

# Kapitel 3: Wirkung und Bekämpfung von Desinformation aus medienpsychologischer Sicht

Autorinnen:

Birte Högden  
Prof. Dr. Nicole Krämer  
Judith Meinert  
Dr. Leonie Schaewitz

## A. *Einleitung*

Medienpsychologische Forschung widmet sich schon seit mehreren Jahrzehnten den Wirkungen von Desinformation. Obwohl das Phänomen eine höhere Aufmerksamkeit erhält seit sich Falschinformationen über Social Networks verbreiten, ist es medienpsychologisch bereits seit langem relevant, zu untersuchen, wie langfristig falsche, über die Medien rezipierte Informationen im Gedächtnis behalten werden und wie sie überschrieben und korrigiert werden können. Der im Rahmen dieser Forschung identifizierte Falschinformationseffekt zeigt dabei, dass eine Korrektur oft nicht gelingt. Während diese Erkenntnisse eher Anlass zur Sorge geben, gibt es andere medienpsychologische Forschungszweige, die insgesamt aufzeigen, dass Menschen sich im digitalen Raum kompetent verhalten und beispielsweise die zugeschriebene Glaubwürdigkeit auf Basis der Kompetenz der Quelle vornehmen. Die in diesem Kapitel vorgenommene Analyse der bisherigen Literatur und Empirie sowie darauf aufbauende weitere eigene Studien sollen somit im Folgenden einen differenzierten Blick darauf ermöglichen, welche Gefahren Falschinformationen darstellen und wie diesen begegnet werden kann.

## B. *Überblick über die Forschungslandschaft*

Medienpsychologische Forschung widmet sich traditionell vor allem der Frage, welche kurz- und langfristigen Auswirkungen Desinformationen auf

die Einstellung der Rezipienten haben. Seit Falschinformationen allerdings nicht mehr nur ein Phänomen sind, das gelegentlich (zum Beispiel aufgrund ungenauer Recherchen) in massenmedialen Kontexten auftaucht, sondern Social Networks betrifft, werden vermehrt auch Auswirkungen auf Verhalten untersucht – im Sinne der Frage, inwieweit Rezipientinnen durch ihr Weiterleitungsverhalten für die Verbreitung von Desinformation sorgen. Seitdem liegt der Fokus der Forschung auf Medien wie Facebook und Twitter. Da die Bekämpfung der Verbreitung von und Einflussnahme durch Desinformation ein Anliegen sowohl der Bevölkerung als auch von Expertenseite ist, wurden verschiedene Interventionsmöglichkeiten und deren Wirksamkeit untersucht. Im Folgenden wird ein Überblick über die bisherige Forschung aus medienpsychologischer Sicht und deren Ergebnisse mit Blick auf die genannten Aspekte gegeben.

## I. Wirkung von Desinformation auf Einstellungen und Verhalten

### 1. Einfluss auf Einstellungen

Die Tatsache, dass jeder Online-Nutzende Inhalte produzieren und potentiell mit einer unbegrenzten Menge an anderen Nutzenden teilen kann, führt zu hohen Mengen an Informationen, mit denen Rezipientinnen konfrontiert werden. Im Regelfall kann als ein Ziel der Internetnutzung angenommen werden, dass Menschen relevante und korrekte Informationen suchen, um Entscheidungen daraus abzuleiten (zu Gesundheitsfragen wie ob bei gegebenen Symptomen ein Arzt aufgesucht werden sollte, zu Erziehungsfragen wie der potenziellen Gefahr durch zu starke Mediennutzung, zu Umweltfragen wie dem tatsächlichen Zustand des Weltklimas und den abzuleitenden Schlussfolgerungen, Co2 vermeidend zu leben, etc.). Ein Faktor bei der Auswahl der zu rezipierenden Nachrichten aus der Vielzahl an Informationen ist dabei, als wie glaubwürdig eine Nachricht empfunden wird.

Bereits in klassischen Studien zur Glaubwürdigkeit und Persuasion konnte identifiziert werden, dass die Quelle der Botschaft einen starken Einfluss auf Einstellungsbildung der Zuhörerschaft nimmt (Hovland, Janis & Kelley, 1953; Hovland & Weiss, 1951). Das liegt daran, dass die Quelle meistens nicht nur der salienteste Hinweisreiz ist, sondern Rezipienten darüber hinaus daran gewöhnt sind, die Quelle einer Information heranziehen, um deren Nützlichkeit einzuschätzen (Sundar, 2008). Diese Erkenntnisse zum Einfluss von Quellenglaubwürdigkeit konnten auch auf den Social-Network-Kontext

transferiert werden, da sich gezeigt hat, dass Nutzende beobachtbare Aspekte der Quelle heranziehen, um die Validität einer Botschaft einzuschätzen (Metzger, Flanagin & Medders, 2010). Dabei zeigt sich, dass Rezipienten durchaus kompetent vorgehen, indem sie die Nachrichten von Social-Network-Autoren (in diesem Fall Bloggerinnen) dann eher zum Lesen auswählen, wenn es Hinweise auf eine Expertise hinsichtlich des gegebenen Themenbereiches gibt (Winter & Krämer, 2012; vgl. auch Reputationsheuristik, Metzger et al., 2010). Dieser Effekt ist bei Personen, die ein hohes Bedürfnis haben, über Sachverhalte gründlich nachzudenken, besonders ausgeprägt.

Allerdings sind Rezipienten aufgrund von limitierten kognitiven Kapazitäten nicht jederzeit in der Lage, jede Information in einer elaborierten Form zu verarbeiten (Lang, 2000). Dies wird beispielsweise nachvollzogen im Rahmen verschiedener zwei-Wege-Modelle der Informationsverarbeitung und Einstellungsänderung. Im Elaboration Likelihood Modell (Petty & Cacioppo, 1986) und im Heuristic Systematic Modell (Chaiken, 1989) werden zwei unterschiedliche Wege beschrieben, über die Informationen rezipiert und weitergehend verarbeitet werden und dementsprechend Einfluss auf Einstellungen und Verhalten nehmen können. In Abhängigkeit von Fähigkeit und Motivation (und thematischem oder situationalem Involvement) wird entweder die zentrale Route, die eine elaborierte Evaluation und Verarbeitung aller eingehenden Informationen beinhaltet, oder die periphere Route gewählt, die eine eher oberflächliche Verarbeitung beschreibt, während der einfache Hinweisreize oder heuristische Regeln Eingang in die Einstellungsbildung finden können.

Im Zuge einer peripheren Verarbeitung erhöht sich die Wahrscheinlichkeit, dass manipulierte oder falsche Informationen, die durch andere Nutzende empfohlen oder beworben werden (beispielsweise auf Social-Network-Plattformen durch eine hohe Anzahl Likes und Shares, die als aggregierte Summe unter jedem Beitrag angezeigt wird), als glaubwürdig empfunden und daraus resultierend inhaltlich angenommen werden. Dieser Effekt begründet sich durch die implizite Annahme, dass „etwas, das von anderen als gut empfunden wird, von einem selbst auch als gut beurteilt werden kann“. Diese Annahme wird als Bandwagon Heuristic (Sundar, 2008) beschrieben, die bereits für die Beurteilung von Online Reviews als einflussreich bestätigt werden konnte (Sundar, Oeldorf-Hirsch & Xu, 2008). Auch für das Vertrauen in inkorrekte Informationen kann dieser Effekt als möglicher Erklärungsansatz herangezogen werden.

