnungen penibel festgehalten

werden. Alle Untersuchungsmaterialien werden im Anschluss einer Prüfung unterzogen. Für einen betrügerischen Forscher aus der Biomedizin ist es also fast unmöglich, bei bestehendem Verdacht, mit seinem Fehlverhalten ungeschoren davon zu kommen. In der Psychologie ist die Prüfung eines Verdachtsfalles dagegen wesentlich schwieriger. Die Studienteilnehmer sind meistens anonym und im Nachhinein unauffindbar. Die einzige Spur, die bei einem psychologischen Experiment von den Ergebnissen zurückverfolgt werden kann, ist der rohe Datensatz und dieser kann bereits im Vorfeld manipuliert werden.

Warum sollte es in den biomedizinischen Wissenschaften, trotz erhöhter Gefahr der Betrugsaufdeckung, also mehr Vorfälle geben als in der Psychologie? Der Grund dafür könnte sein, dass es in der Medizin häufig hohe monetäre Entlohnungen für gefundene Errungenschaften gibt. Der *New York Times* Journalist Brian Deere (2012, zitiert nach Stroebe et al.) enthüllte beispielsweise, dass Wakefield 600 000 Dollar von Anwälten bezahlt bekam, die eine Gerichtsklage gegen ein Pharmaunternehmen vorbereiteten, welches den Impfstoff gegen Masern, Mumps und Röteln herstellte, welcher laut betrügerischer Forschung von Wakefield mit Autismus bei Kindern in Zusammenhang stehen sollte. Wakefield soll diese Gage bereits zwei Wochen vor Erscheinung seines Artikels im *Lancet* Magazin bekommen haben (Braunstein, 2012, zitiert nach Stroebe et al.). Oder Chandra, der 2001 eine Studie in *Nutrition* über den angeblichen Rückgang des Gedächtnisverlustes bei geriatrischen Patienten mithilfe einer von ihm patentierten Multivitaminformel publizierte, soll durch finanzielle Mittel zur Manipulation seiner Daten motiviert worden sein (Kumar, 2008, zitiert nach Stroebe et al. 2012).

Fanelli (2009) schlussfolgert, dass es keine Hinweise darauf gibt, dass es in der Psychologie zu häufigerem wissenschaftlichem Fehlverhalten kommt als in den biomedizinischen Wissenschaften und dass die meisten Betrugsfälle durch interne Personen (derselben Forschungseinrichtung) aufgedeckt werden.

Uri Simonsohn, ein Sozialpsychologe, der an der *University of Pennsylvania* in Philadelphia arbeitet und bereits Stapels Studien als manipuliert identifizierte, überführte 2012 mittels einer eigenen entwickelten statistischen Methode auch zwei weitere Betrugsfälle aus der Sozialpsychologie. Simonsohns Methode beruht auf der Grundlage, gefälschte Daten an zu großen Ähnlichkeiten mit der postulierten Hypothese und einer zu geringen Differenz von anderen Schätzungen zu erkennen (Yong, 2012b).

Dirk Smeesters von der Erasmus Universität Rotterdam in den Niederlanden, der zuletzt untersuchte, welche Auswirkungen Farben auf unser Konsumverhalten haben, trat im Juni 2012 freiwillig zurück, als ein Untersuchungskomitee seiner Universität, welches Simonsohn zu Rate zog, zu dem Schluss kam, dass er die Datensätze mehrerer Studien

manipuliert hatte. Ein weiterer Sozialpsychologe, Lawrence Sanna von der *University of Michigan* in Ann Arbor, der zuvor auch für die *University of North Carolina* in Chapel Hill arbeitete, kündigte im Mai 2012 seine Dienststelle, nachdem Simonsohn unter anderem in einer medialen Interesse weckenden Studie, in der es darum ging, dass sich Personen altruistischer verhalten, wenn sie andere Personen körperlich überragen, beispielsweise auf einer fahrenden Rolltreppe, Auffälligkeiten in den Ergebnissen als Datenmanipulation identifizierte (Yong, 2012c). Neben den Namen Stapel, Smeesters und Sanna soll es auch noch einen vierten Betrugsfall in der Sozialpsychologie geben, den Simonsohn gerade untersucht, der Öffentlichkeit allerdings noch nicht Preis geben möchte. In einem Interview mit der Zeitschrift *Nature* (Yong, 2012b) wurde Simonsohn die Frage gestellt, ob die drei bekannt gewordenen Betrugsfälle ein Indiz für tiefergehende Probleme des Feldes der Sozialpsychologie seien. Er antwortete darauf:

I don't know how systemic the crime is. What's systemic is the lack of defences. Social psychology -- and science in general -- doesn't have sufficient mechanisms for preventing fraud. I doubt that fabrication is any worse in psychology than in other fields. But I'm worried by how easy it was for me to come across these people.

Auch Chris Chambers, ein Psychologe der *Cardiff University* sieht den Ruf der Sozialpsychologie gefährdet. Alok Jha zitiert ihn in einem Bericht des *Guardians* (2012):

In many respects, psychology is at a crossroads – the decisions we take now will determine whether or not it remains a serious, credible, scientific discipline along with the harder sciences.

6.4 Forschungsfrage 4: Welche Rolle spielt die bestehende Forschungskultur der Sozialpsychologie im Betrugsfall Diederik Stapel?

Die Gründe dafür, warum Diederik Stapels Fehlverhalten jahrelang unentdeckt blieb, sind hauptsächlich auch in seiner Arbeitsumgebung, dem System der (psychologischen) Wissenschaft, zu finden. Das Untersuchungskomitee unter der Leitung von Willem J. M. Levelt widmete der Kritik an der psychologischen Forschungskultur sogar ein eigenes Kapitel (Kap. 5: *Research culture: flawed science*, S. 47-54).

Darin schreiben die Autoren, dass Stapel von einer Kultur umgeben war, die grundsätzlich unachtsam, selektiv und unkritisch im Umgang mit Wissenschaft und Datenmaterialien war und dass die gefundenen Fabrikationsfehler nicht nur kleinere „normale“ Unvollkommenheiten im statistischen Prozess, im experimentellen Design oder in der

Durchführung einer Studie sind, sondern eine Verletzung fundamentaler Regeln angemessener wissenschaftlicher Forschung und eine schwerwiegende Schädigung für die nachfolgende Wissenschaftswelt darstellen.

It involved a more general failure of scientific criticism in the peer community and a research culture that was excessively oriented to uncritical confirmation of one's own ideas and to finding appealing but theoretically superficial ad hoc results [...]. These "sloppy science methods" were found on a large scale in the work of Mr Stapel and his co-workers [...]. (S. 47)

Die grundsätzlichen methodologischen Mängel, die vom Untersuchungskomitee im Fall Stapel festgestellt wurden, warfen ihnen die kritische Frage auf, ob die als „sloppy science“ bezeichneten Methoden auch im nationalen und internationalen Feld der Sozialpsychologie verbreitet sind.

Could it be that in general some aspects of this discipline's customary methods should be deemed incorrect from the perspective of academic standards and scientific integrity? (S. 47)

Die Autoren sind sich bewusst, dass es sich bei Stapels Publikationen nicht um eine zufällig gezogene Stichprobe handelt, die man als repräsentativ für die sozialpsychologische Wissenschaft ansehen kann. Das Komitee betont daher, dass es keine Aussagen zur Sozialpsychologie als Ganzes machen kann. Dennoch können sie das Ausmaß des vorliegenden Betrugs nicht als seltene Anomalie betrachten, da Stapel an zu vielen Orten, in zu vielen unterschiedlichen Funktionen und mit zu vielen machtvollen Personen zusammen gearbeitet hat, um von dieser Annahme ausgehen zu können. Basierend auf dieser Tatsache, zwingt sich den Autoren das Bild einer internationalen Forschungsgemeinde auf, in der herkömmliche Forschungsmethoden und damit assoziierte Werte einvernehmlich geteilt werden.

Another clear sign is that when interviewed, several co-authors who did perform the analyses themselves, and were not all from Stapel's 'school', defended the serious and less serious violations of proper scientific method with the words: „That is what I have learned in practice - everyone in my research environment does the same, and so does everyone we talk to at international conferences.“ (S. 48)

Die sich aufdrängende Frage, warum der Betrug und die massive Verletzung der wissenschaftlichen Methoden nicht durch den „normalen“ Evaluationsprozess in der Wissenschaft aufgedeckt wurde, führen die Autoren auf ein gänzliches Scheitern der wissenschaftlichen Kritikfähigkeit zurück. Für die Verantwortlichen des Untersuchungskomitees scheint es schier unfassbar, dass Co-Autoren, die sich intensiv

mit der Datenanalyse beschäftigten und *Reviewer* der international führenden Fachzeitschriften, die als absolute Experten auf ihrem Wissensgebiet bezeichnet werden, daran gescheitert sind, in der Praxis nicht durchführbare Experimente als solche zu identifizieren und unmöglich anmutende statistische Ergebnisse nicht als Warnsignal für Forschungsbetrug zu erkennen.

Die Autoren schlussfolgern daraus:

A 'byproduct' of the Committees' inquiries is the conclusion that, far more than was originally assumed, there are certain aspects of the discipline itself that should be deemed undesirable or even incorrect from the perspective of academic standards and scientific integrity [...]. In the case of the fraud committed by Mr Stapel, the critical function of science has failed on all levels. (S. 54)

Die harsche Kritik der Stapel Investigation an der grundlegenden Forschungskultur der Sozialpsychologie blieb nicht lange unbeantwortet.

Das *Executive Committee* der *European Association of Social Psychology* (EASP) verfasste im Dezember 2012 ein Schreiben, in dem es die Vorwürfe einer Disziplin mit niedrigen akademischen Standards und einer eingeschränkten wissenschaftlichen Integrität vehement von sich wick (The Executive Committee of the European Association of Social Psychology, 2012). Die EASP schätze die mühevollen Analyse Stapels Forschungsarbeit und die Informationen, die aus den ersten vier Kapiteln des Berichts gewonnen werden können zwar sehr, bezeichnet die beiden Kapitel fünf und sechs, in denen die Methodologie der Sozialpsychologie kritisiert wird, als „defamatory attempt to present social psychology as a discipline that has facilitated, through its flawed research culture, Stapel's scientific misconduct“. Weiters kreidet die EASP den Autoren des Endberichts an, dass sie sich in ihren Aussagen wiederholt widersprechen. Obwohl Levelt und seine Kollegen schreiben, dass sie nicht über die Sozialpsychologie als Ganzes urteilen wollen („The Committees are therefore unable to make any statement on these grounds about social psychology as a whole.“ S. 48), betonen sie immer wieder, dass es sich bei Stapels wissenschaftlichem Fehlverhalten nicht um einen Einzelfall handelt, sondern um Praktiken, die in der sozialpsychologischen Wissenschaftsgemeinde weit verbreitet sind („It would nonetheless be simplistic to dismiss the findings given below as merely a local aberration. Mr Stapel worked in too many different places and in too many different capacities to support that view.“ S. 48).

Um überhaupt Generalisierungen auf das gesamte zugrunde liegende Forschungsfeld der Sozialpsychologie treffen zu können, empfiehlt die EASP den Autoren, zunächst Vergleiche zu anderen wissenschaftlichen Sub-Disziplinen der Psychologie anzustellen. Bezüglich der Kritik, dass die (Sozial-) Psychologie aufgrund ihrer „sloppy science“

anfälliger für Verletzungen wissenschaftlicher Integrität sei und dem Umstand, dass Stapels Betrug jahrelang nicht durch das bestehende *Review-System* aufgedeckt wurde, antwortet die EASP mit der bereits erwähnten Publikation von Stroebe et al. (2012), in der aufgezeigt werden konnte, dass Forschungsbetrug in der Psychologie nicht häufiger auftritt als in anderen Wissenschaften und dass die Schwächen des *Peer-Reviewings* nicht nur für die Psychologie, sondern auch für alle anderen Wissenschaften gleichermaßen bestehen. Zuletzt betont das *Executive Committee* (2012):

Likewise, the report misses the opportunity to provide inspiring, not condescending, guidelines and hope to a whole generation of young scientists in social psychology. Their confidence in their discipline has been battered by the "Stapel case" and it needs to be repaired through the provision of considered guidance, not rash generalization. [...] As a community, we repudiate 'sloppy science' in all its forms, and, for that reason, we find the conclusions about social psychology contained in the final Levelt report to be unacceptably flawed.

Auch Stephen Gibson, ein Mitglied der *British Psychological Society's* (BPS) *Social Psychology Section*, stimmt im Namen seiner Organisation vollkommen mit der hervorgebrachten Kritik der EASP überein (Gibson, 2013). Er ist sich dessen bewusst, dass nach Stapels Skandal Handlungen folgen müssen, wie etwa den Druck von Wissenschaftlern zu nehmen, ihre Forschungsergebnisse unbedingt in top Journals publizieren zu müssen oder der Replikation einen höheren Stellenwert einzuräumen. Gibson verweist dabei auf die Arbeiten der Soziologen Harry Collins, Bruno Latour und anderen, die bereits dargestellt haben, dass wissenschaftliches Fehlverhalten in allen Disziplinen anzufinden ist und dass es nicht Aufgabe sein sollte, „schlechte Wissenschaft“ zu identifizieren, sondern stattdessen die Bemühungen der Wissenschaft als soziales System zu fördern. Eine Subdisziplin habe es nämlich nicht verdient, auf Basis einer einzigartigen Mangelerscheinung ihren Ruf gänzlich zu verlieren.

Die Kritik der EASP und der BPS *Social Psychology Section* wurde von den Autoren des Stapel Komitees, Drenth, Levelt und Noort, aufgenommen und in einem Schreiben, das 2013 im *Psychologist* veröffentlicht wurde, beantwortet (Drenth, Levelt & Noort, 2013). Darin betonen die drei nochmals, dass sie Stapels Publikationen nicht als repräsentativ für das Feld der Sozialpsychologie erachten, dass sie deswegen keine grundsätzlichen Schlüsse auf die Sozialpsychologie ziehen möchten und dass sie nicht davon ausgehen, dass schlechte Forschungspraktiken in dieser Disziplin allgemein gehandhabt werden. Allerdings stehen sie weiterhin zu ihrer Kritik, dass es der Wissenschaft so lange nicht möglich war, Stapels Betrug zu erkennen, weswegen sie sich nicht der Frage verwehren können, ob dafür nicht die wissenschaftlichen Praktiken in der Sozialpsychologie zur Verantwortung zu ziehen sind.

[...] The remarkable failure of the relevant national and international peer community to discern, over such a long period, the manifest negligence of scientific standards in the journal publications co-authored by Stapel impelled the Committees to address a more general issue. Could it be that customary procedures of enforcing and monitoring sound scientific practice in social psychology are in need of improvement?

Dass sie mit dieser Meinung in der Wissenschaftswelt nicht allein sind, argumentieren die Autoren damit, dass die Zeitschrift *Perspectives in Psychological Science* im November 2012 einer ganzen Ausgabe mit insgesamt 23 Artikeln dem Thema gewidmet hat, auf welche Weise bestehende wissenschaftliche Praktiken verbessert werden können, unter anderem durch die vermehrte Replikation von Studien (Pashler & Wagenmakers, 2012). In ihrem Vorwort stellen die Herausgeber Harold Pashler und Eric-Jan Wagenmakers die einleitende Frage:

Is there currently a crisis of confidence in psychological science reflecting an unprecedented level of doubt among practitioners about the reliability of research findings in the field? It would certainly appear that there is.

Auf den von der EASP hervor gebrachten Kritikpunkt, dass das Komitee zuerst die Häufigkeit wissenschaftlichen Fehlverhaltes in anderen Subdisziplinen der Psychologie untersuchen solle, bevor sie Rückschlüsse auf die Sozialpsychologie ziehen kann, entgegnet Drenth, Levelt und Noort, dass es nie im Sinn ihrer Untersuchung war, Vergleiche zu anderen Disziplinen anzustellen und falls man dies auf Basis bestehender Literatur tun würde, dann würde sich selbstverständlich zeigen, dass mangelhafte Forschungsmethoden nicht nur in der Sozialpsychologie zu finden sind.

6.4.1 Der Mythos der sich selbst korrigierenden Natur der Wissenschaft

Anknüpfend an die Forschungsfrage, welche Rolle das System der Wissenschaft im Betrugsfall Stapel spielte, müssen zwei entscheidende Mechanismen diskutiert werden, aufgrund derer die Verletzung der wissenschaftlichen Integrität jahrelang unentdeckt bleiben konnte und welche den Schlüssel für spätere Überlegungen zur Verbesserung des Systems im Sinne einer erhöhten Wahrscheinlichkeit der Betrugaufdeckung darstellen.

Philip Handler, der Präsident der *National Academy of Science* erklärte 1981, dass Wissenschaftsbetrug selten vor kommt und wenn er vor kommt, dann „in a system that

operates in an effective, democratic and self-correcting mode“, welches die Aufdeckung unvermeidbar macht (Broad & Wade, 1982, zitiert nach Stroebe et al., 2012).

Diese Mechanismen der Selbstkorrektur gelten heute als *Peer-Review* System und Replikation. Dieser Hypothese folgend, schrieb Kraut (2011) einen Artikel mit dem Titel „Despite Occasional Scandals, Science Can Police Itself“. Crocker und Cooper (2011) meinten in einem Bericht in *Science* über die Replikation von Daten: „Scientists generally trust that fabrication will be uncovered when other scientists cannot replicate findings.“ Und auch Henry Roediger (2012) ist der Meinung, dass Stapels Betrugsaffäre bereits viel früher aufgedeckt hätte werden können, wenn andere Wissenschaftler versucht hätten, seine Ergebnisse zu replizieren.

