



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Wenn Publikationsbias und zu wenig Power in einem
Forschungsfeld vorliegen. Eine meta-metaanalytische
Studie zum decline effect in der Meta-Forschung des
JPSP

verfasst von / submitted by

Stefan Keinrath, BSc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 840

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Psychologie UG2002

Betreut von / Supervisor:

Ass.-Prof. Mag. Dr. Jakob Pietschnig

Mitbetreut von / Co-Supervisor:

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich im Prozess der Erstellung meiner Masterarbeit begleitet und unterstützt haben.

Zuallererst gilt hier mein Dank meinem Erstbetreuer Ass.-Prof. Mag. Dr. Jakob Pietschnig, der mich nicht nur zu dem Thema meiner Masterarbeit inspirierte, sondern mich auch durch seine konstruktive Kritik und seine Hilfestellungen tatkräftig unterstützte. Vor allem möchte ich mich für die allzeits gute Erreichbarkeit per Email bedanken, wodurch eine unkomplizierte Kommunikation auch zu Zeiten der Pandemie möglich war.

Zudem möchte ich mich bei Frau Magdalena Siegel, BSc. MSc. dafür bedanken, dass sie mir den R-Code zur Berechnung des Meta-Levels bereitstellte und somit einen wichtigen Grundbaustein zu meiner statistischen Analyse beigesteuerte.

Mein Dank gilt außerdem meiner Kommilitonin und guten Freundin Junia Eder, die sich einerseits dazu bereit erklärte, die Rolle der zweiten Kodiererin zu übernehmen und mich andererseits, vor allem während meines Kodierungsprozesses, mit ihrem Wissen unterstützte.

Zu guter Letzt möchte ich mich bei meiner Partnerin Hjördis Kasamas bedanken, die mir im Prozess der Erstellung jederzeit zur Seite stand und mich dabei unterstützte, immer fokussiert und motiviert zu bleiben. Zusätzlich dazu bin ich ihr sehr dankbar dafür, dass sie sich die Zeit nahm, meine Arbeit Korrektur zu lesen und mir bei Formulierungsproblemen jederzeit beistand.

Stefan Keinrath

Wien, 14.05.2022

Abstract

Ein Problem vieler empirischer wissenschaftlicher Studien ist die nicht-Replizierbarkeit ihrer Ergebnisse. Ein daraus entstehendes Artefakt ist der sogenannte decline effect, welcher eine Reduktion der Effektstärke über die Zeit abbildet. In der Literatur wird vor allem angenommen, dass der decline effect durch Publikationsbias, Effektverschätzung und Effektverfälschung entsteht. Nach bisherigen Forschungsergebnissen scheint die Sozialpsychologische- und Persönlichkeitsforschung besonders von diesen Faktoren betroffen zu sein, weshalb es das Ziel der hier vorliegenden Studie ist, den decline effect in diesem Forschungsfeld zu erfassen. Anhand meta-metaanalytischer Verfahren wurde die Stärke und Häufigkeit des decline effects erfasst und anhand multipler und einfacher Meta-Metaregressionen mögliche Ursachen für sein Vorliegen untersucht. Die Ergebnisse zeigten, dass decline effects (Crude differences: $N = 25$, $r = .25$; jährliche Effektveränderung: $N = 20$, $r = .01$) im Vergleich zu Effektanstiegen (Crude differences: $N = 11$, $r = .01$; jährliche Effektveränderung: $N = 14$, $r < .01$) häufiger auftraten und stärker ausfielen. Zusätzlich zeigte sich, dass die durchschnittliche Power der Primärstudien bei 46% lag und nur in 5 von 34 untersuchten Metaanalysen keine Anzeichen für Publikationsbias gefunden werden konnten. Anhand der hier vorliegenden Ergebnisse und den Erkenntnissen aus der Meta-Metaregression lässt sich konkludieren, dass der decline effect vor allem auf die Effektüberschätzung und die Überrepräsentation jener Effekte in der Literatur durch den Publikationsbias zurückgeführt werden kann. Durch eine stärkere Fokussierung auf Präregistrierung und einem Umdenken in der Forschungscommunity, bezüglich dem Fokus auf Signifikanz und der Wichtigkeit von Replikationsstudien, könnte diesem Missstand entgegengewirkt werden.

Schlagwörter: decline effect, Publikationsbias, Meta-Metaanalyse, Sozialpsychologie, Replikationskrise

Abstract

A problem of many empirical scientific studies is the non-replicability of their results. One artefact that arises from this is the so-called decline effect, which is defined as a reduction in effect size over time. In the literature it is postulated that the decline effect arises mainly from publication bias, effect overestimation and scientific fraud. According to previous findings, it seems that the research field around social psychology and personality is particularly affected by these factors, which is why the aim of the present study is to capture the decline effect in this research field. Using meta-meta-analytical procedures, the strength and frequency of the decline effect was calculated and possible explanations for its presence were examined using multiple and single meta-meta-regressions. The results showed that decline effects (crude differences: $N = 25$, $r = .25$; annual effect change: $N = 20$, $r = .01$) compared to effect increases (Crude differences: $N = 11$, $r = .01$; annual effect change: $N = 14$, $r < .01$) occurred more frequently and were stronger. In addition, the average power of the primary studies was 46% and no evidence of publication bias was found in only 5 out of 34 meta-analyses. Based on the results presented here and the findings from the meta-meta-regression, it can be concluded that the decline effect can be attributed primarily to the effect overestimation and the overrepresentation of those effects in the literature due to publication bias. A stronger focus on pre-registration and a shift in the research community regarding the focus on significance of effects and the importance of replication studies could counteract this problem.

Keywords: decline effect, publication bias, meta-meta-analysis, social psychology, replication crisis

Einleitung

Die Replikation von publizierten Forschungserkenntnissen stellt einen wichtigen Bestandteil des empirisch wissenschaftlichen Vorgehens dar (Open Science Collaboration, 2015). Denn nur durch das vermehrte replizieren eines Effektes in verschiedenen Populationen und durch unterschiedlicher Forscher*innen ist es möglich, Sicherheit über das Vorliegen eines Populationseffektes zu erlangen (Lamal, 1990). Im Publikationsprozess scheint diese Wichtigkeit für die empirische Forschung jedoch verloren gegangen zu sein (Nosek & Errington, 2020). Stattdessen lässt sich die Annahme aufstellen, dass Journals ihren Fokus hauptsächlich auf originelle, aufregende und vor allem signifikante Ergebnisse zu legen scheinen (Franco et al., 2014; Margraf, 2015). Studien, die signifikante Ergebnisse berichten weisen dabei nicht nur eine höhere Wahrscheinlichkeit auf, publiziert werden, sondern sie durchlaufen den Publikationsprozess im Durchschnitt auch schneller als Studien, die keine signifikante Ergebnisse berichten (Ioannidis, 1998; Franco et al., 2014). Unabhängig von der Richtung oder der Stärke eines Effektes gelten Replikationsstudien im Allgemeinen als uninteressant, da sie keine neuen Erkenntnisse und originellen Ideen berichten und stoßen vor allem bei einflussreichen Journals häufig auf Ablehnung (Franco et al., 2014; Lakens, 2018; Margraf, 2015; Roberts, 2011). Zusätzlich wird vermehrt darüber berichtet, dass Forscherinnen und Forscher unter starken Publikationsdruck leiden, da Fördergelder, Arbeitsplätze und Aufstiegschancen davon abhängen, dass sie möglichst viel und vor allem in möglichst prestigeträchtigen Journals publizieren (Margraf, 2015; Tjldink et al., 2014; Yuan et al. (2013). Um im Wettbewerbskontext zu bestehen, greifen viele Personen zu sogenannten *Questionable Research Practices* (QRP's), wie beispielsweise selektives Berichten von Ergebnissen oder auch der selektive Ausschluss von Datenpunkten, um ein signifikantes Ergebnis zu erzielen (Fiedler & Schwarz, 2016; Franco et al., 2016; John et al., 2012; Margraf, 2015). Aber auch drastischere Schritte wie Datenfälschung finden Anwendung, was aus der steigende Zahl an Studienrückrufen und aus diversen Umfragen abgeleitet werden kann (Fiedler & Schwarz, 2016; John et al., 2012; Margraf, 2015; Tjldink et al., 2014). Zieht man all diese genannten Aspekte in Betracht, so kann dies sehr wohl dazu anregen, die Stabilität des theoretischen Grundgerüsts eines Forschungsbereichs zu hinterfragen.

Unter anderem aus diesem Beweggrund schlossen sich eine Vielzahl von Forscherinnen und Forschern zusammen und führten 100 Replikationen zu verschiedenen Studien aus drei sehr wichtigen Journals im Bereich der Psychologie durch. Je nachdem welche Methodik verwendet wurde um eine Replikation als erfolgreich oder nicht einzustufen (z. B. Effekt der Replikation muss signifikant ausfallen oder Summary effect von

Originalstudie und Replikation muss signifikant ausfallen), konnten etwas mehr als ein Viertel bis etwas mehr als die Hälfte aller Studien repliziert werden (Open Science Collaboration, 2015). Aber nicht nur die Psychologie ist von dieser sogenannten Replikationskrise betroffen, sondern auch andere Forschungsbereiche, wie beispielsweise die Medizin, leiden unter diesem Problem (Ioannidis, 2005; Weir, 2015).

Ein Artefakt, welches aus dieser Problematik heraus entstanden ist, ist der sogenannte *decline effect*. Hierbei handelt es sich um das erfassbare Abbild jenes Problems, welches die Replikationskrise beschreibt, nämlich, dass Effekte und Phänomene nicht in dem Ausmaß vorzuliegen scheinen, wie es in Studien, die diese erstmals beschrieben haben, berichtet wurde. Anders ausgedrückt, beschreibt der decline effect also eine Veränderung der Stärke eines Effektes über die Zeit in Form einer Effektreduktion. Er stellt dadurch ein Maß der Effektverschätzung in einem Forschungsfeld dar und bildet dadurch einen messbaren Wert für die Replizierbarkeit von Effekten in diesem Feld. Untersucht man den Effekt mittels meta-metaanalytischer Verfahren, so kann man anhand der Häufigkeit eines decline effect feststellen, wie viele der berichteten und als geltend angenommenen Effekte nicht replizierbar sind und anhand der Stärke des ausfallenden decline effect kann zusätzlich noch eine Aussage über die Größe der Effektverschätzung getroffen werden. Zusätzlich können anhand von Zusammenhangsanalysen mögliche Ursachen für die Probleme in der Replizierbarkeit der Ergebnisse in einem Forschungsfeld untersucht werden (Protzko & Schooler, 2017).

Es wurden bereits einige Forschungsfelder auf das Vorliegen und die Stärke des vorliegenden decline effect überprüft, Erkenntnisse für die Forschung im Bereich der Sozialpsychologie blieben bisher jedoch aus (Gong & Jiao, 2019; Gong et al., 2021; Jiao et al., 2021; Pietschnig et al., 2019). In der Literatur gibt es aber durchaus einige Hinweise, die darauf schließen lassen, dass die Forschung im Bereich der Sozialpsychologie besonders von einem decline effect betroffen sein könnte (Banks et al., 2016; Fanelli et al., 2017; Franco et al., 2014; Franco et al., 2016). Das Ziel der hier vorliegenden Arbeit ist es dementsprechend, diese Wissenslücke zu füllen und den Forschungsbereich der Sozialpsychologie und Persönlichkeit auf den decline effect zu überprüfen und zusätzlich dazu mögliche Erklärungsansätze für das Vorliegen und die Stärke dieses Effektes zu bieten.

Nomenklatur des Decline Effect:

Um den decline effect besser verstehen zu können ist es in erster Linie wichtig die Annahme abzulegen, dass ein decline effect von vornherein mit Datenfälschung und schlechter Forschungsethik einhergeht (Protzko & Schooler, 2017). Zwar wird er in der Forschung meist gemeinsam mit Publikationsbias und Effektüberschätzungen genannt, es gibt

aber noch eine Vielzahl anderer Gründe, die die Replizierbarkeit von Effekten einschränken können (Fanelli et al., 2017; Gong & Jiao, 2019; Nujiten et al., 2020; Pietschnig et al., 2019; Protzko & Schooler, 2017). Um dies zu verdeutlichen, haben Protzko und Schooler (2017) den decline effect in vier Subgruppen unterteilt, die sich bezüglich ihrer Ursachen zum Teil stark voneinander unterscheiden. So berichten sie zum einen vom *false positive decline effect* und dem *inflated decline effect*. Diese beiden Gruppen sind in ihren angenommenen Auslösern ident, sie werden aber aufgrund der Stärke der erfassten Effektreduktion unterschieden. In beiden Fällen wird angenommen, dass der decline effect dadurch entsteht, dass ein Forschungsfeld zu Beginn hauptsächlich von überschätzten Effekten geprägt ist und erst über den Lauf der Zeit, durch die Publikation von nicht signifikant ausfallenden Replikationsstudien, die Ergebnislage korrigiert wird. Der Unterschied zwischen diesen beiden decline effect Arten ist jedoch, dass beim false positive decline effect gar kein zugrundeliegender Populationseffekt vorzufinden ist, währenddessen der inflated decline effect nur eine Effektüberschätzung eines kleineren Populationseffektes darstellt (Protzko & Schooler, 2017).

Im Gegensatz dazu lässt sich der *underspecified decline effect* in erster Linie nicht durch Effektüberschätzung und Publikationsbias erklären. In diesem Fall liegt nämlich ein Effekt in der Population in derselben Größe vor, wie er durch die Originalstudie berichtet wird, aber Replikationsversuche scheitern, da wichtige Informationen für eine erfolgreiche Replikation fehlen. Aufgrund mangelnder Protokollierung des Forschungsvorgehens in Initialstudien, kann es somit passieren, dass darauffolgende Replikationsstudien fälschlicherweise annehmen, dass der Effekt nicht replizierbar ist, was sich im Gesamtbild als decline effect abbilden lässt (Protzko & Schooler, 2017)

Effekte können jedoch auch aus ihrer Natur heraus beginnen abzunehmen, wodurch sich in Untersuchungen ein decline effect feststellen lässt. Die Ursache für diese Effektreduktion ist jedoch weder auf Effektverschätzungen, noch auf Fehler in der Protokollierung und Durchführung einer Studie oder auf problematische Forschungs- und Publikationspraktiken zurückzuführen, sondern auf tatsächliche, „natürliche“ Veränderungen eines Effektes (Protzko & Schooler, 2017). Protzko und Schooler (2017) bezeichnen diese Kategorie als *genuinely decreasing decline effects*.

Neben diesen vier formulierten Kategorien werden in der Forschung noch einige andere Spezialfälle des decline effects genannt. Im speziellen handelt es sich dabei um die sogenannten *early extremes*, den *US effect* und den *Proteus Effekt* (Fanelli & Ioannidis, 2013; Fanelli et al., 2017; Ioannidis & Trikalinos, 2005; Pietschnig et al., 2019). Bei ersterem

Phänomen wird, im Gegensatz zur Definition des decline effect, angenommen, dass vor allem der berichtete Effekt der Initialstudie einen überschätzten Extremfall in dem Forschungsfeld darstellt, die breite Masse der darauffolgenden Replikationsstudien berichten diesen Effekt jedoch nicht in demselben Ausmaß (Fanelli et al., 2017; Ioannidis & Trikalinos, 2005). Der US effect ist andererseits durch die Annahme gekennzeichnet, dass die verfälschte Darstellung von Effekten vor allem durch Studien aus den USA, wo der Publikationsdruck für Forscher*innen besonders hoch zu sein scheint, entsteht (Fanelli & Ioannidis, 2013). Ebenso wird der sogenannte Proteus Effekt als eine Extremform des decline effect angenommen (Pietschnig et al., 2019). In diesem Fall hat die Initialstudie nämlich nicht nur den Effekt überschätzt, sondern die Wirkrichtung des Populationseffektes falsch erfasst (Pietschnig et al., 2019).

Da es für die Überprüfung des underspecified decline effect und des genuinely decreasing decline effect eine andere methodische Herangehensweise benötigt, als sie hier geplant ist, werden diese beiden Arten des decline effect nicht weiter behandelt. Dementsprechend wird sich die hier vorliegende Arbeit in weiterer Folge vor allem darauf fokussieren, ob durch die gefundenen Ergebnisse, die Annahme gerechtfertigt ist, dass das Forschungsfeld rund um Persönlichkeit und Sozialpsychologie von überschätzten Effekten geprägt ist, die durch problematische oder fehlerhafte Forschungs- und Publikationspraktiken (z.B.: kleine Stichproben, QRP's, etc.) hervorgerufen wurden.

Ursachen für den Decline Effect

Publikationsbias

Ein Faktor, der zwar nicht als unmittelbarer Ursprung für den decline effect angesehen werden kann, jedoch im Wesentlichen zu dessen Entstehung beiträgt, ist der Publikationsbias (Protzko & Schooler, 2017). Es handelt sich dabei um die verzerrte Darstellung des wissenschaftlichen Erkenntnisstandes, welche durch das selektive Publikationsverhalten von Autor*innen und Editor*innen ausgelöst wird (De Bruine & Della Sala, 2015; Ioannidis, 1998; Protzko & Schooler, 2017). In einigen Studien konnte gezeigt werden, dass Arbeiten mit signifikanten und großen Effekte schneller und mit einer höheren Wahrscheinlichkeit publiziert werden als solche, die keine signifikanten Ergebnisse berichten (Ioannidis, 1998; Franco et al., 2014). Dies kann einerseits dadurch erklärt werden, dass originelle neue Ergebnisse als interessanter und dahingehend als publikationswürdiger als Replikationsstudien eingeschätzt werden (De Bruine & Della Sala, 2015; Margraf, 2015). Andererseits wird häufig angenommen, dass durch neue und große Erkenntnisse

Forschungstrends in wissenschaftlichen Feldern losgetreten werden können, von denen sowohl Editor*innen als auch Autor*innen in unterschiedlicher Weise profitieren können (De Bruine & Della Sala, 2015; Gandevia, 2018; Protzko & Schooler, 2017). Um dem vorliegenden Forschungstrend nicht zu schaden, werden nicht signifikante Effekte dementsprechend gar nicht oder erst sehr viel später publiziert (De Bruine & Della Sala, 2015; Protzko & Schooler, 2017). Auf der anderen Seite bekommen Studien, die große und signifikante Effekte berichten, durch den Trend nochmals mehr Aufmerksamkeit und werden von anderen Studien zitiert, wodurch ihr Einflussreichtum noch weiterwächst (Fanelli et al., 2017; Jannot et al., 2013). Dieses Phänomen bezeichnet man auch als *Citation Bias* (Jannot et al., 2013).

Dies führt letztlich zu einer verzerrten Darstellung der Erkenntnislage auf mehreren Ebenen (Cooper et al., 2019). Denn zum einen entsteht durch den Publikationsbias eine Asymmetrie in der Verteilung der Effekte, wobei vor allem die nicht signifikanten Effekte von Studien mit kleinen Stichproben fehlen, zum anderen kommt es in der publizierten Literatur zu einer Überrepräsentation von signifikanten Ergebnissen (Simonsohn et al., 2014; Sterne et al., 2000). Da angenommen wird, dass diese Asymmetrie vor allem dadurch entsteht, dass kleinere, nicht signifikante Studien nicht publiziert werden und somit als graue Literatur gelten, wird diese Bias-Muster häufig als *Grey Literature Bias* bezeichnet (Fanelli et al., 2017). Der publizierte Erkenntnisstand stellt dementsprechend nur eine Teilsumme aller Erkenntnisse dar und ist in Richtung größerer und signifikanter Effekte verzerrt (Fanelli et al., 2017).

Methoden zur Erfassung von Publikationsbias machen sich diese Ungleichverteilung in der Größe der Effekte und der p -Werte zu Nutze, um einen Forschungsbereich auf metaanalytischer Ebene auf das Vorliegen eines Grey Literature Bias zu überprüfen (Cooper et al., 2019). Über die Jahre hinweg wurde eine Vielzahl unterschiedlicher Methoden zur Erfassung von Publikationsbias entwickelt, wobei sie sich grob in drei Gruppen unterteilen lassen (Cooper et al., 2019; Siegel et al., 2021). Je nachdem, wie eine Methode die Daten auf das Vorliegen des Grey Literature Bias überprüft, lässt sie sich entweder den Methoden zuordnen, die die Datenlage auf Asymmetrie in der Verteilung der Effekte überprüfen, die Verteilung der p -Werte untersuchen oder aber einen Selection Model Approach darstellen (Cooper et al., 2019; Siegel et al., 2021). Jede Methode hat zum einen ihre eigenen Voraussetzungen, die erfüllt sein sollten, um eine fehlerfreie Anwendung zu ermöglichen und zum anderen auch ihre eigenen, nach wissenschaftlichen Konventionen festgelegten, Schwellenwerte, anhand derer festgelegt wird, ob von einem Publikationsbias auszugehen ist

oder nicht (Cooper et al., 2019; Siegel et al., 2021). Da manche der Methoden sehr strenge Voraussetzungen vorgeben, schlägt Cooper et al. (2019) vor, nach Triangulation vorzugehen und mehrere unterschiedliche Verfahren anzuwenden. Es wird dabei angenommen, dass man so einen möglichen Powerverlust durch die zusätzliche Anwendung der anderen Erhebungsmethoden ausgleichen kann (Cooper et al., 2019). Für eine genauere Behandlung der einzelnen Methoden ist der Artikel von Siegel et al. (2021), insbesondere dem dazugehörigen Supplement 1, welches auf der OSF-Projektseite der Arbeit zu finden ist, zu empfehlen.

Effektverschätzung

Währenddessen der Publikationsbias vor allem dazu führt, dass signifikante und große Effekte im wissenschaftlichen Feld überrepräsentiert sind, führen die unterschiedlichen Faktoren der Effektverschätzung letztlich in Kombination dazu, dass der publizierte Forschungsstand nicht replizierbar ist, womit es zu einem decline effect kommt (Protzko & Schooler, 2017).

Zu kleine Stichprobe. Eine zu kleine Stichprobe kann aus mehreren Gründen zu einer Effektverschätzung führen und damit das Auftreten eines decline effect begünstigen (Nuijten et al., 2020). Zum einen ist eine Studie mit einer kleineren Stichprobe ungenauer in der Erfassung eines Effektes, wodurch der Sampling Error einen größeren Einfluss auf den Effekt hat (Field, 2016; Nuijten et al., 2020). Anders formuliert bedeutet dies, je kleiner die Stichprobe für eine Studie ausfällt, desto mehr Einfluss haben Ergebnisse jeder einzelnen Person innerhalb der Stichprobe (Field, 2016). Die Ergebnisse sind dahingehend nicht mehr repräsentativ für die Allgemeinbevölkerung und der errechnete Effekt ist dadurch von einer Verzerrung betroffen (Field, 2016). Kommt es durch den Einfluss des Sampling Errors zu einer Überschätzung des untersuchten Effektes, so wird dieses Phänomen als *small-study effect* bezeichnet (Sterne et al., 2000).

Durch eine zu niedrige Teilnehmerzahl einer Studie kann es zudem auch zu einem Powerproblem kommen (Field, 2016). Die Power einer Studie hängt nämlich von der Größe des Populationseffektes, dem gewählten Signifikanzniveau und der Stichprobengröße einer Studie ab. Je kleiner der zugrunde liegende Effekt ist und je niedriger man das Signifikanzniveau einer Studie setzt, desto größer muss die Stichprobe ausfallen (Field, 2016). Fällt die Stichprobengröße nun zu klein aus, so bedeutet dies, dass es aufgrund der mangelnden Präzision, unwahrscheinlicher ist, dass eine Studie einen Effekt erfassen kann und Effekte signifikant ausfallen (Button et al., 2013; Ioannidis, 2005). Falls in einer Studie

doch ein signifikanter Effekt erfasst wurde, so nimmt die Wahrscheinlichkeit, dass dahinter ein zugrundeliegender Populationseffekt vorliegt, immer weiter ab, je kleiner die Stichprobe und die damit einhergehende Power der Studie ist (Button et al., 2013). Letztlich kann eine zu niedrige Power aber auch zu einer Effektüberschätzung führen, wenn tatsächlich ein zugrundeliegender Populationseffekt vorliegt, da die Studie nur dann signifikant ausfällt, wenn ein großer Effekt erfasst wird. Ist die Ausprägung des Effektes in der Gesamtpopulation jedoch nur klein, so wird dadurch der Effekt, durch die jeweilige Studie mit zu wenig Power, überschätzt (Button et al., 2013).

Wenn nun ein Forschungsfeld eine Kombination aus Publikationsbias und zu kleinen Stichproben aufweist, so ist ein decline effect äußerst wahrscheinlich (Protzko & Schooler, 2017). Denn durch das bevorzugte publizieren von signifikanten positiven Ergebnissen, haben auch Studien mit zu niedriger Power oder zu kleiner Stichprobe eine höhere Wahrscheinlichkeit publiziert zu werden, solange sie einen signifikanten Effekt präsentieren (Button et al., 2013).

Die gewählte Effektstärkenmetrik. Ein weiterer Faktor, der zu einer Effektverschätzung führen könnte, ist die Effektstärkenmetrik, die für die Berechnung der Metaanalyse gewählt wurde (Bishara & Hittner, 2015; Magnusson, 2014; Lin, 2018; Rousselet, 2018). Zwar wurde bisher noch kein direkter Zusammenhang zwischen diesem Faktor und dem decline effect hergestellt, einige Simulationsstudien berichten jedoch davon, dass sowohl Person r , als auch Cohen's d und Hedges g dazu neigen, Effekte zu überschätzen, wenn kleine Stichproben vorliegen oder die Daten nicht normalverteilt sind (Bishara & Hittner, 2015; Magnusson, 2014; Lin, 2018; Rousselet, 2018).

Effektverfälschung

Neben dem Problem der Bevorzugung von großen, interessanten und signifikanten Effekten im Publikationsprozess, wird wissenschaftliches Arbeiten immer kompetitiver und hostiler für Wissenschaftler*innen (Gandevia, 2018; Margraf, 2015). Der Druck auf Forscher*innen möglichst viel und in möglichst einflussreichen Journals zu publizieren, wird als die momentan lenkende Kraft im Forschungsprozess angesehen, da Fördergelder, Arbeitsplätze und Aufstiegschancen davon abhängen (Gandevia, 2018; Margraf, 2015; Tjldink et al., 2014; Yuan et al. 2013). Die in der Forschung tätige Person steht somit häufig in einem Interessenskonflikt zwischen korrektem wissenschaftlichen Vorgehen und dem eigenen Karrierefortschritt (Gandevia, 2018). Denn um in einem Forschungsfeld, welches von Publikationsbias und Studien mit zu niedriger Power geprägt ist, möglichst erfolgreich

publizieren zu können, greifen Forscher*innen vermehrt zu sogenannten fragwürdigen Forschungspraktiken (QRP's; Banks et al., 2016b; Gandevia, 2018; John et al., 2012; Protzko & Schooler, 2017).

QRP's stellen per Definition fragwürdige wissenschaftliche Vorgehensweisen dar, die sowohl im Studiendesign, als auch in der Analyse und Berichterstattung der Ergebnisse angewandt werden können, um zugunsten einer wahrscheinlicheren Publikation der Arbeit, möglichst große und signifikante Effekte zu erzielen (Banks et al., 2016a). Eine besondere Aufmerksamkeit in der Literatur wird dem *selektiven Berichten* von signifikanten Ergebnissen, *p-Hacking*, *HARKing* und dem tatsächlichen Verfälschen von Daten zuteil (Banks et al., 2016a). Selektives Berichten bedeutet dabei, dass Forscher*innen nicht alle für eine Studie erhobenen Daten und Ergebnisse berichten, sondern nur jene Resultate publizieren, die tatsächlich auch signifikant ausfallen (Franco et al., 2016). Das *p-Hacking* beschreibt auf der anderen Seite ein Vorgehen, in dem ein Datensatz durch das Hinzufügen und Ausschließen von Datenpunkten so manipuliert wird, dass es zu signifikanten Ergebnissen kommt (Suter, 2020). *HARKing* (oder auch „... Hypothesizing after results are known“, Kerr, 1998, S. 196) steht wiederum für das Formulieren von Hypothesen nachdem bereits Analysen durchgeführt wurden und feststeht, welche Ergebnisse signifikant wurden und welche nicht (Kerr, 1998). Unabhängig davon, welche Methode letztlich gewählt wird, führt jede von ihnen zu einer Verfälschung des erfassten Effektes aufgrund einer Alphafehlerkumulierung und des zu starken Einflusses durch den Standardfehler (Franko et al., 2016; Kerr, 1998; Suter, 2020). Wie sich in einigen Studien gezeigt hat, ist die Anwendung solcher Methoden nicht die Ausnahme und nachdem im Publikationsprozess vor allem signifikante Ergebnisse bevorzugt werden, entsteht so ein Forschungsfeld, dass von Effektüberschätzungen und verfälschten Ergebnissen geprägt ist (Banks et al., 2016b; Franko et al., 2016; Fiedler & Schwarz, 2016; Protzko & Schooler, 2017). Replikationsstudien, die versuchen, jene Effekte erneut zu finden, scheitern dementsprechend, wodurch sich ein decline effect abzeichnet (Protzko & Schooler, 2017).

Der Einfluss von Psi

Neben all den möglichen Verzerrungen des Effektes, die statistisch erklärbar sind, gibt es auch einige Forscher*innen, die einen etwas unkonventionelleren Ansatz verfolgen (Protzko & Schooler, 2017; Rabeyron, 2020). Diese versuchen den decline effect anhand parapsychologischer Phänomene zu erklären, indem sie annehmen, dass entweder durch den Einfluss solcher Phänomene, wie beispielsweise Präkognition, das Auftreten von Effekten so

beeinflusst wird, dass es zu einem decline effect kommt oder aber der decline effect selbst einen Beweis dafür darstellt, dass in dem jeweiligen Forschungsfeld ein Psi-Phänomen erfasst wurde und der berichtete originale Effekt nur ein daraus entstandenes Artefakt darstellt (Rabeyron, 2020). Für einen genaueren Einblick in diese Thematik und alle Probleme, die damit einher gehen ist der Artikel von Rabeyron (2020) sehr empfehlenswert.

Der Decline Effect in der Sozialpsychologie

Der wissenschaftliche Erkenntnisstand zum decline effect wächst stetig und vor allem in den letzten vier Jahren hat die Forschung zu diesem Thema enorm zugenommen. Es gibt bereits eine Vielzahl an spezifischen und forschungsfeldübergreifenden Studien, in welchen, anhand meta-metaanalytischer Methoden, ein decline effect erfasst werden konnte (Fanelli et al., 2017; Gong & Jiao, 2019; Gong et al. 2021; Jiao et al., 2021; Nuijten et al., 2020; Pietschnig et al., 2019; van Aert et al., 2019). Erkenntnisse im Bereich der Sozialpsychologie sind bisher ausgeblieben, wobei es einige Anzeichen dafür gibt, dass auch dieses Forschungsfeld von einem decline effect betroffen sein könnte (Fanelli et al., 2017; Franco et al., 2014; Franco et al., 2016). Im Vergleich zu anderen wissenschaftlichen Feldern zeigten sich die Sozialwissenschaften als besonders von Bias-Mustern, wie dem small-study effect betroffen, aber auch Anzeichen von selektiven Berichten und Grey Literature Bias sind gegeben (Fanelli et al., 2017; Franco et al., 2014; Franco et al., 2016). Zudem wird berichtet, dass aufgrund mangelnder Power, nur ein Viertel aller sozialpsychologischen Primärstudien tatsächlich replizierbar ist (Schimmack, 2015).