Doch auch unabhängig von diesen Tendenzen, sich bei fehlender Fähigkeit oder fehlender Motivation mit den geposteten Nachrichten nicht tiefgehend

auseinanderzusetzen, lassen sich subtile kognitive Effekte nachweisen, denen der Mensch generell unterliegt. In diesem Sinne haben Pennycook, Cannon und Rand (2018) herausgefunden, dass die Tendenz, Desinformationen zu glauben, durch den Eindruck von Vertrautheit infolge einer vorherigen Rezeption gefördert wird. Durch ein gelungenes experimentalpsychologisches Design, in dem Rezipienten manche (Falsch)Nachrichten mehrfach und andere nur einmal präsentiert wurden, konnte nachgewiesen werden, dass falsche Nachrichten dann eher geglaubt werden, wenn man sie mehrfach sieht. Es ist anzunehmen, dass dieser Prozess der Fehlattribution unbewusst stattfindet, da auch Warnmeldungen keine Änderung in der Bewertung der Glaubwürdigkeit von bereits gesehenen Inhalten bewirkten. So wurden Fake-News-Überschriften, die zuvor bereits den Versuchspersonen präsentiert wurden, nach einer zweiten Rezeption als glaubwürdiger eingeschätzt, auch wenn sie von einer Warnung begleitet wurden. Fake-News-Überschriften, die zum ersten Mal und ohne Warnmeldung gezeigt wurden, wurden hingegen als weniger glaubwürdig bewertet.

Ein weiterer kognitiver Fehler, dem alle Nutzenden unterliegen, bezieht sich darauf, dass Menschen nach Konsistenz in ihren Einstellungen, ihrem Verhalten und ihrer Selbstwahrnehmung streben und dadurch Informationen bevorzugen, die sich mit ihrer bereits bestehenden Meinung decken. Werden Nachrichtenkonsumierende vor die Wahl zwischen unterschiedlichen Artikeln und Quellen gestellt, dann entscheiden sie sich für diejenigen, die den eigenen Einstellungen und Ansichten am ehesten entsprechen und somit die geringste Wahrscheinlichkeit einer unangenehmen, Stress verursachenden kognitiven Dissonanz beinhalten (vgl. cognitive dissonance theory, Festinger, 1962). Dieses Phänomen wird auch als Confirmation Bias bezeichnet (Nickerson, 1998). Neben der Tatsache, dass diese Tendenz als eine mögliche Ursache für das Zustandekommen sogenannter Filterblasen diskutiert wird (Lazer et al., 2017), wird auch argumentiert, dass der Confirmation Bias die Wirkung von Falschinformationen verstärken kann – wenn Nutzende eine Voreinstellung haben, die dazu führt, dass sie die Falschinformationen glauben wollen. Der Confirmation Bias wird sogar wirksam, wenn Desinformation als falsch enttarnt wird (Tan & Ang, 2017; Ott, 2017).

## 2. Einfluss auf Verhalten – Weiterleitung von Desinformation

In Bezug auf Desinformationen in Social Networks ist auch ein spezifischer Einfluss auf Verhalten äußerst relevant. Desinformationen verbreiten sich

nämlich dann besonders rasant, wenn sie von Menschen weitergeleitet werden. Durch die Merkmale von Social Networks wie hohe Vernetzung und Einfachheit sowie Schnelligkeit der Kommunikation (bedingt durch einfache Mechanismen wie liken und sharen: Aktionen, die mit einem Klick möglich sind und potenziell eine große Reichweite haben können; vgl. das Konzept der Mass Interpersonal Persuasion von Fogg, 2008) wird begründet, dass auch falsche Nachrichten sich durch Tausende von Likes schnell viral verbreiten. Das wiederum kann im Weiteren eine irreführende Illusion von Vertrauen in bestimmte, vielfach geteilte (und damit, in Anlehnung an gängige Social-Network-Kommunikationskonventionen, empfohlene) Inhalte erzeugen (Elyashar, Bendahan & Puzis, 2017). In diesem Zusammenhang zeigt eine Studie von Vosoughi, Roy und Aral (2018), dass sich Falschmeldungen in Social Networks sogar schneller und häufiger verbreiten als wahre Meldungen, besonders bei politischen Themen. Die Top 1 % der beliebtesten Falschnachrichten erreichten demnach 1.000 bis 100.000 Menschen, während wahre Meldungen selten mehr als 1.000 Menschen erreichten. Die zentrale Erkenntnis der Studie besteht allerdings darin, dass diese Verbreitung eher durch Menschen als durch Bots entsteht. Dabei stellen die Autoren die Hypothese auf, dass insbesondere die Neuartigkeit der Nachrichten ein relevanter Faktor für die Weiterleitung darstellt. Dies sollte allerdings in weiteren Studien geprüft werden, denn weitere Einflussfaktoren könnten etwa der Negativity Bias (die Tatsache, dass negative Nachrichten mehr Aufmerksamkeit auf sich ziehen) oder die Emotionalität der Nachrichten sein. Oftmals werden Falschnachrichten so konzipiert und kommuniziert, dass sie die Emotionen und Gefühle der Rezipientinnen ansprechen anstatt Fakten zu transportieren. Während der US-Wahlen 2016 hat sich beispielsweise gezeigt, dass emotional dargestellte Geschichten gegenüber Fakten bevorzugt wurden (Gross, 2017). Auf der anderen Seite muss gefragt werden, wie viel Einfluss die Inhalte der Nachrichten überhaupt haben. Andere Studien finden nämlich, dass 59 % der Links auf Twitter geteilt werden, ohne aufgerufen worden zu sein (Gabelkov et al., 2016).

Der Einfluss von Falschnachrichten kann sich gegebenenfalls dadurch verstärken, dass Social-Network-Kanäle oftmals als einzige Quelle für den Konsum von News und politischen Inhalten herangezogen werden, ohne dass weitere Quellen wie Zeitungen und Magazine konsultiert werden. Daraus resultierend sind Rezipienten heutzutage (bzw. im Social-Network-Kontext) weitaus anfälliger für Manipulation und falsche Informationen.