Deshalb erscheint es umso merkwürdiger, dass die meisten Betrugsfälle, auch der von Diederik Stapel, eben nicht durch die zwei Selbst-Korrekturmechanismen des *Peer-Reviewings* oder der Replikation aufgedeckt werden, sondern durch interne Verdachtsschöpfungen von Studenten, Kollegen und Mitarbeitern.

6.4.1.1 Forschungsfrage 4.1.: Wie gut funktioniert der „Peer-Review“ Prozess um Betrugsfälle in der Praxis aufzudecken?

Der Hintergedanke des *Peer-Review* Prozesses ist es, Herausgeber von Fachzeitschriften darin zu unterstützen, qualitativ gute Forschung zu identifizieren, die den wissenschaftlichen Fortschritt vorantreiben. Die Aufdeckung von Betrugsfällen ist zwar eigentlich nicht das Hauptziel des *Peer-Reviewings*, da betrügerische Forschung den wissenschaftlichen Fortschritt jedoch massiv behindern kann, sollte es selbstverständlich sein, dass die beauftragten Kritiker Anzeichen eines möglichen Betrugs bei der Evaluation eines Artikels feststellen können. Das System des *Peer Reviewings* wurde bereits häufig kritisiert, da die geringe *Interrater-Reliabilität* mehrerer *Reviewer* eines Manuskripts auf eine schwache Wirksamkeit in der Identifikation guter wissenschaftlicher Forschung hindeutet. Es scheint daher nicht sehr überraschend, dass dieses System weniger gut in der Lage ist, wissenschaftlichen Betrug aufzudecken. Sogar in den top Journals erscheinen immer wieder Artikel, in denen *Reviewer* beinahe unfassbare Inkonsistenzen übersehen. Eine Publikation von Darsee und Heymsfield (1981) im *New England Journal of Medicine* bildet beispielsweise einen Familienstammbaum ab, der einen 17-jährigen Vater mit Kindern im Alter von 8, 7, 5 und 4 zeigt (Diekmann, 2004, zitiert nach Stroebe et al., 2012). Auch im Fall Stapel übersahen die zuständigen *Reviewer* und Journal Herausgeber kuriose Kleinigkeiten und zögerten nicht, diese Artikel zu publizieren.

Stroebe et al. (2012) führen mehrere Punkte als Erklärung dafür an, warum das *Peer-Review System* im Fall Stapel versagt hat:

a) (Bekannter) Forschungsbetrug ist im Vergleich zur Population aller publizierten Forschungsbeiträge immer noch sehr selten. *Reviewer* sehen ihre Hauptaufgabe nicht darin, Betrug aufzudecken, während sie ein Manuskript lesen. Sie wundern sich zwar oder sind erstaunt über starke Effektgrößen einer experimentellen Manipulation, dennoch ist dies kein Grund, ein Manuskript abzuweisen. Wenn das die grundsätzliche Vorgehensweise wäre, dann würden neue überraschende Ergebnisse niemals publiziert werden und der Wissenschaftsfortschritt würde nicht vorankommen.

b) Die Abwesenheit vergleichbarer Studien macht es schwierig, Aussagen über die Angemessenheit einer Effektgröße zu treffen.

c) Im Vergleich zu Untersuchungskomitees, die in Studien spezifisch nach bestimmten Mustern suchen, die Hinweise auf manipulierte Daten liefern könnten, denken *Reviewer* und Herausgeber im Vorfeld nicht daran, dass Betrug eine mögliche Erklärung für herausragende Effekte, kleine Fehler oder Widersprüche sein könnte. Dabei spielt der wohl bekannte „Confirmation Bias“ eine wichtige Rolle, ein Fehler in der Informationssuche und -verarbeitung, der dafür verantwortlich ist, dass man einerseits gezielt nach Informationen sucht, welche die im Vorfeld formulierten Hypothesen bestätigen (z.B. ein Untersuchungskomitee, das gezielt nach Hinweisen für einen vorliegenden Forschungsbetrug sucht), der andererseits aber auch zum genauen Gegenteil führen kann, dem klaren Übersehen von Anzeichen für Betrug, wenn zuvor keine Annahmen darüber verfasst wurden. Sich die Möglichkeit des schlimmsten Falles, wissenschaftliches Fehlverhalten, als Erklärung für Ungereimtheiten ins Bewusstsein zu rufen und bei der Evaluation von Manuskripten im Hinterkopf zu behalten, wäre also bereits ein wichtiger Schritt, um Betrug als *Reviewer* aufdecken zu können.

d) Neben dem „Confirmation-Bias“ spielen noch andere soziale Motive beim *Reviewing* Prozess eine wichtige Rolle. Der „In-Group-Bias“ geht davon aus, dass Mitgliedern der eigenen Gruppe (in diesem Fall Wissenschaftler) als vertrauenswürdiger erachtet werden, als Personen einer externen Gruppe (z.B. nicht-Wissenschaftler). Gruppennormen, welche vorschreiben, wie sich eine Gruppe in bestimmten Situationen zu verhalten hat (z.B. „Betrug ist asozial“), führen dazu, dass ein nicht normgerechtes Verhalten nicht nur als unwahrscheinlich, sondern sogar als undenkbar erachtet wird. In einer Situation, in der *Injunctive Norms* (=anerkannte Regeln einer Gruppe, z.B. „Betrug ist asozial“) auf deskriptive Normen (=tatsächlich gezeigtes Verhalten, z.B. „Betrug ist selten“) stoßen und das beobachtete Verhalten nicht offen zur Sprache gebracht werden kann (z.B. „Betrug ist ein Tabu-Thema“), ist es sehr schwierig, dass ernsthafte Überlegungen zu einem

möglichen Betrugsfall an die Bewusstseinsoberfläche des *Reviewers* gelangen oder überhaupt an eine entscheidungsträchtige Vertrauensperson weitergegeben werden.

e) Ein weiterer Grund, warum *Reviewer* beinahe nie einen Betrug aufdecken, ist, dass die betroffenen Forscher ihre manipulierten Ergebnisse auf der Basis bereits bestehender Forschung sehr plausibel aussehen lassen. Die meisten Effekte, die Stapel berichtete, hörten sich so nachvollziehbar an, dass einige Sozialpsychologen davon ausgehen, dass sie sich wahrscheinlich sogar bestätigt hätten, wenn er sie auf ehrliche Weise getestet hätte.

f) Die wissenschaftliche Integrität verletzende Forscher haben oftmals hoch angesehene Wissenschaftler als Co-Autoren, was den Gedanken an einen möglichen Betrug umso unwahrscheinlicher macht.

6.4.1.2 Forschungsfrage 4.2.: Welche Rolle spielt die Replikation in der Aufdeckung von Betrug in der Praxis?

Obwohl es in der psychologischen Forschung mehrere Gründe gibt, warum ein bestimmtes Forschungsergebnis nicht repliziert werden kann, kann das kontinuierliche Scheitern, mehrere Forschungsergebnisse ein und desselben Wissenschaftlers zu replizieren, ein Warnsignal für Betrug sein. Es ist daher berechtigt, dass die Replikation als mächtige Methode angesehen wird, um Wissenschaftsbetrug aufzudecken. Dennoch ist die Wirksamkeit dieser Selbstkorrekturmaßnahme der Wissenschaft, ähnlich wie beim *Peer Review* Prozess, stark eingeschränkt.

Ein Grund dafür ist, dass es für Wissenschaftler wenig Anreize gibt, eine Studie zu replizieren. Gleichgültig ob Replikationen gelingen oder nicht, sie sind extrem schwer zu publizieren und gelangen grundsätzlich nicht in Fachzeitschriften und somit nicht ins Bewusstsein der Wissenschaftsgemeinde.

Weiters ist ein erfolgloser Replikationsversuch nicht besonders informativ, da es unzählige Gründe gibt, warum ein Untersuchungsergebnis nicht repliziert werden kann. Sei es weil die Untersuchungsbedingungen nicht exakt abgebildet wurden, weil die Experimentatoren Fehler gemacht haben, weil in der Originalstudie nicht alle notwendigen theoretischen Variablen angeführt wurden, welche am Effekt beteiligt sind etc. Betrug ist nur eine von vielen möglichen Erklärungen.

Basierend auf der Tatsache, dass einzelne gescheiterte Replikationsversuche keinen hohen Informationswert bieten, könnte man davon ausgehen, dass Meta-Analysen, die multiple Replikationen statistisch zusammenfassen, eine größere Rolle bei der

Aufdeckung von Betrug spielen. Die Meta-Analyse von Kranke, Apfel, Eberhart, Georgieff und Roewer (2001, zitiert nach Stroebe et al. 2012) zeigte zum Beispiel deutlich, dass die von Fujii publizierten Ergebnisse signifikant von allen anderen Studien abwichen. Die genannten Autoren beschuldigten Fujii zwar nicht direkt des Betrugs, allerdings kann ihre Schlussfolgerung zwischen den Zeilen wohl verstanden werden:

Further, if data of a dominating centre do not appear to be valid for other centres, it may seem advisable to either exclude them from the analysis or perform sub-group analyses so that results without the data from the dominating centre were available.

Die Vorstellung einer sich selbstkorrigierenden Natur der Wissenschaft scheint also nur ein Mythos zu sein.

Ein Hauptgrund dafür ist, dass Wissenschaft auf Vertrauen basiert. Wie bereits erwähnt, erwarten Wissenschaftler nicht, dass ihre Kollegen in Forschungsbetrug verwickelt sind und suchen daher nicht gezielt nach Anzeichen dafür, wenn sie ihre Arbeiten studieren. Das ist auch der Grund, warum das *Peer-Review* System in erster Linie nicht dazu taugt, Betrug aufzudecken. Außerdem müssten die Effekte einer einzelnen Studie schon „too good to be true“ sein, um überhaupt das Misstrauen der Leser zu wecken. Hilfreicher wäre es, sich mehrere Forschungsarbeiten eines Wissenschaftlers dahingehend anzusehen, ob sich die entdeckten Ungereimtheiten einer Studie auch in weiteren finden lassen.

Ein weiterer Hauptgrund für die Bezeichnung der wissenschaftlichen Natur als „Mythos“ ist, dass die Beschuldigung des wissenschaftlichen Betrugs einer Person so schwerwiegend ist, dass sie nicht ohne stichhaltige Beweise getroffen werden kann. Solange man nur den unüberprüften Verdacht hat, dass ein Kollege wissenschaftliche Normen verletzt, scheinen die möglichen Konsequenzen für die betroffene Person und einen selbst viel zu schwerwiegend, um mit der Vermutung an die Öffentlichkeit zu gehen. Und selbst wenn man sich ohne Zweifel sicher sein kann, dass bei einer vorliegenden Studie betrogen wurde, ist es bei einer Forschungsarbeit mit mehreren Autoren schwierig, die Person oder Personen zu ermitteln, welche maßgeblich für den Betrug verantwortlich sind.

Ausgelöst durch den Fall Diederik Stapel, hat es in den letzten 1-2 Jahren viele Diskussionen zum Thema Forschungsethik und Integrität gegeben. Das Hauptaugenmerk aller Autoren, die sich um eine Reaktion bemüht haben, liegt auf der Identifizierung von Maßnahmen, die dabei helfen sollen, das Risiko für Betrug und andere Formen wissenschaftlichen Fehlverhaltens in der Psychologie und in anderen Wissenschaften einzuschränken. In der Welt der Wissenschaft tut sich derzeit viel, um den Ruf einer

vertrauenswürdigen (psychologischen) Disziplin wieder herzustellen, indem versucht wird, die Grundlagen der Wissenschaftspolitik zu überdenken und ihre Methoden zu korrigieren.

7 Forschungsfrage 5: Welche Maßnahmen können getroffen werden, um das Risiko für wissenschaftlichen Betrug in Zukunft zu reduzieren?

7.1 Institutionelle Veränderungen

Bevor darauf eingegangen wird, welche Maßnahmen getroffen werden können, um wissenschaftlichen Betrug einzuschränken, muss zunächst festgehalten werden, dass in den letzten Jahrzehnten bereits viel passiert ist, um Betrugsfällen in der Wissenschaft entgegenzuwirken.

Wissenschaftliches Fehlverhalten war für die Öffentlichkeit nicht wirklich von Interesse, bevor die ersten Fälle in den späten 1970ern und frühen 1980ern ans Tageslicht kamen (Gallup Organization, 2008). Da Forschungsbetrug zumeist auf Universitäten zurückfällt, an denen die Wissenschaftler beschäftigt sind, waren betroffene Institutionen in der Vergangenheit eher zurückhaltend, wenn es darum ging, konkrete Handlungen bei einem bestehenden Verdacht einzuleiten.

Seitdem hat sich die Situation, zumindest in den USA, dramatisch verändert, in denen eigene Institute an Universitäten und auch auf staatlicher Ebene geschaffen wurden, um Verdachtsfällen des Wissenschaftsbetrugs nachzugehen. An vielen Universitäten gibt es Forschungsintegritäts-Komitees, an die sich jede Person wenden kann, die jemand des Betrugs verdächtigt. Diese Vermutungen werden zunächst vom Institut geprüft, an dem der wissenschaftliche Betrug begangen worden sein soll und diese Untersuchung wird wiederum von der ORI beaufsichtigt und geprüft.

Nach dem Wissen der Autoren Stroebe et al. (2012) gibt es keinerlei vergleichbare Einrichtungen in Europa. In den Niederlanden haben die meisten Universitäten Komitees oder bestimmte „Vertrauenspersonen“, an die man sich mit einem Verdacht wenden kann. An der Tilburg Universität, an der Stapel beschäftigt war, war der Rektor selbst diese Vertrauensperson, was eine umso höhere Schwelle für die Studenten darstellte, um ihre Vermutung zu äußern. In Deutschland wurde von der Deutschen Forschungsgemeinschaft (DFG) der „Ombudsman für die Wissenschaft“ eingerichtet, welcher Wissenschaftlern unmittelbar und unabhängig von einer Beteiligung der DFG zur Beratung und Unterstützung in Fragen guter wissenschaftlicher Praxis und ihrer möglichen Verletzung durch wissenschaftliche Unredlichkeit zur Verfügung steht. Dies betrifft insbesondere auch die Beratung von Personen, die einen Hinweis auf wissenschaftliches Fehlverhalten zur Kenntnis gebracht haben („Whistleblower“) und dadurch Sanktionen anderer Einrichtungen ausgesetzt sind. Alle dem Ombudsman

vorgetragenen Anliegen werden grundsätzlich strikt vertraulich behandelt. Personen, Institutionen oder Leitungsgremien, die nicht direkt betroffen sind, werden nicht informiert. Die Gewährleistung der Vertraulichkeit dient dem Schutz aller in ein Verfahren involvierten Personen und sie gilt auch über den Abschluss eines Falles hinaus (<http://www.ombudsman-fuer-die-wissenschaft.de/>).

Stapels Fall wurde von einer „ad hoc Kommission“ untersucht, die nicht Teil der geldgebenden Institutionen NWO und der *Royal Netherlands Academy of Arts and Sciences* war.

Eine weitere positive Entwicklung ist, dass die großen europäischen und amerikanischen Fachzeitschriften 1997 gemeinsam das *Committee on Publication Ethics* (COPE) gegründet haben, welches allen Herausgebern und Verlagen mit Ratschlägen zum Thema wissenschaftliches Fehlverhalten zur Verfügung steht. Der Verhaltenskodex für alle Herausgeber eines Journals lautet beispielsweise:

Editors have a duty to act if they suspect misconduct or if the allegation of misconduct is brought to them. Editors should not simply reject papers that raise concerns about possible misconduct. They are ethically obliged to pursue alleged cases. Editors should first seek a response from those suspected of misconduct. If they are not satisfied with the responses, they should ask the relevant employers, or institutions, or some appropriate body [...] to investigate. (COPE, 2012, zitiert nach Stroebe et al., 2012)

Im September 2011, im selben Monat als Stapels Betrug entdeckt wurde, publizierte die APA (*American Psychological Association*) ein „Handbook of Ethics in Psychology“. Und noch im selben Jahr brachten die *All European Academies* (ALLEA) gemeinsam mit der *European Science Foundation* (ESF) den „Code of Conduct for Scientific Integrity“ heraus, der sich nicht nur auf die Psychologie sondern auf alle Wissenschaften bezieht (Van Lange, Buunk, Ellemers & Wigboldus, 2012).

7.2 Konkrete Maßnahmen nach Van Lange et al. 2012

Ausgelöst durch den Betrugsfall Diederik Stapel, versuchen die Autoren Van Lange et al. (2012) konkrete Schritte zu identifizieren, die dabei helfen sollen, dass Risiko für Betrug und andere Formen wissenschaftlichen Fehlverhaltens in der Psychologie und in anderen Wissenschaften zu reduzieren.

Für die Identifikation von Maßnahmen muss zunächst berücksichtigt werden, dass die Sozialpsychologie nicht auf einem einzigen Wissenschaftsparadigma oder einer alleinigen Forschungsmethode beruht. Obwohl der Forschungsfokus auf der experimentellen Psychologie (meistens im Labor) liegt, können abhängige Variablen auf multi-methodische Weise gemessen werden. Dabei können das menschlichen Verhalten, Gedanken, Entscheidungsprozesse, affektive Prozesse, psycho-physiologische Messungen bis hin zu neurologischen Prozessen erhoben werden. Das Feld der Sozialpsychologie ist zudem ein internationales und immer mehr wachsendes interdisziplinäres, mit Theorien und Methoden, die in der Kognitionsforschung, der Kommunikationsforschung, der Konsumentenforschung, der Forschung von Entscheidungsprozessen, Ökonomien, Evolutionstheorien, in den Gesundheitswissenschaften, in der Unternehmensführung und in den Neurowissenschaften angesiedelt sind.