Zusätzlich dazu steht eines der einflussreichsten Journals im Bereich der Sozialpsychologie, aufgrund fragwürdiger Publikationsentscheidungen sehr häufig in der Kritik (Lakens, 2018; Roberts, 2011).

Das Journal of Personality and Social Psychology unter Kritik

Das *Journal of Personality and Social Psychology* (JPSP) liegt nach einem Ranking des Web of Science auf dem dritten Platz der einflussreichsten Journals im Bereich der Sozialpsychologie (Clarivate, 2020). Umso eine größere Tragweite haben folglich die problematischen Journal Richtlinien, welche Replikationsstudien im Publikationsprozess eher hinten anreihen (Roberts, 2011). Darüber hinaus zeigen sich Tendenzen von *p*-hacking und bereits in mehreren Forschungsarbeiten konnten die nicht-Replizierbarkeit von Studien dieses Journals nachgewiesen werden (Lakens, 2018; Open Science Collaboration, 2015; Schimmack, 2015; Schimmack 2016). Zudem konnte bei einer Analyse der PsychINFO

Datenbank gezeigt werden, dass das JPSP, im Vergleich mit anderen Journals, die meisten Rückrufe von Primärstudien aufwies (Margraf, 2015). In einem speziellen Fall fiel das JPSP, aufgrund seiner Publikationspraktiken, besonders in Kritik (Lakens, 2018; Roberts, 2011). Als Bem (2011) seine Studie zum Thema Präkognitionen, einem parapsychologischen Phänomen, in einem derart einflussreichen Journal publizieren durfte, führte dies zu einem großen Aufruhr in wissenschaftlichen Forschungskreisen (Galak et al., 2012; Bem et al., 2016). Unter anderem aufgrund der skeptischen Haltung vieler Forscher*innen gegenüber parapsychologischen Phänomenen, kam es infolgedessen zu einer Vielzahl an durchgeführten Replikationsstudien zu dieser Arbeit (Rabeyron, 2020). Bem selbst hat diese Replikationsversuche von Beginn an unterstützt und sowohl Instruktionsmanuale, als auch alle nötigen Computerprogramme, für die Durchführung der Studie, gratis zu Verfügung gestellt (Bem et al., 2016). Entgegen den Aussagen einiger Editor*innen des JPSP, die in einem Editorial Comment ihre Hoffnungen auf Replikationsstudien aussprachen, wurden jegliche Versuche, diese im JPSP zu publizieren, abgelehnt (Aldhous, 2011; Judd & Gawronski, 2011; Roberts, 2011). Auf Anfrage, warum Replikationsstudien laufend abgelehnt wurden, antwortete der damalige Editor des JPSP “This journal does not publish replication studies, whether successful or unsuccessful” (Smith, 2011, zitiert in Aldhous, 2011). Hierdurch ist erneut die problematische Einstellung dieses Journals gegenüber Replikationsstudien ersichtlich.

Forschungsfragen und Hypothesen

Anhand des dargestellten theoretischen Hintergrundes lassen sich nun folgende Ziele für die hier vorliegende Arbeit ableiten. Da Erkenntnisse über den decline effect in der sozialpsychologischen und Persönlichkeitsforschung bis zum aktuellen Zeitpunkt ausblieben, in der Literatur jedoch eine Vielzahl an Hinweisen für dessen Vorliegen postuliert werden, ist es das Ziel dieser Studie, das Bestehen des decline effect und dessen Stärke in diesem Forschungsfeld zu erfassen (Fanelli et al., 2017; Franco et al., 2014; Franco et al., 2016). Zusätzlich soll die durchschnittliche Power einer Primärstudie in diesem Forschungsfeld errechnet werden und anhand von 9 unterschiedlichen Publikationsbias Methoden, die Meta-Forschung des JPSP auf Publikationsbias überprüft werden. Anhand verschiedener Merkmale der Metaanalysen und Initialstudien sollen mögliche Gründe für die Erfassung bzw. das Vorliegen des decline effect untersucht werden. Aufgrund dessen, dass sich in einigen Studien zeigte, dass die Heterogenitätsmaße τ^2 und I^2 vor allem durch die Stichprobengröße inkludierter Primärstudien beeinflusst wurden und größer ausfielen, war es ein weiteres Ziel

der hier präsentierten Arbeit, diese Erkenntnisse an den hier vorliegenden Daten zu überprüfen, da eine genaue Schätzung der Heterogenität in den Daten sowohl für die richtige Wahl des Verrechnungsmodells, als auch für die Interpretation der Publikationsbiasmethoden wichtig ist (Bornstein et al., 2009; Inhout et al., 2015; Rücker et al., 2008; Siegel et al., 2021).

Um den decline effect in der hier vorliegenden Studie zu operationalisieren, wird einerseits die Differenz zwischen dem Effekt der Initialstudie und dem Summary Effect (d.s. Crude differences), andererseits die Effektveränderung über die Zeit, berechnet. Dies ermöglicht die Erfassung von early extremes und dem decline effect.

Die folgenden Hypothesen wurden angenommen:

H1: Die durchschnittliche jährliche Effektveränderung zeigt einen decline effect.

H2: Es kommt zu einem Unterschied in der erfassten Stärke der errechneten Crude differences, abhängig davon ob Metaanalysen ihre Effekte als korrelatives Effektmaß oder als Effektmaß berichten, welches auf Mittelwerten basiert.

H3: Wenn Metaanalysen nur auf publizierten Daten aufbauen, führt dies häufiger zu Anzeichen von Publikationsbias, als wenn Metaanalysen zusätzlich auch graue Literatur inkludieren

H4: I^2 weist einen positiven Zusammenhang zur Gesamtstichprobengröße einer Metaanalyse, der durchschnittlichen Stichprobengröße einer Metaanalyse und der Anzahl inkludierter Stichproben auf.

H5: τ^2 weist einen negativen Zusammenhang zur Gesamtstichprobengröße einer Metaanalyse, der durchschnittlichen Stichprobengröße einer Metaanalyse und der Anzahl inkludierter Stichproben auf.

H6a: Der Initialstudieneffekt, der Impact Faktor des publizierenden Journals und die jährliche Zitationszahl der Initialstudie weisen einen positiven Zusammenhang mit den Crude differences auf, die Stichprobengröße der Initialstudie weist einen negativen Zusammenhang mit diesem auf und das Publikationsjahr der Initialstudie zeigt ebenfalls einen Zusammenhang mit den Crude differences.

H6b: Die Anzahl inkludierter Stichproben in einer Metaanalyse und die Inklusion grauer Literatur weisen einen positiven Zusammenhang mit den Crude differences auf und der Summary effect der Metaanalyse, ihre abgedeckte Zeitspanne und ihr Publikationsjahr zeigt einen Einfluss auf die Crude differences, wobei keine Wirkrichtung angenommen wird.

H7a: Der Initialstudieneffekt, der Impact Faktor des publizierenden Journals und die jährliche Zitationszahl der Initialstudie zeigen einen negativen Zusammenhang mit der jährlichen Effektveränderung auf, währenddessen die Stichprobengröße der Initialstudie einen positiven

Zusammenhang aufweist und das Publikationsjahr der Initialstudie einen Zusammenhang ohne angenommener Richtung zeigt.

H7b: Die Inklusion grauer Literatur wird negativ mit der jährlichen Effektveränderung zusammenhängen und der Summary effect der Metaanalyse, ihre abgedeckte Zeitspanne, ihr Publikationsjahr und die Anzahl inkludierter Primärstudien zeigt ebenfalls einen Einfluss darauf, wobei keine Wirkrichtung angenommen wird.

Methodik

Suchstrategie und Inklusions- und Exklusionskriterien

Für die vorliegende Arbeit wurde nach allen Metaanalysen gesucht, die im Journal of Personality and Social Psychology publiziert wurden. Hierfür wurde im Web of Science „SO=(Journal of Personality "AND" Social Psychology) AND TS=(meta-analy* OR research AND syntheses* OR systematic AND review)“ als Suchbegriff angewandt. Da das Journal of Personality and Social Psychology vollständig im Web of Science indexiert ist, wurde diese Suchstrategie als ausreichend betrachtet und keine anderen Datenbanken für die Suche herangezogen. Der oben genannte Suchbegriff wurde ein erstes Mal am 10.05.2021 angewandt und nach einem ersten Screening aller erhaltenen Studien, wurde am 02.10.2021 ein Update des Suchergebnisses durchgeführt. Insgesamt wurden somit 142 Studien mittels der Suchstrategie erfasst und in weiterer Folge auf Inklusionskriterien überprüft.

Um für die geplanten statistischen Analysen als geeignet zu gelten, mussten die einzelnen Studien mehrere Kriterien erfüllen, auf die nun in weiterer Folge eingegangen wird. Als erstes Kriterium galt, dass in einer Studie die Ergebnisse einer Metaanalyse berichtet werden mussten. Diese Metaanalyse musste auf den Daten mehrerer Stichproben, unterschiedlicher Forschungsarbeiten aufbauen und von den Autoren der jeweiligen Arbeit selbst berechnet worden sein. Wurde keine Metaanalyse gerechnet oder lediglich die Ergebnisse einer bereits publizierten Metaanalyse berichtet, kommentiert oder reanalysiert, ohne dabei neue zusätzliche Daten hinzuzufügen, so galt dies als Ausschlussgrund.

Zudem mussten die Effekte der Metaanalysen in einer traditionellen Effektstärkenmetrik berichtet werden, wozu im speziellen Cohen's d , Hedges g , Person r , Spearman ρ , Fisher's z , OR und $logOR$ zählten. Wenn in der jeweiligen Arbeit mit anderen standardisierten Effektstärkemaßen wie beispielsweise Risk Ratios, standardisierten Beta-Koeffizienten oder gar mit unstandardisierten Effekten gearbeitet wurde, so galt dies als Ausschlussgrund.

Als drittes Inklusionskriterium wurde die Verwendung eines metaanalytischen Standardverfahrens zur Synthese und Aggregierung von Effekten gewählt. Dazu zählten die Berechnungsmodelle nach Hedges und Olkin (1985), Hunter und Schmidt (1990), Glass (1976), Rosenthal und Rubin (1986) oder Stouffer (1949). Im Umkehrschluss wurden Forschungsarbeiten, die auf Structural equation Modellen (z. B. Anusic & Schimmack, 2016) oder Multilevel-Modellen (z. B. Fischer & Boer, 2011) aufbauten oder andere non-standard Verfahren in Verbindung mit einer Metaanalyse verwendeten (z. B. Markon et al., 2005), nicht in die Analysen dieser Arbeit inkludiert.

Rein artikelinterne Metaanalysen, in denen nur die Ergebnisse verschiedener Stichproben oder Subgruppen, die im selben Artikel berichtet wurden, metaanalytisch aggregiert wurden, galten ebenfalls als exkludierbar (z.B.: Carson et al., 2003). Im weiteren Verlauf dieser Arbeit, wird diese spezielle Form der Metaanalyse als *Internal Meta-analysis* bezeichnet.

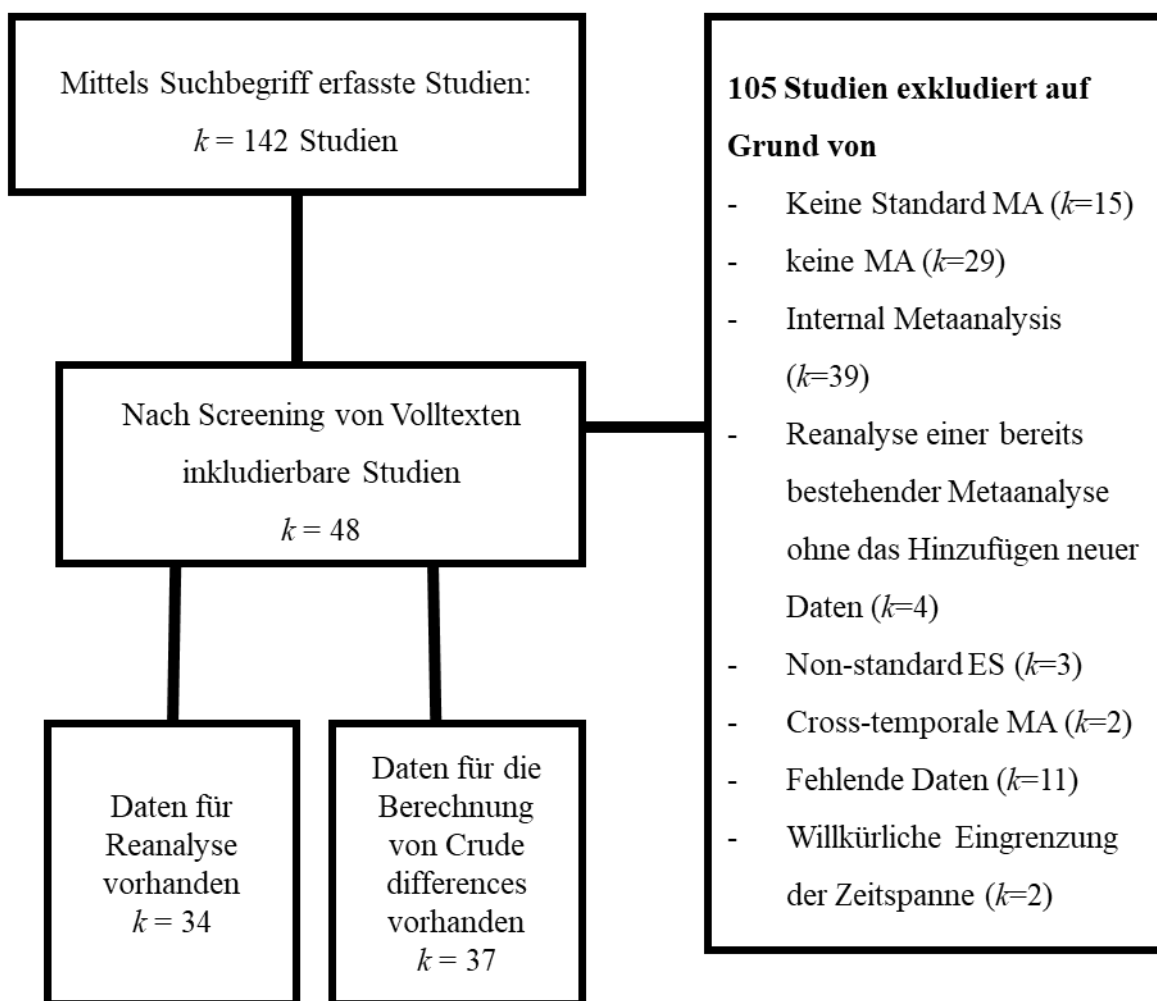
Zudem war es sehr wichtig, dass Metaanalysen eine möglichst große Zeitspanne abdeckten und im Idealfall die Studie, die das jeweilige untersuchte Phänomen als erstes berichtet hat, auch in ihren Analysen inkludiert war. Dementsprechend wurden sämtliche Metaanalysen aus dieser hier vorliegenden Arbeit ausgeschlossen, für die ohne sinnhafte Begründung oder scheinbar willkürlich, eine spezifische Zeitspanne als Inklusionskriterium für ihre Primärstudien angewandt wurde.

Als letztes Kriterium zur Inklusion galt, dass die überprüften Metaanalysen alle Datenpunkte, die für die geplanten Analysen meiner Masterarbeit notwendig waren, berichten. Dementsprechend musste zumindest der berechnete Summary effect, die Gesamtstichprobenzahl, die Anzahl inkludierter Effekte und eine Referenzliste mit allen inkludierten Primärstudien gegeben sein. Zudem konnte zusätzlich eine Reanalyse der jeweiligen Metaanalysen berechnet werden, wenn Effektstärken und Stichprobengrößen der einzelnen inkludierten Studien berichtet wurden. Waren die notwendigen Daten nicht gegeben, so wurden die AutorInnen über einen Zeitraum von einem Monat mittels einer ersten Anfragemail und einer Erinnerungsmail nach zwei und drei Wochen kontaktiert, um so die fehlenden Informationen einzuholen. Waren die AutorInnen nicht erreichbar oder war ihnen der Zugriff auf die jeweils verwendeten Daten nicht mehr möglich, so mussten diese Arbeiten aus den Analysen ausgeschlossen werden.

Wie aus Abbildung 1 zu entnehmen ist, konnten nach Anwendung aller hier genannten Inklusionskriterien 37 Studien in die statistischen Analysen eingeschlossen

Abbildung 1

Flow-Chart zu Darstellung des Selektionsprozesses inklusive aller Ausschlussgründe



Anmerkung. MA = Metaanalyse; ES = Effektstärke

werden. Für 34 Studien waren alle notwendigen Daten gegeben, um eine Reanalyse zu ermöglichen. Für die restlichen 3 Metaanalysen war dies nicht möglich, wodurch diese nur in einem Teil der statistischen Analysen inkludiert wurden. Alle exkludierten Studien sind, nach Ausschlussgrund geordnet, im Appendix genannt

Kodierung und Datenerhebung

Der Kodierungsprozess bestand darin, dass die Volltexte aller 142 Studien auf Inklusionskriterien gescreent wurden und alle notwendigen Daten extrahiert wurden. Dabei wurde dieser Prozess zwei Mal, von zwei unterschiedlichen Personen, unabhängig voneinander durchgeführt und die Ergebnisse miteinander verglichen. Bei Diskrepanzen zwischen den Ergebnissen des Kodierungsprozesses, wurden diese besprochen und abgeglichen. Mittels Cohen's κ wurde die Interrater-Reliabilität vor dem Abgleich der

Kodierungsergebnisse berechnet. Es zeigten sich dabei nur wenig Unterschiede in der Anwendung Inklusions- und Exklusionskriterien, was sich durch ein $\kappa = .91$ verdeutlichen lässt.

Aus jedem inkludierten Artikel wurden sowohl Informationen zur gerechneten Metaanalyse, als auch zur jeweils ältesten inkludierten Studie – im weiteren Verlauf als Initialstudie bezeichnet – extrahiert. Für die Initialstudie wurden (i) die berichtete Effektstärke, (ii) der dazugehörige p -Wert, falls dieser gegeben war, (iii) die Anzahl der Zitationen seit Publikation und (iv) der momentane Impact Factor des Journals, welches die Initialstudie publiziert hat, kodiert. Zusätzlich (v) wurde abgeglichen, ob die Initialstudie und die Metaanalyse bezüglich ihrer Forschungsfragen übereinstimmen und (vi) ob der Summary effect der Metaanalyse und der berichtete Effekt der Initialstudie, dasselbe Vorzeichen aufweisen.

Auf Seiten der Metaanalyse wurde zum einen erfasst, (i) wie viele Metaanalysen in dem Artikel gerechnet wurden, (ii) was die Forschungsfrage des Artikels war, (iii) ob die Metaanalyse auf abhängigen Daten aufgebaut war, (iv) ob ein fixed effect oder ein random effects model gerechnet wurde und (v) ob zur Berechnung des Summary effect die Korrekturen nach Hunter und Schmidt (1990) angewandt wurden oder nicht. Zum anderen wurden aber auch das Publikationsjahr (vi) der ältesten und (vii) der jüngsten inkludierten Studie, (viii) der berichtete Summary effect und der dazugehörige p -Wert, (ix) die verwendete Effektstärkenmetrik, (x) τ^2 und (xi) I^2 , (xii) die Art der Stichprobe, (xiii) die Gesamtstichprobengröße und (xiv) die Anzahl inkludierter Effekte kodiert. Zu guter Letzt wurde noch erhoben, (xv) ob auf Publikationsbias untersucht wurde, (xvi) wie dieser erfasst wurde und (xvii) was die daraus resultierende Schlussfolgerung war.

Falls in einem Artikel die Ergebnisse mehrerer Metaanalysen berichtet wurden, wurde trotz allem immer nur die Analyse extrahiert, welche die älteste Studie inkludiert hatte. Gab es mehrere Analysen mit derselben Initialstudie, so wurde sich für diejenige Metaanalyse entschieden, welche zusätzlich zur Initialstudie, auch auf den meisten Effekten aufbaute. Gab es hierbei wiederum mehrere Möglichkeiten, so wurden diese Metaanalysen durchnummeriert und per Zufallsgenerator ausgewählt. Falls es nicht möglich war eine spezielle Studie als Initialstudie festzulegen, da es mehrere Studien mit demselben Publikationsjahr gab, so wurde zuerst versucht festzustellen, in welchem Monat der jeweilige Artikel publiziert wurde. War dies nicht möglich, so wurden all diese Studien als mögliche Initialstudien kodiert und metaanalytisch zusammengefasst, um so einen einzelnen Effekt für die späteren Berechnungen zu erlangen. Falls die notwendigen Daten der Initialstudie nicht aus der

Metaanalyse zu entnehmen waren, so wurden diese, wenn möglich, aus der Originalstudie extrahiert.

Statistische Analyse

Die statistische Analyse der hier vorliegenden Arbeit lässt sich in zwei Ebenen unterteilen. Auf Meta-Ebene, dem ersten Analyselevel, wurden zum einen verschiedene metaanalytische Verfahren, die im Folgenden genauer behandelt werden, an den Primärdaten der inkludierten Metaanalysen durchgeführt. Zum anderen wurden auf dieser Ebene die sogenannten Crude differences und der Regressionskoeffizient der Effektveränderung über die Zeit, die als Kennwerte zur Erfassung des decline effect dienen, berechnet. Zweiterer Kennwert war nur errechenbar, wenn die Daten aller inkludierten Primärstudien einer Metaanalyse auch gegeben waren. War dies nicht der Fall, so wurden nur die Crude differences berechnet und diese Analyseebene größtenteils übersprungen. Der zweite Analyselevel war die sogenannte Meta-Meta-Ebene. Hierbei lag der Fokus nicht auf den Primärstudien, sondern auf den einzelnen Metaanalysen selbst und den jeweils berechneten Kennwerten für den decline effect, die anhand metaanalytischer Verfahren aggregiert und untersucht wurden.

Statistische Analysen auf Meta-Ebene

Auf der Meta-Ebene wurden zuerst die berichteten Summary effects und Effektstärken der Initialstudien in Fisher's z umgewandelt, um so für die späteren Analysen eine einheitliche Effektstärkenmetrik zu erhalten. Im nächsten Schritt wurden die Crude differences für sämtliche inkludierten Metaanalysen berechnet. Hierbei wurde die Differenz zwischen dem berichteten Summary effect und dem berichteten Effekt der Initialstudie berechnet, indem der Summary effect als Absolutwert vom Effekt der Initialstudie als Absolutwert subtrahiert wurde. Bei einer positiven Differenz, bedeutet dies, dass der Summary effect kleiner ist als der Effekt der Initialstudie, was somit als ein Anzeichen für einen decline effect gilt. Ist die Differenz im negativen Bereich, so ist dies ein Anzeichen für einen Effektanstieg.

Waren die Effektstärken und Stichprobengrößen aller Primärstudien einer Metaanalyse gegeben, so wurden zudem die folgenden Schritte durchgeführt. Zum einen wurden alle Metaanalysen mittels des Restricted Maximum Likelihood Schätzers (REML) metaanalytisch reanalysiert. War die Anwendung dieses Schätzers nicht durchführbar, so wurde stattdessen der Schätzer nach DerSimonian und Laird (1986) verwendet (z. B. Enisman et al., 2021). Zusätzlich wurde für jede Studie eine mixed effects Meta-Regression gerechnet, in der das Publikationsjahr der einzelnen Primärstudien als Prädiktor für die Effektveränderung

herangezogen wurden. Der daraus resultierende Regressionskoeffizient diene als zweiter Kennwert zur Überprüfung des decline effects. Zuvor wurde *Coining* angewandt, wodurch alle Primärstudieneffekte, die in einer Metaanalyse mit einem negativen Summary effect inkludiert waren, mit (-1) multipliziert wurden. Dadurch ergibt sich ein einheitliches Interpretationsmuster des Regressionskoeffizienten, wodurch ein negatives Vorzeichen einen decline effect bedeutet, währenddessen ein positives Vorzeichen ein Anzeichen für einen Effektanstieg darstellt.

Im nächsten Schritt wurden sämtliche Metaanalysen auf Publikationsbias überprüft. Da jede Methode zur Erfassung von Publikationsbias Stärken und Schwächen aufweist, wurde ein Vorgehen nach Triangulation gewählt und insgesamt neun unterschiedliche Verfahren angewandt, die entweder die Verteilung von Effektstärken auf Asymmetrie überprüften, die Verteilung von p -Werten untersuchten oder zu den sogenannten Selection Model Approaches zählen (Cooper et al. 2019, Siegel et al., 2021).

Zum einen wurde Trim-and-Fill (Duval & Tweedie, 2000a; Duval & Tweedie, 2000b). angewandt, um die einzelnen Metaanalysen auf small-study effects zu überprüfen. Hierzu wird der Funnel-Plot auf Asymmetrie in der Verteilung der Effektstärken untersucht. Dafür werden einerseits, in einem iterativen Prozess, Extremwerte in der Verteilung entfernt und zum anderen mögliche fehlende Effektstärken, die zu einer symmetrischen Verteilung beitragen könnten, hinzugefügt. Anschließend wird ein adjustierter Summary effect, basierend auf einer symmetrischen Datenverteilung, errechnet (Cooper et al., 2019). Wie in Siegel et al. (2021) vorgeschlagen wurde, diene sowohl dieser errechnete Effektschätzer, als auch die Anzahl möglicher fehlender Effektstärken der Erfassung von Publikationsbias (siehe Tabelle 1).

Eine weitere Methode, die hier Anwendung fand und für die Überprüfung der Funnel Plot Asymmetrie entwickelt wurde, ist der Rank Correlation Test nach Begg und Mazumdar (1994). Hierbei wird mittels Kendall's τ eine Rangkorrelation zwischen der Effektabweichung der einzelnen Primärstudieneffekte vom Summary effect und der dazugehörigen Varianz berechnet und somit der Zusammenhang zwischen der Präzision einer Primärstudie und ihrer Effektverschätzung untersucht (Cooper et al., 2019). Als Interpretationsgrundlage für das Vorliegen eines Publikationsbias wurde der p -Wert herangezogen, wobei angenommen wurde, dass ein $p < .10$ den Schwellenwert für die Interpretation darstellte (Siegel et al., 2021).

Als dritte Methode wurde der Regressionstest nach Sterne und Egger (2005) angewandt. Bei diesem Verfahren wird der Zusammenhang zwischen der Effektstärke einer

Primärstudie und dem jeweiligen Standardfehler mittels einer gewichteten Regression untersucht. Fällt der Regressionskoeffizient der Steigung signifikant aus, so spricht dies für eine mögliche Asymmetrie in den Daten und dient als Hinweise für das Vorliegen eines small-study effects (Cooper et al., 2019; Siegel et al., 2021). Wie auch beim Rank Correlation Test, wird auch hier ein $p < .10$ als Schwellenwert für die statistische Signifikanz vorgeschlagen (Siegel et al., 2021).

Die vierte Publikationsbiasmethode stellte PET-PEESE (Stanley & Doucouliagos, 2014) dar. Dieses Verfahren baut auf zwei aufeinander folgenden Regressionsanalysen auf und stellt eine Modifikation des ursprünglichen Regressionstest nach Egger et al. (1997) dar. In der ersten Regressionsanalyse, dem sogenannten Precision-Effect Test (PET), wird eine gewichtete Regressionsanalyse durchgeführt und der Zusammenhang zwischen der Primärstudieneffektstärke und ihrem Standardfehler berechnet. Fällt der Intercept dieser Analyse signifikant aus, so wird zusätzlich ein Blick auf die Ergebnisse der zweiten Regressionsanalyse geworfen, die als Precision-Effect Estimate with Standard Error (PEESE) bezeichnet wird. Hierbei wird anstelle des Standardfehlers, die Varianz als Prädiktor in die Regressionsanalyse aufgenommen. In beiden Analysen dient der errechnete Regressionskoeffizient des Intercept zusätzlich als korrigiertes Effektmaß für den Summary effect (Stanley & Doucouliagos, 2014). Berechnet man die Differenz zwischen dem normalen Summary effect und dem nach PET oder PEESE korrigierten Summary effect, so kann man dies für die Identifikation von Publikationsbias nutzen (siehe Tabelle 1; Siegel et al., 2021).

Anhand von p -Curve (Simonsohn et al., 2014), p -Uniform (van Assen et al., 2015), p -Uniform* (van Aert & van Assen, 2018) und dem Test of Excess Significance (TES; Ioannidis & Trikalinos, 2007) wurden die p -Werte auf Anzeichen von p -Hacking und weiteren Bias Mustern innerhalb einer Metaanalyse untersucht (Cooper et al., 2019; Siegel et al., 2021). Der TES stellt dabei ein Verfahren dar, welches überprüft, ob zu viele signifikante Primärstudieneffekte vorliegen. Dies wird mittels eines χ^2 -Test berechnet, indem die Anzahl der beobachteten signifikanten Ergebnisse mit der Anzahl erwarteter signifikanter Effekte verglichen wird, wobei letzterer Wert mittels einer Post-hoc Poweranalyse der einzelnen Primärstudien errechnet wird. Zeigten sich signifikant mehr beobachtete als erwartete signifikante Ergebnisse, so wurde dies als Indikator für das Vorliegen eines Publikationsbias interpretiert, wobei ein $p < .10$ erneut als Signifikanzniveau herangezogen wurde (Ioannidis & Trikalinos, 2007; Siegel et al., 2021).

Mittels p -Curve, p -Uniform und p -Uniform* wurde die Verteilung der p -Werte untersucht (Siegel et al., 2021). Da angenommen wird, dass bei einem tatsächlich

vorliegenden Effekt mehr $p < .02$ vorliegen als Werte, die zwischen $p = 0.2$ und $p = .05$ liegen, sollte sich die Verteilung der p -Werte in diesem Falle rechtsschief präsentieren. Zur Überprüfung der Schiefe der Verteilung dient p -Curve. Kommt es zu einer linksschiefen Verteilung, so gilt dies als Anzeichen für p -Hacking und somit für Publikationsbias. Mittels eines kontinuierlichen Tests wurde sowohl die halbe p -Curve (d.s. alle p -Werte $< .025$) als auch die volle p -Curve (d.s. alle p -Werte $< .05$) darauf überprüft, wie wahrscheinlich es ist, dass ein signifikantes Ergebnis zumindest als Extremwert auftritt. Die daraus resultierenden p -Werte wurden, wie in Tabelle 1 abgebildet, für die Interpretation des Ergebnisses herangezogen, um festzustellen, ob Publikationsbias vorliegt (Simonsohn et al., 2014).