## II. Einfluss von Interventionen

Wie eine aktuelle Bevölkerungsumfrage zeigt, wünschen sich viele Deutsche, dass Desinformationen besser und flächendeckender bekämpft werden (PricewaterhouseCoopers GmbH, 2019). Obwohl die meisten Social-Networks-Nutzende zuerst einmal versuchen, Desinformation selbstständig zu erkennen und sich so vor einem Konsum und entsprechender Beeinflussung zu schützen, wünschen sie sich darüber hinaus Schutz durch externe Quellen zur Erkennung von Falschinformation (vgl. Tandoc et al., 2017). Als geeignete Gegenmaßnahmen werden dabei vor allem eine Löschung von Inhalten durch Netzwerkbetreiber und die Aufklärung der Bevölkerung durch die Regierung genannt (PricewaterhouseCoopers GmbH, 2019). Des Weiteren sehen Bürgerinnen die Medien in der Pflicht, Falschinformationen zu identifizieren und zu korrigieren. Tatsächlich wird seit einigen Jahren verstärkt auch psychologische Forschung zu geeigneten Gegenmaßnahmen und deren Wirksamkeit durchgeführt. In den folgenden Abschnitten wird eine Einordnung von Interventionsmöglichkeiten unter psychologischen Gesichtspunkten vorgenommen sowie jeweils eine Auswahl an konkreten Maßnahmen mit empirischen Ergebnissen zu deren Wirkung gegeben.

### 1. Nachhaltigkeit der Auswirkungen von Desinformation: Falschinformationseffekt

Bereits vor dem Aufkommen von Social Networks existierten Desinformationen, die man mit Hilfe von Korrekturen zu bekämpfen versuchte. In diesem Zusammenhang wurde allerdings schon in den 1970er Jahren in der Psychologie der sogenannte Falschinformationseffekt beschrieben, in dessen Rahmen erkannt wurde, dass sich einmal rezipierte Information nicht einfach korrigieren lässt. Zahlreiche Studien belegen, dass auch nach Aufklärung über die falsche Information und auch unter Bedingungen, in denen Probanden sich eigentlich darüber im Klaren sind, dass die ursprünglich gegebene Information falsch oder wenig vertrauenswürdig ist (Ross, Lepper & Hubbard, 1975) und sie motiviert sind, die neue Information zu glauben, die alte, falsche Information im Gedächtnis bleibt und abgerufen wird. Es handelt sich daher nicht um ein Problem unzureichender Beachtung der neuen Information oder ein motivationales Problem (zum Beispiel der neuen korrekten Information keinen Glauben schenken zu wollen), sondern um einen rein kognitiv erklärbaren Effekt, der auf die Funktionsweise des Gehirns zurückzuführen

ren ist. Die ursprüngliche, falsche Information wird nämlich in bestehende Wissensbestände plausibel eingebaut und mit dieser so vernetzt, dass ein Überschreiben durch die neue Information nur in Ausnahmefällen gelingt (Graesser, Singer & Trabasso, 1994). Insbesondere, wenn Erklärungen für die Sachverhalte gefunden werden, wird an diesen auch noch festgehalten, wenn über die Falschheit informiert wurde (Anderson, Lepper & Ross, 1980). Nur wenn die neue, korrigierte Information die alte Information im gebildeten Modell problemlos ersetzen kann, tritt kein Falschinformationseffekt auf (Johnson & Seifert, 1994). Ein essentielles Ergebnis in der Forschung zum Thema Falschinformationseffekt ist somit, dass die Identifizierung und Korrektur von Falschnachrichten aufgrund von subtilen kognitiven Effekten nicht zwangsläufig zu einer Meinungsänderung in Richtung der korrekten Information führen, da die Veränderung eines bereits gebildeten und mitunter durch ähnliche Informationen bestätigten Eindrucks schwierig ist. Gegenmaßnahmen sind zusätzlich insbesondere im Bereich von Social Networks durch das dynamische Umfeld erschwert, denn die Informationen erscheinen nicht nur an einer Stelle und können gegebenenfalls nicht überall mit der Korrektur versehen werden. Denn sobald eine Information online veröffentlicht ist, kann angenommen werden, dass „sie bereits Beine hatte“ (“the story already had legs”, Berghel, 2017).

## 2. Gegenmaßnahmen

Eine Aufklärung der Gesellschaft über Desinformation stellt den ersten Schritt zur Bekämpfung von Falschinformationsverbreitung dar. So konnte bereits gezeigt werden, dass eine allgemeine Warnung vor Desinformation hilft, um die Verbreitung von Falschinformationen zu mindern (Clayton et al., 2019). Insbesondere ältere Menschen scheinen Bedarf für solche Maßnahmen zu haben, da gezeigt wurde, dass sie auf Social-Network-Plattformen häufiger Falschinformationen teilen als jüngere Bevölkerungsgruppen (Guess, Nagler & Tucker, 2019). Da sich außerdem gezeigt hat, dass die Fähigkeit zum analytischen und strukturierten Denken das Erkennen von Falschinformationen begünstigt (Pennycook & Rand, 2018), sollten Maßnahmen getroffen werden, um diese Art der Informationsverarbeitung zu fördern.

Allerdings ist das Erkennen von Falschinformationen für Laien selbst mit Hintergrundwissen und trotz einer generellen Sensibilisierung schwierig. Daher müssen geeignete Gegenmaßnahmen eingesetzt werden, um Bürgerinnen vor dem Einfluss von Falschinformation zu schützen. Diese Interventionen

bestehen in drei wesentlichen Maßnahmen, die getrennt, aber auch in Kombination angewendet werden können. Zum einen gibt es die Möglichkeit, Desinformationen zu löschen, um sie von einem weiteren Konsum beziehungsweise einer umfassenderen Verbreitung auszuschließen. Zum anderen können Warnhinweise installiert werden, die Inhalte als desinformierend oder zumindest teilweise fehlleitend kennzeichnen. Außerdem kann nach der erfolgreichen Erkennung von Falschinformationen eine Korrektur in Form einer Gegendarstellung vorgenommen werden.

### Löschung von Desinformation

Die erste bekannte Maßnahme, um Fake News aktiv entgegenzuwirken, besteht in der Löschung solcher Meldungen. Natürlich müssen dafür desinformierende Inhalte zuerst einmal erkannt werden. Dann stehen Betreiber von Netzwerkseiten und anderen Social-Network-Portalen in der Pflicht, diese Inhalte zu löschen. Da bei der Menge an Informationen im Netz eine flächendeckende manuelle Sichtung ausgeschlossen ist, werden automatisierte Verfahren erarbeitet (vgl. Kapitel 4). Da eine zweifelsfreie Erkennung von anderen Inhalten wie beispielsweise Satire oder Ironie nicht einfach ist, ist allerdings eine direkte Löschung häufig problematisch.

Zudem kann eine Löschung immer erst dann passieren, wenn ein Beitrag bereits veröffentlicht wurde. Besser wäre es, wenn effizientere und effektivere Maßnahmen verfügbar wären, die die Verbreitung von Falschnachrichten von vornherein verhindern. Wenn Informationen erst einmal im Internet geteilt werden, werden sie zwangsläufig von vielen Menschen konsumiert. Eine anschließende Löschung kann also nicht verhindern, dass Desinformation trotzdem schon von vielen Rezipientinnen gelesen und verinnerlicht wurden.