Wenn man die Interdisziplinarität und internationale Orientierung der Sozialpsychologie und anderen Feldern der Psychologie berücksichtigt, wo unterschiedliche Universitäten unterschiedliche Normen, Strukturen und Gesetze haben, dann können folgende konkrete Empfehlungen in ihrer praktischen Anwendbarkeit womöglich eingeschränkt sein. Dennoch wollen die Autoren mit den aufgeführten Punkten einen Beitrag zur weiteren Diskussion über die Forschungsethik und Integrität liefern und die Wissenschaft dazu anregen, ihre Politik in Bezug auf wissenschaftliches Fehlverhalten und Betrug zu verschärfen:

7.2.1 Niederschwellige Ethikabteilung

Das Bewusstsein um das Vorhandensein und die Erreichbarkeit einer Abteilung für Ethik in einer Universitätsfakultät, die im Falle einer Vermutung oder Vorliegens von Beweisen wissenschaftlichen Fehlverhaltens vertraulich aufgesucht werden kann, sollte vergrößert werden. Bei der Enthüllung eines Betrugs, wie es auch bei Stapel der Fall war, ist es nicht unüblich, dass sich Forschungsmitarbeiter oder Doktorats-Studenten des verdächtigten Professors an eine höhere Autorität wenden. Daher sollte diese Autoritätsperson für Studenten, aber auch für Personen höherer Positionen oder externe Institutionen im In-

und Ausland, mit denen die verdächtige Person eventuell zusammen gearbeitet hat, leicht erreichbar sein. Außerdem sollten für alle folgenden Handlungen der Ethik Abteilung präzise Verhaltensregeln existieren, die den anklagenden Personen ein Gefühl der Sicherheit vermitteln, dass ihrem Verdacht auch wirklich nachgegangen wird und vor allem in welcher Art und Weise dies geschieht.

7.2.2 Datenspeicherung und –veröffentlichung

In jedem Institut, an dem ein bestimmtes Forschungsprojekt bearbeitet wird, sollten Möglichkeiten zur Datenspeicherung zur Verfügung gestellt werden, die dann allen Mitgliedern eines definierten Forschungsteams zugänglich sind. Eine weitere Empfehlung der Autoren lautet, dass es zukünftig übliche Praxis sein sollte, dass die Autoren einer Studie nicht nur die experimentellen Materialien ihres Versuchsdesigns beschreiben, sondern auch den rohen Datensatz, auf dem der Artikel beruht, öffentlich machen, genauso wie eine detaillierte Erläuterung der verwendeten statistischen Methoden und Analysen. Durch diese Zusammenarbeit soll sichergestellt werden, dass zumindest immer zwei verschiedene Mitglieder eines Forschungsteams in eine gemeinsame Aufgabe involviert sind (Datensammlung, Dateneingabe, Datenanalyse, Datenmanagement, Formulieren der Ergebnisse etc.). Auf diese Weise sollen Fehler und das Risiko wissenschaftlichen Fehlverhaltens einer Person, die ansonsten allein die Verantwortung für einen Bereich tragen würde, reduziert werden.

7.2.3 Namentliche Kennzeichnung zuständiger Bereiche

Basierend auf der vorherigen Empfehlung, sollten die Beiträge einzelner Forschungsmitarbeiter zu verschiedenen Bereichen eines Artikels namentlich explizit gekennzeichnet werden (z.B. wer für die Datenerhebung, die statistische Analyse, Graphiken, den Text etc. verantwortlich war). Bei aufkommenden Fehlern oder Inkonsistenzen in einem bestimmten Bereich können die verantwortlichen Personen gezielt befragt werden. Einige Journals haben diese Strategie bereits umgesetzt und die Autoren hoffen darauf, dass es noch viele mehr folgen werden.

7.2.4 Kommunikation mit den Medien

In Bezug auf die Kommunikation von Forschungsergebnissen mit den Medien sollten sich Wissenschaftler an genau formulierte Integritätsnormen halten. Die Autoren empfehlen, dass Forscher nicht die Presse kontaktieren sollten, um ihre Ergebnisse der breiten Öffentlichkeit zu präsentieren, bevor das Manuskript nicht erfolgreich durch das *Peer-Review* System gelaufen ist und vom jeweiligen Journal die Zusage für die Publikation in

einer der nächsten Ausgaben erhalten hat. Falls dies nicht machbar ist (beispielsweise wenn ein Journalist bei einer Pressekonferenz Informationen durchsickern lässt), dann sollte die betroffene Person zumindest explizit darauf hinweisen, dass es sich bei den vorliegenden Informationen um unveröffentlichte Ergebnisse handelt. Diese Empfehlung beruht auf der Tatsache, dass Diederik Stapel den niederländischen Medien eine Studie präsentierte (worin es darum ging, dass Personen, die auf das Essen von Fleisch geprimt wurden, in einem nachfolgenden Experiment eigennütziger Motive verfolgten als die Kontrollgruppe), die von der umworbenen Fachzeitschrift noch nicht zur Publikation akzeptiert worden war. Der Informationsvorschuss sorgte schlussendlich nicht nur für Verwirrung, sondern trübte auch das öffentliche Vertrauen in die psychologische Forschung, als ans Tageslicht kam, dass Stapel auch in dieser Studie den Datensatz manipuliert hatte. Obwohl das *Peer-Review* System, wie bereits erwähnt, keine Garantie für wissenschaftliche Qualität darstellt, ist es immer noch die letzte Instanz einer grundlegenden Prüfung, bevor eine Studie veröffentlicht werden kann.

7.2.5 Deskriptiver Statistik höhere Aufmerksamkeit schenken

Reviewer von Manuskripten, die von Wissenschaftlern zur Publikation eingereicht werden, sollten den deskriptiven Eigenschaften und Vorgehensweisen einer Studie kritischere Aufmerksamkeit schenken, wie etwa der Stichprobenzusammensetzung, der Rekrutierung von Versuchsteilnehmern, dem Umgang mit fehlenden Daten etc. Stapel beging beispielsweise etliche Fehler, indem er vergaß, die Schule zu benennen, in der er angeblich seine Daten gesammelt hatte oder überhaupt das Fehlen einer Angabe von Informationen zum gesamten Datenerhebungsprozess. Natürlich gibt es Ausnahmen, bei denen auf detaillierte Informationen zur Rekrutierung von Versuchspersonen verzichtet werden muss (z.B. wenn eine gesamte Firma, Organisation oder eine bestimmte Patientengruppe erhoben wird, aufgrund deren Nennung auf einzelne Personen rückgeschlossen werden könnte), um die Privatsphäre der Studienteilnehmer nicht zu gefährden. Doch selbst wenn eine Geheimhaltungspflicht besteht, so kann der Autor dies explizit in seinem Artikel erwähnen oder betroffene Studienteilnehmer fragen, ob sie der Weitergabe ihrer Daten einzig an den Herausgeber der Fachzeitschrift zustimmen würden.

7.2.6 Replikationsförderung

Zuletzt sollte das wiederholte Scheitern eines Replikationsversuchs einer Studie im Sinne der Qualitätssicherung guter wissenschaftlicher Forschung und Betrugsentlarvung, Einzug in die wissenschaftliche Literatur finden. Herausgeber von Fachzeitschriften müssten dazu eine größere Bereitschaft zeigen, Replikationen, sowohl erfolglose als auch die

Theorie bestätigende Ergebnisse, zu publizieren. Ebenso sollte die Möglichkeit bestehen, statistisch nicht signifikante (aber dennoch wichtige) Ergebnisse in einem Journal veröffentlichen zu können.

7.3 Präventive Maßnahmen nach Stroebe et al. (2012)

Alle bisher genannten Maßnahmen treten erst dann in Kraft, wenn der Betrug bereits begangen wurde und folglich einer Prüfung unterzogen wird. Um Betrug in der Wissenschaft jedoch präventiv vorbeugen zu können, müssen andere Lösungen überlegt werden. Personen werden in Versuchung geführt, Forschungsbetrug zu begehen, weil der erwartete Gewinn groß, die negativen Konsequenzen unwahrscheinlich und das Risiko entdeckt zu werden gering ist. Stapel beschrieb die bestehende Situation mit einer Metapher: „The cookie jar was on the table without a lid“ (Bhattacharjee, 2013).

7.3.1 Belohnungen

Gute Forscher werden in ihrer Karriere durch Publikationen in Fachzeitschriften mit hohem Impact-Faktor belohnt. Sehr produktive Wissenschaftler können schnell zu akademischen Stars aufsteigen, deren Beschäftigungsdauer an den Universitäten damit nicht nur auf unbestimmte Zeit verlängert wird, sondern wodurch sie auch angesehene Preise und Auszeichnungen für ihre Forschungsarbeit erhalten können. Auch Stapel erhielt während seiner Karriere mehrere Preise und widmete sich bei einer seiner Dankesreden ironischer Weise dem Thema, wie man richtig Forschung betreibe. Diese Auszeichnungen, die Wissenschaftlern verliehen werden, führen schließlich zu Gehaltserhöhungen an den beschäftigten Universitäten und besseren Forschungsbedingungen durch die Unterstützung staatlicher oder universitärer Forschungsfonds.

Um in Fachzeitschriften mit hohem Impact-Faktor publizieren zu können, muss man neue theoretische Hypothesen entwickeln und diese mit guter empirischer Forschung untermauern. Doch das allein reicht nicht. Die vorliegenden Daten müssen die Hypothesen eindeutig und widerspruchsfrei bestätigen. Es ist daher nicht verwunderlich, dass der Impact-Faktor eines Journals signifikant positiv mit der Anzahl zurückgewiesener Artikel korreliert (Fang & Casadevall, 2011, zitiert nach Stroebe et al., 2012).

Unglücklicherweise kann dieses System nur schwer geändert werden, denn es belohnt nicht nur Betrüger, sondern auch brillante Wissenschaftler, die sich diesen herausragenden Erfolg auf ehrliche Weise verdient haben. Außerdem gestaltet sich das System nach den gegebenen Marktbedingungen. Es gibt mehr Wissenschaftler als gut bezahlte Jobs und es werden mehr Artikel produziert als dafür Platz in Journals mit

hohem Impact-Faktor wäre. Es muss also ein Selektionsprozess stattfinden, der bis heute auf dem bewährtesten Gütekriterium, der Qualität einer Studie, beruht.

Um dieses Gütekriterium zu fördern, wäre es sinnvoll, wenn Institutionen junge Wissenschaftler mehr zu ethisch korrekter Forschung anhalten würden. Studien zu betrügerischem Verhalten bei Studenten ergab, dass soziale Normen und das Gefühl der moralischen Verpflichtung starke Prädiktoren für das Ausmaß an betrügerischem Verhalten sind (e.g. Crown & Spiller, 1998, zitiert nach Stroebe et al., 2012). Für angehende Wissenschaftler und bereits tätige sollten klare Verhaltensnormen und Regeln existieren und der Ablauf eines Verfahrens zur Prüfung eines möglichen Betrugsverdachts sollte allen bekannt sein.

7.3.2 Kosten

Die Folgen für Wissenschaftsbetrüger, sofern sie entlarvt werden, sind von Land zu Land verschieden. Grundsätzlich verlieren die Forscher ihre Anstellung und werden zu Außenseitern in der wissenschaftlichen Gemeinde. Die weiteren Kosten sind je nach Fall und Ausmaß des Betruges unterschiedlich.

Stapel gab seinen Dokortitel freiwillig ab, als ihm die Universität von Amsterdam damit drohte, ein Verfahren einzuleiten, um ihm den Titel abzuerkennen. Nach Bekanntwerden des Betrugsskandals um Stapel gab es in den Niederlanden Überlegungen, Wissenschaftsbetrug, genau wie jede andere Art von Betrug, als strafbare Handlung zu verfolgen. Denn Wissenschaftsbetrüger richten mit ihrem Fehlverhalten nicht nur Schaden an, sondern profitieren meistens sogar noch davon (durch die Verlängerung der Anstellung, Gehaltserhöhungen etc.). Außerdem wäre die Gefahr einer Gefängnisstrafe, bei Bekanntwerden des Betrugs, eine sehr starke Abschreckungsmethode für jeden Wissenschaftler. Offen bleibt dennoch die Frage, ob eine solche Konsequenz mehr helfen als die Sachlage verkomplizieren würde. Denn die Beweislast vor Gericht, um jemanden zu einer Freiheitsstrafe zu verurteilen, muss schon sehr groß und wasserdicht sein, im Gegensatz zu einer Schuldsprechung der wissenschaftlichen Gemeinde bei bestehendem Verdacht auf Betrug. Beispielsweise klagte die DFG Herrmann vor Gericht an, dass er die Millionen an Forschungsgeldern zurückzahlen solle, die er von ihnen erhalten habe. Die DFG verlor den Fall, da Herrmann alle Schuld auf seine Kollegin Brach schob, die sich zu diesem Zeitpunkt in den USA befand und für eine Zeugenaussage nicht ins Gericht berufen werden konnte. Das Gericht konnte Herrmanns Behauptung nicht widerlegen und entschied im Zweifel für den Angeklagten (zitiert nach Stroebe et al., 2012).

Um bei bestehendem Tatvorwurf von einem deutschen Gericht verurteilt werden zu können, muss die Staatsanwaltschaft beweisen können, dass Forschungsgelder intendiert

für betrügerische Wissenschaft missbraucht oder Finanzmittel auf andere Weise veruntreut wurden (Jähn, 2009).

In den USA reicht es aus, beweisen zu können, dass bei der Antragstellung auf staatliche Fördermittel manipulierte oder fabrizierte Daten verwendet wurden.

Grundsätzlich macht die Erhöhung der Strafen für Wissenschaftsbetrug nur dann Sinn, wenn die Forscher auch damit rechnen, erwischt zu werden. Die Diskrepanz zwischen der geschätzten Anzahl jährlicher Betrugsfälle von der *Gallup Organization* (2008) und der Anzahl tatsächlich identifizierter Fälle der ORI lässt vermuten, dass die Wahrscheinlichkeit des Betrugs überführt zu werden, gering ist. Es ist daher relativ nachvollziehbar, dass sich Forscher in ihrer Entscheidung, Wissenschaftsbetrug zu begehen, weniger mit möglichen Konsequenzen auseinander setzen (Stroebe et al., 2012).

7.3.3 Erhöhung der Entdeckungswahrscheinlichkeit

Aus vergangener Wissenschaftsgeschichte geht hervor, dass die meisten Betrugsfälle von Skandalaufdeckern („Whistleblower“) enthüllt wurden. Diese Personen sind zumeist mit dem Forschungsthema gut vertraut und besitzen genügend Fachwissen, um einen entscheidenden Beweis als solchen zu erkennen. Obwohl „Whistleblower“ eigentlich nur ein Instrument von vielen sein sollten, um Wissenschaftsbetrug aufzudecken, haben sie sich als effizienteste Strategie bewährt, um Betrüger zur Verantwortung zu ziehen. Daher stellt sich die Frage, wie man Enthüllungspersonen in ihrem Mut unterstützen kann, ohne gleichzeitig eine Forschungskultur vollen gegenseitigen Misstrauens und fälschlicher Beschuldigungen zu schaffen.

Eine der Hauptaufgaben von Forschungsinstituten sollte es sein, das Bewusstsein um das Vorhandensein und die Methoden von Wissenschaftsbetrug zu stärken, damit potentielle Skandalaufdecker 1. nicht das Gefühl haben, sie hätten es dabei mit anormalem Verhalten zu tun, 2. wissen, wonach sie zu suchen haben und 3. sich sicher sein können, dass sowohl sie selbst als auch die beschuldigte Person mit Respekt behandelt werden, wenn sie sich mit ihrem Verdacht an eine Vertrauensperson wenden.

Eine weitere Maßnahme bestünde darin, dass Universitäten bewiesene Betrugsfälle offen legen. Viel zu lange und häufig wurden Details bewiesener Betrugsfälle vor der Öffentlichkeit unter Verschluss gehalten. Dem Bewusstsein, dass Forschungsbetrug existiert, wird nicht geholfen, wenn das begangene Fehlverhalten auf die persönliche Pathologie des Betrügers oder einer Randerscheinung in der Wissenschaft geschoben wird. Beide Formen dieser Attributionen reduzieren die Relevanz des Themas für einen selbst und verringern die eigene Wachsamkeit in Bezug auf Verhaltensweisen der Wissenschaftsgemeinde.