Als siebente Methode wurde p -Uniform angewandt. Ähnlich wie p -Curve, wird auch hier die Verteilung der signifikanten p -Werte untersucht, wobei eine Gleichverteilung dieser Werte angenommen. Entspricht die Verteilung der p -Werte nicht einer Gleichverteilung, so gilt dies als Anzeichen für Publikationsbias. Überprüft wird dies in p -Uniform mittels eines Publikationsbias-Tests, welcher auf der Fisher-Methode basiert. Die vorliegenden signifikanten p -Werte werden anhand dieser Methode in Werte von 0 bis 1 transformiert und auf das Vorliegen einer Gleichverteilung überprüft. Kommt es zu einer signifikanten Abweichung (d.s. $p < .10$) so wird dies als Anzeichen für Publikationsbias interpretiert (Cooper et al., 2019; Van Assen et al., 2015).

p -Uniform* stellt eine Erweiterung von p -Uniform dar, welche als Reaktion auf einige Kritikpunkte gegen p -Curve und p -Uniform entstand (van Aert & van Assen, 2018). p -Uniform* zeigt sich im Gegensatz zu den anderen beiden Methoden robuster gegenüber heterogenen Daten und ermöglicht durch die Inklusion von sowohl signifikanten als auch nicht signifikanten p -Werten einen genaueren korrigierten Summary effect (van Aert & van Assen, 2018). Wie bei p -Uniform, wurde auch für diese Methode ursprünglich der Publikationsbias-Test nach der Fisher-Methode implementiert, letztlich wurde dieser, aufgrund mangelnder statistischer Überprüfungen, von den Autoren aus dem Verrechnungscode entfernt (van Aert, 2022). Dementsprechend wurde stattdessen der korrigierte Summary effect für die Überprüfung von Publikationsbias herangezogen und mit dem Summary effect eines random-effects Modells verglichen. Bei einer Differenz von über 20% wurde von einem Publikationsbias ausgegangen (siehe Tabelle 1; J. Pietschnig, persönliche Unterhaltung, 02. März 2022).

Die letzte angewandte Methode stellte der Selection Model Approach nach Vevea und Woods (2005) dar. Das Vorliegen von Publikationsbias wird hierbei durch den Einsatz von Gewichtungen überprüft. Dabei werden vier unterschiedliche Modelle, mit jeweils anderen

Gewichtungen für die p -Werte der einzelnen Primärstudieneffekte, erstellt und somit. für Publikationsbias korrigierte, Summary effects berechnet. Diese korrigierten Effektschätzer werden in weiterer Folge mit dem unbereinigten Effektschätzer eines random-effects Modells verglichen (Vevea & Woods, 2005). Übersteigt der Unterschied zwischen dem bereinigten und nicht bereinigten Maß 20%, so gilt dies als Anzeichen für Publikationsbias (Siegel et al., 2021).

Alle verwendeten Methoden und Schwellenwerte sind nochmals in Tabelle 1 zu abgebildet. Sämtliche Analysen wurden einmal für einen Datensatz basierend auf allen Daten angewandt und ein zweites Mal nur an Datensätzen mit publizierten Daten. Hierzu wurden zuvor sämtliche Primärdaten, die als graue Literatur kategorisiert werden konnten, aus dem Datensatz entfernt. Zusätzlich wurde für jede Metaanalyse die durchschnittliche Power der inkludierten Primärstudien berechnet. Hierfür wurde berechnet wie viel Power eine Primärstudie aufwies, um den Summary effect der Metaanalyse zu erfassen.

Statistische Analysen auf Meta-Meta-Ebene

Auf der Meta-Meta-Ebene wurden die errechneten Crude Differences und Regressionskoeffizienten metaanalytisch zusammengefasst, um so den gewichteten Durchschnittswert dieser Kennwerte zu erhalten. Zudem wurde mittels mixed-effects Modellen, für beide Kennwerte eine Subgruppenanalyse durchgeführt, mit denen der Unterschied zwischen dem decline effect und dem Effektanstieg untersucht wurde. Diese Subgruppenanalysen wurden einmal mit dem vollständigen Datensatz durchgeführt und ein weiteres Mal nur mit publizierten Daten. Mit einer weiteren Subgruppenanalyse wurde der Einfluss von korrelativen Effektstärken (d.s. Person r und Fisher's z) und Effektstärken basierend auf Mittelwerten (d.s. Cohen's d und Hedges g) auf den decline effect überprüft.

Darauffolgend wurden vier hierarchische multiple mixed-effects Meta-Metaregressionsmodelle gerechnet, um den Einfluss von verschiedenen Charakteristika der Initialstudie bzw. der Metaanalyse auf das Ausmaß des decline effects zu untersuchen. In Modell 1 ging es um Charakteristika der Initialstudie und es wurde schrittweise (i) die Effektstärke, (ii) die Stichprobengröße, (iii) der Impact Factor des publizierenden Journals, (iv) die jährliche Zitationszahl und (v) das Publikationsjahr als Prädiktor in das Modell aufgenommen und deren Einfluss auf die Crude differences überprüft. Modell 2 fokussierte sich auf die Charakteristika der Metaanalyse und es wurde schrittweise (i) der Summary effect, (ii) die Anzahl an inkludierten Stichproben, (iii) die abgedeckte Zeitspanne, (iv) das Publikationsjahr und (v) die Art der inkludierten Studie (d.s. war graue Literatur inkludiert oder nicht) in das Regressionsmodell aufgenommen und erneut der Einfluss auf die Crude

Tabelle 1

Auflistung der verwendeten Publikationsbias-Methoden und Schwellenwerte (adaptiert nach Siegel et al., 2021)

Methode	Schwellenwerte
Trim-and-Fill (Duval & Tweedie, 2000a, 2000b)	Trim-and-Fill zeigt zumindest eine fehlende Studie an. Der Unterschied zwischen dem Summary effect der Metaanalyse und dem errechneten Schätzer beträgt nicht mehr als 20%. Der errechnete Schätzer durch Trim-and-Fill weist eine andere Wirkrichtung auf, als der Summary effect.
Rank Correlation Test (Begg & Mazumdar, 1994)	$p < .10$
Sterne and Egger's Regression Test (Sterne & Egger, 2005)	$p < .10$
PET-PEESE (Stanley & Doucouliagos, 2014)	Der Unterschied zwischen dem Summary effect der Metaanalyse und dem errechneten Schätzer beträgt nicht mehr als 20%. Der PEESE-Schätzer wird interpretiert, wenn der Intercept $p < .10$ aufweist. Andernfalls wird der PET-Schätzer herangezogen.
p – Curve (Simonsohn et al., 2014)	Entweder der p -Wert der halben p -Curve liegt bei $p > .05$ oder die p -Werte der halben und der vollen p -Curve liegen bei $p > .10$.
p Uniform (van Assen et al., 2015)	Publikationsbias-Test zeigt bei einseitigem Testen $p < .10$
p Uniform* (van Aert & van Assen, 2018)	Der Unterschied zwischen dem Summary effect der Metaanalyse und dem errechneten Schätzer beträgt nicht mehr als 20%. ^a
Excess Significance Test (Ioannidis & Trikalinos, 2007)	$p < .10$ und die Anzahl beobachteter signifikanter Effekte ist größer als die erwartete Anzahl signifikanter Effekte

Methode	Schwellenwert
Selection Model Approach (Vevea & Woods, 2005)	Der Unterschied zwischen dem Summary effect der Metaanalyse und einem der errechneten Schätzer beträgt nicht mehr als 20%.

Anmerkung. Anhand der hier aufgelisteten Publikationsbias-Methoden wurden $k = 34$ Metaanalysen, die Daten für eine Reanalyse boten, auf Anzeichen von Bias-Mustern überprüft. Die herangezogenen Schwellenwerte sind aus Siegel et al. (2021) und Cooper et al. (2019) entnommen.

^a Für p Uniform* zeigte sich eine Diskrepanz zwischen den Schwellenwerten, die in Siegel et al. (2021) berichtet wurden und jenen, die hier abgebildet sind, da der geplante Publikationsbias-Test nicht durchführbar war.

differences überprüft. Modell 3 unterschied sich von Modell 1 nur hinsichtlich der abhängigen Variablen, da hier der Einfluss auf die Regressionskoeffizienten überprüft wurde. Dasselbe galt für den Unterschied zwischen Modell 4 und Modell 2. Nach jedem Schritt wurde ein Likelihood Ratio Test durchgeführt um den Model – Fit zu überprüfen. Zusätzlich wurde nach jedem Schritt der Varianzinflationsfaktor (VIF) erfasst, um die Prädiktoren auf Multikollinearität zu überprüfen. War dieser Wert in einem Modell über vier, so wurde der Prädiktor mit dem höchsten VIF-Wert aus dem Modell exkludiert und das Modell nochmals nachgerechnet.

Wie bei einigen Analyseschritten zuvor, wurden auch hier die Analysen ein zweites Mal durchlaufen und nur die Kennwerte exklusive grauer Literatur herangezogen. Schritt 5 aus Modell 2 und 4 wurde im zweiten Durchlauf aufgrund der Redundanz nicht durchgeführt. Zusätzlich sei noch anzumerken, dass für die Modelle 1 und 2, die berichteten Kennwerte der Metaanalysen als Prädiktoren herangezogen wurden, währenddessen für die Modelle 3 und 4 die errechneten Kennwerte genutzt wurden. Diese Entscheidung wurde getroffen, da es in manchen Fällen zu Diskrepanzen zwischen den errechneten und berichteten Kennwerten (d.s. Summary effect, k , total N) kam und die Crude differences aus berichteten Werten errechnet wurden, währenddessen die Regressionskoeffizienten anhand der gegebenen Daten selbst analysiert wurden.

Neben dem mixed-effects Modell wurde zusätzlich die Funktion der linearen Regression für die Berechnung der eben genannten 4 Modelle genutzt. Diese Methode war zwar nicht präregistriert, es wurde sich jedoch dazu entschieden, sie zusätzlich in die

Analysen aufzunehmen, nachdem deutlich wurde, welche Probleme die Anwendung des mixed-effects Modells mit sich bringt (J. Pietschnig, persönliche Unterhaltung, 22.03.2022). Nachdem die Meta-Metaanalyse noch ein relativ neues Verfahren darstellt, gibt es dafür noch keine eigens dafür konzipierten Analysemethoden, weshalb in dieser Arbeit stattdessen normale, metaanalytische Verfahren genutzt wurden. Um möglichst genaue Schätzungen des wahren Wertes zu erlangen, werden in Metaanalysen Effekte der inkludierten Primärstudien nach ihrer Stichprobengröße gewichtet. Auf Meta-Meta-Ebene spielt jedoch nicht nur die Stichprobengröße bei der Berechnung der Varianz eine Rolle, sondern auch die Anzahl inkludierter Effekte innerhalb einer Metaanalyse. Um nun das mixed-effects Modell für die Berechnung heranziehen zu können, musste sich für einen der Faktoren entschieden werden. Mittels der linearen Regression war es möglich eine ungewichtete Analyse der einzelnen Modelle durchzuführen.

Um den Einfluss der einzelnen Prädiktoren auf die Crude differences und Regressionskoeffizienten zu erfassen, wurden zudem einfache Meta-Metaregressionen gerechnet. Auch hier wurde, wie in der Präregistrierung vermerkt, ein mixed-effects Modell für die Berechnung herangezogen, aber zusätzlich auch ein lineares Modell durchgeführt. Für die Analysen wurde immer nur ein Prädiktor ins Modell aufgenommen und die Regression gerechnet. Diese Analysen wurden einmal mit dem vollen Datensatz gerechnet und ein weiteres Mal mit dem Datensatz, der nur publizierte Studien inkludierte.

Danach wurden einzelne lineare und mixed-effects Meta-Metaregressionen gerechnet in denen τ^2 oder I^2 die Kriterien darstellten und (i) die Anzahl an inkludierten Samples innerhalb einer Studie, (ii) die Gesamtstichprobengröße oder (iii) die durchschnittliche Stichprobengröße den Prädiktor darstellten. Das lineare Modell entsprach erneut nicht der Präregistrierung.

In einem letzten Analyseschritt wurde die Häufigkeit von Anzeichen für Publikationsbias in Metaanalysen inklusive und exklusive grauer Literatur verglichen. Anhand der Schwellenwerte in Tabelle 1 wurde überprüft, ob die einzelnen Methoden Hinweise für einen Publikationsbias erfassen und mittels eines χ^2 -Tests wurde darauffolgend überprüft, ob sich Metaanalysen die graue Literatur inkludieren bezüglich ihrer Häufigkeit von Publikationsbias Anzeichen von Metaanalysen unterscheiden, die keine graue Literatur inkludieren. War die Voraussetzung einer Minimumzellenhäufigkeit von Fünf nicht gegeben, so wurde stattdessen der exakte Test nach Fisher durchgeführt.

Zur Auswertung wurde die Version 4.1.2 der Open-Source Software R genutzt. Der verwendete Syntax-Code und alle Datensätze, die zur Verrechnung verwendet wurden, sind auf OSF abrufbar (<https://osf.io/pgybc>)

Resultate

Deskriptive Statistik

Wie aus Abbildung 1 zu entnehmen ist, konnten von den 142 erfassten Studien insgesamt 37 Metaanalysen inkludiert werden. 34 davon lieferten genügend Daten, um eine Reanalyse und gleichzeitig die Berechnung der Regressionskoeffizienten zu ermöglichen. Die deskriptive Statistik zu allen 37 inkludierten Studien ist in Tabelle 2 abgebildet. In Tabelle 3 sind zusätzliche Informationen der durchgeführten Reanalysen zu finden.

Insgesamt baute die hier präsentierte Meta-Metaanalyse, nach Angaben der einzelnen Autor*innen, auf $k = 3,263$ Effekten von $N = 614453$ Personen auf und deckte dabei eine Zeitspanne von 1928 – 2020 ab. In etwas mehr als der Hälfte (57%) aller Studien wurde berichtet, dass graue Literatur in der berechneten Metaanalyse inkludiert wurde. Die Effekte wurden am häufigsten in Person r (59%) berichtet, gefolgt von Hedges g (24%) und Cohen's d (17%). Zur Berechnung der Metaanalyse wurde in den meisten Fällen (78%) die Methode nach Hedges und Olkin (1985) oder andere vergleichbare Methoden, wie die von Glass (1976), Rosenthal und Rubin (1986) oder Stouffer (1949) verwendet. Die Methode nach Hunter und Schmidt (1990) kam in 8 (22%) Fällen vor.

In 17 (46%) Metaanalysen wurde zumindest eine Methode zur Erfassung von Publikationsbias angewandt. In 4 (25%) dieser Analysen wurden, den jeweiligen Autor*innen zufolge, zumindest schwache Hinweise für Publikationsbias gefunden (d.s. Calderon et al., 2020; Leach & Cidam, 2015; Pettigrew & Tropp, 2006; Prinzie et al., 2009). Als Methode wurde hierbei das Fail-Safe N ($k = 9$, 53%) am häufigsten gewählt, gefolgt von der visuellen Inspektion des Funnelplot ($k = 8$, 47%) und Sterne und Egger's Regressions-Test ($k = 7$, 41%). In 6 (35%) Metaanalysen wurden unter anderem auch andere, eher als non-standard einzuschätzende, Methoden verwendet. Darunter fielen Regressions- oder Korrelationsanalysen, die den Zusammenhang zwischen der berichteten Effektstärke einer Studie und ihrem Publikationsstatus, ihrer Stichprobengröße, dem Publikationsjahr oder auch der jeweiligen Forschungsgruppe untersuchen sollten (d.s. Burnham, 2020; Cuadrado et al., 2021; De Dreu et al., 2000; Greenwald et al., 2009; Lipnevich et al., 2017; Pettigrew &

Tabelle 2*Deskriptive Statistik zu berichteten Daten der inkludierten Metaanalysen*

Variablen	Ergebnis
Anzahl inkludierter Metaanalysen	37
Total N	614453
k	3263
Spannweite berichteter k	5 – 712
Abgedeckte Zeitspanne	1928 – 2020
War graue Literatur inkludiert?	
Ja	21
Nein	16
Verwendete Effektstärkenmetrik	
Pearson r	22
Hedges g	9
Cohens d	6
Verwendetes metaanalytisches Modell	
Fixed effect	10
Random effects	27
Metaanalyse – Typ	
Hedges & Olkin	29
Hunter & Schmidt	8
Anzahl der Metaanalysen in denen auf Publikationsbias getestet wurde	17
Hinweise auf PB	2
Teilweise Hinweise auf PB	2
Keine Hinweise auf PB	13
Verwendete Methoden zur Erfassung von PB	
Direkter Vergleich	5
Funnelplot	8
Trim-and-Fill	5
Fail-Safe N	9
Sterne & Egger's Regression	7
p -Curve	1
Andere (non Standard)	6

Anmerkung. PB = Publikationsbias

Tropp, 2006), aber beispielsweise auch eine andere Variante des Fail-Safe N nach Orwin (d.s. Levine et al., 2012) und der General Linear Model Approach (Vevea & Hedges, 1995; d.s. Pettigrew & Tropp, 2006) fanden Anwendung.

Poweranalyse

Um festzustellen, ob die Forschung im Bereich der Sozialpsychologie und Persönlichkeit im Durchschnitt zu wenig Power besitzt, wurde eine Poweranalyse durchgeführt. Hierfür wurde für alle in der Reanalyse inkludierten Metaanalysen separat berechnet, wie hoch die Power einer Primärstudie im Durchschnitt ist, um den Summary Effect zu erfassen.

Wie aus Tabelle 3 zu entnehmen ist, ergab sich eine durchschnittliche Power von $M = 46.04\%$. Zusätzlich zum Mittelwert, wurde auch die Häufigkeit dafür berechnet, dass die errechnete Durchschnittspower unter den, nach wissenschaftlichen Konventionen festgelegten, 80% liegt. Hier zeigte sich, dass die Primärstudien von 26 Metaanalysen (76%) im Schnitt eine Power unterhalb dieses Schwellenwertes aufwiesen. 35% ($k = 9$) davon wiesen eine Power zwischen 30% und 50% auf und 50% ($k = 13$) zeigten sogar einen Wert unterhalb der 30%.

Tabelle 3

Deskriptive Statistik der Ergebnisse nach der Reanalyse der Daten

Variablen	Ergebnis
Anzahl inkludierter Studien	34
Total N	541716
k	2310
Spannweite erfasster k	4 – 712
Durchschnittliche Power einer Primärstudie innerhalb einer Metaanalyse	46.04%
Durchschnittliche Power bei einem decline effect	44.40%
Durchschnittliche Power bei einem Effektanstieg	48.37%
Spannweite der Power einer Primärstudie innerhalb einer Metaanalyse	5.17% – 99.47%
Anzahl an Studien mit zu wenig Power	26/22/13 ^a

Anmerkung. Zur Berechnung der durchschnittlichen Power einer Primärstudie, wurde der Summary effect der jeweiligen Metaanalyse herangezogen.

^a Häufigkeit dafür, dass die durchschnittliche Power < 80%/ < 50%/ < 30% ist.

Der decline effect

Die Ergebnisse der Analysen zeigten, dass anhand der Crude differences $k = 25$ Metaanalysen einen decline effect, $k = 11$ einen Effektanstieg und $k = 1$ einen sogenannten Proteus Effekt aufwiesen. Letzterer wurde aufgrund des mangelnden Konsens darüber, ob es sich bei diesem Effekt um einen Extremfall des decline effect handelt oder eine eigene Art von Disseminationsbias darstellt, aus den weiteren Analysen exkludiert (Ioannidis & Trikalinos, 2005; Pfeifer et al., 2011; Pietschnig et al., 2019). Im Vergleich zu Effektanstiegen ergaben sich also mehr als doppelt so viele decline effects, die zudem auch mehr als doppelt so stark ausfielen wie Effektanstiege, $r = .25, p < .001$; $r = .11, p = .02$. Wie aus Tabelle 4 zu entnehmen ist, stellte der Proteus Effekt auf der anderen Seite nur einen marginal signifikanten Effekt dar, der außerdem unterhalb der Interpretationsgrenzen nach Cohen (1988) lag. Aber nicht nur in den getrennten Gruppen zeigten sich eindeutige Anzeichen für das Vorliegen von early extremes oder decline effects, auch wenn alle Metaanalysen die einen Effektanstieg oder eine -reduktion berichteten zusammengefasst wurden, kam es zu einer signifikanten Reduktion des Effektes, $r = .15, p = .003$.

Tabelle 4

Durchschnittliche Effektabweichung berechnet anhand von Crude differences

	k	$r(\text{RE})$	p	$r - \text{Range}$
Gesamt	36 ^a	.152	.003	.011 – .814
Decline effect	25	.248	< .001	.011 – .814
Anstieg	11	.108	.02	.042 – .178
Proteus effect ^a	1	.030	.06	– ^b

Anmerkung. RE = random effects Modell; $r - \text{Range}$ = Minimum- und Maximumwerte der einzelnen Kennwerte. Anhand von random effects Modellen wurden die errechneten Crude differences der einzelnen Metaanalysen aggregiert und somit ein Durchschnittswert errechnet. Die Effekte sind als Absolutwerte abgebildet. Anhand der Analysen zeigte sich, dass es zu mehr als doppelt so vielen decline effects, im Vergleich zu Effektanstiegen kam und die decline effects im Schnitt stärker ausfielen. Auch über alle Metaanalysen hinweg war ein decline effect feststellbar.

^a Der Proteus Effekt wurde in dieser Analysen nicht inkludiert.

^b Keine Minimum- und Maximumwerte vorhanden, da nur eine Metaanalyse betroffen war.

Tabelle 5*Durchschnittliche Effektabweichung berechnet anhand von Regressionskoeffizienten*

	<i>k</i>	<i>r</i> (RE)	<i>p</i>	<i>r</i> – Range
Gesamt	34	.001	.37	< .001 – .073
Decline	20	.007	< .001	.001 – .073
Effektanstieg	14	.003	.005	< .001 – .013

Anmerkung. RE = random effects Modell; *r* – Range = Minimum- und Maximumwerte der einzelnen Kennwerte. Anhand von random effects Modellen wurden die errechneten Regressionskoeffizienten der jährlichen Effektveränderung aggregiert und somit ein Durchschnittswert errechnet. Die Effekte sind als Absolutwerte abgebildet. Anhand der Analysen zeigte sich, dass es zu mehr als doppelt so vielen decline effects, im Vergleich zu Effektanstiegen kam und die decline effects im Schnitt stärker ausfielen. Auch über alle Metaanalysen hinweg war ein decline effect feststellbar.

Um neben dem visuellen Vergleich von Häufigkeiten und Effektstärken, auch ein inferenzstatistisches Maß für den Unterschied zwischen Effektanstiegen und decline effects zu erhalten, wurde eine Subgruppenanalyse durchgeführt. Das Ergebnis dieser Analyse fiel signifikant aus, was bedeutete, dass sich die Gruppe der Metaanalysen, die einen decline effect aufweisen, signifikant von der Gruppe von Metaanalysen mit Effektanstiegen unterschied, $Q(1) = 42.34$, $p < .001$. Dieselbe Analyse wurde zusätzlich einmal nur mit jenen Metaanalysen, die auf publizierten Daten aufbauen, wiederholt. Auch hier kam es zu einem signifikanten Unterschied zwischen den beiden Subgruppen, $Q(1) = 36.31$, $p < .001$.

Auf Seite der jährlichen Effektveränderung zeigte sich ein vergleichbares Bild, wenn auch weniger eindeutig, als bei den Crude differences. Erneut zeigte sich anhand der Richtung des errechneten Effektes, dass es über alle Studien hinweg ($k = 34$) zu einer jährlichen Effektveränderung in Form eines decline effects kam. Der Effekt fiel jedoch nicht signifikant aus (siehe Tabelle 5). Bei der Aufteilung in Metaanalysen, die einen decline effect oder einen Effektanstieg aufwiesen, ließ sich erkennen, dass $k = 20$ Metaanalysen erstere Effektveränderung aufwiesen und $k = 14$ Metaanalysen zweite. In Tabelle 5 lässt sich zusätzlich erkennen, dass die jährliche Effektveränderung zwar in beiden Fällen unterhalb der Interpretationsgrenzen nach Cohen (1988) liegen, der decline effect jedoch mehr als doppelt so groß ausfällt als der jährliche Effektanstieg. Wie auch bei den Crude differences, kam es also auch hier häufiger zu decline effects und die Effektveränderung fiel stärker aus.

Tabelle 6

Durchschnittliche Effektabweichung berechnet anhand von Crude differences bei nur publizierten Daten

	<i>k</i>	<i>r</i> (RE)	<i>p</i>	<i>r</i> – Range
Gesamt	16	.052	.15	.011 – .814
Decline	10	.127	< .001	.011 – .814
Anstieg	6	.089	< .001	.042 – .178

Anmerkung. RE = random effects Modell; *r* – Range = Minimum- und Maximumwerte der einzelnen Kennwerte. Anhand von random effects Modellen wurden die errechneten Crude differences von Metaanalysen, die nur auf publizierten Daten aufbauten, aggregiert und der durchschnittliche Effektunterschied zwischen dem Summary effect und dem Initialstudieneffekt erfasst.

Für die Regressionskoeffizienten wurde ebenfalls eine Subgruppenanalyse durchgeführt, um zu überprüfen, ob sich die Gruppe der decline effects von der Gruppe der Effektanstiege unterscheiden würde. Auch hier fiel sowohl die Subgruppenanalyse mit allen Daten als auch die Subgruppenanalyse über Metaanalysen, die nur publizierte Studien inkludiert hatten, signifikant aus, $Q(1) = 27.66, p < .001$; $Q(1) = 27.48, p < .001$. Die Ergebnisse ließen dementsprechend darauf schließen, dass bei den Regressionskoeffizienten, auch inferenzstatistisch betrachtet, ein Unterschied zwischen diesen beiden Subgruppen vorlag.

Zusätzlich zur durchgeführten Subgruppenanalyse wurden die errechneten Crude differences und jährlichen Effektveränderungen der Reanalyse mit publizierten Primärstudien ebenfalls gemittelt. Die Ergebnisse dieser Berechnungen sind in Tabelle 6 und 7 abgebildet. Auf Seiten der Crude differences kam es im Vergleich zu den zuvor genannten Ergebnissen zu starken Unterschieden (siehe Tabelle 6). Die durchschnittliche Effektveränderung wies zwar erneut auf einen decline effect hin, aber der Effekt fiel im Gegensatz zur Analyse mit allen Studien nicht mehr signifikant aus, $r = .05, p = .15$. Bei der Aufteilung in Metaanalysen, die einen decline effect oder einen Effektanstieg aufwiesen, zeigte sich zwar erneut, dass der decline effect häufiger auftrat und stärker ausfiel, die Effektreduktion war jedoch, im Vergleich zu den Ergebnissen zuvor, um mehr als $r = .10$ kleiner geworden. Die Effektstärke der jährlichen Effektveränderung blieb durch die Exklusion grauer Literatur unbeeinflusst. Zwei Metaanalysen, die zuvor einen decline effect aufwiesen, zeigten in der Analyse nur mit

Tabelle 7

Durchschnittliche Effektabweichung berechnet anhand von Regressionskoeffizienten nur mit publizierten Daten

	<i>k</i>	<i>r</i> (RE)	<i>p</i>	<i>r</i> – Range
Gesamt	34	.002	.22	< .001 – .174
Decline	18	.009	< .001	.001 – .174
Effektanstieg	16	.003	.017	< .001 – .065

Anmerkung. RE = random effects Modell; *r* – Range = Minimum- und Maximumwerte der einzelnen Kennwerte. Anhand von random effects Modellen wurden die jährliche Effektveränderung von Metaanalysen, die nur auf publizierten Daten aufbauten, aggregiert.

publizierten Primärstudien nun aber einen Effektanstieg. Die Ergebnisse sind in Tabelle 7 abgebildet.

Anhand einer weiteren Subgruppenanalyse wurde überprüft, ob es zu einem Unterschied in der Größe der Crude differences kommt, abhängig davon in welcher Effektstärkenmetrik die Effekte berichtet wurden. Es zeigte sich, dass sich die beiden Subgruppen nicht signifikant voneinander unterschieden und es somit, entgegen der Annahme, keinen Einfluss auf die Crude differences hatte, ob die Effekte in einem korrelativen Effektmaß oder einem auf Mittelwerte basierenden Effektmaß berichtet wurden, $Q(1) = 0.30, p = .58$.

Multiple hierarchische mixed-effects Meta-Metaregression

Wie bereits Anfangs angesprochen, gab es neben der Erfassung der Stärke und Häufigkeit des decline effects, in der Literatur des JPSP, auch die Zielsetzung, mögliche Erklärungsansätze für diesen Effekt zu finden. Hierfür wurden 4 unterschiedliche multiple hierarchische mixed-effects Meta-Metaregressions-Modelle gerechnet, um den Einfluss diverser Studiencharakteristika der Initialstudie und der Metaanalyse auf die Effektveränderung zu überprüfen. Jedes dieser Modelle wurde einmal mit dem vollen Datensatz berechnet und ein weiteres Mal mit den Daten der Metaanalysen, die nur auf publizierten Primärstudien aufbauten.

Modell 1: Der Einfluss von Studiencharakteristika der Initialstudie auf Crude Differences

In insgesamt fünf Schritten wurden die Effektstärke der Initialstudie, ihre Stichprobengröße, ihr Publikationsjahr, die jährliche Anzahl an Zitationen der Initialstudie

Tabelle 8

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen mixed-effects Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Kriterien

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Step 1				
Initialstudien ES	0.922 [0.752, 1.091]	< .001	0.002 [-0.001, 0.005]	.16
Step 2				
Initialstudien ES	0.939 [0.769, 1.110]	< .001	-0.003 [-0.008, 0.002]	.22
Initialstudien <i>N</i>	< -0.001	.43	< 0.001	.01
LRT	$\chi^2(1) = 0.60, p = .44$		$\chi^2(1) = 6.62, p = .01$	
Step 3				
Initialstudien ES	0.863 [0.730, 0.997]	< .001	—	—
Initialstudien <i>N</i>	< -0.001	.22	< 0.001	.15
Publikationsjahr	-0.002 [-0.004, < -0.001]	.03	< -0.001	.24
Initialstudie				
LRT	$\chi^2(2) = 3.25, p = .20$		$\chi^2(1) = 6.51, p = .01$	
Step 4				
Initialstudien ES	0.895 [0.752, 1.037]	< .001	—	—
Initialstudien <i>N</i>	< -0.001	.08	< 0.001	.93
Publikationsjahr	-0.004 [-0.006, -0.001]	.01	< -0.001	.03
Initialstudie				
Jährliche	0.003 [-0.001, 0.006]	.21	< 0.001	.08
Zitationszahl der				
Initialstudie				
LRT	$\chi^2(3) = 4.83, p = .18$		$\chi^2(1) = 3.02, p = .08$	
Step 5				
Initialstudien ES	0.895 [0.752, 1.037]	< .001	—	—
Initialstudien <i>N</i>	< -0.001	.08	—	—
Publikationsjahr	-0.004 [-0.006, -0.001]	.01	< -0.001	.002
Initialstudie				

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Jährliche Zitationszahl der Initialstudie	0.003 [-0.001, 0.006]	.21	< 0.001	.02
Impact Factor 2020	—	—	< -0.001	.41
LRT	— ^c		$\chi^2(2) = 3.67, p = .06$	

Anmerkung. KI = Konfidenzintervall; LRT = Likelihood Ratio Test. Anhand eines mixed-effects Modells wurde sowohl für die Crude differences, als auch für die Regressionskoeffizienten, der Einfluss von Charakteristika der Initialstudie auf den decline effect überprüft. Dazu wurden die einzelnen Prädiktoren schrittweise in das Regressionsmodell aufgenommen und mittels eines LRT auf den Model-Fit überprüft. Wenn Prädiktoren einen Varianzinflationsfaktor > 4 berichteten, so wurden sie aus dem Modell ausgeschlossen. Dies wurde durch leerstehende Zellen gekennzeichnet.