Darüber hinaus erscheint das Bekämpfen von Falschmeldungen am wirksamsten zu sein, wenn eine eindrückliche Gegendarstellung mit wahren Argumenten und Fakten geboten wird (s.u.). Fehlt eine solche, wird die ursprüngliche, fehlerhafte Information nicht überschrieben und bleibt trotzdem im Wissensbestand erhalten. Löschungen sind also nur dann sinnvoll, wenn diese unmittelbar nach der Veröffentlichung eines Postings geschehen. So kann die Anzahl an Rezipienten eingeschränkt und ein Weiterleiten von Falschnachrichten verhindert werden. Besser ist es, wenn Falschmeldungen gar nicht erst veröffentlicht werden oder auch Quellen, die für das Verbreiten von Falschinformationen bekannt sind, blockiert werden (vgl. Lazar et al., 2017).

## Warnhinweise für Falschinformationen

Warnhinweise sind allgegenwärtig und finden Anwendung in verschiedensten Kontexten sowie Disziplinen. So finden sie sich im Straßenverkehr, am Arbeitsplatz oder auf Verpackungen mit gesundheitsschädlichem Inhalt. Sie dienen dazu, die allgemeine Sicherheit zu unterstützen, Informationen über mögliche Gefahren bereitzustellen, das Verhalten von Personen zu beeinflussen oder eine Gefahr ins Gedächtnis zurückzurufen (Laughery & Wogalter, 2006). Eine Voraussetzung für die erfolgreiche Wirkung von Hinweisen ist zuerst einmal, dass sie als solche erkannt werden. Deswegen sollten bestimmte Gestaltungsrichtlinien, die ihren Ursprung in der Wahrnehmungspsychologie haben, eingehalten werden.

Damit Warnungen wirken können, muss deren semantische Bedeutung für alle Menschen klar erkenntlich sein. So ist es in der Regel hilfreich, nicht ausschließlich auf den Einsatz von Symbolen oder Text zu vertrauen, sondern eine Kombination von beidem anzustreben (Haynes, 2016). Ein eindeutiges Icon für die Identifikation von Desinformation könnte also mit einer entsprechenden Erklärung beziehungsweise mit einem einzigen Signalwort (z. B. Gefahr oder Warnung) kombiniert werden (vgl. Laughery & Wogalter, 2006). Außerdem sind eindruckliche Farben (z. B. signalrot) eine gute Ergänzung, da sie Aufmerksamkeit auf sich ziehen und sich von der Umgebung absetzen (Laughery & Wogalter, 2006). Weiterhin ist es wichtig, ein Warnsymbol stets in der unmittelbaren Nähe der Gefahr anzubringen (Wogalter, Conzola & Smith-Jackson, 2002). Noch besser wirken Warnungen, wenn damit interagiert werden muss, sie also aktiv beseitigt werden müssen, um ein Produkt o.ä. zu nutzen (Laughery & Wogalter, 2006). Im Kontext von Digitaler Desinformation könnte das ein Pop-Up Fenster mit einer Warnung vor Falschinformationen sein, das weggeklickt werden muss.

Einige Studien haben sich bereits mit spezifischen Warnhinweisen beschäftigt und diese evaluiert. Die Ergebnisse zu deren Wirkungen sind allerdings teilweise widersprüchlich. So scheinen laut Geary (2017) und Gao, Xiao, Karahalios und Fu (2018) sogenannte Credibility Cues auf Facebook nicht wirksam zu sein. Diese Cues wurden direkt an Desinformations-Postings angehängt, um diese als Falschinformation zu kennzeichnen. Dieses Vorgehen entspricht also den empfohlenen Gestaltungsempfehlungen für Warnhinweise, da Gefahr und Warnung direkt nebeneinander sichtbar waren. Allerdings wurden sie dennoch nicht richtig wahrgenommen, da Nutzende typischerweise schnell durch den Facebook-News-Feed scrollen und so nur wenige der verfügbaren Informationen verarbeiten. Des Weiteren kann das Kennzeich-

nen von Beiträgen dann problematisch sein, wenn nur ein Teil der Postings als falsch gekennzeichnet wird. Dies impliziert, dass der Inhalt der übrigen, nicht gekennzeichneten Beiträge wahr ist, obwohl auch diese falsche Informationen enthalten können (implied truth effect, Pennycook et al., 2019).

Andererseits konnte gezeigt werden, dass Markierungen, die falschen („rated false“) oder zumindest umstrittenen („disputed“) Inhalt anzeigen, wirksam gegen Desinformation sind. Sie sorgen dafür, dass Überschriften desinformierender Artikel als weniger akkurat eingeschätzt werden. Die Wirksamkeit dieser Markierungen ist sogar dann vorhanden, wenn die im kritisierten Artikel ausgedrückte Meinung der der Probandinnen entsprach (Clayton et al., 2019). Außerdem sind sogenannte Truth Scales („Ampeln“, die den Wahrheitsgehalt einer Nachricht anzeigen) laut Amazeen, Thorson, Muddiman und Graves (2018) wirkungsvoll, um Fake News als falsch zu kennzeichnen. Darüber hinaus sorgen sie sogar nachhaltig dafür, dass Falschnachrichten langfristig als solche abgespeichert werden. Allerdings müssen Truth Scales zuerst einmal von Lesern gefunden werden, da sie meist auf speziellen Fact-Checking Webseiten zur Verfügung stehen. Nutzende müssen also aktiv werden, um sich über Falschinformationen zu informieren.

Abschließend ist es wichtig zu bemerken, dass Warnhinweise generell sehr bedacht gestaltet werden sollten. Verschiedene Studien haben gezeigt, dass eine fehlerhafte Gestaltung solcher Hinweise sogar dazu führen kann, dass Falschnachrichten als wahre Nachrichten abgespeichert werden (Skurnik et al., 2005; Wogalter et al., 2002). Sie erregen dann zwar Aufmerksamkeit, korrigieren den Fehleindruck aber nicht ausreichend und verankern den Inhalt der irreführenden Nachricht sogar stärker im Gedächtnis. Deswegen sollte die Wirksamkeit spezifischer Warnhinweise stets empirisch getestet werden.

### Korrekturen und Gegendarstellungen von Falschinformationen

Die letzte und wirksamste Möglichkeit, Fake News zu bekämpfen, besteht darin, diese an geeigneter Stelle zu korrigieren. Das Verhindern einer Verbreitung von Falschinformationen ist zwar das beste Mittel, um Falschnachrichten zu begegnen, doch ist dies nicht immer möglich. Korrekturen können hingegen so gestaltet werden, dass sie eine hohe Chance auf Erfolg haben. Im Folgenden werden verschiedene Möglichkeiten für eine effektive Gestaltung sowie diverse Probleme, die sich bei der Verbreitung von Korrekturen ergeben können, beschrieben.

Bisher konnte gezeigt werden, dass das Eindämmen einer Falschinformationsverbreitung wirksamer ist als jede nachträgliche Korrektur. So gelangen Korrekturen beispielsweise längst nicht an alle Leser der ursprünglichen Falschnachricht. Darüber hinaus verbreiten sich Korrekturen generell deutlich weniger als Desinformationen.