7.4 Alternative Maßnahmen

Im Sinne der Erhöhung der Entdeckungswahrscheinlichkeit gibt es bereits alternative Bewegungen, die wissenschaftlichem Fehlverhalten entgegenwirken sollen:

7.4.1 „Archives of Scientific Policy“

Von der APA (American Psychological Association, 2012) wurde kürzlich das neue *Journal Archives of Scientific Policy* geschaffen, das sich durch fünf neue Gestaltungsprinzipien auszeichnet:

1. Alle Artikel sind für die breite Öffentlichkeit frei zugänglich. Jede Person mit Internetzugang kann sich über aktuelle Forschungsarbeiten informieren.
2. Den „Journal Article Reporting Standards“ (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008) der APA folgend, beinhaltet jeder Artikel eine detaillierte Beschreibung der angewandten Methoden.
3. Die Daten, die für die Berechnung der statistischen Analysen verwendet wurden, sind ebenfalls frei verfügbar. Jeder kann die Datensätze nutzen, um weitere Berechnungen anzustellen oder die dargestellten Ergebnisse zu überprüfen.
4. Da die Artikel sowohl Laien als auch professionellen Wissenschaftlern zur Verfügung stehen, gibt es für jeden dieser zwei Publikumstypen ein eigenes Abstrakt und Methoden Dokument, um den unerfahrenen Leser nicht mit technisch wissenschaftlichen Details zu überfordern.
5. Um die Kommunikation über einen bestimmten Artikel zu erleichtern, werden alle Kommentare von *Peer-Reviewern* und eine mögliche Stellungnahme der Autoren in einem zusätzlichen Dokument zur Verfügung gestellt.

7.4.2 Statistische Analyse

Eine weitere alternative Methode der Betrugsentdeckung stammt von dem bereits erwähnten „Data Detective“ Simonsohn, der mithilfe seiner statistischen Analyse die beiden jüngsten Betrugsfälle der Sozialpsychologie, Smeesters und Sanna, aufdeckte (2012). Das Problem der statistischen Evaluation ist jedoch das Verhältnis, in dem richtig identifizierte Betrugsfälle zu fälschlich identifizierten stehen. Daher sollte diese Methode nur dann angewandt werden, wenn die Anzahl der falsch identifizierten Fälle gering gehalten werden kann.

7.4.3 Replikation

Auf Basis der aufgekommenen Kritik an der zu wenig geförderten Replikation als korrigierende Maßnahme der Wissenschaft, wurde 2012 eine Webseite ins Leben gerufen (Archive of Replication Attempts in Experimental Psychology, 2012), auf der Forscher ihre Replikationsversuche von experimentellen Studien hochladen und sich über andere informieren können. Diese Initiative stellt einen ersten wichtigen Schritt zur Verbesserung hinsichtlich der Replikation als Betrugsdetektor dar.

7.4.4 Sicherung guter wissenschaftlicher Praxis

Bei einem öffentlichen Vortrag am Max-Planck-Institut für Gesellschaftsforschung im April 1999 zu dem Thema „Betrug in der Wissenschaft – Randerscheinung oder wachsendes Problem?“ zog Renate Mayntz folgende Schlussfolgerung:

Die wichtigste Schlussfolgerung scheint mir zu sein, dass wir es wirklich mit einem "Eisberg" zu tun haben - dass aber dieser Eisberg nicht einfach aus Fälschungen besteht. Bewusste Fälschung ist die kleine sichtbare Spitze dieses Eisbergs -- und diese Spitze ist für den praktischen Wert von Wissenschaft nicht besonders gefährlich. Der große, von der Öffentlichkeit vernachlässigte und insofern "unsichtbare" Körper des Eisbergs aber besteht vorwiegend nicht etwa aus unerkannten Fälschungen, sondern aus der Vernachlässigung der Regeln guter Forschungspraxis ohne bewusste Fälschungsabsicht [...].

Zur Prävention von „unbewussten Fälschungen“, die nach Mayntz die Richtigkeit der von Wissenschaftlern getroffenen Aussagen viel stärker beeinträchtigen als die direkte Fälschung, sollten Maßnahmen überlegt werden. Die DFG veröffentlichte 2013 in diesem Sinne eine erweiterte 112-seitige Denkschrift mit dem Titel „Selbstkontrolle in der Wissenschaft“, mit Vorschlägen zur Sicherung guter wissenschaftlicher Praxis (Deutsche Forschungsgemeinschaft, 2013).

8 Umfrage

Um die ausgearbeiteten Forschungsfragen nicht nur durch die theoretische Literaturrecherche beantworten zu können, wurde eine Umfrage durchgeführt, um Reaktionen und Stellungnahmen von Experten im Bereich der Sozialpsychologie zum Fall Diederik Stapel zu erheben.

Mithilfe der Plattform *SoSci Survey* (<https://www.soscisurvey.de/>) wurde ein online-Fragebogen mit 12 Fragen, in direkter Anlehnung an die bearbeitenden Forschungsfragen formuliert. Als Stichprobe wurde die *Deutsche Gesellschaft für Psychologie* ausgewählt, welche eine eigene *Fachgruppe für Sozialpsychologie* beherbergt. Diese Fachgruppe umfasst insgesamt 497 Mitglieder, von denen bei 467 Personen auf der offiziellen Homepage (<http://www.dgps.de/>) eine E-Mail Adresse angegeben war. Im Zeitraum von September bis Oktober 2013 hatten die Mitglieder drei Wochen Zeit, um die Umfrage anonym durchzuführen. Insgesamt wurden 118 Fragebögen ausgefüllt, 104 enthielten einen vollständig ausgefüllten Datensatz und konnten für die Auswertung herangezogen werden. Eine Kopie des Fragebogens findet sich im Anhang.

8.1 Ergebnisse

Die erhaltenen Fragebögen wurden in eine SPSS Datei übertragen und deskriptiv ausgewertet. Die Ergebnisse sind in Häufigkeiten und Prozentsen angegeben.

Frage 1. Wie intensiv haben Sie den Fall Diederik Stapel in den Medien verfolgt?

Tabelle 6: Medieninteresse

Antwort	Häufigkeit	Prozent
gar nicht	9	8.7
ab und zu	36	34.6
des Öfteren	37	35.6
sehr intensiv	22	21.2
gesamt	104	100

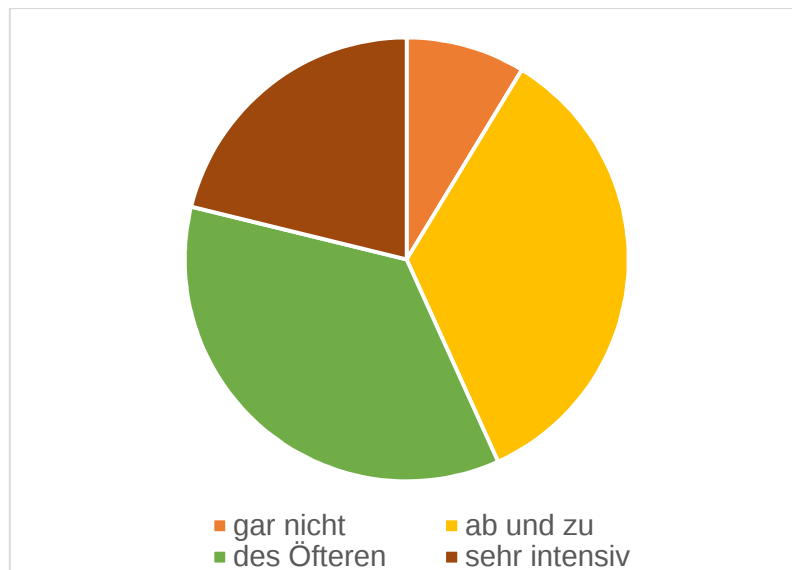


Abbildung 4: Medieninteresse

In Tabelle 6 und Abbildung 4 sind die Antworten der DGP-Mitglieder zu ihrem Medieninteresse tabellarisch und graphisch dargestellt. 8.7% gaben an, den Fall Stapel *gar nicht* in den Medien verfolgt zu haben, 34.6% taten dies immerhin *ab und zu*, 35.6% verfolgten die Berichterstattung *des Öfteren* und 21.2% beschäftigten sich *sehr intensiv* mit der Darstellung des Betrugsfalls in den Medien.

Frage 2. Als Mitglied der Deutschen Gesellschaft für Psychologie, wie sehr hat Sie dieser Fall emotional erregt?

Tabelle 7: Emotionale Betroffenheit

Antwort	Häufigkeit	Prozent
gar nicht	3	2.9
ein wenig	18	17.3
mittelmäßig	36	34.6
sehr	47	45.2
gesamt	104	100

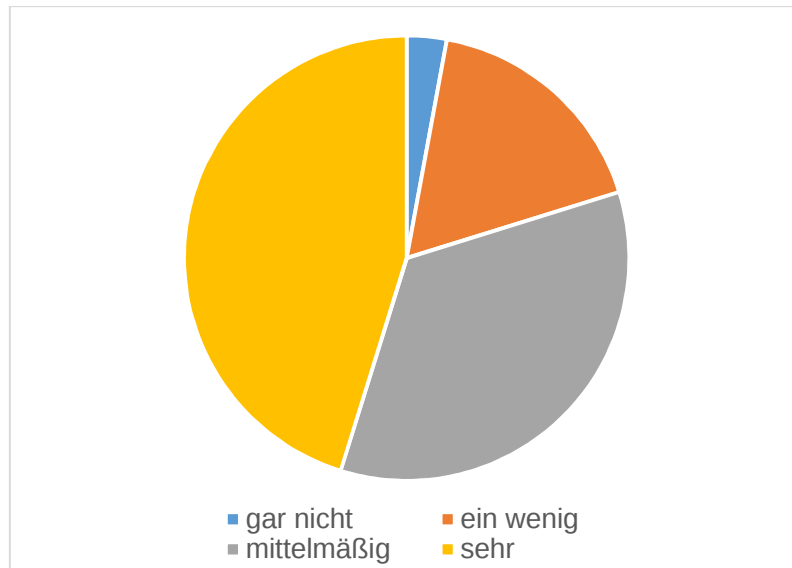


Abbildung 5: Emotionale Betroffenheit

In Tabelle 7 und Abbildung 5 sind die Ergebnisse zur emotionalen Betroffenheit dargestellt. 2.9% der DGP-Mitglieder gaben an, dass sie der Fall Stapel emotional *gar nicht* erregt hat, 17.3% berührte sein Betrug *ein wenig*, 34.6% gaben an, *mittelmäßig* emotional betroffen zu sein und 45.2% erregte Stapels Fehlverhalten emotional *sehr*.

Frage 3. Für wie weit verbreitet halten Sie das Manipulieren und Fabrizieren von Datensätzen in der Wissenschaft?

Tabelle 8: Verbreitung

Antwort	Häufigkeit	Prozent
äußerst selten vorkommend	27	26.0
ab und zu vorkommend	58	55.8
kommt häufig vor	19	18.3
kommt sehr häufig vor	0	0
gesamt	104	100



Abbildung 6: Verbreitung

In Tabelle 8 und Abbildung 6 ist tabellarisch und graphisch dargestellt, für wie weit verbreitet die Umfrageteilnehmer das Manipulieren und Fabrizieren und Datensätzen in der Wissenschaft halten. 26% sind der Meinung, dass dies *äußerst selten* vorkomme, 55.8% gaben an, dass sie glauben, dass ein solches wissenschaftliches Fehlverhalten *ab und zu* vorkomme, 18.3% glauben, dass dieses Verhalten *häufig* vorkommt und keine einzige Person gab an, dass sie die Datenmanipulation für *sehr häufig* vorkommend hält.

Frage 4. Halten Sie die Psychologie als wissenschaftliche Disziplin für anfälliger für Forschungsbetrug als andere wissenschaftliche Disziplinen?

Tabelle 9: Anfälligkeit

Antwort	Häufigkeit	Prozent
ja	19	18.3
nein	80	76.9
weiß nicht	5	4.8
gesamt	104	100

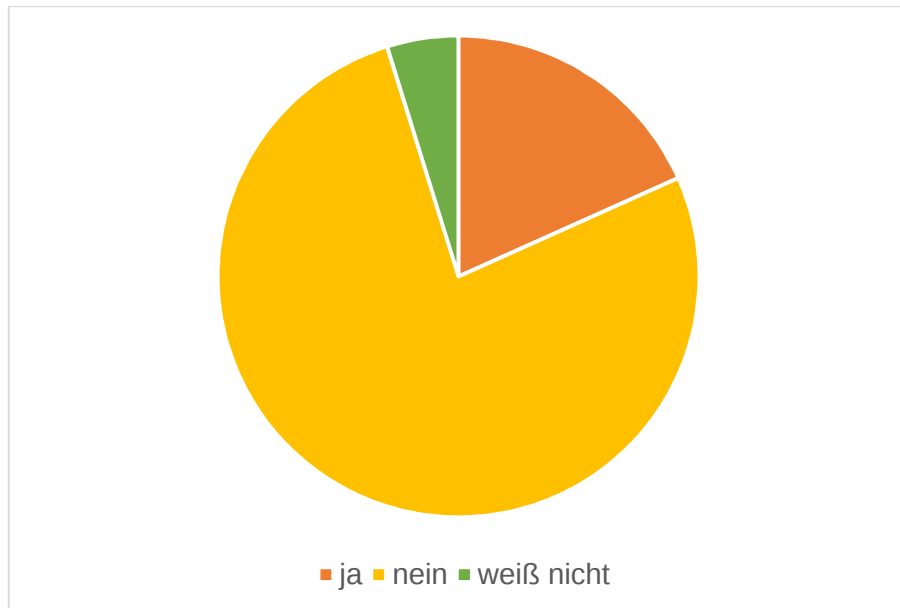


Abbildung 7: Anfälligkeit

Für die Fragen, bei denen die Antwortmöglichkeiten *ja*, *nein* und *weiß nicht* zur Auswahl standen, wurde im Falle einer *ja* beziehungsweise *nein* Antwort ein offenes Textfeld zugeschaltet, in dem die Umfrageteilnehmer mit einigen Worten angeben sollten, inwiefern sie sich für die jeweilige Antwortmöglichkeit entschieden haben.

Bei der 4. Frage gaben 19 Personen (18.3%) an, dass sie die Psychologie für anfälliger für Forschungsbetrug halten als andere Disziplinen, 76.9% verneinten dies und 4.8% waren sich in ihrer Antwort nicht sicher (Tabelle 9 und Abbildung 7).

Auf die Nachfrage, warum sie dieser Überzeugung seien, wurden folgende Antworten formuliert, welche inhaltlich zu 10 Kategorien zusammengefasst wurden:

1. Aufgrund der Jagd nach „sexy“ (=kontraintuitiven) Ergebnissen und absurden Wünschen von Herausgebern nach öffentlichkeitswirksamen Effekten.
2. Aufgrund der schlechten Replikationskultur der Psychologie: die Replikation von psychologischen („sexy“) Ergebnissen ist schwieriger als in anderen Wissenschaften und Nulleffekte werden nicht publiziert.
3. Aufgrund der unzureichenden öffentlichen Bereitstellung der Daten (z.B. in einem Zentralarchiv), damit ist die Überprüfbarkeit eingeschränkt und die Fälschungsanfälligkeit hoch.
4. Wegen zu vielen Freiheitsgraden in der psychologischen Operationalisierung: Stimulusmaterialien weichen oft voneinander ab, das erschwert die exakte Replikation. Es kommt zu einem „subjektiven“ Messen.

5. Weil andere Wissenschaften stärker formalisiert sind, z.B. ist eine exakte Dokumentation von der Versuchsplanung bis zur Auswertung in naturwissenschaftlichen Disziplinen, wie der Chemie oder in der Medizin, gänglicher Usus.
6. Es gibt keine von Fachzeitschriften unabhängigen Datenbanken zur (Pre-) Registrierung und Publikation von Studien.
7. Aufgrund der schlechteren methodischen und wissenschaftlichen Ausbildung der Psychologen.
8. In der Psychologie als „Nicht-Lebenswissenschaft“ geht es, im Vergleich zur Medizin, wo jedes Patientenleben zählt, nicht täglich um Leben und Tod – darum ist gerade die Sozialpsychologie aufgrund mangelnder praktischer Anwendbarkeit mehr gefährdet.
9. Sozial- und Geisteswissenschaften sind anfälliger als Naturwissenschaften aufgrund des „weicheren“ Gegenstandes und der stärkeren Konstruierbarkeit.
10. In der Psychologie gibt es weniger internationale Großprojekte (z.B. „CERN“), bei denen durch die internationale Zusammenarbeit mehrerer Forscherteams die Bereitschaft und Durchführung von Forschungsbetrug erschwert wird.

Frage 5. Sehen Sie den Betrug Stapels als Einzelfall in der Psychologie an?