^a *k* = 27

^b *k* = 24

^c Step 5 nach Ausschluss des Impact Factor als Prädiktor ident mit Step 4, wodurch auch das Ergebnis des LRT ident ist.

und der Impact Factor des Journals, in dem die Initialstudie publiziert wurde, nacheinander als Prädiktor in das Modell aufgenommen. Zur Berechnung wurde ein mixed-effects Modell mit einem Maximum-Likelihood-Schätzer durchgeführt und der Model-Fit wurde nach jedem Schritt mittels eines Likelihood-Ratio-Tests (LRT) überprüft. Da die jährliche Anzahl an Zitationen einer Initialstudie und der Impact Factor ihres Journals nicht immer feststellbar war, mussten jene Metaanalysen, für die diese Daten nicht erfassbar waren, auch wenn für die anderen Prädiktoren Daten vorlagen, aus dem gesamten Modell exkludiert werden, um den LRT durchführen zu können. Der endgültige Datensatz für diese Analyse bestand aus *k* = 27 Metaanalysen, welcher für die Berechnung mit Daten, die ausschließlich publizierter Primärstudien umfassen sollte, nochmals auf *k* = 14 reduziert wurde. Um zu überprüfen, ob es durch die Exklusion der Metaanalysen zu einer nennenswerten Veränderung der Datenlage gekommen ist, wurde für die Schritte eins bis vier, ebenfalls die Analyse ohne Exklusion von Daten, durchgeführt. Die Ergebnisse des ungefilterten Datensatzes sind in Tabelle S1 und S2

Tabelle 9

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen mixed-effects Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable, nur anhand publizierter Daten berechnet

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Step 1				
Initialstudien ES	0.494 [-0.266, 1.255]	.20	-0.005 [-0.017, 0.008]	.45
Step 2				
Initialstudien ES	0.495 [-0.365, 1.355]	.26	-0.007 [-0.019, 0.006]	.28
Initialstudien <i>N</i>	< 0.001	1.00	< 0.001	.19
LRT	$\chi^2(1) = 0.00, p = 1.00$		$\chi^2(1) = 1.70, p = .19$	
Step 3				
Initialstudien ES	0.472 [-0.351, 1.295]	.26	-0.013 [-0.024, -0.01]	.01
Initialstudien <i>N</i>	< 0.001	.61	< 0.001	.07
Publikationsjahr Initialstudie	-0.003 [-0.007, 0.001]	.17	< -0.001	.01
LRT	$\chi^2(2) = 1.15, p = .56$		$\chi^2(2) = 7.55, p = .02$	
Step 4				
Initialstudien ES	0.472 [-0.351, 1.295]	.26	-0.011 [-0.024, 0.002]	.09
Initialstudien <i>N</i>	< 0.001	.61	< 0.001	.50
Publikationsjahr Initialstudie	-0.003 [-0.007, 0.001]	.17	< -0.001	.01
Jährliche Zitationszahl der Initialstudie	—	—	< 0.001	.57
LRT	— ^c		$\chi^2(1) = 0.30, p = .59$	
Step 5				
Initialstudien ES	0.456 [-0.487, 1.399]	.34	-0.009 [-0.024, 0.005]	.21

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95%KI]	<i>p</i>	<i>b</i> [95%KI]	<i>p</i>
Initialstudien <i>N</i>	< 0.001	.72	< 0.001	.40
Publikationsjahr	-0.003 [-0.008, 0.002]	.19	< -0.001	.02
Initialstudie				
Jährliche	—	—	< 0.001	.64
Zitationszahl				
der Initialstudie				
Impact Factor	0.003 [-0.080, 0.085]	.95	< -0.001	.64
2020				
LRT	$\chi^2(3) = 1.15, p = .76$		$\chi^2(1) = 0.51, p = .77$	

Anmerkung. KI = Konfidenzintervall; LRT = Likelihood Ratio Test. Anhand eines mixed-effects Modells wurde anhand eines Datensatzes, welcher nur auf publizierten Primärstudien basierte, sowohl für die Crude differences, als auch für die Regressionskoeffizienten, der Einfluss von Charakteristika der Initialstudie auf den decline effect überprüft. Dazu wurden die einzelnen Prädiktoren schrittweise in das Regressionsmodell aufgenommen und mittels eines LRT auf den Model-Fit überprüft. Wenn Prädiktoren einen Varianzinflationsfaktor > 4 berichteten, so wurden sie aus dem Modell ausgeschlossen. Dies wurde durch leerstehende Zellen gekennzeichnet.

^a *k* = 14

^b *k* = 26

^c Step 4 nach Ausschluss der jährlichen Zitationszahl als Prädiktor ident mit Step 3, wodurch auch das Ergebnis des LRT ident ist.

im Appendix abgebildet. Die Resultate der Regressionsanalyse basierend auf *k* = 27 Metaanalysen sind aus Tabelle 8 zu entnehmen.

Anhand der Ergebnisse der einzelnen LRTs zeigte sich, dass für das Modell 1 der erste Schritt, mit der Effektstärke der Initialstudie als Prädiktor, im Vergleich zu allen anderen Schritten den besten Model-Fit aufwies und somit die Heterogenität in den Daten durch diesen einen Prädiktor schon ausreichend gut erklärt werden konnte (statistische Kennwerte in Tabelle 9 abgebildet). Das Modell konnte $R^2 = 94.04\%$ der Heterogenität in den Daten erklären und fiel signifikant aus, $\chi^2(1) = 113.58, p < .001$. Die Effektstärke der Initialstudie wies einen signifikanten positiven Zusammenhang mit dem decline effect auf, $b = 0.92, p <$

.001. Je höher der berichtete Initialstudieneffekt also ausfiel, desto stärker fielen auch die Crude differences in Richtung eines decline effect aus.

Bei der Reanalyse, basierend auf dem Datensatz mit ausschließlich publizierten Primärstudien ($k = 14$), ergab sich ein ähnliches Bild. Auch hier erwies sich das Modell mit dem Initialstudieneffekt als alleinigen Prädiktor, als das Modell mit dem besten Model-Fit im Vergleich zu allen anderen Prädiktoren (siehe Tabelle 9). Das Modell fiel nicht signifikant aus und der Anteil an erklärter Heterogenität war kleiner, $R^2 = 19.25\%$, $\chi^2(1) = 1.62$, $p = .20$. Der Prädiktor zeigte dieselbe Wirkrichtung, wie in der Regressionsanalyse, die auch graue Literatur inkludierte, der Zusammenhang war jedoch nicht signifikant, $b = 0.49$, $p = .20$.

Modell 2: Der Einfluss von Studiencharakteristika der Metaanalyse auf Crude Differences

In Modell 2 wurde der berichtete Summary effect der Metaanalyse, die Anzahl der inkludierten Stichproben, die abgedeckte Zeitspanne durch die Metaanalyse, die Art der inkludierten Studie (ob graue Literatur inkludiert war oder nur publizierte Primärstudien) und das Publikationsjahr der Metaanalyse einzeln, schrittweise in das Modell eingefügt. Nach jedem Schritt wurde ein LRT zur Überprüfung des Model-Fit durchgeführt. In der Analyse zu Modell 2 gab es keine fehlenden Werte, weshalb der volle Datensatz mit $k = 36$ Metaanalysen verwendet werden konnte. Für die Regressionsanalyse mit nur publizierten Daten wurden $k = 16$ Studien inkludiert. Die Ergebnisse der beiden Regressionsmodelle sind in Tabelle 10 und 11 ersichtlich.

Anhand der Resultate der LRTs ließ sich erkennen, dass der zweite Schritt der multiplen hierarchischen Meta-Meta-Regression, im Vergleich zu allen anderen Schritten, den besten Model-Fit aufwies (siehe Tabelle 10). Auch verglichen mit dem Intercept-Modell konnte sich der zweite Schritt durchsetzen, $R^2 = 51.18\%$, $\chi^2(2) = 17.37$, $p < .001$. Der Summary effect als Prädiktor wies keinen signifikanten Zusammenhang mit den Crude differences auf ($b = -.03$, $p = .88$), die Anzahl inkludierter Stichproben, auf der anderen Seite, erwies sich als signifikant positiv zusammenhängend mit der Stärke des decline effect, $b < .01$, $p < .001$.

Die Ergebnisse der Reanalyse dieses Modells, mit nur publizierten Daten, zeigten ein ähnliches Bild (siehe Tabelle 11). Es kam aber zu einer kleineren Anpassung des Modells, nachdem vier statt fünf Schritte durchgeführt wurden. Dies war darauf zurückzuführen, dass die Art der inkludierten Studien innerhalb einer Metaanalyse in dieser Berechnung nicht variieren konnte, da nur publizierte Studien inkludiert wurden. Dementsprechend wurde dieser Prädiktor für diese Analyse aus dem Modell entfernt. Die Ergebnisse der Analyse haben gezeigt, dass erneut der zweite Schritt den besten Model-Fit im Vergleich zu allen

Tabelle 10

Ergebnisse der Modelle 2 und 4 der theoriegeleiteten hierarchischen multiplen mixed effects Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariablen

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Step 1				
Summary effect	-0.338 [-1.114, 0.438]	.39	0.011 [-0.004, 0.026]	.14
Step 2				
Summary effect	-0.049 [-0.706, 0.607]	.88	0.008 [-0.007, 0.022]	.29
<i>k</i>	0.001 [< 0.001 , 0.001]	$< .001$	< 0.001	$< .001$
LRT	$\chi^2(1) = 11.00, p < .001$		$\chi^2(1) = 0.00, p = 1.00$	
Step 3				
Summary effect	-0.013 [-0.685, 0.659]	.97	0.004 [-0.011, 0.018]	.49
<i>k</i>	0.001 [< 0.001 , 0.001]	$< .001$	< 0.001	$< .001$
Abgedeckte	0.001 [-0.003, 0.005]	.66	< -0.001	.005
Zeitspanne				
LRT	$\chi^2(1) = 0.20, p = .66$		$\chi^2(2) = 7.05, p = .03$	
Step 4				
Summary effect	-0.044 [-0.687, 0.599]	.89	0.003 [-0.013, 0.018]	.74
<i>k</i>	0.001 [< 0.001 , 0.001]	$< .001$	< 0.001	$< .001$
Abgedeckte	0.001 [- 0.003, 0.005]	.63	< -0.001	.008
Zeitspanne				
Art inkludierter	0.140 [-0.014, 0.294]	.08	0.001 [-0.003, 0.005]	.63
Studie				
LRT	$\chi^2(2) = 3.12, p = .21$		$\chi^2(1) = 0.24, p = .63$	
Step 5				
Summary effect	-0.085 [-0.739, 0.569]	.80	0.003 [-0.013, 0.018]	.74
<i>k</i>	< 0.001	$< .001$	< 0.001	$< .001$
Abgedeckte	0.001 [-0.003, 0.005]	.67	< -0.001	.008
Zeitspanne				
Art inkludierter	0.146 [-0.009, 0.300]	.07	0.001 [-0.003, 0.005]	.63
Studie				

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Publikationsjahr	-0.003 [-0.012, 0.006]	.52	–	–
Metaanalyse				
LRT	$\chi^2(3) = 3.54, p = .32$		– ^c	

Anmerkung. KI = Konfidenzintervall; LRT = Likelihood Ratio Test. Anhand eines mixed-effects Modells wurde sowohl für die Crude differences, als auch für die Regressionskoeffizienten, der Einfluss von Charakteristika der Metaanalyse auf den decline effect überprüft. Dazu wurden die einzelnen Prädiktoren schrittweise in das Regressionsmodell aufgenommen und mittels eines LRT auf den Model-Fit überprüft. Wenn Prädiktoren einen Varianzinflationsfaktor > 4 berichteten, so wurden sie aus dem Modell ausgeschlossen. Dies wurde durch leerstehende Zellen gekennzeichnet.

^a *k* = 36

^b *k* = 34

^c Step 5 nach Ausschluss des Impact Factor als Prädiktor ident mit Step 4, wodurch auch das Ergebnis des LRT ident ist.

anderen Schritten aufweist (siehe Tabelle 11). Der Summary effect der Metaanalyse als zeigte dabei keine signifikanten Einflüsse, die Anzahl inkludierter Stichproben jedoch schon, $b = 0.06, p = .90$; $b < 0.01, p = .05$. Es konnte ein signifikanter Anteil der Heterogenität in den Daten durch dieses Modell erklären, $R^2 = 100.00\%, \chi^2(2) = 11.39, p < .003$.

Modell 3: Der Einfluss von Studiencharakteristika der Initialstudie auf Regressionskoeffizienten

Modell 3 entsprach in seinem Aufbau Modell 1, mit dem einzigen Unterschied, dass nun der Regressionskoeffizient der jährlichen Effektveränderung als Zielvariable diente. Wie auch bei Modell 1 kam es in Modell 3 zu fehlenden Werten in den Prädiktoren „jährliche Anzahl an Zitationen“ und „Impact Factor“. Um einen Vergleich der einzelnen Modelle mittels des LRTs zu ermöglichen, mussten jene Metaanalysen, die fehlende Werte berichteten, erneut aus der Analyse ausgeschlossen werden. Die nun folgende multiple hierarchische Meta-Regressionsanalyse baute sowohl für die Analyse mit also auch ohne grauer Literatur dementsprechend auf $k = 26$ Metaanalysen auf. Zu jenen Schritten, die mehr Daten lieferten wurde die Analyse wiederholt, um zu überprüfen, ob es durch die Exklusion der Metaanalysen aus den einzelnen Schritten zu einer nennenswerten Veränderung der

Tabelle 11

Ergebnisse der Modelle 2 und 4 der theoriegeleiteten hierarchischen multiplen mixed-effects Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable, nur anhand publizierter Daten berechnet

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Step 1				
Summary effect	-0.676 [-1.159, -0.192]	.006	0.012 [-0.010, 0.036]	.33
Step 2				
Summary effect	0.058 [-0.817, 0.928]	.90	0.012 [-0.011, 0.035]	.31
<i>k</i>	0.001 [<0.001, 0.001]	.05	< 0.001	.20
LRT	$\chi^2(1) = 3.90, p = .05$		$\chi^2(1) = 1.68, p = .19$	
Step 3				
Summary effect	-0.433 [-1.232, 0.366]	.29	0.012 [-0.012, 0.036]	.32
<i>k</i>	—	—	< 0.001	.23
Abgedeckte	0.002 [-0.004, 0.008]	.46	< 0.001	.92
Zeitspanne				
LRT	$\chi^2(1) = 0.56, p = .46$		$\chi^2(2) = 1.68, p = .43$	
Step 4				
Summary effect	-0.287 [-1.131, 0.558]	.51	0.010 [-0.014, 0.033]	.43
<i>k</i>	—	—	< 0.001	.32
Abgedeckte	0.004 [-0.003, 0.011]	.25	< 0.001	.84
Zeitspanne				
Publikationsjahr	0.006 [-0.005, 0.018]	.29	< -0.001	.39
Metaanalyse				
LRT	$\chi^2(2) = 1.66, p = .44$		$\chi^2(3) = 2.47, p = .48$	

Anmerkung. KI = Konfidenzintervall; LRT = Likelihood Ratio Test. Anhand eines mixed-effects Modells wurde anhand eines Datensatzes, welcher nur auf publizierten Primärstudien basierte, sowohl für die Crude differences, als auch für die Regressionskoeffizienten, der Einfluss von Charakteristika der Metaanalyse auf den decline effect überprüft. Dazu wurden die einzelnen Prädiktoren schrittweise in das Regressionsmodell aufgenommen und mittels eines LRT auf den Model-Fit überprüft. Wenn Prädiktoren einen Varianzinflationsfaktor > 4 berichteten, so wurden sie aus dem Modell ausgeschlossen. Dies wurde durch leerstehende Zellen gekennzeichnet.

^a $k = 16$
^b $k = 34$

Datenlage gekommen ist. Die Ergebnisse dieser Analysen sind in Tabelle S1 und S2 im Appendix abgebildet.

Aus Tabelle 8 sind die Ergebnisse der Analyse zu entnehmen, die auch graue Literatur inkludiert hat. Dabei zeigte sich, dass über den Verlauf hinweg die Initialstudien Effektstärke und ihre Stichprobengröße, aufgrund von Anzeichen von Multikollinearität, als Prädiktor aus dem Modell entfernt werden mussten. Dies führte dazu, dass durch den LRT kein direkter Vergleich zwischen Schritt Zwei und Drei durchgeführt werden konnte, da sich die Anzahl an Prädiktoren im Modell nicht verändert hatte. Stattdessen wurden beide Modelle mit Schritt 1 verglichen und überprüft, welches Modell letztlich das niedrigere AIC und AICc aufwies.

Sowohl Schritt Zwei als auch Schritt Drei erklärten das Regressionsmodell besser als Modell Eins, aber das AIC und AICc fiel für den zweiten Schritt ($AIC = -150.56$, $AICc = -148.65$) niedriger aus als für den dritten Schritt ($AIC = -150.45$, $AICc = -148.54$). Schritt Zwei wies daher einen bessern Model-Fit auf, als Schritt Drei, aber auch im Vergleich zu den anderen Modellschritten, schnitt der zweite Schritt am besten ab (siehe Tabelle 8). Im Omnibustest erwies sich das Modell als signifikant und es konnten $R^2 = 100.00\%$ der Heterogenität in den Daten erklärt werden, $\chi^2(2) = 8.59$, $p = .01$. Auf Seiten der Prädiktoren ließ sich erkennen, dass die Stichprobengröße der Initialstudie ($b < .01$, $p = .01$) einen signifikant positiven Zusammenhang zur jährlichen Effektveränderung aufwies, währenddessen die Effektstärke der Initialstudie ($b < -0.01$, $p = .22$) einen nicht signifikanten negativen Zusammenhang aufwies.

Wenn die analysierten Metanalysen nur auf publizierten Primärstudien aufbauten, kam es erneut zu etwas anderen Ergebnissen. Wie aus Tabelle 9 zu entnehmen ist, wies hier der dritte Schritt des multiplen hierarchischen Meta-Metaregressionsmodells, im Vergleich zu allen anderen Modellen, den besten Model-Fit auf. Sowohl die Effektstärke der Initialstudie ($b = -0.01$, $p = .01$) als auch ihr Publikationsjahr ($b < -0.01$, $p = .01$) wiesen signifikante negative Zusammenhänge mit der jährlichen Effektveränderung auf. Die Stichprobengröße der Initialstudie erwies sich nicht als signifikanter Prädiktor in diesem Modell, $b < 0.01$, $p = .07$. Alle drei Prädiktoren trugen zu einer erklärten Heterogenität von $R^2 = 55.67\%$ bei und das Modell fiel signifikant aus, $\chi^2(3) = 8.70$, $p = .03$.

Modell 4: Der Einfluss von Studiencharakteristika der Metaanalyse auf Regressionskoeffizienten

Modell 4 entsprach in seinem Aufbau dem von Modell 2, mit dem einzigen Unterschied, dass nun der Regressionskoeffizient der jährlichen Effektveränderung als Zielvariable diente. Nachdem keiner der Prädiktoren dieses Modells fehlende Werte aufwies, baute diese Analyse auf $k = 34$ Metaanalysen auf.

Wie in Tabelle 10 erkennbar ist, stellte sich für Modell 4 der dritte Schritt als das Modell mit dem besten Model-Fit heraus. Sowohl die Anzahl inkludierter Samples ($b < .01$, $p < .001$) als auch die abgedeckte Zeitspanne einer Metaanalyse ($b < -0.01$, $p = .005$) erwiesen sich als signifikante Prädiktoren in dem Modell, wobei Ersterer einen positiven Zusammenhang und Zweiterer einen negativen Zusammenhang mit der jährlichen Effektveränderung aufwies. Der Summary effect zeigte einen nicht signifikanten positiven Zusammenhang mit der Zielvariable, $b < 0.01$, $p = .49$. Das Modell fiel im Model-Fit-Vergleich zum Intercept-Modell signifikant aus und es konnte einen hohen Anteil der Heterogenität in den Daten erklären, $R^2 = 100.00\%$, $\chi^2(3) = 37.40$, $p < .001$.

In der Reanalyse dieses Modells, mit nur publizierten Primärstudien innerhalb einer Metaanalyse, konnte keiner der Schritte im Vergleich mit dem ersten Schritt einen signifikant besseren Model-Fit erzielen (siehe Tabelle 11). Der Summary effect einer Metaanalyse stellte dabei den einzigen inkludierten Prädiktor im Modell dar und zeigte einen nicht signifikanten positiven Zusammenhang mit der jährlichen Effektveränderung, $b = 0.01$, $p = .33$. Es konnte nur ein geringer Anteil der Heterogenität durch dieses Modell erklärt werden, $R^2 < 0.01$, $\chi^2(1) = 0.96$, $p = .33$.

Multiple hierarchische lineare Meta-Metaregression

Wie bereits im Methodikteil angesprochen, wurde zusätzlich zur Berechnung der Regressionsmodelle, mittels eines mixed-effects Modells, auch das lineare Modell für die Berechnung dieser Modelle herangezogen, um eine Analyse dieser ohne Einfluss von Gewichtung zu ermöglichen.

Modell 1: Der Einfluss von Studiencharakteristika der Initialstudie auf Crude Differences

Wie auch beim mixed effects Modell musste für die Analyse ein gefilterter Datensatz genutzt werden, da die Variablen „Impact Factor“ und „jährliche Anzahl an Zitationen“ fehlende Werte aufwiesen. Dies führte zu einem Datensatz von $k = 28$ bzw. $k = 14$ Metaanalysen für den Datensatz, welcher nur publizierte Daten inkludierte (Ergebnisse der Berechnungen ohne Filter sind in Tabelle S3 und S4 im Appendix abgebildet).

Tabelle 12

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen linearen Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable.

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Step 1				
Initialstudien ES	0.926 (0.074)	< .001	-0.016 (0.012)	.18
Step 2				
Initialstudien ES	0.921 (0.075)	< .001	-0.015 (0.012)	.22
Initialstudien <i>N</i> ANOVA	< -0.001 (< 0.001) $F(1, 24) = 0.268, p = .61$	.59	< 0.001 (<0.001) $F(1, 23) = 0.827, p = .37$	.42
Step 3				
Initialstudien ES	0.912 (0.080)	< .001	-0.026 (0.011)	.03
Initialstudien <i>N</i>	< -0.001 (< 0.001)	.63	< 0.001 (<0.001)	.23
Publikationsjahr	< -0.001 (0.001)	.71	< -0.001 (<0.001)	.01
Initialstudie ANOVA	$F(1, 23) = .131, p = .72$		$F(1, 22) = 7.519, p = .01$	
Step 4				
Initialstudien ES	0.915 (0.086)	< .001	-0.025 (0.012)	.04
Initialstudien <i>N</i>	< -0.001 (< 0.001)	.60	< 0.001 (< 0.001)	.98
Publikationsjahr	< -0.001 (< 0.001)	.65	< -0.001 (< 0.001)	.008
Initialstudie Jährliche Zitationszahl der Initialstudie ANOVA	0.001 (0.002) $F(1, 22) = 0.102, p = .75$	.75	< 0.001 (< 0.001) $F(1, 21) = 0.658, p = .43$	.41
Step 5				
Initialstudien ES	0.911 (0.086)	< .001	-0.025 (0.012)	.05
Initialstudien <i>N</i>	< -0.001 (< 0.001)	0.59	< -0.001 (<0.001)	.98

Prädiktor	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Publikationsjahr	-0.001 (0.001)	.65	< -0.001 (< 0.001)	.01
Initialstudie				
Jährliche	0.001 (0.002)	.74	< 0.001 (<0.001)	.43
Zitationszahl				
der Initialstudie				
Impact Factor	0.002 (0.011)	.88	< 0.001 (0.002)	.99
2020				
ANOVA	$F(1, 21) = 0.024, p = .88$		$F(1, 20) < 0.001, p = .99$	

Anmerkung. Anhand eines linearen Modells wurden die Prädiktoren schrittweise in das formulierte Modell aufgenommen und mittels einer ANOVA auf den Model-Fit geprüft. In Modell 1 und Modell 3 wurde der Einfluss von verschiedenen Charakteristika der Initialstudie auf den decline effect überprüft.

^a $k = 27$

^b $k = 24$

Für die Reanalyse von Modell 1 zeigte der erste Schritt im Vergleich zu allen anderen Modellen den besten Model-Fit (siehe Tabelle 12). Die Effektstärke der Initialstudie war in diesem Schritt der einzige Prädiktor im Modell und wies einen signifikant positiven Zusammenhang mit dem decline effect auf, $b = 0.926, p < .001$. Das Modell selbst fiel ebenfalls signifikant aus und konnte einen großen Anteil der Heterogenität in den Daten erklären, $R^2 = 86.58\%, F(1, 25) = 158.55, p < .001$.

Auch wenn die untersuchten Metaanalysen nur publizierte Studien inkludierten, stellte sich der erste Schritt als der mit dem besten Model-Fit heraus (siehe Tabelle 13). Erneut zeigte sich ein signifikanter positiver Zusammenhang zwischen der Effektstärke der Initialstudie und der Stärke des decline effect, $b = 0.48, p = .01$. Das Modell konnte $R^2 = 37.79\%$ der Heterogenität erklären und fiel signifikant aus, $F(1, 12) = 8.90, p = .01$.

Modell 2: Der Einfluss von Studiencharakteristika der Metaanalyse auf Crude Differences

Für Modell 2, welches auf $k = 36$ und $k = 16$ Metanalysen aufbaute, stellte sich bei der Regressionsanalyse der zweite Schritt, im Vergleich zu allen anderen Schritten, als der mit dem besten Model-Fit heraus (siehe Tabelle 14). Der Summary effect einer Metaanalyse ($b = -0.16, p = .35$) stellte sich nicht als signifikanter Prädiktor heraus, die Anzahl der Stichproben, auf die eine Metaanalyse aufbaute ($b < .01, p = .04$), zeigte jedoch einen signifikant positiven

Tabelle 13

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen linearen Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Kriterien, nur anhand publizierter Daten berechnet.

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Step 1				
Initialstudien ES	0.476 (0.160)	.01	-0.054 (0.025)	.04
Step 2				
Initialstudien ES	0.459 (0.165)	.02	-0.051 (0.026)	.06
Initialstudien <i>N</i> ANOVA	< 0.001 (< 0.001) $F(1, 11) = 0.518, p = .49$	.48	< 0.001 (< 0.001) $F(1, 23) = 0.843, p = .37$	.38
Step 3				
Initialstudien ES	0.457 (0.173)	.02	-0.069 (0.026)	.01
Initialstudien <i>N</i>	< 0.001 (< 0.001)	.59	< 0.001 (< 0.001)	.26
Publikationsjahr Initialstudie ANOVA	< 0.001 (0.002) $F(1, 10) = 0.047, p = .83$	.84	-0.001 (< 0.001) $F(1, 22) = 3.873, p = .06$	.054
Step 4				
Initialstudien ES	0.464 (0.191)	.04	-0.071 (0.026)	.01
Initialstudien <i>N</i>	< 0.001 (< 0.001)	.60	< 0.001 (< 0.001)	.28
Publikationsjahr Initialstudie Jährliche Zitationszahl der Initialstudie ANOVA	< 0.001 (0.002) $F(1, 9) = 0.020, p = .89$	.83	-0.001 (< 0.001) $F(1, 21) = 0.292, p = .60$	.11
Step 5				
Initialstudien ES	0.377 (0.183)	.07	-0.073 (0.027)	.01

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Initialstudien <i>N</i>	< 0.001 (< 0.001)	.51	< 0.001 (< 0.001)	.36
Publikationsjahr	< -0.001 (0.002)	.83	-0.001 (< 0.001)	.11
Initialstudie				
Jährliche	-0.004 (0.004)	.40	-0.001 (0.001)	.59
Zitationszahl				
der Initialstudie				
Impact Factor	0.028(0.017)	.14	0.002 (0.003)	.61
2020				
ANOVA	$F(1, 8) = 2.742, p = .14$		$F(1, 20) = 0.274, p = .61$	

Anmerkung. Anhand eines linearen Modells wurde nur mit einem Datensatz, basierend auf nur publizierten Daten, der Einfluss diverser Initialstudien Charakteristika untersucht. Die Prädiktoren wurden dazu schrittweise in das Modell aufgenommen und mittels einer ANOVA auf ihren Model-Fit überprüft.

^a $k = 14$

^b $k = 26$

Zusammenhang mit der Stärke des decline effect. Im Vergleich zu dem Modell, welches nur den Intercept inkludiert hatte, fiel das Modell 2 jedoch nicht signifikant aus, $R^2 = 9.16\%$, $F(2, 33) = 2.76, p = .08$.

In der Reanalyse, die nur die publizierten Daten untersuchte, wurde von fünf auf vier Schritte reduziert, nachdem die Art der inkludierten Primärstudie innerhalb einer Metaanalyse als Prädiktor aus diesem Modell entfernt wurde. Es konnte bei der Analyse keiner der Schritte einen signifikanten Model-Fit aufweisen. Schritt Eins wies zwar, wie aus Tabelle 15 zu entnehmen ist, einen besseren Model-Fit im Vergleich zu allen anderen Schritten des multiplen hierarchischen Modells auf, im Modellvergleich mit einem Modell ohne Prädiktoren, kam es jedoch zu keinem signifikanten Ergebnis. Das errechnete R^2 erreichte einen negativen Wert, welcher, wie in Bakeman und Robinson (2005) vorgeschlagen, als Null interpretiert wurde, $R^2 = 0.00\%$, $F(1, 12) = 0.12, p = .73$.

Modell 3: Der Einfluss von Studiencharakteristika der Initialstudie auf Regressionskoeffizienten

Wie auch bei dem zuvor berechneten mixed-effects Modell, baute die Reanalyse des Modells 3, aufgrund fehlender Werte beim „Impact Factor“ und der „jährlichen Anzahl der

Tabelle 14

Ergebnisse der Modelle 2 und 4 der theoriegeleiteten hierarchischen multiplen linearen Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariablen.