Da hauptsächlich Menschen - und nicht etwa Social Bots - für die Verbreitung von Falschnachrichten verantwortlich sind (Vosoughi et al., 2018), sollte eine Weiterleitung von Falschnachrichten durch Social-Networks-Nutzende bekämpft werden. Eine Korrektur des Beitrags ist nur dann sinnvoll, wenn diese den ursprünglichen Link zum desinformierenden Nachrichtenartikel nicht enthält. Algorithmen auf sozialen Netzwerken erkennen, wenn ein Link weiterhin geteilt wird, und empfehlen entsprechende Inhalte an andere Nutzende weiter. Ein viel gelesener Beitrag taucht so beispielsweise vermehrt im Facebook-News-Feed auf. In ähnlicher Weise ist es auch nicht empfehlenswert, eine Information als fehlleitend zu markieren und auf dem eigenen Social-Network-Profil zu teilen, da sie von anderen Social-Network-Nutzenden trotzdem als wahr abgespeichert werden könnte. Social-Network-Nutzende nehmen Informationen häufig nur am Rande wahr und könnten so einen Hinweis auf Falschnachrichten übersehen. Besonders hilfreich scheint es hingegen zu sein, wenn Freunde in Social Networks darauf hinweisen, wenn jemand Falschnachrichten (auch ohne bewusste Intention) geteilt hat. Die Akzeptanz einer solchen Korrektur ist höher als die einer unbekannt-Quelle. Außerdem werden Richtigstellungen durch Quellen, die mit der eigenen politischen Meinung übereinstimmen, eher angenommen und verinnerlicht (Lazer et al., 2017). Bemerkenswert ist trotzdem, dass es sich generell schwierig gestaltet, eine Falschinformation zu überschreiben, vor allem, wenn diese mit vorherigen Einstellungen einer Person übereinstimmt (Nyhan & Reifler, 2010; vgl. Falschinformationseffekt, Kapitel 3, B.I.).

Daher ist eine Aufgabe für Fact-Checking Organisationen oder Journalistinnen, ansprechende Narrative für ihre Korrekturen zu finden. Diese können dazu beitragen, dass sich korrigierende Informationen schnell und weit verbreiten, da sie das Interesse und die Aufmerksamkeit von Lesern wecken. Darüber hinaus kann eine involvierende Geschichte, in der korrekte Alternativerklärungen und Hintergrundinformationen geboten werden, dafür sorgen, dass sich richtige Information fester in den vorhandenen Wissensstand einfügen.

Bei der Widerlegung von Fake News sollten generelle Regeln des sogenannten „Debunking“ eingesetzt werden, um diese wirksam zu gestalten (Cook & Lewandowsky, 2011). Das Berichten von Fakten sollte bei einer

Korrekturmeldung generell im Vordergrund stehen, komplizierte Formulierungen sollten vermieden werden und Alternativerklärungen für berichtete Ereignisse sollten enthalten sein. Nyhan und Reifler (2015) konnten zeigen, dass es nicht ausreicht, eine Desinformation über das Auftreten eines Ereignisses lediglich zu negieren. Leserinnen benötigen eine kausale Begründung, um die korrekte Information abzuspeichern. Außerdem sollten unnötige Wiederholungen von Falschnachrichten vermieden werden, um sie nicht im Wissensnetz zu festigen. Allerdings wurde gezeigt, dass durch eine explizite Bezugnahme und Wiederholung der ursprünglichen Falschinformation innerhalb einer korrigierenden Gegendarstellung das Vertrauen in die Falschinformation signifikant stärker gesenkt werden konnte als durch eine Korrekturdarstellung ohne Wiederholung (Ecker, Hogan & Lewandowsky, 2017). Das heißt, dass solange beim Verfassen einer Gegendarstellung der Fokus auf der Bestätigung von Tatsachen statt auf der Widerlegung von Mythen liegt, eine Wiederholung der ursprünglichen Falschinformation ihre Wirkung sogar zu reduzieren vermag (Ecker, Hogan & Lewandowsky, 2017).

Teilweise kann eine Korrektur auch dann besonders wirksam sein, wenn sie in Verbindung mit Warnhinweisen oder in einer speziellen Darstellungsform veröffentlicht wird. Grafische Wahrheitsskalen (Truth Scales, s.o.) können korrigierende Informationen unterstützen und verfestigen (Amazeen et al., 2018). Auch erklärende Texte sind in der Lage Fehlinformationen erfolgreich zu widerlegen (Amazeen et al., 2018). Des Weiteren können Videos als Medium eingesetzt werden, um die Aufmerksamkeit für korrekte Informationen zu steigern und eine mögliche Verwirrung über Ereignisse zu verringern (Young, Jamieson, Poulsen & Goldring, 2017).

### *C. Zusammenfassung eigener Forschung*

Aufbauend auf den oben zusammengefassten Erkenntnissen haben wir durch eigene Studien zu einer Verbreiterung der Wissensbasis beigetragen. Zuerst wurde untersucht unter welchen Bedingungen Falschinformationen einen Einfluss haben können. Hierzu wurden Merkmale der Nachricht, der Quelle und der Rezipientinnen betrachtet. Außerdem wurde erforscht, wie Interventionen gestaltet werden sollten, um eine möglichst wirksame Bekämpfung von Falschinformationen zu erreichen.

## I. Unter welchen Bedingungen ist Desinformation einflussreich?

Bevor die Gestaltung verschiedener Interventionsmöglichkeiten und deren Wirkungsweise beleuchtet werden, wurden Grundlagen für die Implementierung solcher Gegenmaßnahmen untersucht. Dazu wurden vor allem spezifische Eigenschaften von Nachrichten und deren Einfluss auf Einstellungen untersucht. Ein weiterer Fokus lag auf Charakteristika der Quellen, die Falschnachrichten verbreiten, sowie auf den Persönlichkeitsmerkmalen von Rezipienten.

### 1. Einfluss von Nachrichtenmerkmalen auf resultierende Einstellungen

Um zu adressieren und zu überprüfen, inwieweit verschiedene Nachrichtenmerkmale von Falschinformationen die wahrgenommene Glaubwürdigkeit von fabrizierten Online-Nachrichten und die Weiterverbreitungsintentionen beeinflussen, wurde eine experimentelle Studie konzipiert und durchgeführt. In dieser wurde den Studienteilnehmenden ein Online-Nachrichtenartikel präsentiert, der eines der folgenden fünf Merkmale von Desinformation aufwies: reißerische Formulierungen, Inkonsistenzen, Subjektivität, unglaubliche Quelle oder manipuliertes Bild. In einer Kontrollbedingung wurde der Basis-Text ohne eine dieser Merkmalsvariationen präsentiert, sodass es insgesamt sechs verschiedene Versuchsbedingungen gab. Um Erkenntnisse zu verschiedenen Themenbereichen zu gewinnen, wurden zwei verschiedene Nachrichtenartikel (Kriminalität und Pflege) verwendet. Gemessen wurde, als wie glaubwürdig und akkurat die Teilnehmenden den Artikel einschätzten und mit welcher Wahrscheinlichkeit sie ihn weiterleiten würden. Insgesamt nahmen 294 Personen an der Studie teil. Die Ergebnisse des Experiments zeigen, dass keines der untersuchten Nachrichtenmerkmale (z. B. reißerische Formulierungen) in besonderem Maße dazu führt, dass Desinformation weniger glaubwürdig bewertet wird oder eher weitergeleitet wird. Generell begegneten die Probandinnen den Nachrichten aber mit eher großer Skepsis.