Tabelle 10: Einzelfall

Antwort	Häufigkeit	Prozent
Auf jeden Fall	8	7.7
Ich bin mir nicht sicher	43	41.3
Es gibt noch ein paar mehr Betrugsfälle	46	44.2
Es gibt noch wesentlich mehr Betrugsfälle	7	6.7
gesamt	104	100

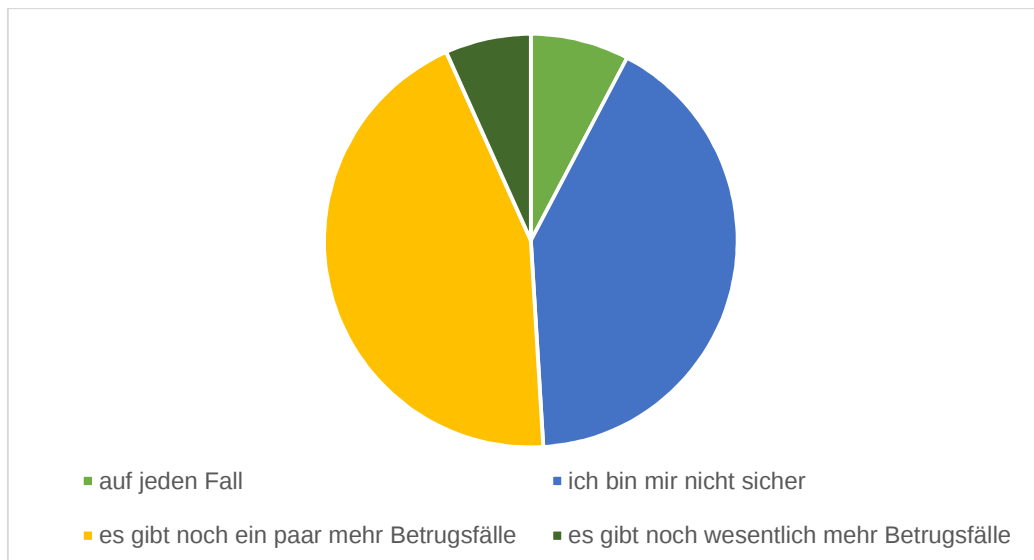


Abbildung 8: Einzelfall

In Tabelle 10 und Abbildung 8 sind die Antworten der Umfrageteilnehmer auf die Frage, ob sie Stapels Betrug als Einzelfall in der Psychologie ansehen, dargestellt. 7.7% sind der Meinung, dass Stapel auf *jeden Fall* eine Ausnahme darstellt, 41.3% sind sich dessen *nicht sicher*, 44.2% glauben, dass es noch *ein paar mehr Betrugsfälle* gibt und 6.7% waren der Überzeugung, dass es noch *wesentlich mehr Betrugsfälle* in der Psychologie gibt.

Frage 6. Wo sehen Sie hauptsächlich die Gründe dafür, dass Diederik Stapels wissenschaftliches Fehlverhalten jahrelang unentdeckt blieb?

Tabelle 11: Gründe für das Unentdecktbleiben des Betrugs

Antwort	Häufigkeit	Prozent	Prozent der Fälle
Stapels Persönlichkeitsfaktoren	32	13.7	30.8
Forschungsmethoden der Psychologie	65	27.9	62.5
Stapels Macht- & Autoritätsposition	50	21.5	48.1
System der Wissenschaft	86	36.9	82.7
gesamt	233	100	224

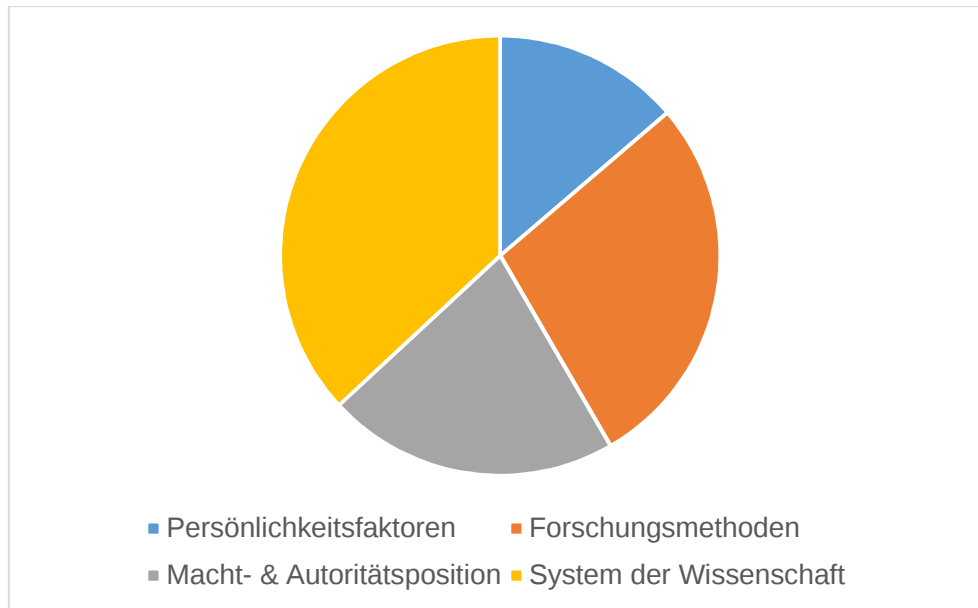


Abbildung 9: Gründe für das Unentdecktbleiben des Betrugs

Tabelle 11 und Abbildung 9 stellen die Antworten zu den Gründen für das lange Unentdecktbleiben Stapels Betrugs tabellarisch und graphisch dar. Zur Beantwortung dieser Frage wurden Mehrfachantworten zugelassen. Die vierte Spalte in Tabelle 11 stellt daher die Anzahl aller Fälle in Prozent dar, mit der eine Antwortmöglichkeit angeklickt wurde und übersteigt daher insgesamt die 100 Prozent. Zu den angegebenen Antwortmöglichkeiten in der ersten Spalte in Tabelle 11 wurden bei der Umfrage jeweils Beispiele in Klammern angegeben. In 82.7% der Fälle gaben die Umfrageteilnehmer an, dass sie das bestehende System der Wissenschaft (z.B. die fehlende Förderung von Replikationen, Mängel im bestehenden Review-Prozess) für das lange Unentdecktbleiben Stapels Betrugsfall verantwortlich machen. 62.5% waren der Meinung, dass den Forschungsmethoden der Psychologie (z.B. geringe Kontrollen in der Datenerhebung, keine öffentliche Bereitstellung der Daten etc.) die Schuld zugeschoben werden kann, 48.1% machten Stapels hohe Macht- und Autoritätsposition (als Universitätsprofessor, Dekan und anerkannter Wissenschaftler) verantwortlich und in 30.8% aller Fälle wurden Stapels zugeschriebene Persönlichkeitsfaktoren (z.B. charismatisch, vertrauenswürdig, ehrgeizig, charmant, freundlich, engagiert etc., welche lange kein wissenschaftliches Fehlverhalten vermuten ließen) angeklickt.

Frage 7. Wie sehr denken Sie, hat der Fall Diederik Stapel den Ruf der Sozialpsychologie geschädigt?

Tabelle 12: Rufschädigung

Antwort	Häufigkeit	Prozent
gar nicht	2	1.9
ein wenig	26	25.0
mittelmäßig	38	36.5
sehr	38	36.5
gesamt	104	100

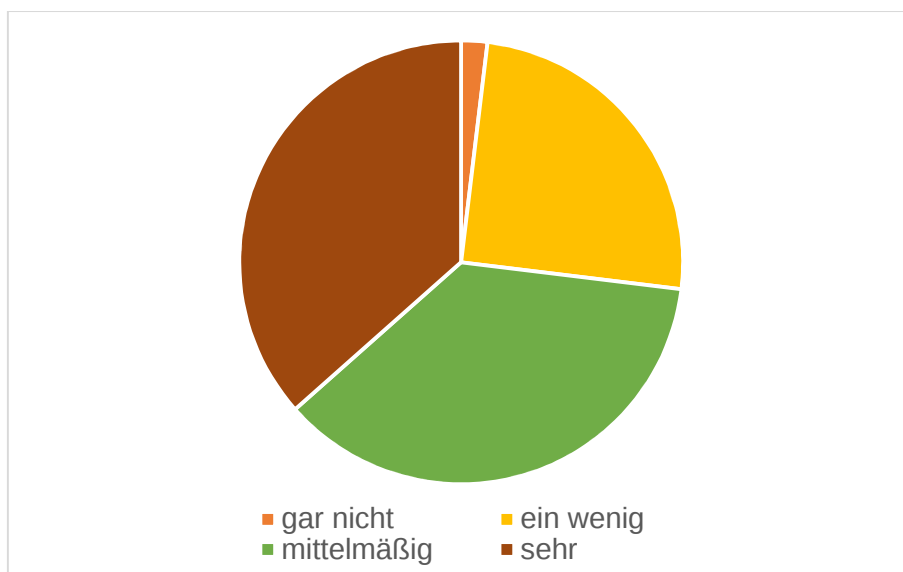


Abbildung 10: Rufschädigung

In Tabelle 12 und Abbildung 10 sind die Antworten der DGP-Mitglieder zur Frage dargestellt, wie sehr sie denken, dass der Betrugsfall Stapel den Ruf der Sozialpsychologie geschädigt hat. 1.9% glauben, dass Stapel den Ruf *gar nicht* geschädigt hat, 25% meinen, dass er *ein wenig* zur Rufschädigung beigetragen hat und jeweils 36.5% waren der Meinung, dass er den Ruf der Sozialpsychologie *mittelmäßig* bzw. *sehr* geschädigt hat.

Frage 8. Mussten Sie sich als Mitglied der DGP in Ihrem beruflichen oder privaten Umfeld seit Bekanntwerden des Betrugsfalls jemals für die angewandten Methoden in der (sozial-) psychologischen Forschung und in der Folge für Ihre persönliche berufliche Tätigkeit rechtfertigen?

Tabelle 13: Persönliche Rechtfertigung

Antwort	Häufigkeit	Prozent
ja	18	17.3
nein	76	73.1
weiß nicht	9	8.7
gesamt	104	100

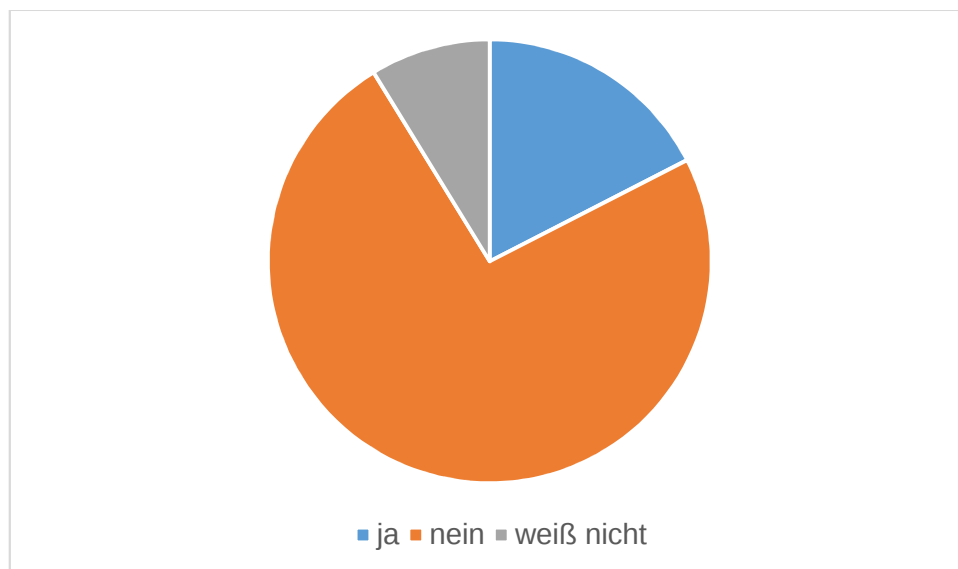


Abbildung 11: Persönliche Rechtfertigung

Auf die Frage 8 antworteten 18 Personen (17.3%), dass sie sich bereits einmal persönlich rechtfertigen mussten, 73.1% gaben an, dass sie dies noch nie tun mussten und 8.7% waren sich dessen nicht sicher (Tabelle 13 und Abbildung 11). Die Umfrageteilnehmer, die mit „ja“ geantwortet hatten, sollten daher angeben, inwiefern sie sich bereits persönlich rechtfertigen mussten. Dabei wurden folgende Antworten formuliert:

- Es gab Rechtfertigungen für die Forschungsmethoden der Sozialpsychologie, die Glaubbarkeit der Ergebnisse und Generalisierbarkeit von Laborergebnissen wurde in Frage gestellt.
- Die Aussage "Ich wusste doch, dass man Euch nicht trauen kann" wurde mehrfach geäußert, gegen die ich mich und das Fach verteidigen musste.
- Andeutungen über die Einfachheit des Betrugs in der sozialpsychologischen Forschung.
- Vorbehalte bei Studienteilnehmern und Probanden.
- Rechtfertigung auf internationalen Kongressen.
- Die Berufswahl wurde in Frage gestellt. Durch psychologische Forschung werden sowieso nur Forschungsgelder verschwendet.
- Es gab die Unterstellung, dass wir das doch bestimmt alle machen, dass das eben in weichen Wissenschaften nicht anders ginge.
- Diskussionen über den Publikationsdruck.
- Es wurden Aussagen getätigt, wie: „Die Studien sind Unsinn - alles nur gefälscht“.
- Dumme Sprüche von Kollegen, z.B.: „Wo arbeiten Sie?“ „Am Institut für Psychologie“ „Na hoffentlich nicht für Sozialpsychologie...!“.
- Laborinterne Diskussionen führten zur besseren Formalisierung von Rohdaten, so dass Änderungen am Rohdatensatz technisch unmöglich gemacht wurden, da jede Änderung im Datensatz in einer Kopie des „Files“ mit neuem Datum führt.
- Ich musste eine Erklärung abgeben, wie ich selbst arbeite.
- In vielen Gesprächen mit Naturwissenschaftlern wurde die Psychologie als unseriös bezeichnet.
- Rechtfertigungen bei interdisziplinären Diskussionen im Freundeskreis: „Warum macht ihr das so? Wieso werden nicht alle Experimente veröffentlicht? Muss immer alles signifikant sein für eine Veröffentlichung?“.
- In der Psychologie gäbe es mehr „wow“ als „how“ Effekte.

Frage 9. Welche langfristigen Konsequenzen denken Sie, hat der Fall Stapel für die wissenschaftliche Disziplin der Sozialpsychologie?

Zur Frage 9 wurde ein offenes Texteingabefeld hinzugefügt, in dem die Umfrageteilnehmer ihre Antworten frei verschriftlichen konnten. Die Antworten wurden in positive und negative Konsequenzen kategorisiert, wobei die positiven häufig als Wünsche formuliert wurden (Tabelle 14).

Tabelle 14: Langfristige Konsequenzen für die Sozialpsychologie

positive	negative
Änderung der Publikationspolitik	Weniger Glaubwürdigkeit psychologischer Forschungsergebnisse
Pre-registrierte Studien	Mehr Misstrauen unter Kollegen
Kritischere Datenkontrolle	Der Fall Stapel wird im allgemeinen Misstrauen gegenüber der Wissenschaft „untergehen“
Klarere Forschungsdokumentation	Extreme Rufschädigung für betroffene Forscher und der gesamten Sozialpsychologie
Publikationsqualität statt -quantität	Es wird kaum Konsequenzen geben: das „Publish or Perish“-Prinzip bleibt bestehen
Mehr „open-Journals“	Bürokratisierung von wissenschaftlichen Prozessen
Sensibilisierung gegenüber den Grauzonen in den Forschungsmethoden	Verlust von Drittmitteln für psychologische Forschung
Stärkere Absicherung von „Whistleblowern“	„Pseudo-Kontrollmaßnahmen“ für alle, statt richtige Maßnahmen bei Verdächtigen
Erstreckung des „Peer-Reviewings“ auch auf die Datensätze	Ausuferung in einer „Hexenjagd“ nach weiteren Betrügern innerhalb der Wissenschaftsgemeinde
Verstärkte Selbstüberprüfung: Glaube ich an die Replizierbarkeit meiner Ergebnisse?	
Abschreckungseffekt auf andere Forscher durch die extreme Ächtung Stapels in den Medien und unsachlicher medialer Ausschlachtung	
Mehr Aufmerksamkeit für die Möglichkeit einer Datenfälschung	

Frage 10. Hat Sie der Betrugsfall Stapel zu konkreten Verhaltensänderungen in Ihrer Lehre/Forschungstätigkeit bewegt?