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Step 1				
Summary effect	-0.290 (0.087)	.43	0.006 (0.025)	.81
Step 2				
Summary effect	-0.160 (0.347)	.64	0.007 (0.024)	.77
<i>k</i>	0.001 (< 0.001)	.04	< 0.001 (< 0.001)	.17
ANOVA	<i>F</i> (1, 33) = 5.089, <i>p</i> = .03		<i>F</i> (1, 31) = 1.994, <i>p</i> = .17	
Step 3				
Summary effect	-0.175 (0.357)	.63	0.012 (0.024)	.61
<i>k</i>	0.001 (< 0.001)	.04	< 0.001 (< 0.001)	.49
Abgedeckte	-0.001 (0.002)	.80	< 0.001 (< 0.001)	.14
Zeitspanne				
ANOVA	<i>F</i> (1, 32) = 0.068, <i>p</i> = .80		<i>F</i> (1, 30) = 2.290, <i>p</i> = .14	
Step 4				
Summary effect	-0.230 (0.347)	.51	0.017 (0.024)	.50
<i>k</i>	0.001 (< 0.001)	.06	< 0.001 (< 0.001)	.34
Abgedeckte	-0.001 (0.003)	.71	< 0.001 (< 0.001)	.12
Zeitspanne				
Art inkludierter	0.162 (0.090)	.08	-0.008 (0.007)	.23
Studie				
ANOVA	<i>F</i> (1, 31) = 3.306, <i>p</i> = .08		<i>F</i> (1, 29) = 1.460, <i>p</i> = .24	
Step 5				
Summary effect	-0.313 (0.350)	.38	0.014 (0.026)	.59
<i>k</i>	0.001 (< 0.001)	.08	< 0.001 (< 0.001)	.37
Abgedeckte	-0.001 (0.003)	.70	< 0.001 (< 0.001)	.13
Zeitspanne				
Art inkludierter	0.165 (0.089)	.07	-0.008 (0.007)	.24
Studie				

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Publikationsjahr	-0.005 (0.004)	.22	< 0.001 (< 0.001)	.76
Metaanalyse				
ANOVA	$F(1, 30) = 1.537, p = .22$		$F(1, 28) = 0.099, p = .76$	

Anmerkung. Um den Einfluss von Merkmalen der Metaanalyse zu überprüfen, wurden in Modell 2 und 4 jeweils ein lineares Modell für die Berechnung herangezogen und schrittweise jeder der Prädiktoren in das Modell eingefügt und der Einfluss auf die beiden Indizes für den decline effect untersucht.

^a $k = 36$

^b $k = 34$

Zitationen“, auf $k = 26$ Metaanalysen auf (Ergebnisse der Berechnungen ohne Filter sind in Tabelle S3 und S4 im Appendix abgebildet). Es stellte sich der dritte Schritt, mit der Initialstudieneffektstärke, ihrer Stichprobengröße und ihrem Publikationsjahr als Prädiktor, im Vergleich zu allen anderen Schritten, als das Modell mit dem besten Model-Fit heraus (siehe Tabelle 13). Sowohl die Initialstudieneffektstärke als auch das Publikationsjahr wiesen dabei signifikante Zusammenhänge mit einer Effektreduktion auf, $b = -0.03, p = .03$; $b < -0.01, p = .01$. Die Stichprobengröße zeigte keine signifikanten Einflüsse, $b < 0.01, p = .23$. Verglichen mit dem Intercept-Modell kam es ebenfalls zu einem signifikanten Ergebnis, $R^2 = 24.80\%, F(3, 22) = 3.75, p = .03$.

Wenn nur publizierte Daten für die Analyse inkludiert waren, konnte keiner der Schritte die Daten signifikant besser als Schritt Eins erklären (siehe Tabelle 13). Zusätzlich zeigte sich auch ein besserer Model-Fit im Vergleich zum Modell ohne Prädiktor, $R^2 = 12.27\%, F(1, 24) = 5.50, p = .04$. Die Effektstärke der Initialstudie wies dabei einen signifikanten negativen Zusammenhang mit der jährlichen Effektveränderung auf, $b = -0.05, p = .04$.

Modell 4: Der Einfluss von Studiencharakteristika der Metaanalyse auf Regressionskoeffizienten

Im vierten und letzten formulierten Regressionsmodell zeigte sich bei der Analyse anhand des linearen Modells, dass keiner der vorformulierten Schritte signifikant zur Erklärung der Daten beitragen konnte. Zwar wies Schritt Eins, wie aus Tabelle 14 zu entnehmen ist, einen besseren Model-Fit als alle anderen Schritte des Modells auf, der F -Test auf Gesamtsignifikanz des Modells fiel jedoch nicht signifikant aus. Erneut fiel das R^2 negativ

Tabelle 15

Ergebnisse der Modelle 2 und 4 der theoriegeleiteten hierarchischen multiplen linearen Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable, nur anhand publizierter Daten berechnet

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Step 1				
Summary effect	-0.097 (0.275)	.73	0.051 (0.050)	.32
Step 2				
Summary effect	0.004 (0.330)	.89	0.052 (0.051)	.30
<i>k</i>	< 0.001 (<0.001)	.24	< 0.001 (< 0.001)	.38
ANOVA	<i>F</i> (1, 11) = 1.798, <i>p</i> = .21		<i>F</i> (1, 31) = 0.886, <i>p</i> = .35	
Step 3				
Summary effect	0.034 (0.296)	.91	0.060 (0.050)	.24
<i>k</i>	0.001 (< 0.001)	.17	< 0.001 (< 0.001)	.70
Abgedeckte	-0.003 (0.004)	.32	< 0.001 (< 0.001)	.28
Zeitspanne				
ANOVA	<i>F</i> (1, 10) = 0.786, <i>p</i> = .40		<i>F</i> (1, 30) = 1.317, <i>p</i> = .26	
Step 4				
Summary effect	0.179 (0.286)	.55	0.091 (0.051)	.08
<i>k</i>	0.001 (< 0.001)	.11	< 0.001 (< 0.001)	.55
Abgedeckte	-0.003 (0.004)	.47	< 0.001 (< 0.001)	.25
Zeitspanne				
Publikationsjahr	0.005 (0.003)	.13	0.001 (0.001)	.06
Metaanalyse				
ANOVA	<i>F</i> (1, 9) = 2.791, <i>p</i> = .13		<i>F</i> (1, 29) = 3.783, <i>p</i> = .06	

Anmerkung. Um den Einfluss von Merkmalen der Metaanalyse zu überprüfen, wurden in Modell 2 und 4 jeweils ein lineares Modell für die Berechnung herangezogen und schrittweise jeder der Prädiktoren in das Modell eingefügt und der Einfluss auf die beiden Indizes für den decline effect untersucht. Für die Analyse wurde hierbei ein Datensatz genutzt, der nur auf publizierten Primärstudien aufbaute.

^a *k* = 16

^b *k* = 34

aus und wurde dementsprechend als Null interpretiert, $R^2 = 0.00\%$, $F(1, 32) = 0.06$, $p = .81$. Der Summary effect einer Metaanalyse, welcher den Prädiktor im Modell darstellte, zeigte einen nicht signifikanten, positiven Zusammenhang zur jährlichen Effektveränderung, $b = 0.01$, $p = .81$.

Wie bei Modell 2 wurde auch für die Reanalyse von Modell 4 mit nur publizierten Daten ein Schritt weniger durchgeführt, nachdem die Art der inkludierten Primärstudie in dieser Analyse konstant gehalten wurde. Die Ergebnisse der ANOVA ließen darauf schließen, dass auch hier der erste Schritt, im Vergleich zu allen anderen Schritten, die Daten am besten erklärte (siehe Tabelle 15). Der Summary effect zeigte jedoch keinen signifikanten Einfluss auf die jährliche Effektveränderung, $b = 0.05$, $p = .32$. Zusätzlich zeigte sich, dass es auch im Vergleich zum Intercept-Modell zu keinem signifikanten Ergebnis kam und somit das Modell ohne Prädiktoren einen besseren Model-Fit aufzeigte, als die Modelle die Prädiktoren inkludiert hatten, $R^2 = 0.11\%$, $F(1, 32) = 1.04$, $p = .32$.

Tabelle 16

Ergebnisse der einfachen mixed-effects Meta-Metaregression mit Crude differences als Kriterium

Prädiktoren	b	p	95% KI		R^2
			UG	OG	
ES der Initialstudie	.907	< .001	.813	1.002	100.00%
N der Initialstudie	< -0.001	.57	-0.001	< 0.001	0.00%
Publikationsjahr der Initialstudie	-0.003	.12	-0.007	0.001	9.03%
Jährliche Zitationszahl der Initialstudie	-0.002	.30	-0.007	0.002	0.77%
Impact Factor 2020 Initialstudie	0.036	.16	-0.014	0.085	10.13%
Summary effect	-0.336	.41	-1.129	0.457	0.00%
k	0.001	< .001	< 0.001	0.001	47.18%
Publikationsjahr der Metaanalyse	-0.004	.12	-0.001	0.009	9.58%
Graue Literatur inkludiert	0.153	.14	-0.052	0.357	3.94%
Abgedeckte Zeitspanne	-0.004	.53	-0.015	0.008	0.00%

Anmerkung. KI = Konfidenzintervall. UG = Untere Grenze. OG = Obere Grenze. ES = Effektstärke

Einfache Meta-Metaregressionen

Um den Zusammenhang der einzelnen Prädiktoren mit den beiden decline effect Indizes isoliert betrachten zu können, wurde neben den multiplen hierarchischen Meta-Metaregressionsmodellen auch einfache Meta-Metaregressionen gerechnet. Für die Berechnung wurde hierbei zum einen, wie in der Präregistrierung vermerkt, ein mixed-effects Modell verwendet, zusätzlich wurde erneut ein lineares Modell genutzt. Darüber hinaus wurden sämtliche Regressionsanalysen einmal inklusive grauer Literatur gerechnet und ein weiteres Mal ausschließlich mit publizierten Studien.

Einfache mixed effects Meta-Metaregression

Crude differences. Die Ergebnisse der einzelnen Analysen inklusive und exklusive grauer Literatur, mit Crude differences als Zielvariable, sind aus Tabelle 16 und 17 zu entnehmen. Wenn alle Daten für die Analyse inkludiert waren, erwiesen sich die Initialstudieneffektstärke und die Anzahl inkludierter Studien in einer Metaanalyse, wie bereits in der multiplen hierarchischen Meta-Metaregression, als die beiden Prädiktoren mit

Tabelle 17

Ergebnisse der einfachen mixed-effects Meta-Metaregression mit Crude differences als Kriterium nur auf publizierten Primärstudien basierend

Prädiktoren	<i>b</i>	<i>p</i>	95% KI		<i>R</i> ²
			<i>UG</i>	<i>OG</i>	
ES der Initialstudie	0.491	.11	-0.115	1.097	100.00%
<i>N</i> der Initialstudie	< 0.001	.62	-0.001	0.001	0.00%
Publikationsjahr der Initialstudie	-0.003	.06	-0.006	< 0.001	100.00%
Jährliche Zitationszahl der Initialstudie	< 0.001	.95	-0.007	0.007	0.00%
Impact Factor 2020 Initialstudie	0.020	.42	-0.028	0.069	0.00%
Summary effect	-0.529	.13	-1.213	0.156	-
<i>k</i>	0.001	< .001	< 0.001	0.001	-
Publikationsjahr der Metaanalyse	0.002	.76	-0.009	0.013	0.00%
Abgedeckte Zeitspanne	0.004	.21	-0.002	0.009	-

Anmerkung. KI = Konfidenzintervall; UG = Untere Grenze; OG = Obere Grenze; ES = Effektstärke. In einigen Fällen war es in den Analysen nicht möglich, einen Wert für *R*² zu erhalten. Diese Fälle sind durch leerstehende Zellen gekennzeichnet.

signifikanten Einfluss auf die Crude differences. Beide Prädiktoren wiesen dabei einen positiven Zusammenhang mit der Stärke des decline effect auf, $b = 0.91$, $p < .001$; $b < 0.01$, $p < .001$. Alle anderen Prädiktoren konnten das Signifikanzniveau nicht erreichen.

Wenn die Analyse nur mit publizierten Daten durchgeführt wurde, zeigte sich ein etwas anderes Bild. Denn die Effektstärke der Initialstudie ($b = 0.49$, $p = .11$) zeigte nun keinen signifikant Zusammenhang mehr, aber das Publikationsjahr der Initialstudie wies einen marginal signifikanten negativen Zusammenhang mit dem decline effect, $b < -0.01$, $p = .06$. Die Anzahl inkludierter Stichproben, auf die eine Metaanalyse aufbaute, blieb auch in dieser Analyse signifikant und wies dabei einen positiven Zusammenhang mit dem decline effect auf, $b < .01$, $p < .001$.

Jährliche Effektveränderung. Auf Seiten des Regressionskoeffizienten der jährlichen Effektveränderung zeigte sich bei der einfachen mixed effects Meta-Meta-Regression, dass vor allem die Stichprobengröße der Initialstudie und das Publikationsjahr der

Tabelle 18

Ergebnisse der einfachen mixed-effects Meta-Metaregression mit dem Regressionskoeffizienten als Kriterium

Prädiktoren	b	p	95% KI		R^2
			UG	OG	
ES der Initialstudie	-0.007	.23	-0.017	0.004	0.00%
N der Initialstudie	< 0.001	.01	< 0.001	< 0.001	77.28%
Publikationsjahr der Initialstudie	< -0.001	.39	< -0.001	< 0.001	0.00%
Jährliche Zitationszahl der Initialstudie	< 0.001	.26	< -0.001	< 0.001	0.00%
Impact Factor 2020 Initialstudie	< -0.001	.98	-0.002	0.001	0.00%
Summary effect	0.011	.23	-0.007	0.029	50.61%
k	< 0.001	.24	< -0.001	< 0.001	0.00%
Publikationsjahr der Metaanalyse	< -0.001	.02	< -0.001	< -0.001	86.18%
Graue Literatur inkludiert	0.003	.30	-0.003	0.009	0.00%
Abgedeckte Zeitspanne	< -0.001	.70	< -0.001	< 0.001	61.48%

Anmerkung. KI = Konfidenzintervall. UG = Untere Grenze. OG = Obere Grenze. ES = Effektstärke

Tabelle 19

Ergebnisse der einfachen mixed-effects Meta-Metaregression mit Regressionskoeffizienten als Kriterium nur auf publizierten Primärstudien basierend

Prädiktoren	<i>b</i>	<i>p</i>	95% KI		<i>R</i> ²
			<i>UG</i>	<i>OG</i>	
ES der Initialstudie	-0.005	.45	-0.017	0.007	0.00%
<i>N</i> der Initialstudie	< 0.001	.20	< -0.001	< 0.001	0.00%
Publikationsjahr der Initialstudie	< -0.001	.42	< -0.001	< 0.001	0.00%
Jährliche Zitationszahl der Initialstudie	< 0.001	.39	< -0.001	< 0.001	0.00%
Impact Factor 2020 Initialstudie	< -0.001	.68	-0.002	0.001	0.00%
Summary effect	0.013	.32	-0.012	0.038	0.00%
<i>k</i>	< 0.001	.24	< -0.001	< 0.001	0.00%
Publikationsjahr der Metaanalyse	< -0.001	.24	-0.001	< 0.001	4.10%
Abgedeckte Zeitspanne	< 0.001	.75	< -0.001	< 0.001	0.00%

Anmerkung. KI = Konfidenzintervall. UG = Untere Grenze. OG = Obere Grenze. ES = Effektstärke

Metaanalyse eine Rolle zu spielen scheint. Ersterer Prädiktor wies dabei einen signifikanten, positiven Zusammenhang mit der Zielvariable auf, $b < 0.01$, $p = .01$. Das Publikationsjahr wiederum hing mit der jährlichen Effektveränderung signifikant negativ zusammen, $b < -0.01$, $p = .02$. Für alle anderen Prädiktoren konnten keine signifikanten Ergebnissen erzielt werden (siehe Tabelle 18).

Bei der Reanalyse dieser Berechnungen, ohne graue Literatur zu inkludieren, kam es im Gegenzug zu keinen signifikanten Ergebnissen. Diese Resultate sind in Tabelle 19 zu finden.

Einfache lineare Meta-Metaregression

Crude differences. Wurde zur Berechnung der Meta-Metaregression ein lineares Modell herangezogen, so zeigte sich ein ähnliches Bild. Wie in Tabelle 20 zu erkennen ist, stellten die Effektstärke der Initialstudie und die Anzahl inkludierter Primärstudien in einer Metaanalyse erneut signifikante Prädiktoren für die Crude differences dar. Beide Prädiktoren wiesen dabei einen positiven Zusammenhang mit der Zielvariable auf, ein Anstieg dieser

Tabelle 20

Ergebnisse der einfachen linearen Meta-Metaregression mit Crude differences als Kriterium

Prädiktor	<i>b</i> (SE)	<i>p</i>	<i>R</i> ²
ES der Initialstudie	0.856 (0.073)	< .001	79.52%
<i>N</i> der Initialstudie	< -0.001 (< 0.001)	.30	0.29%
Publikationsjahr der Initialstudie	-0.002 (0.002)	.28	0.60%
Jährliche Zitationszahl der Initialstudie	-0.002 (0.002)	.32	0.14%
Impact Factor 2020 Initialstudie	0.018 (0.025)	.48	0.00%
Summary effect	-0.290 (0.361)	.43	0.00%
<i>k</i>	0.001 (< 0.001)	.03	11.26%
Publikationsjahr der Metaanalyse	-0.005 (0.005)	.27	0.75%
Graue Literatur inkludiert	0.182 (0.091)	.053	7.94%
Abgedeckte Zeitspanne	0.002 (0.003)	.47	0.00%

Tabelle 21

Ergebnisse der einfachen lineare Meta-Metaregression mit Crude differences als Zielvariable, nur anhand publizierten Daten berechnet

Prädiktor	<i>b</i> (SE)	<i>p</i>	<i>R</i> ²
ES der Initialstudie	0.493 (0.146)	.004	41.11%
<i>N</i> der Initialstudie	< 0.001 (< 0.001)	.44	0.00%
Publikationsjahr der Initialstudie	0.001 (0.002)	.59	0.00%
Jährliche Zitationszahl der Initialstudie	0.003 (0.004)	.38	0.00%
Impact Factor 2020 Initialstudie	0.032 (0.013)	.03	27.98%
Summary effect	-0.085 (0.263)	.75	0.00%
<i>k</i>	< 0.001 (< 0.001)	.15	7.85%
Publikationsjahr der Metaanalyse	0.004 (0.003)	.19	5.31%
Abgedeckte Zeitspanne	0.001 (0.003)	.70	0.00%

Prädiktoren ging mit einem Anstieg in der Stärke des decline effect einher, $b = 0.86$, $p < .001$; $b < .01$, $p = .03$.

In Tabelle 21 sind die Ergebnisse der Analyse mit ausschließlich publizierten Daten abgebildet. Hierbei erwies sich die Effektstärke der Initialstudie abermals als signifikanter, positiver Prädiktor für die Stärke des decline effect, $b = 0.49$, $p = .004$. Zusätzlich wies der Impact Factor einen signifikanten positiven Zusammenhang mit den Crude differences auf, $b = 0.03$, $p = .03$. Alle anderen Zusammenhänge fielen nicht signifikant aus.

Jährliche Effektveränderung. Wurde die jährliche Effektveränderung als Zielvariable in der Analyse angewandt, so zeigte die Initialstudieneffektstärke und das Publikationsjahr der Initialstudie einen signifikanten Einfluss, $b = -0.02$, $p = .04$; $b < -0.01$, $p = .05$. Alle anderen Zusammenhänge fielen nicht signifikant aus (siehe Tabelle 22). Wurden nur publizierte Daten untersucht, zeigte sich durch keinen der Prädiktoren ein signifikanter Einfluss auf die Zielvariable (siehe Tabelle 23).

Tabelle 22

Ergebnisse der einfachen linearen Meta-Metaregression mit Regressionskoeffizienten als Kriterium

Prädiktor	$b(SE)$	p	R^2
ES der Initialstudie	-0.023 (0.011)	.04	9.80%
N der Initialstudie	< 0.001 (< 0.001)	.16	3.26%
Publikationsjahr der Initialstudie	< -0.001 (< 0.001)	.05	8.83%
Jährliche Zitationszahl der Initialstudie	< 0.001 (< 0.001)	.43	0.00%
Impact Factor 2020 Initialstudie	< - 0.001 (0.002)	.91	0.00%
Summary effect	0.006 (0.025)	.81	0.00%
k	< 0.001 (< 0.001)	.17	2.86%
Publikationsjahr der Metaanalyse	< -0.001 (< 0.001)	.58	0.00%
Graue Literatur inkludiert	-0.005 (0.007)	.50	0.00%
Abgedeckte Zeitspanne	< 0.001 (< 0.001)	.06	7.77%

Tabelle 23

Ergebnisse der einfachen linearen Meta-Metaregression mit Regressionskoeffizienten als Kriterium nur anhand publizierter Daten berechnet

Prädiktor	$b(SE)$	p	R^2
ES der Initialstudie	-0.034 (0.023)	.14	3.71%
N der Initialstudie	< 0.001 (< 0.001)	.28	0.60%
Publikationsjahr der Initialstudie	< -0.001 (< 0.001)	.63	0.00%
Jährliche Zitationszahl der Initialstudie	< 0.001 (< 0.001)	.67	0.00%
Impact Factor 2020 Initialstudie	0.001 (0.004)	.77	0.00%
Summary effect	0.051 (0.050)	.32	0.11%
k	< 0.001 (< 0.001)	.40	0.00%
Publikationsjahr der Metaanalyse	0.001 (0.001)	.18	2.49%
Abgedeckte Zeitspanne	< 0.001 (< 0.001)	.24	1.39%

Tabelle 24

Ergebnisse der einfachen mixed-effects Meta-Metaregression mit τ^2 als Kriterium

Prädiktoren	b	p	95% KI		R^2
			UG	OG	
Anzahl der Stichproben	< 0.001	.26	< - 0.001	0.010	1.91%
Total N	< 0.001	.56	< -0.001	< 0.001	0.00%
Durchschnittliches N einer Primärstudie	< -0.001	.49	< -0.001	< 0.001	0.00%

Anmerkung. $k = 34$. KI = Konfidenzintervall. UG = Untere Grenze. OG = Obere Grenze

Der Einfluss von k und N auf die Heterogenitätsmaße τ^2 und I^2

Die Analyse baute auf den Daten von $k = 34$ Metaanalysen auf und überprüfte den Einfluss der gesamten Stichprobengröße, der durchschnittlichen Stichprobengröße einer Primärstudie und der Anzahl inkludierter Stichproben auf die beiden Heterogenitätsmaße τ^2 und I^2 . Zur Berechnung wurde einerseits ein mixed-effects Modell herangezogen, andererseits wurde zusätzlich ein lineares Modell mit denselben Prädiktoren berechnet. Die Ergebnisse der einzelnen Analysen sind in den Tabellen 24 bis 27 abgebildet. Weder auf Seiten der einfachen mixed-effects Meta-Meta-Regressionen noch bei den Berechnungen mittels eines linearen

Tabelle 25*Ergebnisse der einfachen mixed-effects Meta-Metaregression mit I^2 als Kriterium*

Prädiktoren	b	p	95% KI		R^2
			UG	OG	
Anzahl der Stichproben	< 0.001	.48	-0.001	.001	0.00%
Total N	< 0.001	.30	< -0.001	< .001	0.00%
Durchschnittliches N einer Primärstudie	.001	.19	< -0.001	.001	3.31%

Anmerkung. $k = 34$. KI = Konfidenzintervall. UG = Untere Grenze. OG = Obere Grenze**Tabelle 26***Ergebnisse der einfachen linearen Meta-Metaregression mit τ^2 als Kriterium*

Prädiktor	$b(SE)$	p	R^2
Anzahl der Stichproben	< -0.001 (< 0.001)	.98	0.00%
Total N	< -0.001 (< 0.001)	.85	0.00%
Durchschnittliches N einer Primärstudie	< -0.001 (< 0.001)	.30	0.35%

Anmerkung. $k = 34$.

Modells, konnten signifikante Zusammenhänge zwischen den genannten Prädiktoren und den beiden Heterogenitätsmaßen gefunden werden.

Publikationsbias

Im letzten geplanten Analyseschritt dieser Arbeit wurde untersucht, wie häufig es zu Anzeichen von Publikationsbias in der Literatur der einzelnen Metaanalysen kam und welchen Einfluss die Inklusion bzw. Exklusion von grauer Literatur auf das Ergebnis dieser Methoden hatte. Dafür wurden auf Meta-Ebene an insgesamt $k = 34$ Metaanalysen neun unterschiedliche Methoden zur Erfassung diverser Bias-Muster angewandt (eine genaue Aufschlüsselung der einzelnen Verfahren und ihre Schwellenwerte sind aus Tabelle 1 zu entnehmen). Auf Meta-Meta-Ebene wurden die Ergebnisse anhand wissenschaftlich anerkannter Schwellenwerten interpretiert.

Die Ergebnisse zeigten, dass lediglich $k = 5$ Metaanalysen (15%) keinerlei Bias-Muster in den Daten aufwiesen. Im Mittel wiesen $Md = 3$ unterschiedliche Verfahren auf

Tabelle 27*Ergebnisse der einfachen linearen Meta- Meta-Regression mit I^2 als Kriterium*

Prädiktor	$b(SE)$	p	R^2
Anzahl der Stichproben	$< 0.001 (< 0.001)$	.44	0.00%
Total N	$< 0.001 (< 0.001)$	.27	0.72%
Durchschnittliches N einer Primärstudie	$0.001 (< 0.001)$	.14	3.67%

Anmerkung. $k = 34$.

Verzerrungen in der Literatur hin, wobei in einer Metaanalyse sieben unterschiedliche Verfahren den notwendigen Schwellenwert überschritten hatten. Die Häufigkeit für ein Bias-positives Ergebnis, getrennt für jede einzelne Methode, ist aus Tabelle 28 zu entnehmen.

Unterteilte man die einzelnen Verfahren und Schwellenwerte nun nach methodischer Ähnlichkeit in Gruppen, wie in Cooper et al. (2019) und Siegel et al. (2021) vorgeschlagen, zeigte sich, dass in $k = 12$ Studien (35%) zumindest eine Methode aus jeder dieser Gruppen auf Verzerrungsmuster in der Literatur hinwies.

Der Einfluss von Grey-Literature Bias

Um zu überprüfen, ob es durch die Exklusion von grauer Literatur aus einer Metaanalyse zu Unterschieden in der Häufigkeit von erfassten Bias-Mustern kam, wurden zum einen, wie in der Präregistrierung vermerkt, χ^2 -Tests durchgeführt. Zum anderen wurde der exakte Fisher Test angewandt. Dies war zwar nicht präregistriert worden, stellte sich aber nach Überprüfung der Voraussetzungen als notwendig heraus, da nur für p -Uniform*, PET-PEESE und dem Selection Model Approach von Vevea und Woods (2005) die, für den χ^2 -Test notwendige, Zelloberhäufigkeit ≥ 5 gegeben war. Die Ergebnisse der beiden Analysen sind in Tabelle 29 und 30 abgebildet.

Entgegen der getroffenen Annahme, dass die Inklusion grauer Literatur zur Reduktion von Publikationsbias führt, zeigten sich keine signifikanten Unterschiede zwischen den beiden untersuchten Gruppen. Metaanalysen, die graue Literatur inkludiert haben, zeigten also nicht, wie erwartet, weniger Hinweise für Publikationsbias Muster als jene, die nur auf publizierten Studien aufbauten.

Explorative Datenanalyse

Anlässlich des hohen Anteils an Metaanalysen, die Hinweise auf Publikationsbias boten, wurde in einem letzten explorativen Analyseschritt, der Zusammenhang zwischen der

Tabelle 28*Häufigkeit von Publikationsbias aufgeteilt nach Methode*

	Häufigkeiten
Überprüfung auf Asymmetrie	
TF missing	10
TF Schätzerdifferenz	6
TF-Schätzer hat anderes Vorzeichen	1
Rank Correlation Test	8
Sterne & Egger's Regression	9
PET-PEESE	16
Überprüfung der p -Werte	
Excess Significance Test	5
p -Curve	0
p -Uniform	2
p -Uniform*	13
Selection Model approach	18

Anmerkung. $k = 34$. TF = Trim and Fill. Jede Metaanalyse, die eine Reanalyse erlaubt hat, wurde anhand 9 unterschiedlicher Methoden auf Publikationsbias überprüft. Die zusammengefassten Ergebnisse dieser Analyse sind hier abgebildet.

Zahl an Publikationsbias-Mustern in einer Metaanalyse und den Maßen für Effektveränderung, mittels einer linearen Meta-Metaregression, überprüft. Es wurde dabei der Forschungsfrage nachgegangen, ob Metaanalysen, die mehr Bias-Muster aufweisen, auch stärker vom decline effect betroffen sind. Diese explorative Datenanalyse wurde nicht präregistriert, da diese Forschungsfrage erst durch die vorliegende Datenlage entstanden ist.

Für diese Analyse musste in einem ersten Schritt die Prädiktorvariable berechnet werden. Hierzu wurden die einzelnen Methoden zuerst, nach methodischer Ähnlichkeit, in drei unterschiedliche Gruppen geteilt (Gruppenunterteilung ist in Tabelle 29 abgebildet). Im

Tabelle 29

Ergebnisse des Fisher exakt Test zum Unterschied in der Häufigkeit von Publikations-Bias zwischen Metaanalysen mit und ohne grauer Literatur

	<i>p</i>
Trim & Fill (TF)	
Mindestens eine Studie fehlt	.08
Differenz zwischen TF-Schätzer und Summary effect > 20%	.14
TF-Schätzer hat anderes Vorzeichen	.64
Rank Correlation Test	
$p < .10$	.56
Sterne & Egger's Regression	
$p < .10$	.37
Excess Significance Test	
$p < .10$ und mehr beobachtete als erwartete signifikante Effekte	.80
<i>p</i> -Curve	
$p < .10$	.63
<i>p</i> -Uniform	
$p < .10$	.57

Anmerkung. $k = 34$. Da die Voraussetzungen für die Durchführung eines χ^2 -Tests für eine Vielzahl der geplanten Analysen nicht erfüllt war, wurde sich für den Fisher exakt Test als Alternative entschieden

nächsten Schritt wurde für jede der Gruppen eine Dummy-Variable erstellt, mit den Werten 0 = „keine der Methoden zeigt Hinweise für Bias“ und 1 = „mindestens eine Methode aus der Gruppe erfasst ein Bias-Muster“. Im Anschluss daran wurden die erstellten Dummy-Variablen aufsummiert und für den Anteil an Bias-positiven Methoden, anhand folgender Formel, korrigiert:

$$\text{Korrigierte Summe der Bias-Muster} = a + \frac{\sum_{i=1}^k x_i}{6} + b + \frac{\sum_{i=1}^k y_i}{4} + c$$

Die Variablen a, b und c stellen dabei die Dummy-Variablen der drei formulierten Gruppen dar, währenddessen x_i die Summe aller Methoden darstellt, die Asymmetrie in den Daten erfasst haben und y_i die Summe aller Methoden darstellt, die Bias-Muster im p -Wert erfasst haben.