In zwei weiteren experimentellen Studien wurde die Valenz der Nachricht (positiv / negativ) auf die Wirkung einer politischen Desinformation untersucht. Dafür wurde eine positive oder eine negative Falschnachricht über einen deutschen Politiker gezeigt. Dazu wurden verschiedene positive (z. B. Knochenmarkspende, Ausrichtung einer Benefizgala) und negative (z. B. Steuerhinterziehung, Schlägerei) Falschnachrichten generiert, vorgetestet und randomisiert im Experiment präsentiert. Gemessen wurde die Bewer-

tung des Politikers, vor und nachdem die Probanden die Information erhielten, dass die Nachricht erfunden war. In der ersten Studie wurde diese Korrektur durch die Versuchsleiterin in Form einer mündlichen Aufklärung vorgenommen. Im zweiten Szenario bekamen Probandinnen eine Richtigstellung als Meldung einer Fact-Checking Organisation zu sehen.

Die Ergebnisse zeigen, dass positive Falschinformationen glaubwürdiger bewertet werden als negative, negative allerdings brisanter wahrgenommen werden. Die Bewertung des Politikers wird marginal durch die Valenz der Nachricht beeinflusst, sodass eine negative Nachricht zu einer schlechteren Bewertung führt. Es gibt somit Hinweise darauf, dass negative Falschmeldungen folgenschwerer sein könnten, dies muss aber in weiteren Studien geprüft werden.

## 2. Einfluss von Attributen der Quelle

Des Weiteren wurde der Einfluss der Informationsquelle einer in sozialen Medien geteilten Nachricht untersucht. Der genaue Fokus lag hierbei auf dem Einfluss dem Vermittler einer Botschaft mit besonderer Berücksichtigung seiner Beziehung zur Rezipientin. Um die spezifische Umgebung von Social-Networks-Websites bei der Frage nach der Wirkung online geteilter Falschinformationen zu berücksichtigen, wurde untersucht, ob es einen Unterschied macht, wenn eine Falschinformation von einer guten Freundin oder einem entfernten Bekannten auf Facebook geteilt wurde. Dazu wurden den Studienteilnehmenden individualisierte Nachrichten-Posts gezeigt, die augenscheinlich von einem ihrer engen oder losen Facebook-Freundinnen geteilt wurden. Dabei zeigte sich, dass die Beziehungsstärke zum teilenden Freund keinen Einfluss auf die Bewertung der Nachricht oder die Bewertung des Politikers hatte.

## 3. Einfluss von Personenmerkmalen

Um den Einfluss der Merkmale des Empfängers einer Botschaft zu untersuchen, wurden verschiedene Personenvariablen, wie z. B. die Tendenz zu analytischem Denken (Need for Cognition) und die Einstellung zum Thema der Nachricht, im Rahmen einer experimentellen Studie erfasst. Weiterhin wurde untersucht, als wie glaubwürdig und akkurat verschiedene Nachricht-

tenartikel von Studienteilnehmenden in Abhängigkeit ihrer Ausprägung der Persönlichkeitsvariablen bewertet wurden.

Individuelle Prädispositionen scheinen entscheidend für die Glaubwürdigkeitswahrnehmung sowie für die Einschätzung der Akkuratheit von Desinformation zu sein. Die Ergebnisse deuten darauf hin, dass insbesondere Personen mit einer ausgeprägten Neigung zum analytischen Denken (Need for Cognition) Desinformation als weniger akkurat einstufen und dass die eigene Meinung zum Thema der Nachricht ein starker Prädiktor für die wahrgenommene Glaubwürdigkeit von Falschnachrichten ist.

## II. Wirkung von Gegenmaßnahmen

Nachdem Einflussfaktoren für die Wirkung von Falschinformationen evaluiert und analysiert wurden, wurden konkrete Interventionsmechanismen entwickelt. Dazu wurde zunächst untersucht, wie die Gestaltung von Warnhinweisen aussehen sollte. Anschließend wurde die Wirksamkeit von Falschmeldungskorrekturen ergründet sowie der Einfluss von Reaktionen anderer Nutzender erforscht.

### 1. Studien zu Warnhinweisen

Bezüglich der Wirkung von Warnhinweisen wurde überprüft, wie Desinformation verarbeitet wird, wenn sie als fehlleitend markiert wird. Um sich dieser Frage zu nähern, wurde zunächst eine Voruntersuchung zur Wahrnehmung von Warnhinweisen durchgeführt. Dabei wurden den Teilnehmenden in einem Online-Experiment verschiedene Arten von Warnhinweisen präsentiert und anschließend erhoben, wie die Nützlichkeit wahrgenommen wird. Variiert wurde einerseits die Grafik des Hinweises (Warndreieck vs. Prozentuale Skala) und andererseits die textuelle Formulierung („als falsch bewertet“ vs. „umstritten“, vgl. Clayton et al., 2019). Die Ergebnisse der Studie zeigen, dass ein „einfaches“ Warndreieck nützlicher und glaubwürdiger bewertet wurde als eine prozentuale Warnskala und zu einer höheren Akzeptanz führte. Die textuelle Variation hatte keinen Effekt auf die wahrgenommene Nützlichkeit oder Glaubwürdigkeit der Warnung. Als positive Effekte von Warnhinweisen merkten die Teilnehmenden an, dass sie zu einer intensiveren Beschäftigung mit dem Thema der Meldung führen würden und bestimmte Personengruppen, wie Kinder und Jugendliche, schützen können.

Negativ hervorgehoben wurde, dass es keine Quellenangabe zum Warnhinweis gab, sodass der Ursprung der Bewertung unklar blieb und das Vertrauen dadurch möglicherweise negativ beeinträchtigt wurde. Ein möglicher Grund für die schlechtere Bewertung der prozentualen Warnskala könnte sein, dass diese bei den Lesenden Unklarheit darüber auslöst, ob es einen richtigen und einen falschen Teil der Meldung gibt und welcher Anteil der Meldungen falsch und welcher richtig ist.