Tabelle 15: Verhaltensänderung

Antwort	Häufigkeit	Prozent
ja	36	34.6
nein	65	62.5
weiß nicht	3	2.9
gesamt	104	100

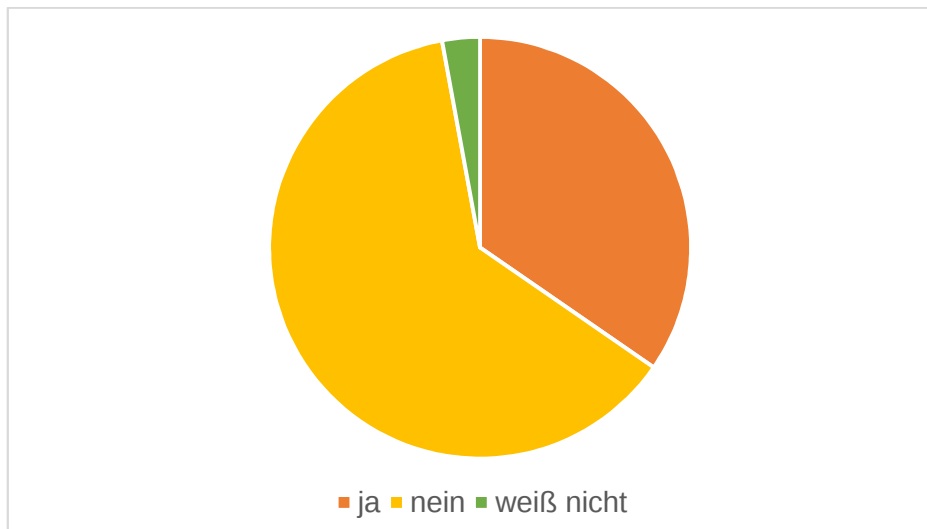


Abbildung 12: Verhaltensänderung

Bei Frage 10 gaben 36 Personen (34.6%) an, dass sie der Betrugsfall Stapel zu konkreten Verhaltensänderungen in ihrer Lehr- und/oder Forschungstätigkeit bewegt hat, 62.5% verneinten dies und 2.9% waren sich dessen nicht sicher (Tabelle 15 und Abbildung 12). Auf die Nachfrage, zu welchen Verhaltensänderungen die „ja“ antwortenden DGP-Mitglieder konkret bewogen wurden, wurden folgende Angaben gemacht:

- Thematisierung des Falls in Lehrveranstaltungen und in der Ausbildung von Doktorierenden und „Post Docs“, verstärktes Briefing von studentischen Hilfskräften.
- Thematisierung wissenschaftlicher Standards, Forschungsethik und Grauzonen mit Studenten (der Fall hat auch zu viel Unsicherheit bei Studierenden und Doktoranden gesorgt).
- Eine stärkere Kontrolle der Datenerhebung meiner Masterarbeiten.
- Vermehrte Aufmerksamkeit bei der eigenen Datenerhebung und -auswertung, häufigeres Überprüfen der Daten und lückenlose Dokumentation.
- Eine grundsätzlich höhere Skepsis gegenüber publizierten Befunden, insbesondere bei spektakulären.
- In meinem Lehrbuch habe ich die Bedeutsamkeit von Replikationsstudien anderer Forscher noch deutlicher herausgestellt.
- In meiner Arbeitsgruppe werden Studien wesentlich stärker dokumentiert (explorative Studien werden von hypothesen-testenden unterschieden). Außerdem werden mehr Replikationen durchgeführt und es werden größere Stichproben erhoben.
- Ich ziehe seither öfters „open access Journals“ in Betracht, die als Bedingung voraussetzen, dass Daten öffentlich zugänglich gemacht werden. Zudem überlege ich, ob auf unseren Webseiten auch solche Untersuchungen präsentiert werden sollen, die keine hochspannenden Ergebnisse erbringen oder Replikationen sind.
- Ich leite unsere Studierenden dazu an, kritisch reflektiert wissenschaftliche Studien zu lesen.
- Wir speichern jetzt auch die „log Files“ bei Computeruntersuchungen.
- Abgesehen davon, dass ich vermeide, Stapel zu zitieren, bin ich mir nicht so sicher. Ich glaube, mein Publikationswille hat etwas nachgelassen und ich dränge mit mehr Selbstbewusstsein darauf, sich weniger von aufmerksamkeitserregenden Studien blenden zu lassen und stattdessen mehr Gewicht auf Prozessforschung zu legen, auch wenn es schwieriger ist. Aber ob das was mit Stapel zu tun hat? Keine Ahnung.
- „Data-Sharing“ als Pflicht in der Arbeitsgruppe.
- Volle Transparenz hinsichtlich des Ausschlusses von Versuchspersonen.

- Ich habe als *Editor/Reviewer* eventuell leichter den Verdacht, dass etwas „faul“ ist. Ich habe kürzlich als *Editor* im „Action Letter“ "durch die Blume" einen solchen Verdacht geäußert und ins Schwarze getroffen (der Seniorautor gab zu, dass der Erstautor ihm gegenüber den Betrug gestanden hat).
- Ich sage noch deutlicher und öfter was ich meinen DoktorandInnen und AbsolventInnen schon immer sage: „Jedes Ergebnis ist willkommen. Jedes Ergebnis deiner Studien ist gleich gut. Es gibt keine Studie mit einem "schlechten" Ergebnis.“
- Eine erhöhte Sensibilität bei der eigenen Forschung und bei den Forschungsergebnissen von KollegInnen.
- Die Entwicklung eines abteilungsinternen Systems zur Speicherung und Einsehbarkeit von Daten.
- Ich kontrolliere bereits Abschlussarbeiten auf Plagiate, spreche mit meinen Absolvierenden und DoktorandInnen über gute wissenschaftliche Praxis, schicke sie zu Vorträgen und Workshops zum Thema "Ethik in der Wissenschaft" und bin noch kritischer bei *Peer-Reviews*.
- Ich vergebe für Empirie-Praktika und andere Arbeiten verstärkt Themen, die bekannte Arbeiten replizieren.

Frage 11. Wären Sie bereit, aufkommende Veränderungsvorschläge zur Prävention wissenschaftlichen Betrugs (z.B. durch die Veröffentlichung und Bereitstellung eigener Datensätze) aktiv zu unterstützen?

Tabelle 16: Prävention

Antwort	Häufigkeit	Prozent
ja	83	79.8
nein	8	7.7
weiß nicht	13	12.5
gesamt	104	100

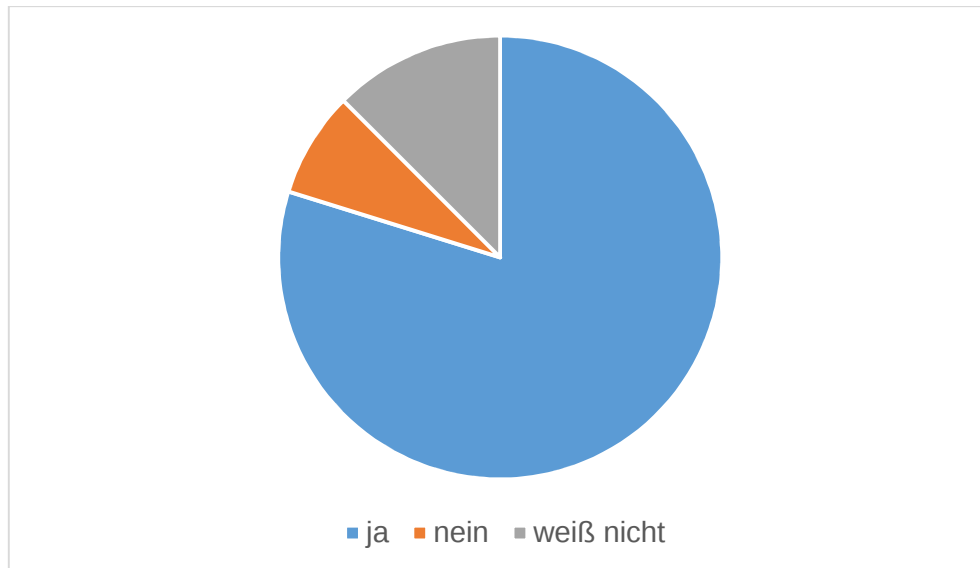


Abbildung 13: Prävention

In Tabelle 16 und Abbildung 13 sind die Ergebnisse auf die Frage dargestellt, ob die Umfrageteilnehmer aufkommende Veränderungsvorschläge zur Prävention wissenschaftlichen Betrugs aktiv unterstützen würden. 79.8% gaben an, dass sie dies tun würden, 12.5% waren sich dessen nicht sicher und 8 Personen (7.7%) meinten, dass sie aufkommende Veränderungsvorschläge nicht aktiv unterstützen würden. Auf die Nachfrage, warum sie dies nicht tun würden, nannten die betreffenden Personen folgende Gründe:

- Durch das Bereitstellen von Daten besteht für andere die Möglichkeit, diese ohne meine Beteiligung zu nutzen.
- Ich sehe es kritisch, der Öffentlichkeit selbst akquirierte Datensätze zur Verfügung zu stellen, diese verschiebt doch nur die Problematik. Wie soll geprüft werden, dass die SPSS-Datei echt ist? Im Zeitalter von Online-Erhebungen stellt dies für mich keine echte Alternative dar, solche Betrugsfälle zu vermeiden.
- Zusätzliche Kontrollmechanismen verhindern Betrug nicht, sondern machen für alle das Forscherleben unnötig schwieriger.
- Die Förderung und Forderung von mehr Transparenz als Substitut für Vertrauen und individuelle ethische Verantwortung finde ich falsch und würde es nicht unterstützen.
- Nicht wenn Persönlichkeits- und Unternehmensschutz (Rechte Dritter) vorgeht.
- Nicht wenn umfangreiche Datensätze aufwendig erhoben werden, andere aber die Meriten dafür „einheimsen“ wollen.

- Das Vermeiden von Betrug ist nicht das, worum es in der Wissenschaft geht. Und das ehrliche und kleinkarierte Sammeln von sauberen Befunden, bringt die Psychologie auch nicht weiter und macht sie kein bisschen interessanter. Insgesamt muss die Psychologie damit leben, dass es wohl schwer sein wird Betrug zu verhindern. Vielleicht sollte die Wissenschaft offener für „non replications“ werden und ihnen eine offene Chance zur Publikation geben, aber ohne dass dadurch dem nicht replizierten Autor gleich Betrug unterstellt wird.

Frage 12. Auf welche Art und Weise sollten mögliche Veränderungen im wissenschaftlichen System umgesetzt werden?

Tabelle 17: Systemänderung

Antwort	Häufigkeit	Prozent
Pflichtbewusstsein durch allgemein gültige Regeln	33	31.7
Wissenschaftliche Standards als Publikationsvoraussetzung	47	45.2
keine Veränderungen	5	4.8
weiß nicht	19	18.3
gesamt	104	100

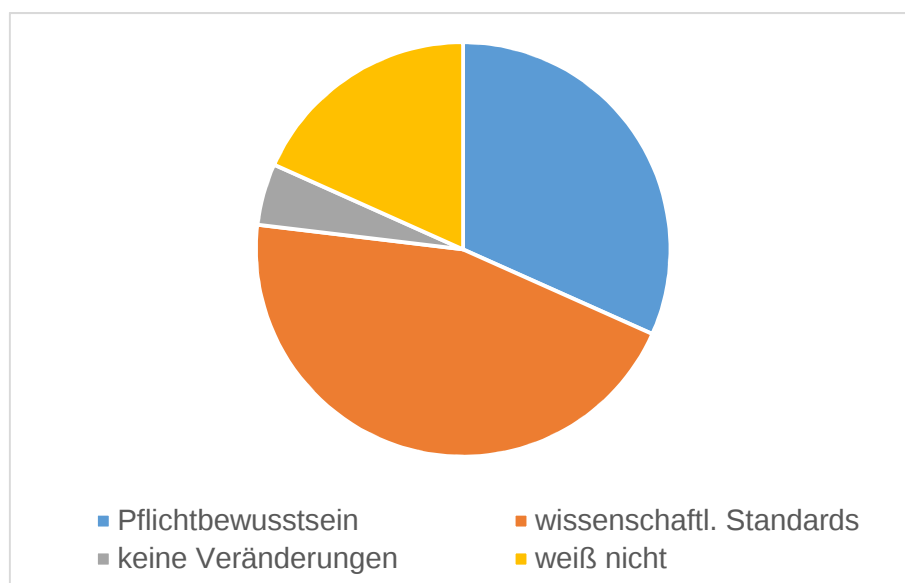


Abbildung 14: Systemänderung

Die letzte Frage der DGP-Umfrage bezieht sich auf eine Studie nach Fuchs, Jenny und Fiedler, die 2012 ebenfalls eine Umfrage mit 1292 Psychologen durchgeführt haben. Die Autoren schreiben in ihrem Artikel, dass sich die Art und Weise ändern muss, in der psychologische Forschung durchgeführt und berichtet wird und sie beziehen sich für ihre Umfrage auf eine Studie von Simmons, Nelson und Simonsohn (2011), welche sechs Bedingungen für die Datensammlung und das Berichten von Forschungsprozessen und vier Richtlinien für Reviewer aufgestellt haben, die den Publikationsprozess verbessern sollen. Die Versuchspersonen wurden für jede dieser sechs Änderungsvorschläge und vier Richtlinien gefragt, wie sehr sie diese unterstützen würden und zwar als „Standards of good practice“ oder auch als Publikationsbedingung von Fachzeitschriften. Bei 5 von 6 Vorschlägen gaben sie an, dass sie durch den Appell an die guten wissenschaftlichen Standards umgesetzt werden sollten und bei nur 3 von 6 Vorschlägen stimmten sie einer Publikationsvoraussetzung zu.

In der Umfrage dieser Diplomarbeit wurde die grundsätzliche Unterscheidung zwischen diesen beiden Umsetzungsmöglichkeiten getroffen und in einer Fußnote einige der Änderungsvorschläge nach Simmons et al. (2011) angegeben. 45.2% waren dafür, dass wissenschaftliche Standards als Publikationsvoraussetzung umgesetzt werden und 31.7% entschieden sich für die Veränderung durch die Bestärkung des Pflichtbewusstseins. 4.8% meinten, dass gar keine Veränderungen umgesetzt werden sollen und 18.3% entschieden sich für die Antwort „weiß nicht“ (Tabelle 17 und Abbildung 14).

8.2 Kritik an der Umfrage

Insgesamt wurde die höfliche Aufforderung zur Teilnahme an einer Umfrage zu den Auswirkungen des Forschungsbetrügers Diederik Stapel von den Mitgliedern der DGP sehr gut aufgenommen. Von einigen Teilnehmern wurde im Anschluss in einer persönlichen E-Mail Kritik am Aufbau und der Gestaltung der Umfrage geübt, allerdings wurden alle Punkte als konstruktive Verbesserungsvorschläge formuliert, welche durchaus nachvollziehbar sind.

Bezüglich der Fragestellung von Frage 11 wurde kritisiert, dass keine genaue Unterscheidung zwischen mehreren Veränderungsvorschlägen zur Prävention wissenschaftlichen Betrugs getroffen wurde, so dass sich einige Teilnehmer nicht zu einer pauschalen „ja“ oder „nein“ Antwort hinreißen ließen, sondern stattdessen aus Mangel an Alternativen die Antwort „weiß nicht“ anklickten.

Am häufigsten wurde jedoch die Antwortauswahl bei Frage 12 kritisiert. Einige Teilnehmer fühlten sich in ihrer Freiheit eingeschränkt, nur eine der beiden Antwortmöglichkeiten zur aktiven Veränderung des wissenschaftlichen Systems anklicken zu können. Die Gestaltung der Antwortalternativen wurde jedoch bewusst an die Studie von Fuchs et al. (2012) angelehnt, so dass sich die Umfrageteilnehmer, im Falle einer Veränderungsbereitschaft, entweder nur für den Appell an das Pflichtbewusstsein, sich an allgemein gültige Regeln zu halten oder für den erweiterten Standpunkt, wissenschaftlicher Standards als Publikationsvoraussetzung, entscheiden sollten. Zweitens wurde bei Frage 12 kritisiert, dass an dieser Stelle kein offenes Textfeld zur Verfügung stand, in dem die Teilnehmer eigene Veränderungsvorschläge anbringen hätten können. Dies wurde mit der vorangehenden Überlegung deswegen nicht umgesetzt, da Frage 9, mit den langfristigen Konsequenzen für die Disziplin der Sozialpsychologie, bereits als offenes Antwortformat vorgegeben wurde und sich die Antworten hierzu mit den möglichen offenen Antworten, im Falle eines freien Textfeldes bei Frage 12, inhaltlich überschneiden hätten.

9 Diskussion

Um ein Gesamtfazit aus den recherierten Konsequenzen des Forschungsbetrügers Diederik Stapel ziehen zu können, werden die Auswirkungen nach Stroebe et al. (2012) in einem Überblick zusammengefasst.

9.1 Behinderung des wissenschaftlichen Fortschritts

9.1.1 Verschwendung von finanziellen Ressourcen

Durch den Endbericht der Untersuchungskommission und aus dem Schreiben der NWO Pressesprecherin Margot Custers ist bekannt, dass Diederik Stapel Forschungsförderungsgelder in der Höhe von etwa 2.2 Millionen Euro für seine auf Betrug basierende Wissenschaft verschwendet hat. Diese Summe ist für jede durchschnittliche verdienende Person eine Menge Geld. Zieht man jedoch das jährlich zur Verfügung stehende Budget der NWO in der Höhe von 625 Millionen in Betracht, erscheint diese Zahl bereits weit weniger gewichtig, zumindest was den Schaden für die niederländische Organisation betrifft.

Diederik Stapel konnte von der niederländischen Staatsanwaltschaft kein Missbrauch der Forschungsgelder nachgewiesen werden, da er die Förderungen angeblich nicht zur persönlichen Bereicherung nutzte. Stattdessen wurden in einer außergerichtlichen Einigung 120 Sozialstunden festgelegt, eine Strafe, die angesichts der Stunden, die Stapel jahrelang damit verbracht hat, Daten zu fälschen, Mitarbeiter zu belügen und sich neue aufregende Experimente zu überlegen und der Anzahl von Arbeitsstunden, die seine Kollegen und Doktoranden in die Auswertung seiner manipulierten Daten gesteckt haben, sehr gering erscheint. Die vermutlich höchste Strafe erhielt Stapel jedoch ohnehin durch die mediale Ächtung und zur Schaustellung seiner Person, wodurch seine zukünftige wissenschaftliche Karriere für immer beendet sein wird.

9.1.2 Verschwendung von zeitlichen Ressourcen

Bis auf die vor kurzem veröffentlichte Replikation bekannter Experimente zum Thema „Intelligence Priming“ (Shanks, 2013), bei denen auch Stapel mitwirkte, gab es während der gesamten wissenschaftlichen Karrierelaufbahn des Niederländers keine einzige Studie, in der versucht wurde, seine stets für Aufregung sorgenden Ergebnisse zu replizieren. Diese Tatsache bedeutet zwar, dass dadurch immerhin keine zeitlichen Ressourcen durch immer wieder scheiternde Replikationsversuche verschwendet wurden,

allerdings ist die Nichtexistenz von Replikationen mit Sicherheit einer der Hauptgründe, warum Stapels jahrelanger Betrug für so lange Zeit unentdeckt bleiben konnte.