Tabelle 30

Ergebnisse des χ^2 -Test zum Unterschied in der Häufigkeit von Publikations-Bias zwischen Metaanalysen mit und ohne grauer Literatur

	$\chi^2(df)$	p
Selection Model approach		
Differenz zwischen berechneten Schätzer und Summary effect > 20%	0.05(1)	.82
PET-PEESE		
Differenz zwischen PET-Schätzer oder PEESE-Schätzer und Summary effect > 20%.	0.02(1)	.89
p -Uniform ^{*a}		
Differenz zwischen p -Uniform*-Schätzer und Summary effect > 20%	0.60(1)	.44

Anmerkung. Der Selection Model Approach, PET-PEESE und p Uniform* erfüllten die Voraussetzung von einer Zellohligkeit von mindestens fünf, wodurch ein χ^2 -Test durchgeführt werden konnte.

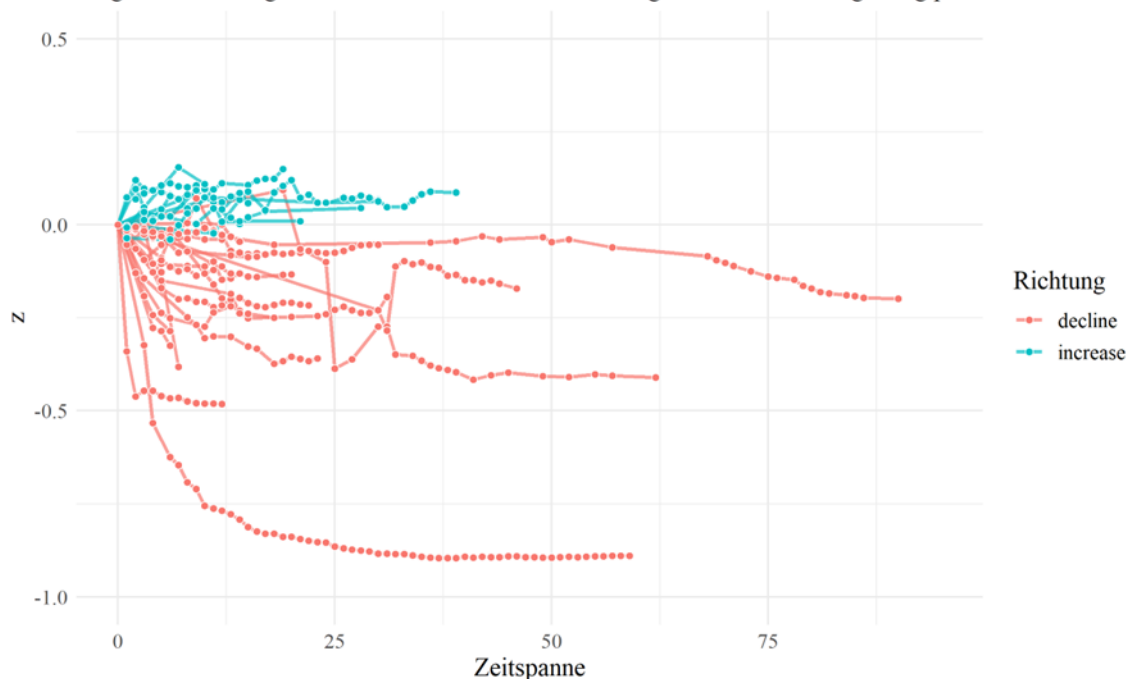
Nutzte man diese berechnete Variable als Prädiktor in einer Regressionsanalyse mit $k = 34$ Metaanalysen, so zeigte sich ein marginal signifikanter positiver Zusammenhang zu den Crude differences, $b = 0.07$, $p = .06$, $R^2 = 8.04\%$. Im linearen Meta-Meta-Regressionsmodell mit der jährlichen Effektveränderung als Zielvariable ergab sich ein negativer und ebenfalls nicht signifikanter Zusammenhang, $b < -0.01$, $p = .24$, $R^2 = 1.21\%$.

Diskussion

Die hier vorliegende meta-metaanalytische Studie stellt die erste Arbeit dar, die sich mit der systematischen Effektveränderung im Bereich der sozialpsychologischen und Persönlichkeitsforschung befasst hat. Die Ergebnisse implizieren, dass die Metaforschung des JPSP im Allgemeinen einen decline effect aufzeigt. Darüber hinaus stellte sich auf Seiten der Crude differences heraus, dass decline effects im Vergleich zu Effektsteigerungen doppelt so häufig auftreten und stärker ausfallen. Zusätzlich dazu wiesen auch mehr als die Hälfte aller Metaanalysen eine signifikante jährliche Effektreduktion auf. Verdeutlicht wird dies ebenfalls durch den Lightning plot in Abbildung 2. Wie in dieser Arbeit hypothetisiert, lässt sich dementsprechend aus den Ergebnissen ableiten, dass der größte Teil der

Abbildung 2

Darstellung der Verteilung von decline effects und Effektanstiegen anhand eines Lightning plots



Anmerkung. Der Lightning plot veranschaulicht die Effektveränderungen der einzelnen inkludierten Metaanalysen über die Zeit. Die einzelnen Punkte stellen dabei die Differenz der Effektstärke, zwischen der abgebildeten Primärstudie und der davor dargestellten Primärstudie, dar.

sozialpsychologischen und Persönlichkeitsforschung von einem decline effect betroffen ist und die fundamentalen Effekte, auf welchen sowohl die Forschung als auch die Praxis aufbauen, nicht in dem Ausmaß vorzuliegen scheinen, wie bisher angenommen wurde. Interessant ist hierbei auch, dass sich die Ergebnisse zwar mit den Erkenntnissen bisheriger Forschungen decken, der decline effect aber nicht, wie aus der Literatur abzuleiten ist, im Bereich der sozialpsychologischen Forschung stärker ausfällt (Gong & Jiao, 2019; Fanelli et al., 2017; Franco et al. 2014; Franco et al., 2016; Nuijten et al., 2020; Pietschnig et al., 2019). So zeigte sich beispielsweise, dass die anhand der Crude differences berechneten decline effects aus Metaanalysen des JPSP, im Vergleich zu denen aus der Intelligenzforschung, stärker ausfallen, dafür aber die jährliche Effektreduktion im Vergleich dazu, schwächer und weniger häufig auftritt (Pietschnig et al., 2019).

Die Ergebnisse der vorliegenden Analysen deuten darauf hin, dass der decline effect im Forschungsfeld der Sozialpsychologie und Persönlichkeit vor allem durch Effektüberschätzung und Grey Literature Bias zustande kommt. So zeigte sich unter anderem, dass die Forschung in diesem Bereich, mit einem durchschnittlichen Wert von etwa 46%, eine zu niedrige Power aufweist, um die tatsächlich dahinterliegenden Effekte zu erfassen. Dieses Ergebnis ist zwar weit unterhalb des notwendigen Schwellenwertes von 80%, stellt man

dieses jedoch mit anderen wissenschaftlichen Bereichen in Relation, wird deutlich, dass eine zu niedrige Power eher die Norm, als die Ausnahme ist (Button et al., 2013; Pietschnig et al., 2019). Für die Intelligenzforschung wurde beispielsweise eine durchschnittliche Power von 50% berichtet und im Bereich der Neurowissenschaften lag sie sogar bei 20% (Button et al., 2013; Pietschnig et al., 2019). Wie bereits in einigen Studien postuliert wird, erweist sich eine zu niedrige Power einer Primärstudie als besonders problematisch für ein Forschungsfeld. Zum einen sinkt damit die Wahrscheinlichkeit den „wahren“ Effekt zu erfassen, berichtete Effekte stellen dementsprechend häufig gar keine echten Effekte dar und falls ein Studie tatsächlich einen zugrundeliegenden Effekt erfasst haben sollte, so wird dieser durch die mangelnde Power überschätzt (Button et al., 2013; Ioannidis, 2005). Als Folge daraus wird ein decline effect angenommen (Protzko & Schooler, 2017). Dies wurde ebenfalls durch die vorliegenden Ergebnisse deutlich, da sich zeigen ließ, dass neben der allgemein niedrigen Power einer Primärstudie, diese zusätzlich nochmals kleiner ausfiel, wenn man die durchschnittlicher Power nur für Metaanalysen, die eine jährliche Effektreduktion aufwiesen, berechnete.

Auch die Ergebnisse der linearen Meta-Metaregression wiesen darauf hin, dass die Effektüberschätzung von Primärstudien eine große Rolle für den decline effect im JPSP zu spielen scheint. So zeigte sich sowohl bei den multiplen hierarchischen Meta-Metaregressionsmodellen, als auch bei den einfachen Meta-Metaregressionen, dass die Effektstärke der Initialstudie eine signifikante moderierende Rolle für den decline effect zu spielen scheint. Je größer der berichtete Effekt einer Initialstudie ausfiel, desto größer war auch die letztliche Effektreduktion. Dieser moderierende Einfluss zeigte sich sowohl bei den Crude differences als auch bei der jährlichen Effektveränderung und lässt somit darauf schließen, dass vor allem die großen berichteten Effekte von Initialstudien mit hoher Wahrscheinlichkeit nur kleine oder sogar gar keine zugrundeliegenden Effekte erfasst haben. Diese Erkenntnisse stehen hiermit in Einklang mit jenen aus der Studie von Pietschnig et al. (2019) und unterstützen die eben postulierten Annahmen, die aus der Interpretation der Ergebnisse der Poweranalyse geschlossen werden konnten.

Eine weitere interessante Erkenntnis der linearen Meta-Metaregression war, dass unter der Bedingung ausschließlich publizierter Daten innerhalb einer Metaanalyse, der Impact Factor des Journals, in welchem die Initialstudie publiziert wurde, ebenfalls einen positiven Einfluss auf die Stärke des decline effect nahm. Die ursprüngliche Annahme war, dass ein höherer Impact Factor des publizierenden Journals im Allgemeinen mit dem decline effect in Zusammenhang steht und nicht nur ausschließlich, wenn lediglich publizierte Primärstudien

in einer Metaanalyse inkludiert sind. Dies ist insofern interessant, da bereits in einigen Arbeiten der momentane Publikationsprozess kritisiert wurde (de Bruine & Della Sala, 2015; Margraf, 2015; Protzko & Schooler, 2017). Es wird häufig darüber berichtet, dass eine zu große Fokuslegung auf signifikante und große Effekte gelegt wird, Replikationsstudien und nicht signifikante Ergebnisse kleinerer Studien gelten jedoch als uninteressant und werden nicht oder nur viel später publiziert (Ioannidis, 1998; Franco et al., 2014; Franco et al., 2016; Margraf, 2015). Vor allem in prestigeträchtigeren Journals, wie beispielsweise dem JPSP, zeigten sich dabei bereits seit einigen Jahren problematische Journal Richtlinien, wie unter anderem die Verweigerung, Replikationsstudien zu publizieren (Aldhous, 2011; Roberts, 2011). Dementsprechend war die Annahme naheliegend, dass es, für die Publikation der eigenen Forschungsarbeit in einem Journal mit hohem Impact Factor, vor allem wichtig ist, große, signifikante und dahingehend interessanten Effekte zu berichten, unabhängig davon, wie robust diese Effekte letztendlich sind. In einer meta-metaanalytischen Untersuchung des Journals *Intelligence* zeigte sich bereits ein positiver Zusammenhang zwischen dem Impact Factor und dem decline effect (Pietschnig et al., 2019). Im JPSP konnte dieser jedoch nur unter der Bedingung gefunden werden, dass ausschließlich publizierte Studien in einer Metaanalyse inkludiert waren. Dies lässt darauf schließen, dass der Impact Factor, zumindest bei dem hier vorliegenden Datensatz, im Allgemeinen keinen Einfluss auf den decline effect zu haben scheint und überschätzte Effekte sowohl in prestigeträchtigeren als auch in weniger prestigeträchtigen Journals publiziert werden. Es scheint jedoch für die Publikationswahrscheinlichkeit von nicht erfolgreichen nachfolgenden Studien eine Rolle zu spielen, wie bedeutend das Journal ist, indem die jeweilige Initialstudie publiziert wurde. Der präsentierte Zusammenhang mit den errechneten Crude differences lässt nämlich darauf schließen, dass mit ansteigendem Impact Factor des Journals, der Unterschied zwischen dem Summary effect der Metaanalyse und dem Effekt der Initialstudie größer wird, was folglich auf einen decline effect hinweist. Der Zusammenhang mit der jährlichen Effektveränderung fiel jedoch nicht signifikant aus, wodurch davon auszugehen ist, dass in diesem Fall die Initialstudie ein sogenanntes early extreme darstellt und die restlichen Forschungsergebnisse einheitlich kleinere Effekte zu berichten scheinen. Dies verdeutlicht erneut den eben genannten Standpunkt. Wenn eine Studie, wie die vorliegenden Erkenntnisse erwarten lassen, eine zu niedrige Power aufweist, einen daraufhin überschätzten Effekt berichtet und in einem einflussreichen Journal publiziert wird, so weist der größte Teil der restlichen publizierten Forschung keine unterstützenden Ergebnisse auf. Da dieser Zusammenhang nur in dem Fall vorliegt, wenn nur publizierte Studien untersucht wurden, lässt sich daraus weiter ableiten,

dass es scheinbar einfacher ist, widerlegende Ergebnisse zu publizieren, wenn die widerlegte Arbeit aus einem Journal mit einem hohen Impact Factor stammt. Dies könnte daher kommen, weil von diversen Editor*innen und Reviewer*innen angenommen wird, dass es mehr Aufsehen erregen könnte, wenn die Erkenntnisse einer einflussreichen und bedeutenden Arbeit widerlegt werden (de Bruine & Della Sala, 2015). Dies kann als Hinweis für das Vorliegen eines Publikationsbias interpretiert werden.

Neben dieser Schlussfolgerung aus den Daten, gibt es jedoch noch zahlreiche weiterer Beweise für Anzeichen von Publikationsbias in der sozialpsychologischen und Persönlichkeitsforschung. Denn in nur 5 der 34 reanalysierten Metaanalysen konnten, anhand der neun angewandten Methoden zur Erfassung von Publikationsbias, keine Anzeichen für diesen erfasst werden. Bei jenen Metaanalysen, die von Publikationsbias betroffen waren, wurden im Durchschnitt drei der, in Forschungen postulierten Schwellenwerte überschritten und es zeigten sowohl das Selection Model, als auch die diversen Methoden zur Überprüfung der p -Werte und der Asymmetrie in den Daten Hinweise für Publikationsbias (Siegel et al., 2021). Dies divergiert stark von den berichteten Ergebnissen der einzelnen Metaanalysen. Im Kodierungsprozess ergab sich nämlich, dass nur in vier Fällen von Hinweisen auf Publikationsbias berichtet wurde. Ein möglicher Erklärungsansatz für diese Divergenz der Ergebnisse könnte in der Verwendung veralteter bzw. ungenauer Methoden zur Erfassung von Publikationsbias liegen. Daraus resultiert das Problem, dass Erkenntnisse zum Publikationsbias im Forschungsfeld nicht repräsentativ sind und der Erkenntnisstand durch Primärstudien als weniger verfälscht darstellt, als er tatsächlich ist. Wie in Protzko und Schooler (2017) postuliert, kann dies als Erklärung für den decline effect in dieser Studie herangezogen werden, da durch den Publikationsbias, wie bereits oben beschrieben, vor allem signifikante und überschätzt Effekte bevorzug werden.

Eine weitere wesentliche Erkenntnis aus den Analysen war hier zusätzlich, dass es keinen Unterschied in der Erfassung von Publikationsbias machte, ob in einer Metaanalyse nur publizierte Studien oder auch graue Literatur inkludiert war. Keine der angewandten Methoden zeigte jedoch einen solchen Unterschied und Publikationsbias kam dementsprechend in beiden Gruppen im selben Ausmaß vor. Dieses Ergebnis verdeutlicht ein großes Problem, das mit grauer Literatur einhergeht. Aufgrund dessen, dass diese Arbeiten nicht publiziert sind, ist es nicht möglich die gesamte graue Literatur zu einem Forschungsthema zu erfassen, weshalb die eigene Analyse nur auf einem kleinen Anteil der tatsächliche vorliegenden Erkenntnisse aufbaut. Manche Arbeiten, wie beispielsweise Dissertationen, sind in spezifischen Datenbanken abgespeichert und können von dort

abgerufen werden, in vielen weiteren Fällen ist es jedoch notwendig, Autor*innen persönlich zu kontaktieren und nach möglichen „File-Drawer“-Studien zu fragen (Aloia & Naughton, 2017). Die Kontaktaufnahme mit Forscher*innen kann sich als sehr zeitintensiv herausstellen und ist dementsprechend vor allem für kleinere Forschungsteams eventuell weniger ansprechend. In vielen Fällen werden daher häufig nur eigene File-Drawer-Studien in die Metaanalyse inkludiert, was jedoch nur einen verfälschten Anteil der tatsächlichen Datenlage darstellen könnte, da persönliche Motivationen durch die Forscher*innen hier nicht ausgeglichen werden können (Franco et al., 2016). Ebenso könnte die Problematik bestehen, dass auch graue Literatur methodische Schwächen aufweisen kann, Effekte überschätzt und von zu wenig Power oder zu kleine Stichproben geprägt sein könnte (Joobert et al., 2012). Dementsprechend sollte man bei der Berechnung einer Metaanalyse weniger darauf achten, dass man graue Literatur in seiner Arbeit inkludiert, sondern es sollte der Fokus viel mehr darauf gelegt werden, dass möglichst viel graue Literatur von unterschiedlichen Autor*innen und Forschungsteams inkludiert ist, um so den Publikationsbias auszugleichen.

Wie in der Arbeit von Protzko und Schooler (2017) die sogenannte *Regression zur Mitte* postuliert wird, hat sich auch in der hier vorliegenden Studie gezeigt, dass die Anzahl inkludierter Effekte in eine Metaanalyse eine Rolle für die Erfassung des decline effect zu spielen scheint. Die Regression zur Mitte beschreibt dabei das Phänomen, dass man sich durch wiederholtes Testen eines Effektes, dem dahinterliegenden wahren Wert immer weiter annähert (Barnett et al., 2004). Am konkreten Beispiel der vorliegenden Ergebnisse erklärt, zeigte sich, dass durch die Inklusion von mehr Effekten in eine Metaanalyse, der daraus entstandene Summary effect, einen genaueren Schätzwert des wahren Wertes darstellte und dadurch mit jedem inkludierten Effekt kleiner ausfiel. Die Differenz zwischen dem Summary effect und dem Effekt der Initialstudie deutete dadurch auf einen immer stärker werdenden decline effect hin. Interessanterweise zeigte sich dieser Zusammenhang zwischen der Anzahl inkludierter Effekte in einer Metaanalyse und dem decline effect jedoch nicht, wenn nur publizierte Primärstudien in den Analysen inkludiert waren. Dieser Unterschied könnte wiederum als Hinweis für den Einfluss von Grey Literature Bias auf den decline effect interpretiert werden. Denn währenddessen bei Metaanalysen, mit publizierten und nicht publizierten Daten, der decline effect mit ansteigender Zahl an Effekten innerhalb der Analysen ebenfalls anstieg, blieb er bei Metaanalysen, die keine graue Literatur inkludierten, unbeeinflusst davon. Wie in der Literatur häufig angenommen, könnte dieses Ergebnis dahingehend interpretiert werden, dass erst durch die Inklusion von grauer Literatur, die im Gegensatz zu publizierten Arbeiten weniger große und nicht signifikante Effekte berichtet, die

eigentliche Datenlage besser dargestellt wird, wodurch sich letztlich ein decline effect abzeichnet (Protzko & Schooler, 2017). Zusätzlich untermauert dies die im letzten Punkt genannte Schlussfolgerung, dass mit steigender Anzahl an inkludierter grauer Literatur, der decline effect stärker wird.

Ein Faktor, der entgegen der formulierten Hypothese keinen Einfluss auf den decline effect zeigte, war die gewählte Effektstärkenmetrik zur metaanalytischen Verrechnung. Zwar legten die Ergebnisse mehrerer Simulationsstudien die Annahme nahe, dass sowohl durch Cohen's d und Hedges g , als auch durch Pearson r im speziellen, Effektüberschätzungen entstehen können, wenn Stichproben zu klein sind oder die Daten nicht normalverteilt sind, doch zeigte sich zwischen den beiden Effektmaßfamilien in den hier vorliegenden Analysen kein Unterschied bezüglich der Größe des decline effect (Bishara & Hittner, 2015; Magnusson, 2014; Lin, 2018; Rousselet, 2018). Daraus lässt sich schließen, dass die beiden Effektmaßfamilien die Stärke des decline effect entweder im selben Ausmaß beeinflussen oder die Effektstärkenmetrik tatsächlich doch zu keiner Effektüberschätzung führt.

Neben der Erfassung und Begründung des decline effect in der Meta-Forschung des JPSP, wurde des Weiteren der Fragestellung nachgegangen, wie stark die Heterogenitätsmaße τ^2 und I^2 durch die Stichprobengröße und Anzahl inkludierter Stichproben beeinflusst wurden. Da die Heterogenität der Daten ein wichtiges Maß für das methodisch korrekte Vorgehen in einer Metaanalyse darstellt und zudem für die Interpretation von Publikationsbias-Methoden herangezogen werden sollte, war es wesentlich festzustellen, ob diese Maße durch diverse Moderatoren beeinflusst werden können (Bornstein et al., 2009; Siegel et al., 2021). Bisherige Erkenntnisse ließen darauf schließen, dass sowohl I^2 als auch τ^2 durch die Stichprobengröße beeinflusst werden könnte (Inhout et al., 2015; Rücker et al., 2008). Diese Annahme ließ sich anhand der vorliegenden Ergebnisse jedoch nicht unterstützen. Weder die Anzahl inkludierter Stichproben, noch die gesamte oder durchschnittliche Stichprobengröße zeigten signifikante Zusammenhänge mit den beiden Heterogenitätsmaßen. Anders als in einigen Arbeiten postuliert, stellen dementsprechend beide Effektmaße einen robusten Messwert zur Erfassung der Heterogenität dar (Bornstein et al., 2009; Inhout et al., 2015; Rücker et al., 2008).

Was kann gegen diese Entwicklung getan werden?

Wie die Ergebnisse dieser und einiger weiterer Studien zeigen, ist die empirische Forschung von Effektüberschätzungen, zu niedriger Power und Publikationsbias geprägt (Fanelli et al., 2017; Gong & Jiao, 2019; Nuijten et al., 2020; Pietschnig et al., 2019). Zusätzlich wird immer wieder davon berichtet, dass Forscher*innen häufig zu fragwürdigen

Forschungspraktiken wie selektivem Berichten von Ergebnissen oder auch *p*-hacking oder HARKing greifen, um bessere Chancen für eine Publikation zu bekommen (Fiedler & Schwarz, 2016; Franco et al., 2016; John et al., 2012; Margraf, 2015). Ausgelöst wurde dies durch die jahrelange Fokuslegung von Journals, hauptsächlich große und signifikante Effekte als interessant und publikationswürdig zu betrachten (Ioannidis, 1998; Franco et al., 2014; Margraf, 2015). Die empirische Forschung befindet sich dementsprechend gerade in einer schwierigen Ausgangslage, die, wie es in vielen Arbeiten berichtet wird, einiges an Veränderung verlangt (Bolibaugh et al., 2021; Button et al., 2013; Jooper et al., 2012; Protzko & Schooler, 2017).

Ein wichtiger Schritt wäre zum einen, den Fokus nicht mehr auf einzelne, möglicherweise überschätzte Effekte von Primärstudien zu legen, sondern viel mehr auf die Ergebnisse groß angelegter Replikationsstudien und Metaanalysen zu achten (Schooler, 2014). Zum anderen wäre es in diesem Zusammenhang auch wichtig, dass es nicht, wie in Studien berichtet, zu Unterschieden in der Publikationshäufigkeit kommt, abhängig davon, ob Ergebnisse signifikant sind oder nicht (Franco et al., 2014; Ioannidis, 1998). Dies würde in weiterer Folge bedeuten, dass große Journals, wie beispielsweise das JPSP, ihre Einstellung zu Replikationen ändern müssten und die Publikation von Replikationsstudien ermöglichen, auch wenn diese keine signifikanten Effekte berichten (Aldhous, 2011; Lakens, 2018; Roberts, 2011).

Um diesen aktuellen Mängeln im wissenschaftlichen Forschungsfeld entgegenzuwirken, wäre es ebenfalls ein wichtiger Schritt, die Durchführung einer Präregistrierung oder eine andere Art der Studienprotokollierung zum Status-Quo des wissenschaftlichen Vorgehens zu machen, um so einerseits die *scientific degrees of freedom* einzuschränken und andererseits auch die mögliche Suche nach grauer Literatur zu erleichtern (Protzko & Schooler, 2017; Jooper et al., 2012; Bolibaugh et al., 2021). Wie aus den Ergebnissen der hier vorliegenden Arbeit geschlossen werden konnte, scheint zu wenig graue Literatur in Metaanalysen inkludiert worden zu sein, um dem vorliegenden Publikationsbias entgegenzuwirken. Ein möglicher Erklärungsgrund dafür könnte sein, dass graue Literatur für Autor*innen teilweise nur schwer erfasst werden kann, da sie in vielen Fällen nicht dokumentiert ist und somit nur durch direkte Kontaktaufnahme mit den jeweiligen Forscher*innen erfragt werden kann (Aloia & Naughton, 2017; Jooper et al., 2012). Würden Forscher*innen nun ihre Arbeiten in der Planungsphase bereits präregistrieren, so würden diese in einer Onlinedatenbank aufliegen und für andere somit auffindbar sein (Jooper et al., 2012; Schooler, 2011). Würde die durchgeführte Studie dann in weiterer Folge nicht

publiziert werden, so hätte man beispielsweise auf der Webseite des Open Science Framework, zusätzlich zur Präregistrierung auch die Möglichkeit, seine erhobenen Daten dort hochzuladen und für andere Forscher zur Verfügung zu stellen (Bolibaugh et al., 2021). Zusätzlich dazu dient die Präregistrierung und Offenlegung der Daten als Präventionsmaßnahme gegen die Anwendung von diversen fragwürdigen Forschungspraktiken, da anhand dieser abgeglichen werden kann, ob die Autor*innen einer Arbeit tatsächlich methodisch so vorgegangen sind, wie sie es zu Beginn des Forschungsprozesses präregistriert haben (Bolibaugh et al., 2021; Button et al., 2013). Ergebnisse, wie die von Schäfer und Schwarz (2019), untermauern diese Annahme, nachdem sich in ihrer Studie zeigte, dass publizierte präregistrierte Studien weniger häufig signifikant ausfielen und im Schnitt kleinere Effekte berichteten, als publizierte Studien ohne Präregistrierung.

Wie in einigen Studien postuliert wird, ist auch die Anwendung metaanalytischer Verfahren ein hilfreiches Werkzeug gegen die falsche Darstellung von Effekten in einem Forschungsgebiet (Protzko & Schooler, 2017; Schooler, 2014). Wie sich in der hier vorliegenden Arbeit gezeigt hat, sollte man als Autor*in einer Metaanalyse zum einen darauf achten, dass die Analyse auf einer möglichst großen Anzahl an Effekten aufbaut, zum anderen sollte aber auch ein spezifischer Fokus darauf gelegt werden, dass der Anteil inkludierter grauer Literatur möglichst hoch ist. Dies ermöglicht es, auch bei Primärstudien die zu wenig Power aufweisen und dahingehend stark zu Effektverschätzung tendieren, ein möglichst präzises Abbild des wahren Wertes zu erlangen, da durch die Inklusion von grauer Literatur im Vergleich zu den publizierten Daten, auch das andere Spektrum der Effektverschätzung inkludiert wird (Schooler, 2011).

Abschließend soll hier noch auf zwei interessante Open-Science-Projekte eingegangen werden, die beide anhand unterschiedlicher Ansätze versuchen, die wissenschaftliche Fokuslegung auf Replikationen zu lenken und so der Fehldarstellung von Effekten in der publizierten Literatur entgegenzuwirken.

Zum einen soll hierbei das sogenannte Collaborative Replication Education Project (CREP) genannt werden. Dabei handelt es sich um ein Multi-Lab Replikationsprojekt mit dem Ziel, Studien zu replizieren, die häufig zitiert und in Journals mit hohem Impact Factor publiziert wurden. Die dahinterliegende Idee ist, dass die Replikationsstudien nicht von einzelnen Forschergruppen, sondern von Studierendengruppen unter Anleitung ihrer Lehrveranstaltungsleiter*innen, in Lehrveranstaltungen an Universitäten rund um den Globus durchgeführt werden. Neben dem Replizieren einflussreicher Studien, hat sich das CREP-

Team zusätzlich das Ziel gesetzt, Studierende in den Prozess des Replizierens zu inkludieren und den Student*innen somit zum einen Übung in der Durchführung richtiger Studien zu bieten und zum anderen auch die Wichtigkeit von Replikationsstudien deutlich zu machen. Studierende sollen unter Anleitung ihren Lehrveranstaltungsleiter*innen die geplante Replikation präregistrieren, eine Stichprobe generieren und letztlich die eigenen Ergebnisse statistisch auswerten und interpretieren. Zusätzlich werden die erfassten Daten an das CREP-Team weitergeleitet und nachdem das Projekt nicht national oder an eine gewisse Universität gebunden ist, wird so eine große und internationale Stichprobe generiert. Mittels einer Metaanalyse werden die Erkenntnisse aller teilnehmenden Studierendengruppen zum Abschluss des Projektes aggregiert und die tatsächliche Replizierbarkeit des überprüften Effektes verdeutlicht (Collaborative Replications and Education Project, 2013).

Bei zweitem handelt es sich um das sogenannte „Check me – Replicate me“-Projekt von Gilad Feldman (o.D.). Hierbei wird versucht, wie es auch Button et al. (2013) in ihrer Arbeit vorschlagen, einzelnen Forscher*innen Anreize für die Durchführung von Replikationsstudien zu bieten. Feldman (o. D.) versucht dies im speziellen so zu lösen, indem er zum einen seinen Forschungsprozess möglichst offen und für andere Forscher*innen zugänglich dokumentiert und zum anderen aktiv um die Replikation und Überprüfung seiner Ergebnisse bittet. Zusätzlich bietet er Forscher*innen finanzielle Entschädigungen von 5\$ bis 50\$ an, wenn diese Fehler in seinen Arbeiten feststellen konnten. Vor allem durch seinen speziellen Wunsch, repliziert und auf Fehler aufmerksam gemacht zu werden, stellt Feldman (o. D.) hier ein positives Beispiel für den Forschungs- und Replikationsprozess dar.

Spezifische Anmerkungen für das JPSP

Zusätzlich zu den eben genannten Änderungsvorschlägen für die Forschung im Allgemeinen, sollten auch die Probleme, die in der Meta-Forschung des JPSP vorliegen, in weiterer Folge erwähnt werden.