Aufbauend auf diesen ersten Erkenntnissen wurde ein Laborexperiment konzipiert, in dem mittels Videoaufzeichnung des Klickverhaltens von Teilnehmenden untersucht werden sollte, welche Wirkung Warnhinweise auf das Selektions- und Leseverhalten von Online-Nachrichtenartikeln haben. Dazu wurde eine klickbare Version einer Nachrichtenplattform erstellt, auf der den Teilnehmenden eine Auswahl verschiedener Nachrichtenüberschriften präsentiert wurde, von denen die Hälfte mit einem Warnhinweis (rotes Warndreieck) markiert war. Die Teilnehmenden wurden instruiert, innerhalb von vier Minuten die Nachrichten auf der Seite zum genaueren Lesen auszuwählen, die sie sich näher ansehen wollten. Um zu überprüfen, ob als falsch markierte Nachrichten besonders häufig ausgewählt werden, wenn der Warnhinweis salient ist, wurde zudem variiert, ob entweder nur die Hälfte der Überschriften mit einem Hinweis (roter Warnhinweis, als falsch markiert) versehen wurden oder ob auch die restlichen Überschriften einen Hinweis (grüner Hinweis, als wahr markiert) enthielten. Außerdem sollte untersucht werden, wie Warnhinweise in Kombination mit Hinweisen auf Bewertungen von anderen Nutzenden (Anzahl von Likes) wahrgenommen werden. Daher wurde zusätzlich variiert, ob die Überschriften eine hohe (ca. 9000) oder niedrige (ca. 100) Anzahl von Likes aufwiesen. Die Analyse der Daten zeigt, dass Artikel mit Warnhinweisen generell seltener ausgewählt wurden als Artikel ohne Warnhinweis. Wenn alle Artikel markiert waren (grüne und rote Hinweise), wurden häufiger als falsch markierte Nachrichten ausgewählt als wenn ausschließlich rote Warnhinweise verfügbar waren. Besonders saliente Hinweise, die fehlleitende Artikel gegenüber wahren Nachrichten speziell abgrenzen, scheinen also abschreckender zu sein, als wenn alle Artikel gekennzeichnet sind. Die Anzahl an Likes anderer Nutzender hatte wiederum keinen systematischen Einfluss auf die Häufigkeit, mit der bestimmte Artikel zum Lesen selektiert wurden.

## 2. Wirkung von Korrekturen

Um zu untersuchen wie Korrekturen (inhaltlich) gestaltet und Online-Nutzenden präsentiert werden sollen, um die Wirkung von Falschinformation zu reduzieren, wurde aufbauend auf den bisherigen Kenntnissen und systematischen Literaturrecherchen zu Wirkungen von Korrekturdarstellungen eine experimentelle Studie konzipiert. Um den Einfluss der Art der Korrektur zu untersuchen, wurde den Teilnehmenden eine Falschnachricht und entweder eine kurze, faktenbasierte Korrektur oder eine detaillierte Korrektur mit plausibler Begründung präsentiert. Außerdem wurde variiert, ob die Versuchsteilnehmenden beide Meldungen (die Falschnachricht und die Korrektur) gleichzeitig präsentiert bekommen oder die Korrektur einige Minuten verzögert gezeigt wird.

Die Ergebnisse dieser Studie zeigen, dass eine detailliertere Gegendarstellung nicht dafür sorgt, dass diese Korrektur häufiger geglaubt wird als die hierdurch widerlegte Falschnachricht. Allerdings konnte herausgefunden werden, dass die in der Korrektur enthaltenen Fakten zur wahren Begebenheit besser erinnert wurden und somit fester im Wissensnetz der Versuchspersonen verankert waren als wenn nur eine oberflächliche Korrektur verfügbar war. Zudem wurde gezeigt, dass eine detaillierte Korrektur eher dazu führt, dass wahre Fakten nachhaltig erinnert werden, wenn sie auf der gleichen Seite zu sehen ist wie die Falschnachricht, also zeitgleich konsumiert wird. Diese Resultate deuten darauf hin, dass es sinnvoll ist, eine Korrektur in Form eines plausiblen Gegennarrativs zur ursprünglich falschen Information zu gestalten. Darüber hinaus sollte sie unmittelbar in der Nähe der Desinformation dargestellt werden, damit Konsumierende von online Nachrichten wahre Details besser erinnern.

Auch die in Kapitel 3, B.I. aufgeführten Studien zur Wirkung von positiven und negativen Nachrichten lassen Rückschlüsse darauf zu, welche Wirkungen unterschiedliche Korrekturen bzw. unterschiedliche Korrekturquellen erzielen. Resultate beider Studien illustrieren, dass die Bewertung der Person nach der Aufklärung, dass es sich um eine falsche Information gehandelt hat, steigt, wenn die Nachricht negativ war. Bei positiven Nachrichten sind die Ergebnisse weniger eindeutig. Die erste Studie (in der die Information, dass es sich um eine erfundene Meldung handelte, durch die Versuchsleiterinnen gegeben wurde) zeigt, dass es keinen Effekt auf die Personenbewertung hat, wenn eine positive Nachricht korrigiert wird. Ergebnisse der zweiten Studie, in der die Korrektur scheinbar durch eine Fact-Checking-Organisation vorgenommen wurde, verdeutlicht hingegen, dass die Korrektur einer positiven Nachricht

dazu führen kann, dass die Person im Nachhinein negativer bewertet wird. In diesem Setting wird offensichtlich angenommen, dass der Politiker selbst für die Verbreitung der positiven Nachrichten gesorgt hat. Die Ergebnisse erweitern die Erkenntnisse zur Wahrnehmung und Wirkung von positiver und negativer Desinformation über Politikerinnen und zeigen insbesondere starke Effekte von Korrekturdarstellungen auf Social-Networks-Websites. Zusammengefasst deuten die Ergebnisse der beiden Experimente darauf hin, dass die Art und Weise, wie eine Korrektur präsentiert wird und von wem sie kommt, ein entscheidender Faktor dafür ist, dass der Einfluss von Desinformation reduziert werden kann.

### 3. Wirkung von Nutzerkommentaren und Ratings

Zusätzlich zu den Untersuchungen zur Wirkung von Warnhinweisen haben wir in einer Studie im Rahmen einer Masterarbeit untersucht, inwieweit Online-Nutzende selbst als Ressource zur Bekämpfung von Desinformation dienen können. Dazu wurden auf Basis eines Online-Experiments die Effekte von Nutzer-Kommentaren und numerischen Nutzer-Bewertungen der Artikelglaubwürdigkeit auf die Wahrnehmung von irreführenden Nachrichtenartikeln untersucht. Zudem wurde überprüft, inwieweit die Kommentare und Bewertungen anderer Nutzenden die Bereitschaft, den Artikel weiterzuleiten, beeinflussen. Die Ergebnisse zeigen, dass die Nutzer-Bewertungen und -Kommentare insgesamt wenig Einfluss auf die Glaubwürdigkeitswahrnehmung und Weiterleitungswahrscheinlichkeit von irreführenden Nachrichten haben. Negative Kommentare, die Zweifel gegenüber einem Online-Artikel ausdrücken, können jedoch die Glaubwürdigkeit von Falschnachrichten reduzieren, insbesondere für Personen, die sich selbst fähig fühlen, relevante Informationen in Social Networks zu finden. Außerdem zeigte sich, dass Personen weniger bereit dazu waren, den Artikel privat über Social-Messenger-Services weiterzuleiten, wenn negative Kommentare anderer Nutzende angezeigt wurden (Kluck et al., 2019).