Auch wenn die erste Replikation Stapels Studien erst 2013 veröffentlicht wurde, so statuiert die Arbeit von Shanks et al. ein ganz wichtiges Exempel für den zukünftigen Umgang mit „sexy“ kontraintuitiven Befunden, so wie die Experimente zum sozialen Priming lange Zeit wirkten. Kahnemans Lösungsvorschlag, dass sich das sozialpsychologische Feld umgehend rehabilitieren würde, sofern die bedeutensten Priming Effekte erfolgreich repliziert werden würden, stellt einen ersten Schritt in die richtige Richtung dar und könnte auch als Lösungsansatz für die Wiederherstellung des durch Stapel verursachten Schadensausmaßes dienen.

Bei der durchgeführten Literaturanalyse der vorliegenden Arbeit wurde befunden, dass 77 Artikel weiterer Autoren explizit auf den Erkenntnissen des niederländischen Forschungsbetrügers und seinen unwissenden Kollegen aufbauen. Bezieht man die Tatsache mit ein, dass diese Stapel zitierenden Artikel wiederum von weiteren Autoren zitiert wurden, welche ihre Studien auf der Grundlagenforschung dieser Forscher stützen, so kann nur davon ausgegangen werden, dass Stapels Betrug sehr weite, nicht mehr überblickbare Wellen in der sozialpsychologischen Wissenschaft geschlagen hat.

Als Lösungsansatz bietet sich dadurch nur an, Stapels Studien zukünftig nicht mehr zu zitieren und auch Folgeartikel, die ihren Theorieteil auf den Erkenntnissen des Niederländers stützen, zu vermeiden. Die sicherste Lösung für Wissenschaftler, die auf den publizierten Assimilations- und Kontrasteffekten innerhalb der sozialen Urteilsforschung aufbauen wollen, wäre es, die grundlegenden Basiseffekte selbst zu replizieren, bevor von deren Existenz und Gültigkeit ausgegangen werden kann. Der dadurch zumeist minimal erhöhte Versuchsaufwand würde einen wesentlichen Beitrag zur Förderung von Replikationen und der Steigerung wissenschaftlicher Qualität anstatt Quantität leisten. Ist eine eigene Grundlagenforschung durch fehlende Ressourcen nicht möglich, so empfiehlt es sich immerhin, Stapels fundamentale Annahmen durch eine gründliche Literaturrecherche und den Vergleich mit den Ergebnissen anderen Forscherteams auf dem Gebiet der sozialen Kognitionsforschung anzustellen.

Diederik Stapels Beiträge in (Lehr-) Büchern sind höchst unerfreulich, da durch die Verbreitung betrugshafter Forschungsergebnisse falsches Wissen unter Studierende gebracht wird. Von 48 stichprobenartig untersuchten Büchern enthalten 10 Bücher einen oder mehrere Artikel Stapels, davon wurden insgesamt 13 Artikel von der Untersuchungskommission (Stapel Investigation, 2012) als „fraudulent“ erklärt. Der Schaden hält sich diesbezüglich also in Grenzen, vor allem wenn man bedenkt, dass sich Stapels Zitierungen in Büchern thematisch fast ausschließlich auf die Phänomene der

Assimilation und des Kontrasts in der sozialen Urteilsforschung beschränken. Davon abgesehen, dass es erstens sehr hilfreich wäre, wenn der Betrugsfall Stapel in sozialpsychologischen Lehrveranstaltungen an den Universitäten thematisiert wird und Studierende über die betrugshaften Lehren aufgeklärt werden und somit ein Auge auf Zitierungen in Büchern oder Fachzeitschriften haben, wäre es zweitens angebracht, wenn bei Neuauflagen von Lehrbüchern, in denen als Betrüger entlarvte Wissenschaftler zitiert werden, im Vorwort eine Anmerkung dazu abgegeben wird, dass das folgende Buch auf Artikel verweist, die offiziell von den jeweiligen Fachzeitschriften zurückgewiesen wurden.

9.1.3 Persönlicher Schaden

Das Leid, das Diederik Stapel seiner Familie, seinen engsten Freunden, seinen Studierenden, Doktoranden, Arbeitskollegen und Co-Autoren zugefügt hat, lässt sich weder erheben noch in Zahlen messen und wiegt dennoch am schwersten. Zu hoffen bleibt, dass sich seine Doktoranden und Kollegen von der Stigmatisierung erholen, jahrelang mit einem Betrüger Hand in Hand zusammen gearbeitet zu haben und ihre Glaubwürdigkeit durch ihre ehrliche Arbeit und ihren Fleiß wiederherstellen können.

9.1.4 Rufschädigung

Dass der Ruf der zugrundeliegenden Wissenschaftsdiziplin durch Diederik Stapels Fehlverhalten massiv geschädigt wurde, wurde in der vorliegenden Arbeit durch die Zitierung zahlreicher Medienberichte und Stellungnahmen von Forschern der Sozialpsychologie in online Zeitschriften und Portalen dargestellt. Auch die Umfrageteilnehmer der Deutschen Gesellschaft für Psychologie antworteten jeweils mit 36.5%, dass sie glauben, dass der Betrugsfall Stapel die Sozialpsychologie mittelmäßig beziehungsweise sehr geschädigt habe. Auf die Frage, welche langfristigen Konsequenzen dieser Fall für die sozialpsychologische Wissenschaft haben werde, antworteten immerhin einige DGP-Mitglieder, dass dieser mediale Skandal im allgemeinen Misstrauen gegenüber der psychologischen Forschung „untergehen“ werde, beziehungsweise dass Diederik Stapel, genauso wie andere Betrugsfälle aus der Vergangenheit, irgendwann in Vergessenheit geraten wird.

Was diesen konkreten Fall jedoch von vielen anderen wissenschaftlichen Betrugsfällen unterscheidet, ist, dass das Betrugsausmaß und die „Dreistheit“ Stapels jahrelangen Vorgehens eine unglaublich hohe Welle an Diskussionen zu Veränderungsvorschlägen des wissenschaftlichen Systems, der psychologischen Forschungsmethoden und präventiver Maßnahmen ausgelöst haben, die mit Sicherheit nicht allzu schnell

verstummen werden und daher auch als große Chance zur Veränderung angesehen werden können.

Was schlussendlich bleibt, ist der bittere Geschmack der Kritiker, die die Psychologie aufgrund ihres „weicheren“ Gegenstandes im Vergleich zu den „harten“ Naturwissenschaften ohnehin schon belächelt haben und gegen die sich das (sozial-)psychologische Feld in Zukunft umso mehr behaupten muss. Inwiefern die Psychologie dies bewerkstelligen wird, wird sich weisen. Die Weichen für eine Veränderung und ein womöglich neues „wissenschaftliches Zeitalter“ sind gelegt, die Frage bleibt, wie viele Menschen diesem Wandel folgen werden.

Angesichts dessen hat ein einzelner Mann und niederländischer Forscher, Diederik Alexander Stapel, eine Menge bewegt.

**„Wenn der Wind der Veränderung weht,
bauen die einen Mauern und die anderen Windmühlen“**

Chinesisches Sprichwort

10 Literaturverzeichnis

Abbott, A. (2013). Disputed results a fresh blow for social psychology. *Nature News*, 497, 16.

American Psychological Association (2012). Archives of scientific psychology [online]. URL: <http://www.apa.org/pubs/newsletters/access/2012/08-21/first-journal.aspx>

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63, 839-851.

Archive of Replication Attempts in Experimental Psychology (2012). URL: PsychFileDrawer.org

Bargh, J. A. (2012). The natural unconscious: Priming effects replicate just fine, thanks. *Psychology Today* [online]. URL: <http://www.psychologytoday.com/blog/the-natural-unconscious/201205/priming-effects-replicate-just-fine-thanks>

Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.

Bargh, J. A. & Huang, J. Y. (2009). The selfish goal. In G. B. Moskowitz & H. Grant (Eds.), *The psychology of goals* (pp. 127-150). New York: Guilford.

Bhattacharjee, Y. (2013). The mind of a con man. *The New York Times Magazin* [online]. URL: http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html?pagewanted=all&_r=0

Brüne, M. (2008). Soziale Kognition - Psychologie. In T. Kirchler & S. Gauggel (Hrsg.), *Neuropsychologie der Schizophrenie* (S. 347-356). Heidelberg: Springer.

Carey, B. (2011). Fraud case seen as a red flag for psychology research. *The New York Times* [online]. URL: <http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html>

Campbell, H. (2011). Diederik Stapel: Another world class psychology fraud. *Science 2.0* [online]. URL: http://www.science20.com/science_20/blog/diederik_stapel_another_world_class_psychology_fraud-84171

Cesario, J., Plaks, J. E. & Higgins, E. T. (2006). Automatic social behaviour as motivated preparation to interact. *Journal of Personality and Social Psychology*, 90, 893-910.

Crocker, J. & Cooper, M. L. (2011). Addressing scientific fraud. *Science*, 334 (6060), 1182.

Deutsche Gesellschaft für Psychologie [online]. URL: <http://www.dgps.de/>

Deutsche Forschungsgemeinschaft (2013). Vorschläge zur Sicherung guter wissenschaftlicher Praxis [online]. URL:

http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf

Dijksterhuis, A. & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology*, 74, 865–877.

Dijksterhuis, A. & van Knippenberg, A. (2000). Behavioral indecision: Effects of selffocus on automatic behavior. *Social Cognition*, 18, 55–74.

Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D. A., Koomen, W. et al. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology* 75, 862–871.

Doyen, S., Klein, O., Pichon, C., Cleeremans, A. (2012). Subliminal behavioral priming: It is all in the brain, but whose brain? *PLoS One*, 7 (1).

Drenth, P., Levelt, W. & Noort, E. (2013). Flawed Science? *The Psychologist*, 26, 80-81.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4, 1–11.

Fuchs, H. M., Jenny, M. & Fiedler, S. (2012). Psychologists are open to chance, yet wary of rules. *Perspectives of Psychological Science*, 7, 639-642.

Gallup Organization (2008). Observing and reporting suspected misconduct in biomedical research [online]. URL: http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf

Gibson, S. (2013). Flawed psychology [online]. URL: <http://www.timeshighereducation.co.uk/comment/letters/flawed-psychology/2001550.article>

Gordijn, E. H. & Stapel, D. A. (2006). Behavioural effects of automatic interpersonal versus intergroup social comparison. *British Journal of Social Psychology*, 45, 717–729.

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, 51, 1106-1115.

Highly Cited (2013). Highly cited research from Thomson Reuters [online]. URL: <http://highlycited.com/>

Highly Cited Liste (2012). URL: <http://community.thomsonreuters.com/t5/InCites-Customer-Forum/Preliminary-publication-of-new-lists-of-Highly-Cited-Researchers/td-p/36685>

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569-16572.

Hull, J., Slone, L., Metayer, K. & Matthews, A. (2002). The nonconsciousness of self-consciousness. *Journal of Personality and Social Psychology*, 83 (2), 406.

Jähn, C. (2009). Die Reputation zu verlieren ist die „härteste Strafe“. *Zeit online* [online]. URL: <http://www.zeit.de/online/2005/52/forscher>

Jha, A. (2012). False positives: Fraud and misconduct are threatening scientific research. *The Guardian* [online]. URL: <http://www.guardian.co.uk/science/2012/sep/13/scientific-research-fraud-bad-practice>

John, L. K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524-532.

Kahneman (2012). A proposal to deal with questions about priming effects [online]. URL: http://www.nature.com/polopoly_fs/7.6716.1349271308!/supinfoFile/Kahneman%20Letter.pdf

Kraut, A. (2011). Despite occasional scandals, science can police itself. *The chronicle of higher education* [online]. URL: <http://chronicle.com/article/Despite-Occasional-Scandals/129997/>

LeBoeuf, R. A. & Estes, Z. (2004). "Fortunately, I'm no Einstein": Comparison relevance as a determinant of behavioral assimilation and contrast. *Social Cognition*, 22, 607–636.

Mayntz, R. (1999). Betrug in der Wissenschaft: Randerscheinung oder wachsendes Problem? Öffentlicher Vortrag am Max-Planck-Institut für Gesellschaftsforschung 1999 in Köln [online]. URL: <http://www.mpifg.de/pu/workpap/wp99-4/wp99-4.html>

National Institute of Health. (1993). Final findings of scientific misconduct. *NIH Guide*, 22 (23) [online]. URL: <http://grants.nih.gov/grants/notice-files/not93-177.html>

National Science Foundation (2001). New research misconduct policies [online]. URL: <http://www.nsf.gov/oig/session.pdf>

Netherlands Organisation for Scientific Research (2013). What does NWO do? [online]. URL: <http://www.nwo.nl/en/about-nwo/what+does+nwo+do>

Newsletter Studenten (2010). Diederik Stapel succeeds Theo Verhallen as dean of the Tilburg School of Social and Behavioral Sciences. *Sociale Wetenschappen* [online]. URL: http://uvtapp.uvt.nl/fsw/spits.NPC.ShowArticleN?v_article_id=7425424320095932&v_newletter_id=8823896128229022&v_language=gb

Office of Research Integrity. (2001). Annual Report 2001: Findings of scientific misconduct [online]. URL: http://ori.hhs.gov/annual_reports

Ombudsman für die Wissenschaft [online]. URL: <http://www.ombudsman-fuer-die-wissenschaft.de/index.php?id=6091>

Pashler, H. & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530.

Retraction Watch [online]. URL: <http://retractionwatch.com>

Roediger, H. (2012). Psychology's woes and a partial cure: The value of replication. *Observer*, 25 (2).

- Shanks, D. R., Newell, B. R., Hee Lee, E., Balakrishnan, D., Ekelund, L., Cenac, Z., Kavvadia, F. & Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *Plos One*, 8 (4).
- Simmons, J. P., Nelson, L. D. & Simonsohn, U. (2011). False-Positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.
- Simonsohn, U. (2012). Just post it: The lesson from two cases of fabricated data detected by statistics alone [online]. URL: <http://www.ssrn.com/>
- Social Psychology Network [online] URL: <http://www.socialpsychology.org/texts.htm>
- Stapel Investigation (2012). Flawed Science: The fraudulent research practices of social psychologist Diederik Stapel [online]. URL: <https://www.commissielevelt.nl/>
- Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Stroebe, W., Postmes, T. & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7, 670-688.
- The Executive Committee of the European Association of Social Psychology (2012). Statement EASP on Levelt [online]. URL: http://www.easp.eu/news/Statement%20EASP%20on%20Levelt_December_%202012.pdf
- Van Noorden, R. (2013). Dutch psychology fraudster avoids trial. *Nature News* [online]. URL: <http://blogs.nature.com/news/2013/06/dutch-psychology-fraudster-avoids-trial.html>
- Van Lange, P. A. M., Buunk, A. P., Ellemers, N., & Wigboldus, D. H. J. (2012). Sharpening scientific policy after Stapel [online]. URL: <http://www.sociale-psychologie.nl>
- Wikipedia (2013). H-Index [online]. URL: <http://de.wikipedia.org/wiki/H-Index>
- Yong, E. (2012a). Nobel laureate challenges psychologists to clean up their act. *Nature*, 489.
- Yong, E. (2012b). The data detective. *Nature*, 487, 18-19.
- Yong, E. (2012c). Uncertainty shrouds psychologist's resignation. *Nature News* [online]. URL: <http://www.nature.com/news/uncertainty-shrouds-psychologist-s-resignation-1.10968>
- Yong, E. (2012d). A failed replication draws a scathing personal attack from a psychology professor. *Discover Magazine* [online]. URL: <http://blogs.discovermagazine.com/notrocketscience/2012/03/10/failed-replication-bargh-psychology-study-doyen/>

11 Tabellenverzeichnis

Tabelle 1: Zitationsreport des Autors Diederik A. Stapel	10
Tabelle 2: h-Index Vergleich	13
Tabelle 3: Anzahl durch Forschungsgelder geförderte betrügerische Publikationen.....	13
Tabelle 4: Ergebnisse der Literaturrecherche	34
Tabelle 5: Stapel Literatur in (Lehr-) Büchern	115
Tabelle 6: Medieninteresse	160
Tabelle 7: Emotionale Betroffenheit	161
Tabelle 8: Verbreitung.....	162
Tabelle 9: Anfälligkeit	163
Tabelle 10: Einzelfall	165
Tabelle 11: Gründe für das Unentdecktbleiben des Betrugs	166
Tabelle 12: Rufschädigung.....	168
Tabelle 13: Persönliche Rechtfertigung.....	169
Tabelle 14: Langfristige Konsequenzen für die Sozialpsychologie	171
Tabelle 15: Verhaltensänderung	172
Tabelle 16: Prävention.....	174
Tabelle 17: Systemänderung	176

12 Abbildungsverzeichnis

Abbildung 1: Anzahl der Publikationen von 1994 – 2013	11
Abbildung 2: Anzahl der Zitierungen durch andere von 1996 – 2013	11
Abbildung 3: Literaturecherche.....	31
Abbildung 4: Medieninteresse	161
Abbildung 5: Emotionale Betroffenheit	162
Abbildung 6: Verbreitung	163
Abbildung 7: Anfälligkeit	164
Abbildung 8: Einzelfall	166
Abbildung 9: Gründe für das Unentdecktbleiben des Betrugs	167
Abbildung 10: Rufschädigung.....	168
Abbildung 11: Persönliche Rechtfertigung.....	169
Abbildung 12: Verhaltensänderung.....	172
Abbildung 13: Prävention	175
Abbildung 14: Systemänderung	176

13 Anhang

13.1 Anhang 1: (Lehr-) Bücher Sozialpsychologie

- Argyle, M. (1995). *Social Psychology*. London: Longman.
- Aronson, E., Wilson, T. D. & Akert, R. M. (2008). *Sozialpsychologie*. München: Pearson Studium.
- Aronson, E., Wilson, T. D. & Akert, R. M. (2012). *Social Psychology*. Harlow: Pearson.
- Auhagen, A. E. & Bierhoff, H. W. (2003). *Angewandte Sozialpsychologie: Das Praxishandbuch*. Weinheim: Beltz Verlag.
- Baron, R. A. & Byrne, D. (2004). *Social Psychology*. Boston: Allyn & Bacon.
- Baron, R. A., Branscombe, N. R. & Byrne, D. (2013). *Social Psychology*. Harlow: Pearson Education Limited.
- Baumeister, R. F. & Vohs, K. D. (2007). *Encyclopedia of Social Psychology*. Thousand Oaks: Sage Publications.
- Baumeister, R. F. & Vohs, K. D. (2012). *New Directions in Social Psychology*. London: Sage Publications.
- Bierbrauer, G. (2005). *Sozialpsychologie*. Stuttgart: Kohlhammer.
- Bierhoff, H. W. (2006). *Sozialpsychologie: Ein Lehrbuch*. Stuttgart: Kohlhammer.
- Bless, H., Fiedler, K. & Strack, F. (2004). *Social Cognition: How Individuals construct social reality*. New York: Psychology Press.
- Brehm, S. S. & Kassin, S. M. (1996). *Social Psychology*. Boston: Houghton Mifflin.
- Buunk, B. P. & Van Vugt, M. (2007). *Applying Social Psychology: From problems to solutions*. Thousand Oaks: Sage Publications.
- Jonas, K., Stroebe, W., Hewstone, M. R. C. & Leberherz, C. (2007). *Sozialpsychologie: Eine Einführung*. Heidelberg: Springer.
- Myers, D. G. (2012). *Social Psychology*. New York: McGraw-Hill.
- Myers, D. G. (2013). *Psychology*. New York: Worth Palgrave Macmillan.
- Myers, D. G., Wahl, S., Hoppe-Graff, S. & Keller, B. (2008). *Psychologie*. Heidelberg: Springer.
- Bierhoff, H. W. & Frey, D. (2006). *Handbuch der Psychologie: Handbuch der Sozialpsychologie und Kommunikationspsychologie*. Göttingen: Hogrefe Verlag.
- Delamater, J. (2006). *Handbook of Social Psychology*. New York: Springer.
- Dovidio, J. F. (2010). *The Sage handbook of prejudice, stereotyping and discrimination*. Los Angeles: Sage.