Wie nämlich in den vorliegenden Analysen festgestellt werden konnte, scheint es eine große Diskrepanz zwischen dem hier erfassten und dem in den einzelnen Metaanalysen berichteten Problem bezüglich dem Vorliegen von Publikationsbias zu geben. Denn während nur drei Metaanalysen davon berichteten, Hinweise für Publikationsbias erfasst zu haben, wiesen bei den hier berichteten Ergebnissen lediglich sechs Metaanalysen keinerlei Anzeichen für diesen auf. Zusätzlich dazu wurde nur in etwa 50% aller Studien überhaupt auf Publikationsbias geprüft, wobei die zwei am häufigsten verwendeten Methoden das Fail-Safe-N nach Rosenthal (1979) und die visuelle Überprüfung des Funnelplots darstellten. Dies ist

wiederum problematisch, da von der Anwendung dieser beiden Methoden inzwischen abgeraten wird, nachdem ersteres Verfahren methodisch und zweiteres hauptsächlich aufgrund der subjektiven Ergebnisinterpretation kritisiert wurde (Siegel et al., 2021). Zwar weisen auch andere Methoden zur Erfassung von Publikationsbias Schwächen auf, um diese auszugleichen wird in der Literatur dementsprechend Triangulation, wie es beispielsweise in Pietschnig et al. (2019) oder auch der hier vorliegenden Studie gehandhabt wurden, vorgeschlagen (Cooper et al., 2019)

Einen weiteren Kritikpunkt stellen die sogenannten *Internal Meta-analyses*. Diese stellten den mit Abstand häufigsten Grund für die Exklusion einer Studie, aus den hier vorliegenden Analysen, dar.

Häufig werden Internal Meta-analyses von Forscher*innen angewandt, um einem möglichen Powerproblem entgegenzuwirken und dadurch präzisere, robustere und genauere Effekte zu berichten (Carson et al., 2003; Chen et al., 2018; DelPriore et al., 2018). Jedoch wird häufig vernachlässigt, dass auch metaanalytische Verfahren von zu niedriger Power betroffen sein können, insbesondere wenn die Anzahl inkludierter Effekte innerhalb einer Metaanalyse gering ausfällt (Cafri et al., 2010). Darüber hinaus entscheiden sich Forscher*innen einer solchen Analyse des Öfteren dazu, aufgrund der methodischen Ähnlichkeit der eigenen Studien, ein fixed-effect Modell für die Berechnung heranzuziehen, ohne dabei die Heterogenität in den Daten zu beachten (DelPriore et al., 2018). Eine Fehleinschätzung des Verrechnungsmodells kann jedoch zu dem Problem führen, dass der metaanalytische Effektschätzer präziser ausfällt, als er es in Wirklichkeit wäre (Cafri et al., 2010; Schmidt et al., 2009). Darüber hinaus ist aber auch davon auszugehen, dass der errechnete Summary effect dieser speziellen Art der Metaanalyse verzerrt sein könnte, da die inkludierten Primärdaten ausschließlich von einem einzigen Forschungsteam stammen und somit auch persönliche Motivationen einen Einfluss haben könnten (Franco et al., 2016; Pietschnig et al., 2010). Zusammenfassend lässt sich also daraus schließen, dass Internal Meta-analyses, entgegen der Annahme der Autor*innen, ebenfalls zu unpräzisen und verzerrten Effekten führen können und somit keine Alternative zu Primärstudien darstellen.

Forschungsausblick

Eine Idee für zukünftige Forschungsarbeiten wäre nun, eine Meta-Metaanalyse durchzuführen, die sich nicht, wie die hier vorliegende Arbeit, auf ein spezifisches Journal fokussiert, sondern die Metaforschung im Gesamtbereich der Sozialpsychologie auf den

decline effect überprüft. Dies könnte dabei helfen, generalisierbarere Aussagen über den decline effect in diesem Forschungsfeld treffen zu können.

Ein weiterer Ansatz für die Untersuchung des decline effect wäre, den Fokus von Metaanalysen aus dem JPSP, auf Primärstudien, die in diesem Journal publiziert wurden, zu legen. In anderen Worten sollte als Inklusionskriterium für die Analysen, nicht das Publikationsjournal der Metaanalyse eine Rolle spielen, sondern ausschließlich jenes, in welchem die Initialstudie publiziert wurde. Dies würde es ermöglichen, die Annahme zu überprüfen, ob Journals mit einem hohen Impact Factor mit einem verstärkten Berichten überschätzter Effekte einhergehen.

Im Zuge der Durchführung einer Meta-Metaanalyse, wäre es ebenfalls interessant zu überprüfen, in welchem Zusammenhang der decline effect mit den Ergebnissen der einzelnen Publikationsbias Methoden steht. In einer explorativen Datenanalyse der vorliegenden Studie konnte nur ein geringer Zusammenhang zwischen dem decline effect und dem Anschlagen mehrerer Publikationsbias Methoden unterschiedlicher Untersuchungsfamilien festgestellt werden. Eine mögliche Erklärung für diese Ergebnisse könnte einerseits sein, dass die vorliegende Analyse zu wenig Power aufwies, um einen tatsächlichen Zusammenhang zu erfassen. Andererseits könnte es aber ebenfalls darauf zurückgeführt werden, dass es tatsächlich keine Rolle spielt, ob mehrere Methoden aus unterschiedlichen Untersuchungsfamilien anschlagen, sondern dass entweder allgemein Publikationsbias erfasst wird oder eine spezifische Art von Bias-Mustern vorliegt.

Zu guter Letzt wäre es für die Meta-Meta-Forschung von großem Nutzen, wenn hierfür spezifische Auswertungsmethoden entwickelt werden würden. Denn so wie es die hier vorliegende Arbeit verdeutlichen konnte, trifft man bei der Anwendung metaanalytischer Verfahren auf Meta-Meta-Ebene auf Komplikationen, wenn es um die Wahl des Präzisionsmaßes geht. Da diese auf Meta-Meta-Ebene nämlich eine Gewichtung nach der Anzahl inkludierter Studien in Abhängigkeit der jeweiligen Stichprobengröße verlangen würde, bieten metaanalytische Verfahren nur die Möglichkeit zur Gewichtung nach einem Maß. Die beste Wahl stellen dementsprechend zum momentanen Zeitpunkt Verfahren dar, die keinerlei Gewichtungen vornehmen.

Limitationen

Aufgrund der Fokuslegung dieser Arbeit auf Metaanalysen des JPSP ist eine Generalisierung der Ergebnisse auf das gesamte Forschungsfeld der Sozialpsychologie nur bedingt möglich. Solange es dementsprechend keine größer angelegte Meta-Metaanalyse zu

diesem Forschungsbereich gibt, welche die hier vorliegenden Ergebnisse noch weiter untermauern, sollten die vorliegenden Ergebnisse nur unter dem Vorbehalt interpretiert werden, dass es sich hierbei explizit um die Abbildung der publizierten Meta-Forschung im JPSP handelt. Es ist zum momentanen Zeitpunkt nicht abzuschätzen, welchen Einfluss die, in der Literatur kritisierten, Journal Richtlinien und Einstellungen des JPSP auf die publizierte Meta-Forschung in diesem Journal hat (Lakens, 2018; Roberts, 2011). Es könnte möglich sein, dass dadurch eine Über- oder Unterschätzung des decline effect im Forschungsfeld der Sozialpsychologie und Persönlichkeit dargestellt wird.

Zusätzlich muss kritisch angemerkt werden, dass die hier angewandten analytischen Verfahren nicht für die vorliegende Analyseebene entwickelt wurden und folglich zu Effektverschätzungen in den Daten geführt haben könnten. Dies verdeutlicht nochmals die Wichtigkeit der Entwicklung passender analytischer Methoden. Im Speziellen ist dies in der hier vorliegenden Arbeit anhand der multiplen hierarchischen mixed-effects Meta-Metaregressionen zu erkennen. Aufgrund des bereits angesprochenen Gewichtungproblems wurde sich hier entgegen der Präregistrierung entschieden und zusätzlich noch eine ungewichtete Methode zur Berechnung herangezogen, welche letztlich auch für die Interpretation der Ergebnisse genutzt wurde. Aber auch das lineare Modell könnte zumindest von einem Powerproblem betroffen sein, denn wie sich anhand der Tabellen im Appendix ableiten lässt, kommt es durch den Ausschluss von $k = 9$ bzw. $k = 8$ Metaanalysen, die keinen Impact Factor für die Initialstudie erfassen ließen, bereits zu Veränderungen in den erfassten Wirkrichtungen der Effekte und Signifikanzniveaus. Dies betraf sowohl die Ergebnisse, die anhand eines mixed-effects Modells berechnet wurden, als auch diejenigen, die anhand eines linearen Modells erfasst wurden. Eine Reanalyse der hier formulierten Modelle mit einem größeren Datensatz wäre dementsprechend ratsam.

Conclusio

Zusammenfassend lässt sich aus der vorliegenden Arbeit ableiten, dass auch das Forschungsgebiet der Sozialpsychologie und Persönlichkeitsforschung von einem Publikationsbias betroffen ist und damit einhergehend die dargebotene Ergebnislage nicht die Wirklichkeit repräsentiert. Durch den Fokus von Editor*innen und Autorinnen auf hauptsächlich signifikante und große Effekte werden so vor allem überschätzte und falsche Effekte publiziert, was zu einem decline effect im Forschungsfeld führt (Protzko & Schooler, 2017). Diesem Umstand könnte durch eine Veränderung der Richtlinien der Journals und durch ein Umdenken bezüglich dem Mehrwert von Replikationsstudien entgegengewirkt

werden (Roberts, 2011; Schooler, 2014). Auch könnte durch eine größere Fokuslegung auf Präregistrierung oder Protokollierung von Primärstudien die Anwendung von QRPs reduziert werden (Suter, 2020). Darüber hinaus könnte sich durch eine verpflichtende Präregistrierung von Studien dem Problem der Unauffindbarkeit von grauer Literatur entgegengewirkt werden (Bolibaugh et al., 2021). Positiv ist jedoch anzumerken, dass dieser Missstand bei einigen Forscher*innen bereits mehr ins Bewusstsein gerückt zu sein scheint und Anstrengungen unternommen werden, diesem entgegenzuwirken. Allein die Tatsache, dass ein decline effect in der Forschung vorliegt, ist bereits ein Hinweis dafür, dass die Wissenschaft beginnt, sich selbst zu korrigieren.

Literaturverzeichnis

- Aldhous, P. (2011, 5. Mai). *Journal rejects studies contradicting precognition*. New Scientist. Abgerufen am 19. April 2022, von <https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition/#ixzz7EeuzO7rE>
- Aloia, D., & Naughton, R. (2017). The GreyLit Report: Understanding the Challenges of Finding Grey Literature. *Grey Journal*, 12(2), 75–80. <https://doi.org/10.17026/dans-2z8-x27y>
- Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, 110(5), 766–781. <https://doi.org/10.1037/pspp0000066>
- Bakeman, R., & Robinson, B. F. (2005). *Understanding Statistics In The Behavioral Sciences*. Psychology Press.
- Barnett, A. G. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220. <https://doi.org/10.1093/ije/dyh299>
- Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, 50(4), 1088–1101. <https://doi.org/10.2307/2533446>
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407–425. <https://doi.org/10.1037/a0021524>
- Bem, D., Tressoldi, P. E., Rabeyron, T., & Duggan, M. (2016). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research*, 4. <https://doi.org/10.12688/f1000research.7177.2>
- Bishara, A. J., & Hittner, J. B. (2014). Reducing Bias and Error in the Correlation Coefficient Due to Nonnormality. *Educational and Psychological Measurement*, 75(5), 785–804. <https://doi.org/10.1177/0013164414557639>
- Bolibaugh, C., Vanek, N., & Marsden, E. J. (2021). Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research. *Bilingualism: Language and Cognition*, 24(5), 801–806. <https://doi.org/10.1017/s1366728921000535>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (1. Aufl.). Wiley.

- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A Meta-Meta-Analysis: Empirical Review of Statistical Power, Type I Error Rates, Effect Sizes, and Model Selection of Meta-Analyses Published in Psychology. *Multivariate Behavioral Research*, 45(2), 239–270. <https://doi.org/10.1080/00273171003680187>
- Calderon, S., Mac Giolla, E., Ask, K., & Granhag, P. A. (2020). Subjective likelihood and the construal level of future events: A replication study of Waksak, Trope, Liberman, and Alony (2006). *Journal of Personality and Social Psychology*, 119(5), e27–e37. <https://doi.org/10.1037/pspa0000214>
- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2003). Decreased Latent Inhibition Is Associated With Increased Creative Achievement in High-Functioning Individuals. *Journal of Personality and Social Psychology*, 85(3), 499–506. <https://doi.org/10.1037/0022-3514.85.3.499>
- Chen, C. Y., Rossignac-Milon, M., & Higgins, E. T. (2018). Feeling distressed from making decisions: Assessors' need to be right. *Journal of Personality and Social Psychology*, 115(4), 743–761. <https://doi.org/10.1037/pspp0000181>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Routledge.
- Collaborative Replications and Education Project. (2013, 30. Mai). *Collaborative Replications and Education Project (CREP)*. OSF. Abgerufen am 19. April 2022, von <https://osf.io/wfc6u/>
- Cuadrado, D., Salgado, J. F., & Moscoso, S. (2021). Personality, intelligence, and counterproductive academic behaviors: A meta-analysis. *Journal of Personality and Social Psychology*, 120(2), 504–537. <https://doi.org/10.1037/pspp0000285>
- DelPriore, D. J., Proffitt Leyva, R., Ellis, B. J., & Hill, S. E. (2018). The effects of paternal disengagement on women's perceptions of male mating intent. *Journal of Personality and Social Psychology*, 114(2), 286–302. <https://doi.org/10.1037/pspi0000113>
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
- Duval, S., & Tweedie, R. (2000a). A Nonparametric „Trim and Fill“ Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, 95(449). <https://doi.org/10.2307/2669529>

- Duval, S., & Tweedie, R. (2000b). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>
- Enisman, M., Shpitzer, H., & Kleiman, T. (2021). Choice changes preferences, not merely reflects them: A meta-analysis of the artifact-free free-choice paradigm. *Journal of Personality and Social Psychology*, 120(1), 16–29. <https://doi.org/10.1037/pspa0000263>
- Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, 7(1), 45–52. <https://doi.org/10.1177/1948550615612150>
- Field, A. (2016). *Discovering Statistics Using IBM SPSS Statistics* (4. Aufl.). SAGE Publications Ltd.
- Fischer, R., & Boer, D. (2011). What is more important for national well-being: Money or autonomy? A meta-analysis of well-being, burnout, and anxiety across 63 societies. *Journal of Personality and Social Psychology*, 101(1), 164–184. <https://doi.org/10.1037/a0023663>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. <https://doi.org/10.1126/science.1255484>
- Franco, A., Malhotra, N., & Simonovits, G. (2016). Underreporting in Psychology Experiments. *Social Psychological and Personality Science*, 7(1), 8–12. <https://doi.org/10.1177/1948550615598377>
- Frey, D., & Schmalzried, L. (2013). Der Kritische Rationalismus. *Philosophie der Führung*, 257–286. https://doi.org/10.1007/978-3-642-34439-8_14
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103(6), 933–948. <https://doi.org/10.1037/a0029709>
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3–8. <https://doi.org/10.3102/0013189x005010003>
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <https://doi.org/10.1037/a0015575>
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.

- Hunter, J., & Schmidt, F. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage Publications, Inc.
- IntHout, J., Ioannidis, J. P., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: a meta-meta-analysis. *Journal of Clinical Epidemiology*, 68(8), 860–869. <https://doi.org/10.1016/j.jclinepi.2015.03.017>
- Ioannidis, J. P. A. (1998). Effect of the Statistical Significance of Results on the Time to Completion and Publication of Randomized Efficacy Trials. *JAMA*, 279(4), 281. <https://doi.org/10.1001/jama.279.4.281>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4(3), 245–253. <https://doi.org/10.1177/1740774507079441>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Joober, R., Schmitz, N., Annable, L., & Boksa, P. (2012). Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry & Neuroscience*, 37(3), 149–152. <https://doi.org/10.1503/jpn.120065>
- Judd, C. M., & Gawronski, B. (2011). Editorial comment. *Journal of Personality and Social Psychology*, 100(3), 406. <https://doi.org/10.1037/0022789>
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Lakens, D. (2018, 19. März). *The Journal of Personality and Social Psychology: The Good, The Bad, and The Ugly*. The 20% Statistician. Abgerufen am 19. April 2022, von <http://daniellakens.blogspot.com/2018/03/the-journal-of-personality-and-social.html>
- Lamal, P. A. (1990). On the importance of replication. *Journal of Social Behavior and Personality*, 5(4), 31–36. <https://www.proquest.com/scholarly-journals/on-importance-replication/docview/1292298971/se-2?accountid=14682>
- Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology*, 103(4), 584–605. <https://doi.org/10.1037/a0029544>
- Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLOS ONE*, 13(9), e0204056. <https://doi.org/10.1371/journal.pone.0204056>

- Lipnevich, A. A., Credè, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the Big Five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology*, 112(3), 491–509. <https://doi.org/10.1037/pspp0000099>
- Magnusson, K. (2014, 27. Januar). *Expected overestimation of Cohen's d under publication bias*. R-Bloggers. Abgerufen am 19. April 2022, von <https://www.r-bloggers.com/2014/01/expected-overestimation-of-cohens-d-under-publication-bias-2/>
- Margraf, J. (2015). Zur Lage der Psychologie. *Psychologische Rundschau*, 66(1), 1–30. <https://doi.org/10.1026/0033-3042/a000247>
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the Structure of Normal and Abnormal Personality: An Integrative Hierarchical Approach. *Journal of Personality and Social Psychology*, 88(1), 139–157. <https://doi.org/10.1037/0022-3514.88.1.139>
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783. <https://doi.org/10.1037/0022-3514.90.5.751>
- Pfeiffer, T., Bertram, L., & Ioannidis, J. P. A. (2011). Quantifying Selective Reporting and the Proteus Phenomenon for Multiple Datasets with Similar Bias. *PLoS ONE*, 6(3), e18362. <https://doi.org/10.1371/journal.pone.0018362>
- Popper, K. (2009). *Gesammelte Werke: Band 10: Vermutungen Und Widerlegungen. Das Wachstum Der Wissenschaftlichen Erkenntnis* (H. Keuth, Hrsg.). Mohr Siebeck.
- Prinzie, P., Stams, G. J. J. M., Deković, M., Reijntjes, A. H. A., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: A meta-analytic review. *Journal of Personality and Social Psychology*, 97(2), 351–362. <https://doi.org/10.1037/a0015823>
- Rabeyron, T. (2020). Why Most Research Findings About Psi Are False: The Replicability Crisis, the Psi Paradox and the Myth of Sisyphus. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.562992>
- Roberts, B. W. (2011, 6. November). *The Online Newsletter for Personality Science*. Association for Research in Personality. Abgerufen am 19. April 2022, von http://www.personality-arp.org/html/newsletter06/problem.html#_ftnref1

- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99(3), 400–406. <https://doi.org/10.1037/0033-2909.99.3.400>
- Rousselet, G. A. (2018, 4. April). *Cohen's d is biased*. Basic Statistics. Abgerufen am 19. April 2022, von <https://garstats.wordpress.com/2018/04/04/dbias/>
- Rücker, G., Schwarzer, G., Carpenter, J. R., & Schumacher, M. (2008). Undue reliance on I² in assessing heterogeneity may mislead. *BMC Medical Research Methodology*, 8(1). <https://doi.org/10.1186/1471-2288-8-79>
- Schäfer, T., & Schwarz, M. A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.00813>
- Schimmack, U. (2015, 13. August). *Replicability rankings of 26 psychology journals*. Replicability-Index. Abgerufen am 19. April 2022, von <https://replicationindex.com/2015/08/13/replicability-ranking-of-26-psychology-journals/>
- Schimmack, U. (2016, 26. Januar). *2015 Replicability Ranking of 100+ Psychology Journals*. Replicability-Index. Abgerufen am 19. April 2022, von <https://replicationindex.com/2016/01/26/2015-replicability-ranking-of-100-psychology-journals/>
- Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62(1), 97–128. <https://doi.org/10.1348/000711007x255327>
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470(7335), 437. <https://doi.org/10.1038/470437a>
- Schooler, J. W. (2014). Turning the Lens of Science on Itself. *Perspectives on Psychological Science*, 9(5), 579–584. <https://doi.org/10.1177/1745691614547878>
- Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2022). Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices, prevalence rates, and predictors in industrial/organizational psychology. *Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000991>

- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547.
<https://doi.org/10.1037/a0033242>
- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60–78.
<https://doi.org/10.1002/jrsm.1095>
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton & M. Bornstein (Hrsg.), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* (S. 99–110). Wiley, Ltd.
- Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis. *Journal of Clinical Epidemiology*, 53(11), 1119–1129.
[https://doi.org/10.1016/s0895-4356\(00\)00242-0](https://doi.org/10.1016/s0895-4356(00)00242-0)
- Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., & Williams, R. M. J. (1949). *Studies in Social Psychology in World War II: The American Soldier - Adjustment during Army Life Vol. 1* (Bd. 1). Princeton Univ. Press.
- Suter, W. N. (2020). Questionable Research Practices: How to Recognize and Avoid Them. *Home Health Care Management & Practice*, 32(4), 183–190.
<https://doi.org/10.1177/1084822320934468>
- Tijdsink, J. K., Verbeke, R., & Smulders, Y. M. (2014). Publication Pressure and Scientific Misconduct in Medical Scientists. *Journal of Empirical Research on Human Research Ethics*, 9(5), 64–71. <https://doi.org/10.1177/1556264614552421>
- van Aert, R. C. M. (2022, 21. März) *R/puni_star.R*. rdrv.io. Abgerufen am 13.05.2022, von https://rdrv.io/cran/puniform/src/R/puni_star.R.
- van Aert, R. C. M., & van Assen, M. A. L. M. (2018, 2. Oktober). *Correcting for Publication Bias in a Meta-Analysis with the P-uniform* Method*. MetaArXiv Preprints. Abgerufen am 19. April 2022, von <https://doi.org/10.31222/osf.io/zqjr9>
- van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLOS ONE*, 14(4), e0215052. <https://doi.org/10.1371/journal.pone.0215052>
- van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, 20(3), 293–309. <https://doi.org/10.1037/met0000025>

- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435.
<https://doi.org/10.1007/bf02294384>
- Vevea, J. L., & Woods, C. M. (2005). Publication Bias in Research Synthesis: Sensitivity Analysis Using A Priori Weight Functions. *Psychological Methods*, 10(4), 428–443.
<https://doi.org/10.1037/1082-989x.10.4.428>
- Yuan, H. F., Xu, W. D., & Hu, H. Y. (2013). Young Chinese doctors and the pressure of publication. *The Lancet*, 381(9864), e4. [https://doi.org/10.1016/s0140-6736\(13\)60174-9](https://doi.org/10.1016/s0140-6736(13)60174-9)

Appendix

Liste aller Studien, die in die Analysen inkludiert wurden ($k = 37$)

- Amit, A., Mentser, S., Arieli, S., & Porzycki, N. (2021). Distinguishing deliberate from systematic thinking. *Journal of Personality and Social Psychology*, 120(3), 765–788. <https://doi.org/10.1037/pspp0000284>
- Burnham, B. R. (2020). Are liberals really dirty? Two failures to replicate Helzer and Pizarro's (2011) study 1, with meta-analysis. *Journal of Personality and Social Psychology*, 119(6), e38–e42. <https://doi.org/10.1037/pspa0000238>
- Calderon, S., Mac Giolla, E., Ask, K., & Granhag, P. A. (2020). Subjective likelihood and the construal level of future events: A replication study of Wakslak, Trope, Liberman, and Alony (2006). *Journal of Personality and Social Psychology*, 119(5), e27–e37. <https://doi.org/10.1037/pspa0000214>
- Cohn, L. D., & Westenberg, P. M. (2004). Intelligence and maturity: Meta-analytic evidence for the incremental and discriminant validity of Loevinger's measure of ego development. *Journal of Personality and Social Psychology*, 86(5), 760–772. <https://doi.org/10.1037/0022-3514.86.5.760>
- Connor-Smith, J. K., & Flachsbart, C. (2007). Relations between personality and coping: A meta-analysis. *Journal of Personality and Social Psychology*, 93(6), 1080–1107. <https://doi.org/10.1037/0022-3514.93.6.1080>
- Crawford, J. T., Brandt, M. J., Inbar, Y., & Mallinas, S. R. (2016). Right-wing authoritarianism predicts prejudice equally toward “gay men and lesbians” and “homosexuals”. *Journal of Personality and Social Psychology*, 111(2), e31–e45. <https://doi.org/10.1037/pspp0000070>
- Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, 113(3), 492–511. <https://doi.org/10.1037/pspp0000102>
- Cuadrado, D., Salgado, J. F., & Moscoso, S. (2021). Personality, intelligence, and counterproductive academic behaviors: A meta-analysis. *Journal of Personality and Social Psychology*, 120(2), 504–537. <https://doi.org/10.1037/pspp0000285>
- De Dreu, C. K. W., Weingart, L. R., & Kwon, S. (2000). Influence of social motives on integrative negotiation: A meta-analytic review and test of two theories. *Journal of*

- Personality and Social Psychology*, 78(5), 889–905. <https://doi.org/10.1037/0022-3514.78.5.889>
- Dittmar, H., Bond, R., Hurst, M., & Kasser, T. (2014). The relationship between materialism and personal well-being: A meta-analysis. *Journal of Personality and Social Psychology*, 107(5), 879–924. <https://doi.org/10.1037/a0037409>
- Donahue, M. J. (1985). Intrinsic and extrinsic religiousness: Review and meta-analysis. *Journal of Personality and Social Psychology*, 48(2), 400–419. <https://doi.org/10.1037/0022-3514.48.2.400>
- Elfenbein, H. A., & Eisenkraft, N. (2010). The relationship between displaying and perceiving nonverbal cues of affect: A meta-analysis to solve an old mystery. *Journal of Personality and Social Psychology*, 98(2), 301–318. <https://doi.org/10.1037/a0017766>
- Enisman, M., Shpitzer, H., & Kleiman, T. (2021). Choice changes preferences, not merely reflects them: A meta-analysis of the artifact-free free-choice paradigm. *Journal of Personality and Social Psychology*, 120(1), 16–29. <https://doi.org/10.1037/pspa0000263>
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103(6), 933–948. <https://doi.org/10.1037/a0029709>
- Gordon, R. A. (1996). Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology*, 71(1), 54–70. <https://doi.org/10.1037/0022-3514.71.1.54>
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <https://doi.org/10.1037/a0015575>
- Hicks, L. L., McNulty, J. K., Faure, R., Meltzer, A. L., Righetti, F., & Hofmann, W. (2021). Do people realize how their partners make them feel? Relationship enhancement motives and stress determine the link between implicitly assessed partner attitudes and relationship satisfaction. *Journal of Personality and Social Psychology*, 120(2), 335–369. <https://doi.org/10.1037/pspi0000247>
- Jetten, J., Spears, R., & Postmes, T. (2004). Intergroup Distinctiveness and Differentiation: A Meta-Analytic Integration. *Journal of Personality and Social Psychology*, 86(6), 862–879. <https://doi.org/10.1037/0022-3514.86.6.862>

- Joshi, P. D., Waksalak, C. J., Appel, G., & Huang, L. (2020). Gender differences in communicative abstraction. *Journal of Personality and Social Psychology*, 118(3), 417–435. <https://doi.org/10.1037/pspa0000177>
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4), 681–706. <https://doi.org/10.1037/0022-3514.65.4.681>
- Koestner, R., Lekes, N., Powers, T. A., & Chicoine, E. (2002). Attaining personal goals: Self-concordance plus implementation intentions equals success. *Journal of Personality and Social Psychology*, 83(1), 231–244. <https://doi.org/10.1037/0022-3514.83.1.231>
- Leach, C. W., & Cidam, A. (2015). When is shame linked to constructive approach orientation? A meta-analysis. *Journal of Personality and Social Psychology*, 109(6), 983–1002. <https://doi.org/10.1037/pspa0000037>
- Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology*, 103(4), 584–605. <https://doi.org/10.1037/a0029544>
- Lipnevich, A. A., Credè, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the Big Five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology*, 112(3), 491–509. <https://doi.org/10.1037/pspp0000099>
- Lucas, R. E., & Fujita, F. (2000). Factors influencing the relation between extraversion and pleasant affect. *Journal of Personality and Social Psychology*, 79(6), 1039–1056. <https://doi.org/10.1037/0022-3514.79.6.1039>
- Marcus-Newhall, A., Pedersen, W. C., Carlson, M., & Miller, N. (2000). Displaced aggression is alive and well: A meta-analytic review. *Journal of Personality and Social Psychology*, 78(4), 670–689. <https://doi.org/10.1037/0022-3514.78.4.670>
- Mullen, B. (1985). Strength and immediacy of sources: A meta-analytic evaluation of the forgotten elements of social impact theory. *Journal of Personality and Social Psychology*, 48(6), 1458–1466. <https://doi.org/10.1037/0022-3514.48.6.1458>
- Orth, U., Dapp, L. C., Erol, R. Y., Krauss, S., & Luciano, E. C. (2021). Development of domain-specific self-evaluations: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, 120(1), 145–172. <https://doi.org/10.1037/pspp0000378>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of*

- Personality and Social Psychology*, 105(2), 171–192.
<https://doi.org/10.1037/a0032734>
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.
<https://doi.org/10.1037/0022-3514.90.5.751>
- Prinzle, P., Stams, G. J. J. M., Deković, M., Reijntjes, A. H. A., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: A meta-analytic review. *Journal of Personality and Social Psychology*, 97(2), 351–362.
<https://doi.org/10.1037/a0015823>
- Roisman, G. I., Holland, A., Fortuna, K., Fraley, R. C., Clausell, E., & Clarke, A. (2007). The Adult Attachment Interview and self-reports of attachment style: An empirical rapprochement. *Journal of Personality and Social Psychology*, 92(4), 678–697.
<https://doi.org/10.1037/0022-3514.92.4.678>
- Saragovi, C., Koestner, R., Di Dio, L., & Aubé, J. (1997). Agency, communion, and well-being: Extending Helgeson's (1994) model. *Journal of Personality and Social Psychology*, 73(3), 593–609. <https://doi.org/10.1037/0022-3514.73.3.593>
- Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology*, 89(3), 395–406. <https://doi.org/10.1037/0022-3514.89.3.395>
- Thielmann, I., Hilbig, B. E., & Zettler, I. (2020). Seeing me, seeing you: Testing competing accounts of assumed similarity in personality judgments. *Journal of Personality and Social Psychology*, 118(1), 172–198. <https://doi.org/10.1037/pspp0000222>
- Van Dessel, P., De Houwer, J., Gast, A., Roets, A., & Smith, C. T. (2020). On the effectiveness of approach-avoidance instructions and training for changing evaluations of social groups. *Journal of Personality and Social Psychology*, 119(2), e1–e14.
<https://doi.org/10.1037/pspa0000189>
- Weber, B., & Hertel, G. (2007). Motivation gains of inferior group members: A meta-analytical review. *Journal of Personality and Social Psychology*, 93(6), 973–993.
<https://doi.org/10.1037/0022-3514.93.6.973>