#### *D. Zusammenfassung*

Insgesamt lässt sich vor dem Hintergrund des Forschungsstandes konstatieren, dass sich aus psychologischer Sicht einige Gefahren von Desinformation feststellen lassen, dass aber auch in verschiedener Hinsicht Grund zur Hoff-

nung besteht. So lässt sich nicht nur für einen vermutlich großen Teil der Bevölkerung feststellen, dass sie Desinformation vermeiden möchten, sondern auch, dass Studien zeigen, dass Personen größtenteils kompetent sind in der Auswahl von Informationen aus dem Internet. Dagegen stehen allerdings Befunde, dass Nachrichten oft auch ohne genauere Prüfung weitergeleitet werden und dass ein hoher Einfluss der Voreinstellung erkennbar ist. Auf Basis der Voreinstellung werden eben auch Nachrichten, die zwar den Verdacht erwecken, Falschinformationen zu sein, aber in das eigene Weltbild passen, rezipiert und gegebenenfalls verbreitet. Während es noch keine Lösungen für diese Auswirkungen des Confirmation Bias gibt, zeigen erste Studie zu Interventionen, dass es wirksame Möglichkeiten gibt, zumindest diejenigen, die keine gefestigte Meinung in Bezug auf die Falschinformation haben, davon abgehalten werden können, die Nachricht zu lesen, in die eigenen Einstellungen zu integrieren oder weiterzuleiten. Sowohl Warnmeldungen als auch Korrekturen können hilfreich sein und erste Studien geben Hinweise darauf, dass es wirksamere und weniger wirksame Interventionen gibt.



## Literaturverzeichnis zu Kapitel 3

- Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats. *Journalism & Mass Communication Quarterly*, 95(1), 28–48. <https://doi.org/10.1177/1077699016678186>
- Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of personality and social psychology*, 39(6), 1037.
- Berghel, H. (2017). Lies, damn lies, and fake news. *Computer*, 50(2), 80–85. <https://doi.org/10.1109/MC.2017.56>
- Clayton, K., Blair, S., Busam, J. A., Forstner, S., Gance, J., Green, G., . . . Nyhan, B. (2019). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, 38(2), 173. <https://doi.org/10.1007/s11109-019-09533-0>
- Cook, J., & Lewandowsky, S. (2012). *The debunking handbook* (Version 2). St. Lucia, Australia: University of Queensland.
- Czerwinski, M., Lund, A., & Tan, D. (Eds.) 2008. *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. New York, NY: ACM.
- Ecker, U. K.H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185–192. <https://doi.org/10.1016/j.jarmac.2017.01.014>
- Elyashar, A., Bendahan, J., & Puzis, R. (2017). Has the online discussion been manipulated? Quantifying online discussion authenticity within online social media. Retrieved from <http://arxiv.org/pdf/1708.02763v2>
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). California, USA: Stanford University Press.
- Gao, M., Xiao, Z., Karahalios, K., & Fu, W.-T. (2018). To label or not to label. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–16. <https://doi.org/10.1145/3274324>
- Geary, L. (2017). *Spread of false news stories on Facebook: An assessment of credibility cues and personality* (Master's Thesis), Morgantown, West Virginia.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, 101(3), 371.
- Gross, M. (2017). The dangers of a post-truth world. *Current Biology*, 27(1), R1-R4. <https://doi.org/10.1016/j.cub.2016.12.034>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances*, 5(1), eaau4586.
- Haynes, K. (2016). *Investigating the methodology of warning symbol design* (Master's Thesis), Auburn, Alabama.

- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Communication and persuasion: psychological studies of opinion change. New Haven, CT, US: Yale University Press.
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4), 635. <https://doi.org/10.1086/266350>
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1420.
- Kluck, J. P., Rösner, L., & Krämer, N. C. (in press) (2019) Doubters are more convincing than advocates - The impact of user comments and ratings on credibility perceptions of false news stories on social media. *Studies in Communication and Media*, 8(4).
- Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, 50(1), 46–70. <https://doi.org/10.1111/j.1460-2466.2000.tb02833.x>
- Laughery, K. R., & Wogalter, M. S. (2006). Designing effective warnings. *Reviews of Human Factors and Ergonomics*, 2(1), 241–271. <https://doi.org/10.1177/1557234X0600200109>
- Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). Combating fake news: An agenda for research and action.
- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413–439. <https://doi.org/10.1111/j.1460-2466.2010.01488.x>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Nyhan, B., & Reifler, J. (2015). Displacing misinformation about events: An experimental test of causal corrections. *Journal of Experimental Political Science*, 2(1), 81–93. <https://doi.org/10.1017/XPS.2014.22>
- Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1), 59–68. <https://doi.org/10.1080/15295036.2016.1266686>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880. <https://doi.org/10.1037/xge0000465>
- Pennycook, G., & Rand, D. G. (2018). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking.
- Pennycook, G., Bear, A., Collins, E., & Rand, D. G. (2019). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. <http://dx.doi.org/10.2139/ssrn.3035384>
- PricewaterhouseCoopers GmbH. (2019). Fake News: Ergebnisse einer Bevölkerungsumfrage. Retrieved from <https://www.pwc.de/de/technologie-medien-und-telekommunikation/PwC%20190420%20Berichtsband%20Fake%20News.pdf>
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. *Journal of personality and social psychology*, 32(5), 880.

- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, 31(4), 713–724. <https://doi.org/10.1086/426605>
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital media, youth, and credibility*, 73100.
- Sundar, S. S., Oeldorf-Hirsch, A., & Xu, Q. (2008). The bandwagon effect of collaborative filtering technology. In M. Czerwinski, A. Lund, & D. Tan (Eds.), *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3453–3458). New York, NY: ACM. <https://doi.org/10.1145/1358628.1358873>
- Tan, E. E. G., & Ang, B. (2017). Clickbait: Fake News and Role of the State. *RSIS Commentaries*, 026-17. Retrieved from <https://dr.ntu.edu.sg/bitstream/10220/42108/1/CO17026.pdf>
- Tandoc, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., & Zheng Wei, L. (2018). Audiences' acts of authentication in the age of fake news: A conceptual framework. *New Media & Society*, 20(8), 2745–2763. <https://doi.org/10.1177/1461444817731756>
- Winter, S., & Krämer, N. C. (2012). Selecting science information in web 2.0: How source cues, message sidedness, and need for cognition influence users' exposure to blog posts. *Journal of Computer-Mediated Communication*, 18(1), 80–96. <https://doi.org/10.1111/j.1083-6101.2012.01596.x>
- Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied Ergonomics*, 33(3), 219–230. [https://doi.org/10.1016/S0003-6870\(02\)00009-1](https://doi.org/10.1016/S0003-6870(02)00009-1)
- Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, 95(1), 49–75. <https://doi.org/10.1177/1077699017710453>