- Fischer, L. & Wiswede, G. (2009). *Grundlagen der Sozialpsychologie*. München: Oldenbourg.
- Frey, D. (2002). *Theorien der Sozialpsychologie. 3. Motivations-, Selbst- und Informationsverarbeitungstheorien*. Bern: Huber.
- Gerrig, R. J. & Zimbardo, P. G. (2008). *Psychologie*. München: Pearson Studium.
- Gilbert, D. T., Fiske, S. T. & Lindzey, G. (2010). *Handbook of Social Psychology*. New York: Wiley.
- Gilovich, T., Keltner, D., Chen, S. & Nisbett, R. E. (2012). *Social Psychology*. New York: W. W Norton & Company.
- Gollwitzer, M. & Schmitt, M. (2006). *Sozialpsychologie*. Weinheim: Beltz Verlag.
- Güttler, P. O. (2003). *Sozialpsychologie: Soziale Einstellungen, Vorurteile, Einstellungsänderungen*. München: Oldenbourg.
- Hartung, J. (2010). *Sozialpsychologie*. Stuttgart: Kohlhammer.
- Herkner, W. (2001). *Lehrbuch Sozialpsychologie*. Bern: Verlag Hans Huber.
- Hewstone, M. (2008). *Introduction to Social Psychology: A European perspective*. Maiden: Blackwell.
- Hogg, M. A. & Vaughan, G. M. (2011). *Social Psychology*. Harlow: Pearson Prentice Hall.
- Hogg, M. A. (2003). *The Sage Handbook of Social Psychology*. Thousand Oaks: Sage Publications.
- Jones, J. M. (1997). *Prejudice and Racism*. New York: McGraw-Hill.
- Kassin, S., Fein, S. & Markus, H. R. (2010). *Social Psychology*. Boston: Houghton Mifflin.
- Kessler, T. & Fritsche, I. (2012). *Sozialpsychologie*. Wiesbaden: Verlag für Sozialwissenschaften.
- Kruglanski, A. W. & Higgins, E. T. (2007). *Social Psychology: Handbook of Basic Principles*. New York: Guilford Press.
- Kunda, Z. (1999). *Social Cognition: Making Sense of People*. Cambridge: MIT Press.
- Maderthaner, R. (2008). *Psychologie*. Stuttgart: UTB.
- Mann, L. (2001). *Sozialpsychologie*. Weinheim: Beltz Verlag.
- Myers, D. (2012). *Social Psychology*. New York: McGraw-Hill Higher Education.
- Sadava, S. W. & McCreary, D. R. (1997). *Applied Social Psychology*. New York: Prentice Hall.
- Schachinger, H. (2005). *Das Selbst, die Selbsterkenntnis und das Gefühl für den eigenen Wert*. Bern: Verlag Hans Huber.
- Semin, G. R. (1996). *Applied Social Psychology*. London: Sage Publications.
- Smith, E. R. & Mackie, D. M. (2007). *Social Psychology*. New York: Taylor & Francis.

Stainton-Rogers, W. (2011). *Social Psychology*. New York: Open University Press.

Stürmer, S. (2009). *Sozialpsychologie*. München: Ernst Reinhardt Verlag.

Taylor, S. E., Peplau, A. L. & Sears, D. O. (2006). *Social Psychology*. New York: Prentice Hall.

Werth, L. & Mayer, J. (2007). *Sozialpsychologie*. Heidelberg: Springer.

13.2 Anhang 2: E-Mail Margot Custers (2013)

Gesendet: Montag, 07. Oktober 2013 um 14:08 Uhr
Von: "Custers, M. [Margot]" <m.custers@nwo.nl>
An: "'a0707669@unet.univie.ac.at'" <a0707669@unet.univie.ac.at>
Betreff: RE: Information about Diederik A. Stapel

Dear Lisa,

Thank you for your questions. Please find the answers below.
In case you want to be reassured of a correct display of NWO's standpoint, we are more than happy to 'check' your text when finished.
If you have any further questions, please contact me.

Kind regards,
Margot Custers

Diederik Stapel received 1.4 million euros in science funding from NWO.

Dutch authorities opened an investigation into Stapel in October 2012. In June 2013 the public prosecutor announced that there was no abuse of the grants. The funds went mainly to salaries of staff (PhD students), who did the work they were paid for — even though that work was partly based on fabricated data. Stapel didn't use the money for personal enrichment.

In a pre-trial settlement, Stapel has agreed to 120 hours of community service. In the settlement, Stapel also agreed not to make a claim on 18 months of half-pay salary that he might have been entitled to contest under Dutch law.

NWO follows the decision made by the public prosecutor, and decided not to take legal action against Stapel.

Since the fraud of Diederik Stapel came to light, NWO has taken action to further stimulate maintaining scientific integrity.

- NWO asks all researchers who receive funding from NWO to commit to the principles of the Netherlands Code of Conduct on Scientific Conduct and to act accordingly.

- NWO has opened a research integrity desk where complaints about research integrity can be filed.

Anyone can file a complaint about research integrity against a researcher who is or was funded by NWO or about an applicant who is still in the procedure. NWO believes that such a research integrity desk will be a more accessible way for someone for filing a complaint than at their own university, since most of the researchers funded by NWO are employed at the Dutch universities.

Margot Custers | Nederlandse Organisatie voor Wetenschappelijk Onderzoek | Afdeling Voorlichting en Communicatie | Telefoon: 070 344 07 41 | m.custers@nwo.nl | Postadres:

Postbus 93138, 2509 AC Den Haag | Bezoekadres: Laan van Nieuw Oost-Indië 300, 2593
CE Den Haag | Werkdagen: ma, di, wo, do | Woordvoering buiten kantooruren: 070 344
0729

13.3 Anhang 3: Fragebogen – Umfrage: Deutsche Gesellschaft für Psychologie

Seite 01

Sehr geehrtes Mitglied der Deutschen Gesellschaft für Psychologie!

Vielen Dank, dass Sie sich Zeit für die Beantwortung meiner Umfrage nehmen.

Im Rahmen meiner Diplomarbeit an der Fakultät für Psychologie der Universität Wien, beschäftige ich mich mit dem Thema: „**Die Auswirkungen des Betrugsfalls Diederik Stapel auf die Wissenschaft**“.

Zur Erinnerung möchte ich die wichtigsten Fakten noch einmal kurz zusammenfassen:

Der niederländische Sozialpsychologe, Universitätsprofessor, Wissenschaftler und Dekan, **Diederik A. Stapel**, veröffentlichte im Laufe seiner Karriere insgesamt 137 Publikationen in den renommiertesten Fachzeitschriften, wobei sich herausstellte, dass er bei mindestens 55 Artikeln, in denen er Erst- oder Co-Autor war, manipuliertes Datenmaterial verwendete. Im November 2011 wurde Stapels jahrelanger Betrug durch einige seiner StudentInnen aufgedeckt und dessen Ausmaß von einer Untersuchungskommission geprüft.



Diederik A. Stapel

Die Deutsche Gesellschaft für Psychologie, insbesondere mit Ihnen als Mitglied der Fachgruppe für Sozialpsychologie, stellt für mein Forschungsinteresse eine ideale Plattform dar, um mittels dieses Fragebogens, Reaktionen und Stellungnahmen von ExpertInnen im Bereich der Sozialpsychologie zum Fall Stapel zu erheben.

Ihre Angaben sind dabei vollkommen anonym, Ihre Antworten können nach Beendigung des Fragebogens nicht mit Ihrer IP-Adresse in Verbindung gebracht werden. Es gibt keine richtigen oder falschen Antworten, ich bitte Sie lediglich, ihre persönliche Meinung kundzutun.

Die Befragung beinhaltet 12 Fragen und wird ca. 10 Minuten in Anspruch nehmen.

Ich bedanke mich vielmals für Ihre Unterstützung!

Lisa Meisner

Wenn Sie über die Ergebnisse der Umfrage informiert werden wollen, dann senden Sie bitte eine E-Mail an die unten genannte Adresse.

1. Wie intensiv haben Sie den Fall Diederik Stapel in den Medien verfolgt?

- ☐ gar nicht
- ☐ ab und zu
- ☐ des Öfteren
- ☐ sehr intensiv

2. Als Mitglied der Deutschen Gesellschaft für Psychologie, wie sehr hat Sie dieser Fall emotional erregt?

- ☐ gar nicht
- ☐ ein wenig
- ☐ mittelmäßig
- ☐ sehr

3. Für wie weit verbreitet halten Sie das Manipulieren und Fabrizieren von Datensätzen in der Wissenschaft?

- ☐ äußerst selten vorkommend
- ☐ ab und zu vorkommend
- ☐ kommt häufig vor
- ☐ kommt sehr häufig vor

4. Halten Sie die Psychologie als wissenschaftliche Disziplin für anfälliger für Forschungsbetrug als andere wissenschaftliche Disziplinen?

☐ ja

Warum?

☐ nein

☐ weiß nicht

5. Sehen Sie den Betrug Diederik Stapels als Einzelfall in der Psychologie an?

- ☐ auf jeden Fall
- ☐ ich bin mir nicht sicher
- ☐ es gibt noch ein paar mehr Betrugsfälle
- ☐ es gibt noch wesentlich mehr Betrugsfälle

6. Wo sehen Sie hauptsächlich die Gründe dafür, dass Diederik Stapels wissenschaftliches Fehlverhalten jahrelang unentdeckt blieb?

Bei dieser Frage sind Mehrfachnennungen möglich

Stapels zugeschriebenen Persönlichkeitsfaktoren, die kein wissenschaftliches Fehlverhalten vermuten ließen (z.B.: charismatisch, vertrauenswürdig, ehrgeizig, charmant, freundlich, engagiert etc.)

- ☐ Stapels hoher Macht- und Autoritätsposition als Universitätsprofessor, Dekan und anerkannter Wissenschaftler
- ☐ den bestehenden Forschungsmethoden in der Psychologie (z.B.: geringe Kontrollen in der Datenerhebung, keine öffentliche Bereitstellung der Daten etc.)
- ☐ dem System der Wissenschaft (z.B.: fehlende Förderung von Replikationen, Mängel im bestehenden Peer-Review Prozess etc.)

7. Wie sehr denken Sie, hat der Fall Diederik Stapel den Ruf der Sozialpsychologie geschädigt?

- ☐ gar nicht
- ☐ ein wenig
- ☐ mittelmäßig
- ☐ sehr

8. Mussten Sie sich als Mitglied der DGP in Ihrem beruflichen oder privaten Umfeld seit Bekanntwerden des Betrugsfalls jemals für die angewandten Methoden in der (sozial-) psychologischen Forschung und in der Folge für Ihre persönliche berufliche Tätigkeit rechtfertigen?

☐ ja

inwiefern?

☐ nein

☐ weiß nicht

Seite 04

9. Welche langfristigen Konsequenzen denken Sie, hat der Fall Stapel für die wissenschaftliche Disziplin der Sozialpsychologie?

10. Hat Sie der Betrugsfall Stapel zu konkreten Verhaltensänderungen in Ihrer Lehre/Forschungstätigkeit bewegt?

☐ ja

welche?

☐ nein

☐ weiß nicht

11. Wären Sie bereit, aufkommende Veränderungsvorschläge zur Prävention wissenschaftlichen Betrugs (z.B. durch die Veröffentlichung und Bereitstellung eigener Forschungsdatensätze) aktiv zu unterstützen?

☐ ja

☐

nein

warum nicht?

☐ weiß nicht

12. Auf welche Art und Weise sollten mögliche Veränderungen im wissenschaftlichen System umgesetzt werden?

- ☐ Wissenschaftler sollen sich durch die Veranschaulichung allgemeiner Regeln guter wissenschaftlicher Praxis dazu verpflichtet fühlen, diese Regeln einzuhalten
- ☐ Die strikte Umsetzung wissenschaftlicher Standards* sollte Voraussetzung für den Antrag auf Publikation eines Artikels bei einer Fachzeitschrift sein
- ☐ Es sollen keine Veränderungen umgesetzt werden
- ☐ Weiß nicht

*z.B.: das Berichten aller Variablen in einer Studie, die a-priori Festlegung einer Regel für das Beenden der Datensammlung, das Berichten aller experimentellen Versuchsbedingungen inklusive fehlgeschlagener Manipulationen, das Berichten von Ergebnissen sowohl vor als auch nach dem Ausschluss von Versuchspersonen/Daten, das Berichten von Ergebnissen mit und ohne der Berücksichtigung einer möglichen Kovariate etc.

Letzte Seite

Danke für Ihre Teilnahme!

Ich möchte mich ganz herzlich für Ihre Mithilfe bedanken.

Für weitere Fragen stehe ich Ihnen gerne unter der folgenden Adresse zur Verfügung:

a0707669@unet.univie.ac.at

Fenster schließen

Lisa Meisner, a0707669@unet.univie.ac.at, Universität Wien

14 Lebenslauf



Persönliche Daten

NAME: Lisa Carolin Meisner

GEBURTSDATEN: 16.07.1989, Zwettl, Niederösterreich

FAMILIENSTAND: ledig

NATIONALITÄT: Österreich

Schulbildung

1995 – 1999 Volksschule Zwettl

1999 – 2007 Bundesgymnasium und Bundesrealgymnasium Zwettl
(Absolvierung der Reifeprüfung 2007)

2007 – 2008 Bakkalaureatsstudium „Bildungswissenschaften“
(Abbruch aus persönlichen Gründen 2008)

seit 2007 Psychologiestudium an der Universität Wien
(Absolvierung des 1. Abschnitts 2010)

ab 2014 Postgraduelle Ausbildung zur Klinischen- und Gesundheitspsychologin

Bisherige Berufserfahrung

2005 – 2007 Ferialpraktika bei lokalen Instituten
(Kindernest Hilfswerk Zwettl, private Kinderbetreuung, Malerei Hobauer)

2010 – 2011 Geringfügige Beschäftigung bei MANPOWER
(Bereich Service und Eventgastronomie)

2012 6 – wöchiges Pflichtpraktikum BBRZ Leopoldau
(ambulantes Rehabilitationszentrum für seelische Gesundheit)

2013 – 2014 Geringfügige Beschäftigung bei der Piaristenprovinz St. Thekla und Maria Treu Wien
(Hortbetreuung)

Besondere Kenntnisse

Sprachen:	Deutsch – Muttersprache Englisch – fließend in Wort und Schrift (europäisches Niveau B2) Französisch – Grundkenntnisse (europäisches Niveau A2)
IT-Kenntnisse:	Microsoft Office Paket Windows7/Vista/XP SPSS
Weiteres:	Führerschein Klasse B Erste Hilfe Kurs NOEVV Volleyball Schiedsrichter Lizenz C

Hobbies

Tätigkeiten:	Mitglied des Sparkasse Zwettl Volleyballteams
Interessen:	Sport (Volleyball, Kraft- und Ausdauersport, Thai-Boxen) Filme und Kino Lesen Kochen und Essen Reisen Unternehmungen im Freundeskreis