**Aus den Analysen ausgeschlossen, da keine Standardmetaanalyse durchgeführt wurde
 (k = 15)**

- Affleck, G., Tennen, H., Urrows, S., & Higgins, P. (1994). Person and contextual features of daily stress reactivity: Individual differences in relations of undesirable daily events

- with mood disturbance and chronic pain intensity. *Journal of Personality and Social Psychology*, 66(2), 329–340. <https://doi.org/10.1037/0022-3514.66.2.329>
- Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, 110(5), 766–781. <https://doi.org/10.1037/pspp0000066>
- Ayanian, A. H., Tausch, N., Acar, Y. G., Chayinska, M., Cheung, W. Y., & Lukyanova, Y. (2021). Resistance in repressive contexts: A comprehensive test of psychological predictors. *Journal of Personality and Social Psychology*, 120(4), 912–939. <https://doi.org/10.1037/pspi0000285>
- Essien, I., Calanchini, J., & Degner, J. (2021). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology*, 120(5), 1204–1230. <https://doi.org/10.1037/pspi0000302>
- Fischer, R., & Boer, D. (2011). What is more important for national well-being: Money or autonomy? A meta-analysis of well-being, burnout, and anxiety across 63 societies. *Journal of Personality and Social Psychology*, 101(1), 164–184. <https://doi.org/10.1037/a0023663>
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, 117(3), 522–559. <https://doi.org/10.1037/pspa0000160>
- Gaertner, L., Sedikides, C., Vevea, J. L., & Iuzzini, J. (2002). The „I,“ the „we,“ and the „when“: A meta-analysis of motivational primacy in self-definition. *Journal of Personality and Social Psychology*, 83(3), 574–591. <https://doi.org/10.1037/0022-3514.83.3.574>
- Lange, J., Weidman, A. C., & Crusius, J. (2018). The painful duality of envy: Evidence for an integrative theory and a meta-analysis on the relation of envy and schadenfreude. *Journal of Personality and Social Psychology*, 114(4), 572–598. <https://doi.org/10.1037/pspi0000118>
- Lucas, R. E., & Baird, B. M. (2004). Extraversion and Emotional Reactivity. *Journal of Personality and Social Psychology*, 86(3), 473–485. <https://doi.org/10.1037/0022-3514.86.3.473>

- Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology*, 102(3), 592–615. <https://doi.org/10.1037/a0025948>
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the Structure of Normal and Abnormal Personality: An Integrative Hierarchical Approach. *Journal of Personality and Social Psychology*, 88(1), 139–157. <https://doi.org/10.1037/0022-3514.88.1.139>
- Miles, D. R., & Carey, G. (1997). Genetic and environmental architecture on human aggression. *Journal of Personality and Social Psychology*, 72(1), 207–217. <https://doi.org/10.1037/0022-3514.72.1.207>
- Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., Mortensen, E. L., Colodro-Conde, L., & Jang, K. L. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, 117(4), e35–e50. <https://doi.org/10.1037/pspp0000202>
- Sedikides, C., Gaertner, L., & Vevea, J. L. (2005). Pancultural self-enhancement reloaded: A meta-analytic reply to Heine (2005). *Journal of Personality and Social Psychology*, 89(4), 539–551. <https://doi.org/10.1037/0022-3514.89.4.539>
- Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology*, 84(1), 205–220. <https://doi.org/10.1037/0022-3514.84.1.205>

Aus den Analysen ausgeschlossen, da keine Metaanalyse durchgeführt wurde (k = 29)

- Anooshian, L. J., & Seibert, P. S. (1996). Conscious and unconscious retrieval in picture recognition: A framework for exploring gender differences. *Journal of Personality and Social Psychology*, 70(3), 637–645. <https://doi.org/10.1037/0022-3514.70.3.637>
- Bélanger, J. J., Lafrenière, M. A. K., Vallerand, R. J., & Kruglanski, A. W. (2013). When passion makes the heart grow colder: The role of passion in alternative goal suppression. *Journal of Personality and Social Psychology*, 104(1), 126–147. <https://doi.org/10.1037/a0029679>
- Bornstein, R. F., & D’Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, 63(4), 545–552. <https://doi.org/10.1037/0022-3514.63.4.545>
- Cioffi, D., & Holloway, J. (1993). Delayed costs of suppressed pain. *Journal of Personality and Social Psychology*, 64(2), 274–282. <https://doi.org/10.1037/0022-3514.64.2.274>

- Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, 97(6), 931–945.
<https://doi.org/10.1037/a0017452>
- Davis, J. L., & Rusbult, C. E. (2001). Attitude alignment in close relationships. *Journal of Personality and Social Psychology*, 81(1), 65–84. <https://doi.org/10.1037/0022-3514.81.1.65>
- Dolinski, D. (2000). On inferring one's beliefs from one's attempt and consequences for subsequent compliance. *Journal of Personality and Social Psychology*, 78(2), 260–272. <https://doi.org/10.1037/0022-3514.78.2.260>
- Dubois, D., Rucker, D. D., & Galinsky, A. D. (2015). Social class, power, and selfishness: When and why upper and lower class individuals behave unethically. *Journal of Personality and Social Psychology*, 108(3), 436–449.
<https://doi.org/10.1037/pspi0000008>
- Eagly, A. H. (1992). Uneven progress: Social psychology and the study of attitudes. *Journal of Personality and Social Psychology*, 63(5), 693–710. <https://doi.org/10.1037/0022-3514.63.5.693>
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1), 5–17. <https://doi.org/10.1037/0022-3514.84.1.5>
- Frey, K. P., & Eagly, A. H. (1993). Vividness can undermine the persuasiveness of messages. *Journal of Personality and Social Psychology*, 65(1), 32–44.
<https://doi.org/10.1037/0022-3514.65.1.32>
- Gallo, I. S., Keil, A., McCulloch, K. C., Rockstroh, B., & Gollwitzer, P. M. (2009). Strategic automation of emotion regulation. *Journal of Personality and Social Psychology*, 96(1), 11–31. <https://doi.org/10.1037/a0013460>
- Gorassini, D. R., & Olson, J. M. (1995). Does self-perception change explain the foot-in-the-door effect? *Journal of Personality and Social Psychology*, 69(1), 91–105.
<https://doi.org/10.1037/0022-3514.69.1.91>
- Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of Personality and Social Psychology*, 74(3), 622–628. <https://doi.org/10.1037/0022-3514.74.3.622>
- Hilmert, C. J., Kulik, J. A., & Christenfeld, N. J. S. (2006). Positive and negative opinion modeling: The influence of another's similarity and dissimilarity. *Journal of*

- Personality and Social Psychology*, 90(3), 440–452. <https://doi.org/10.1037/0022-3514.90.3.440>
- Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2013). Like Schrödinger's cat, the impact bias is both dead and alive: Reply to Wilson and Gilbert (2013). *Journal of Personality and Social Psychology*, 105(5), 749–756. <https://doi.org/10.1037/a0034340>
- Marsh, H. W., & Byrne, B. M. (1991). Differentiated additive androgyny model: Relations between masculinity, femininity, and multiple dimensions of self-concept. *Journal of Personality and Social Psychology*, 61(5), 811–828. <https://doi.org/10.1037/0022-3514.61.5.811>
- Meltzer, A. L., McNulty, J. K., Jackson, G. L., & Karney, B. R. (2014). Men still value physical attractiveness in a long-term mate more than women: Rejoinder to Eastwick, Neff, Finkel, Luchies, and Hunt (2014). *Journal of Personality and Social Psychology*, 106(3), 435–440. <https://doi.org/10.1037/a0035342>
- Nikles, C. D., Brecht, D. L., Klinger, E., & Bursell, A. L. (1998). The effects of current-concern- and nonconcern-related waking suggestions on nocturnal dream content. *Journal of Personality and Social Psychology*, 75(1), 242–255. <https://doi.org/10.1037/0022-3514.75.1.242>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108(4), 562–571. <https://doi.org/10.1037/pspa0000023>
- Scherer, K. R., Mortillaro, M., Rotondi, I., Sergi, I., & Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion inference. *Journal of Personality and Social Psychology*, 114(3), 358–379. <https://doi.org/10.1037/pspa0000107>
- Sohn, D. (1980). Critique of Cooper's meta-analytic assessment of the findings on sex differences in conformity behavior. *Journal of Personality and Social Psychology*, 39(6), 1215–1221. <https://doi.org/10.1037/h0077717>
- Stacy, A. W., Newcomb, M. D., & Bentler, P. M. (1991). Personality, problem drinking, and drunk driving: Mediating, moderating, and direct-effect models. *Journal of Personality and Social Psychology*, 60(5), 795–811. <https://doi.org/10.1037/0022-3514.60.5.795>

- Suedfeld, P., & Bluck, S. (1993). Changes in integrative complexity accompanying significant life events: Historical evidence. *Journal of Personality and Social Psychology*, 64(1), 124–130. <https://doi.org/10.1037/0022-3514.64.1.124>
- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, 66(1), 21–36. <https://doi.org/10.1037/0022-3514.66.1.21>
- Tackman, A. M., Sbarra, D. A., Carey, A. L., Donnellan, M. B., Horn, A. B., Holtzman, N. S., Edwards, T. S., Pennebaker, J. W., & Mehl, M. R. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology*, 116(5), 817–834. <https://doi.org/10.1037/pspp0000187>
- Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology*, 73(5), 941–959. <https://doi.org/10.1037/0022-3514.73.5.941>
- Wehrle, T., Kaiser, S., Schmidt, S. & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78(1), 105–119. <https://doi.org/10.1037/0022-3514.78.1.105>
- Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84(3), 608–618. <https://doi.org/10.1037/0022-3514.84.3.608>

Aus den Analysen ausgeschlossen, da eine Internal Metaanalysis durchgeführt wurde (k = 39)

- Ballinger, T., & Crocker, J. (2021). Understanding Whites' perceptions of multicultural policies: A (non)zero-sum framework? *Journal of Personality and Social Psychology*, 120(5), 1234–1260. <https://doi.org/10.1037/pspi0000315>
- Bentley, S. V., Greenaway, K. H., Haslam, S. A., Cruwys, T., Steffens, N. K., Haslam, C., & Cull, B. (2020). Social identity mapping online. *Journal of Personality and Social Psychology*, 118(2), 213–241. <https://doi.org/10.1037/pspa0000174>
- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2003). Decreased Latent Inhibition Is Associated With Increased Creative Achievement in High-Functioning Individuals.

- Journal of Personality and Social Psychology*, 85(3), 499–506.
<https://doi.org/10.1037/0022-3514.85.3.499>
- Chen, C. Y., Rossignac-Milon, M., & Higgins, E. T. (2018). Feeling distressed from making decisions: Assessors' need to be right. *Journal of Personality and Social Psychology*, 115(4), 743–761. <https://doi.org/10.1037/pspp0000181>
- Chu, C., Walker, K. L., Stanley, I. H., Hirsch, J. K., Greenberg, J. H., Rudd, M. D., & Joiner, T. E. (2018). Perceived problem-solving deficits and suicidal ideation: Evidence for the explanatory roles of thwarted belongingness and perceived burdensomeness in five samples. *Journal of Personality and Social Psychology*, 115(1), 137–160.
<https://doi.org/10.1037/pspp0000152>
- DelPriore, D. J., Proffitt Leyva, R., Ellis, B. J., & Hill, S. E. (2018). The effects of paternal disengagement on women's perceptions of male mating intent. *Journal of Personality and Social Psychology*, 114(2), 286–302. <https://doi.org/10.1037/pspi0000113>
- Dunlop, P. D., Bourdage, J. S., de Vries, R. E., Hilbig, B. E., Zettler, I., & Ludeke, S. G. (2017). Openness to (reporting) experiences that one never had: Overclaiming as an outcome of the knowledge accumulated through a proclivity for cognitive and aesthetic exploration. *Journal of Personality and Social Psychology*, 113(5), 810–834.
<https://doi.org/10.1037/pspp0000110>
- Dupree, C. H., & Fiske, S. T. (2019). Self-presentation in interracial settings: The competence downshift by White liberals. *Journal of Personality and Social Psychology*, 117(3), 579–604. <https://doi.org/10.1037/pspi0000166>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–34. <https://doi-org.uaccess.univie.ac.at/10.1136/bmj.315.7109.629>
- Everett, J. A. C., Clark, C. J., Meindl, P., Luguri, J. B., Earp, B. D., Graham, J., Ditto, P. H., & Shariff, A. F. (2021). Political differences in free will belief are associated with differences in moralization. *Journal of Personality and Social Psychology*, 120(2), 461–483. <https://doi.org/10.1037/pspp0000286>
- Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, 97(6), 1097–1114.
<https://doi.org/10.1037/a0016786>
- Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. C. P., Egloff, B., Denissen, J. J. A., & Back, M. D. (2017). Puffed-up but shaky selves: State self-esteem level and

- variability in narcissists. *Journal of Personality and Social Psychology*, 112(5), 769–786. <https://doi.org/10.1037/pspp0000093>
- Girme, Y. U., Overall, N. C., Simpson, J. A., & Fletcher, G. J. O. (2015). “All or nothing”: Attachment avoidance and the curvilinear effects of partner support. *Journal of Personality and Social Psychology*, 108(3), 450–475. <https://doi.org/10.1037/a0038866>
- Hall, W., Schmader, T., Aday, A., Inness, M., & Croft, E. (2018). Climate control: The relationship between social identity threat and cues to an identity-safe culture. *Journal of Personality and Social Psychology*, 115(3), 446–467. <https://doi.org/10.1037/pspi0000137>
- Hone, L. S. E., McCauley, T. G., Pedersen, E. J., Carter, E. C., & McCullough, M. E. (2021). The sex premium in religiously motivated moral judgment. *Journal of Personality and Social Psychology*, 120(6), 1621–1633. <https://doi.org/10.1037/pspp0000296>
- Hornsey, M. J., Wohl, M. J. A., Harris, E. A., Okimoto, T. G., Thai, M., & Wenzel, M. (2020). Embodied remorse: Physical displays of remorse increase positive responses to public apologies, but have negligible effects on forgiveness. *Journal of Personality and Social Psychology*, 119(2), 367–389. <https://doi.org/10.1037/pspi0000208>
- Jackson, J. C., & Gray, K. (2019). When a good god makes bad people: Testing a theory of religion and immorality. *Journal of Personality and Social Psychology*, 117(6), 1203–1230. <https://doi.org/10.1037/pspp0000206>
- Kim, J. Y., Campbell, T. H., Shepherd, S., & Kay, A. C. (2020). Understanding contemporary forms of exploitation: Attributions of passion serve to legitimize the poor treatment of workers. *Journal of Personality and Social Psychology*, 118(1), 121–148. <https://doi.org/10.1037/pspi0000190>
- Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology*, 113(5), 730–752. <https://doi.org/10.1037/pspi0000103>
- Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., Bremner, R., Moser, J., & Ayduk, O. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of Personality and Social Psychology*, 106(2), 304–324. <https://doi.org/10.1037/a0035173>
- Lammers, J., & Baldwin, M. (2018). Past-focused temporal communication overcomes conservatives’ resistance to liberal political ideas. *Journal of Personality and Social Psychology*, 114(4), 599–619. <https://doi.org/10.1037/pspi0000121>

- Lowe, M. L., & Haws, K. L. (2019). Confession and self-control: A prelude to repentance or relapse? *Journal of Personality and Social Psychology*, 116(4), 563–581.
<https://doi.org/10.1037/pspi0000152>
- Miketta, S., & Friese, M. (2019). Debriefed but still troubled? About the (in)effectiveness of postexperimental debriefings after ego threat. *Journal of Personality and Social Psychology*, 117(2), 282–309. <https://doi.org/10.1037/pspa0000155>
- Murray, S. L., Lamarche, V. M., Gomillion, S., Seery, M. D., & Kondrak, C. (2017). In defense of commitment: The curative power of violated expectations. *Journal of Personality and Social Psychology*, 113(5), 697–729.
<https://doi.org/10.1037/pspi0000102>
- Orth, U., & Luciano, E. C. (2015). Self-esteem, narcissism, and stressful life events: Testing for selection and socialization. *Journal of Personality and Social Psychology*, 109(4), 707–721. <https://doi.org/10.1037/pspp0000049>
- Orth, U., Robins, R. W., Meier, L. L., & Conger, R. D. (2016). Refining the vulnerability model of low self-esteem and depression: Disentangling the effects of genuine self-esteem and narcissism. *Journal of Personality and Social Psychology*, 110(1), 133–149. <https://doi.org/10.1037/pspp0000038>
- Oyserman, D., Sorensen, N., Reber, R., & Chen, S. X. (2009). Connecting and separating mind-sets: Culture as situated cognition. *Journal of Personality and Social Psychology*, 97(2), 217–235. <https://doi.org/10.1037/a0015850>
- Park, Y., Impett, E. A., MacDonald, G., & Lemay, E. P. (2019). Saying “thank you”: Partners’ expressions of gratitude protect relationship satisfaction and commitment from the harmful effects of attachment insecurity. *Journal of Personality and Social Psychology*, 117(4), 773–806. <https://doi.org/10.1037/pspi0000178>
- Raj, M., Wiltermuth, S. S., & Adams, G. S. (2020). The social costs of forgiving following multiple-victim transgressions. *Journal of Personality and Social Psychology*, 119(2), 344–366. <https://doi.org/10.1037/pspi0000215>
- Rattan, A., Savani, K., Komarraju, M., Morrison, M. M., Boggs, C., & Ambady, N. (2018). Meta-lay theories of scientific potential drive underrepresented students’ sense of belonging to science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, 115(1), 54–75.
<https://doi.org/10.1037/pspi0000130>

- Schaerer, M., Schweinsberg, M., & Swaab, R. I. (2018). Imaginary alternatives: The effects of mental simulation on powerless negotiators. *Journal of Personality and Social Psychology*, 115(1), 96–117. <https://doi.org/10.1037/pspi0000129>
- Selcuk, E., Zayas, V., Günaydin, G., Hazan, C., & Kross, E. (2012). Mental representations of attachment figures facilitate recovery following upsetting autobiographical memory recall. *Journal of Personality and Social Psychology*, 103(2), 362–378. <https://doi.org/10.1037/a0028125>
- Sutin, A. R., Luchetti, M., Stephan, Y., Robins, R. W., & Terracciano, A. (2017). Parental educational attainment and adult offspring personality: An intergenerational life span approach to the origin of adult personality traits. *Journal of Personality and Social Psychology*, 113(1), 144–166. <https://doi.org/10.1037/pspp0000137>
- Tesser, A., & Beach, S. R. H. (1998). Life events, relationship quality, and depression: An investigation of judgment discontinuity in vivo. *Journal of Personality and Social Psychology*, 74(1), 36–52. <https://doi.org/10.1037/0022-3514.74.1.36>
- van Dick, R., Wagner, U., Pettigrew, T. F., Christ, O., Wolf, C., Petzel, T., Castro, V. S., & Jackson, J. S. (2004). Role of Perceived Importance in Intergroup Contact. *Journal of Personality and Social Psychology*, 87(2), 211–227. <https://doi.org/10.1037/0022-3514.87.2.211>
- Visser, P. S., & Krosnick, J. A. (1998). Development of attitude strength over the life cycle: Surge and decline. *Journal of Personality and Social Psychology*, 75(6), 1389–1410. <https://doi.org/10.1037/0022-3514.75.6.1389>
- Webb, C. E., Coleman, P. T., Rossignac-Milon, M., Tomasulo, S. J., & Higgins, E. T. (2017). Moving on or digging deeper: Regulatory mode and interpersonal conflict resolution. *Journal of Personality and Social Psychology*, 112(4), 621–641. <https://doi.org/10.1037/pspp0000131>
- White, M. H., & Crandall, C. S. (2017). Freedom of racist speech: Ego and expressive threats. *Journal of Personality and Social Psychology*, 113(3), 413–429. <https://doi.org/10.1037/pspi0000095>
- Zee, K. S., Bolger, N., & Higgins, E. T. (2020). Regulatory effectiveness of social support. *Journal of Personality and Social Psychology*, 119(6), 1316–1358. <https://doi.org/10.1037/pspi0000235>
- Zee, K. S., Cavallo, J. V., Flores, A. J., Bolger, N., & Higgins, E. T. (2018). Motivation moderates the effects of social support visibility. *Journal of Personality and Social Psychology*, 114(5), 735–765. <https://doi.org/10.1037/pspi0000119>

Aus den Analysen ausgeschlossen, da nur eine Reanalyse einer Metaanalyse durchgeführt wurde ($k = 4$)

- Gallucci, M. (2003). I Sell Seashells by the Seashore and My Name Is Jack: Comment on Pelham, Mirenberg, and Jones (2002). *Journal of Personality and Social Psychology*, 85(5), 789–799. <https://doi.org/10.1037/0022-3514.85.5.789>
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, 108(4), 553–561. <https://doi.org/10.1037/pspa0000016>
- Hall, J. A. & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, 77(2), 350–359. <https://doi.org/10.1037/0022-3514.77.2.350>
- Wilson, T. D., & Gilbert, D. T. (2013). The impact bias is alive and well. *Journal of Personality and Social Psychology*, 105(5), 740–748. <https://doi.org/10.1037/a0032662>

Aus den Analysen ausgeschlossen, da keine Standard-Effektstärken verwendet wurde ($k = 3$)

- Carswell, K. L. & Finkel, E. J. (2018). Can you get the magic back? The moderating effect of passion decay beliefs on relationship commitment. *Journal of Personality and Social Psychology*, 115(6), 1002–1033. <https://doi.org/10.1037/pspi0000147>
- Carswell, K. L., Finkel, E. J., & Kumashiro, M. (2019). Creativity and romantic passion. *Journal of Personality and Social Psychology*, 116(6), 919–941. <https://doi.org/10.1037/pspi0000162>
- Whitley, B. E. (1983). Sex role orientation and self-esteem: A critical meta-analytic review. *Journal of Personality and Social Psychology*, 44(4), 765–778. <https://doi.org/10.1037/0022-3514.44.4.765>

Aus den Analysen ausgeschlossen, da eine Cross-temporale Metaanalyse durchgeführt wurde ($k = 2$)

- Twenge, J. M. (2000). The age of anxiety? The birth cohort change in anxiety and neuroticism, 1952–1993. *Journal of Personality and Social Psychology*, 79(6), 1007–1021. <https://doi.org/10.1037/0022-3514.79.6.1007>

Twenge, J. M. (2001). Changes in women's assertiveness in response to status and roles: A cross-temporal meta-analysis, 1931–1993. *Journal of Personality and Social Psychology*, 81(1), 133–145. <https://doi.org/10.1037/0022-3514.81.1.133>

**Aus den Analysen ausgeschlossen, da notwendige Daten für die Analyse nicht vorlagen
(*k* = 11)**

Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait–multimethod approach. *Journal of Personality and Social Psychology*, 102(2), 408–426. <https://doi.org/10.1037/a0025559>

Cooper, H. M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology*, 37(1), 131–146. <https://doi.org/10.1037/0022-3514.37.1.131>

Eastwick, P. W., Neff, L. A., Finkel, E. J., Luchies, L. B., & Hunt, L. L. (2014). Is a meta-analysis a foundation, or just another brick? Comment on Meltzer, McNulty, Jackson, and Karney (2014). *Journal of Personality and Social Psychology*, 106(3), 429–434. <https://doi.org/10.1037/a0034767>

Effron, D. A., Bryan, C. J., & Murnighan, J. K. (2015). Cheating at the end to avoid regret. *Journal of Personality and Social Psychology*, 109(3), 395–414. <https://doi.org/10.1037/pspa0000026>

Foster, C. A., Witcher, B. S., Campbell, W. K., & Green, J. D. (1998). Arousal and attraction: Evidence for automatic and controlled processes. *Journal of Personality and Social Psychology*, 74(1), 86–101. <https://doi.org/10.1037/0022-3514.74.1.86>

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology*, 83(3), 693–710. <https://doi.org/10.1037/0022-3514.83.3.693>

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All? *Journal of Personality and Social Psychology*, 86(1), 148–161. <https://doi.org/10.1037/0022-3514.86.1.148>

Suls, J. & Wan, C. K. (1989). The relation between Type A behavior and chronic emotional distress: A meta-analysis. *Journal of Personality and Social Psychology*, 57(3), 503–512. <https://doi.org/10.1037/0022-3514.57.3.503>

- Sweeney, P. D., Anderson, K., & Bailey, S. (1986). Attributional style in depression: A meta-analytic review. *Journal of Personality and Social Psychology*, 50(5), 974–991.
<https://doi.org/10.1037/0022-3514.50.5.974>
- Urban, L. M., & Miller, N. (1998). A theoretical analysis of crossed categorization effects: A meta-analysis. *Journal of Personality and Social Psychology*, 74(4), 894–908.
<https://doi.org/10.1037/0022-3514.74.4.894>
- Yeager, D. S., Krosnick, J. A., Visser, P. S., Holbrook, A. L., & Tahk, A. M. (2019). Moderation of classic social psychological effects by demographics in the U.S. adult population: New opportunities for theoretical advancement. *Journal of Personality and Social Psychology*, 117(6), e84–e99. <https://doi.org/10.1037/pspa0000171>

Aus den Analysen ausgeschlossen, aufgrund willkürlicher Eingrenzungen der Zeitspanne ($k = 2$)

- Ferguson, C. J. (2010). A meta-analysis of normal and disordered personality across the life span. *Journal of Personality and Social Psychology*, 98(4), 659–667.
<https://doi.org/10.1037/a0018770>
- Harris, M. A., & Orth, U. (2020). The link between self-esteem and social relationships: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, 119(6), 1459–1477. <https://doi.org/10.1037/pspp0000265>

Auflistung der Tabellen S1 – S4 der multiplen hierarchischen Meta-Metaregression

Tabelle S1

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen mixed-effects Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable ohne Anwendung des Filters

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Step 1				
Initialstudien ES	0.912 [0.764, 1.061]	< .001	0.003 [0.001, 0.005]	.007
Step 2				
Initialstudien ES	0.934 [0.785, 1.084]	< .001	–	–
Initialstudien <i>N</i>	< -0.001	.30	< 0.001	< .001
Step 3				
Initialstudien ES	0.880 [0.749, 1.011]	< .001	–	–
Initialstudien <i>N</i>	< -0.001	.22	< 0.001	< .001
Publikationsjahr Initialstudie	-0.002 [-0.004, < - 0.001]	.03	< 0.001	.34
Step 4				
Initialstudien ES	0.903 [0.763, 1.043]	< .001	–	–
Initialstudien <i>N</i>	< -0.001	.11	< 0.001	.001
Publikationsjahr Initialstudie	-0.003 [-0.006, < - 0.001]	.03	< -0.001	.82
Jährliche Zitationszahl der Initialstudie	0.002 [-0.002, 0.006]	.29	< 0.001	.73

Anmerkung. KI = Konfidenzintervall. LRT = Likelihood Ratio Test.

^a *k* = 36

^b *k* = 34

Tabelle S2

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen mixed-effects Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable, nur anhand publizierter Daten ohne Anwendung des Filters berechnet

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> [95% KI]	<i>p</i>	<i>b</i> [95% KI]	<i>p</i>
Step 1				
Initialstudien ES	0.510 [-0.207, 1.227]	.16	-0.004 [-0.015, 0.007]	.45
Step 2				
Initialstudien ES	0.536 [-0.271, 1.343]	.19	-0.007 [-0.018, 0.004]	.22
Initialstudien <i>N</i>	< -0.001	.89	< 0.001	.10
Step 3				
Initialstudien ES	0.519 [-0.257, 1.294]	.19	-0.010[-0.021, 0.001]	.08
Initialstudien <i>N</i>	< 0.001	.64	< 0.001	.04
Publikationsjahr Initialstudie	-0.003[-0.007, 0.001]	.13	< -0.001	.11
Step 4				
Initialstudien ES	0.501 [-0.406, 1.408]	.27	-0.009 [-0.022, 0.005]	.20
Initialstudien <i>N</i>	< 0.001 [-0.001, 0.001]	.65	< 0.001	.33
Publikationsjahr Initialstudie	-0.003 [-0.007, 0.001]	.18	< -0.001	.08
Jährliche Zitationszahl der Initialstudie	—	—	< 0.001	.82

Anmerkung. KI = Konfidenzintervall. LRT = Likelihood Ratio Test.

^a *k* = 16

^b *k* = 34

Tabelle S3

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen linearen Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable, ohne Anwendung des Filters

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Step 1				
Initialstudien ES	0.856 (0.073)	< .001	- 0.023 (0.011)	.04
Step 2				
Initialstudien ES	0.805 (0.071)	< .001	-0.024 (0.011)	.04
Initialstudien <i>N</i>	< -0.001 (< 0.001)	.07	< 0.001 (< 0.001)	.16
Step 3				
Initialstudien ES	0.845 (0.072)	< .001	-0.029 (0.010)	.005
Initialstudien <i>N</i>	< -0.001 (< 0.001)	.09	< 0.001 (< 0.001)	.06
Publikationsjahr Initialstudie	< -0.001 (0.001)	.62	< -0.001 (< 0.001)	.004
Step 4				
Initialstudien ES	0.906 (0.076)	< .001	-0.022 (0.011)	.07
Initialstudien <i>N</i>	< -0.001 (< 0.001)	.56	< 0.001 (< 0.001)	.83
Publikationsjahr Initialstudie	-0.001 (0.001)	.45	< -0.001 (< 0.001)	.02
Jährliche Zitationszahl der Initialstudie	0.001 (0.002)	.72	< 0.001 (< 0.001)	.53

^a *k* = 36

^b *k* = 34

Tabelle S4

Ergebnisse der Modelle 1 und 3 der theoriegeleiteten hierarchischen multiplen linearen Meta-Metaregression mit Crude Differences und Regressionskoeffizienten als Zielvariable, nur anhand publizierter Daten ohne Anwendung des Filters berechnet

Prädiktoren	Crude differences ^a		Regressionskoeffizienten ^b	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Step 1				
Initialstudien ES	0.493 (0.145)	.004	-0.034 (0.023)	.14
Step 2				
Initialstudien ES	0.482 (0.150)	.007	-0.033 (0.023)	.15
Initialstudien <i>N</i>	< 0.001 (< 0.001)	.55	< 0.001 (< 0.001)	.30
Step 3				
Initialstudien ES	0.482 (0.157)	.01	-0.038 (0.023)	.11
Initialstudien <i>N</i>	< 0.001 (< 0.001)	.59	< 0.001 (< 0.001)	.26
Publikationsjahr Initialstudie	< -0.001 (0.001)	.98	< -0.001 (< 0.001)	.34
Step 4				
Initialstudien ES	0.459 (0.187)	.03	-0.066 (0.025)	.02
Initialstudien <i>N</i>	< 0.001 (< 0.001)	.56	< 0.001 (< 0.001)	.22
Publikationsjahr Initialstudie	< -0.001 (0.002)	.98	-0.001 (< 0.001)	.14
Jährliche Zitationszahl der Initialstudie	-0.001 (0.004)	.90	< -0.001 (0.001)	.51

^a *k* = 16

^b *k* = 34