7. Untersuchungsanlage und Methoden

In Kapitel 7.1 wird die Gesamtanlage der Untersuchung vorgestellt. Zwei Methoden sind, dies wird in diesem Kapitel deutlich, für diese Untersuchung zentral. Die Methode der Inhaltsanalyse und das Rezeptionsexperiment mit der Erhebungsmethode Befragung werden in den darauffolgenden Kapiteln erläutert.

7.1 Gesamtanlage der Untersuchung

Um die ersten drei Forschungsfragen dieser Untersuchung zur Evidenzdarstellung in TV-Wissenschaftsbeiträgen zu beantworten, wurde als Methode eine Inhaltsanalyse gewählt. Früh (2011) definiert die Inhaltsanalyse als „empirische Methode zur systematischen, intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen" (S. 27). Bei der Inhaltsanalyse in dieser Untersuchung handelt es sich um eine standardisierte Inhaltsanalyse.[65] Standardisierte Verfahren reduzieren komplexe Zusammenhänge auf wenige, ausgesuchte Merkmale, die auf zahlenmäßig breiter Basis untersucht werden (Brosius, Haas & Koschel, 2012). So können die dargestellten Indikatoren für die externe und interne Evidenz der präsentierten Evidenzquellen in einer großen Anzahl von TV-Wissenschaftsbeiträgen erfasst werden. Durch die Inhaltsanalyse können des Weiteren größere strukturelle Zusammenhänge der inhaltlichen Variablen erkennbar gemacht werden (Früh, 2011).

Mit Hilfe einer Inhaltsanalyse ist es möglich aus den Beiträgen, die für diese Untersuchung wesentlichen Aspekte herauszufiltern und verallgemeinerbare Aussagen über die dargestellte Evidenz in TV-Wissenschaftsbeiträgen zu formulieren. Die Komplexität des Inhalts der TV-Wissenschaftsbeiträge kann durch die Inhaltsanalyse unter dieser forschungsleitenden Perspektive reduziert, hinsichtlich der theoretisch relevanten Merkmale der dargestellten Evidenz selektiert, klassifiziert und gruppiert

65 Die Identifizierung eines inhaltlichen Merkmales im einzelnen TV-Wissenschaftsbeitrag ist an sich ein qualitativer Analyseakt. Dieses wird dann zählend-quantifizierend weiterverarbeitet. Die Bezeichnung als quantitative Inhaltsanalyse ist deswegen irreführend; es handelt sich um eine *standardisierte* Inhaltsanalyse. Die Bedingungen dafür, dass die generierte Datenmenge mit Hilfe statistischer Verfahren weiterverarbeitet werden kann, sind durch die messend-quantifizierende Vorgehensweise erfüllt. (Früh, 2011)

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

werden. So können durch die Inhaltsanalyse die Framebestandteile erfasst werden (Matthes & Kohring, 2004). Um die zentralen Muster der dargestellten Evidenz (Evidenzframes) zu ermitteln, wurde in dieser Untersuchung das empfohlene manuell-dimensionierende Verfahren verwendet (vgl. Matthes, 2007; Matthes & Kohring, 2004). Die Frames werden nicht als Ganzes bestimmt. Die mit Hilfe der ETDS berechneten Evidenzmaße, welche sich aus der internen und externen Evidenz jeder präsentierten Evidenzquelle im Beitrag ergeben, wurden in dieser Untersuchung als Frameelemente als erstes einzeln induktiv erfasst, um anschließend die Frames reliabel mit Hilfe einer Clusteranalyse identifizieren zu können (vgl. Matthes & Kohring, 2004).

Um Wirkungseffekte zu untersuchen, reicht eine Inhaltsanalyse von TV-Wissenschaftsbeiträgen allein nicht aus. Im Anschluss an die Inhaltsanalyse wurde ein Rezeptionsexperiment mit Befragung als Methode gewählt, um die vierte und fünfte Forschungsfrage dieser Untersuchung zu beantworten. Im Rezeptionsexperiment wurde Versuchspersonen ein prototypischer Beitrag eines Frames gezeigt. Vor und nach der Präsentation eines ausgewählten Stimulusbeitrags wurden die Überzeugungen der Rezipienten untersucht, damit die Wirkung der Stimuli nachgewiesen werden kann.

Wissenschaftliche Experimente sind Versuchsanordnungen, mit denen Kausalzusammenhänge überprüft werden (Brosius et al., 2012). Der Einfluss der unabhängigen Variablen, die präsentierten Evidenzframes, auf die abhängigen Variablen, die Überzeugungen der Rezipienten, sollte in dieser Untersuchung gemessen werden. Zur Untersuchung der Überzeugungsbildung und -änderung wurde als Datenerhebungsmethode die schriftliche, quantitative Befragung ausgewählt, da hier zahlreiche Versuchspersonen systematisch, nach festgelegten Regeln zu relevanten Merkmalen befragt werden können (Scheufele & Engelmann, 2009). Die Überzeugungsurteile der Rezipienten wurden durch Selbsteinschätzung schriftlich und direkt mittels einer Befragung erfasst. Die Befragung gilt als das einzig sinnvolle Verfahren, wenn kognitive Inhalte interessieren und Überzeugungen und/oder Einstellungen zu ermitteln sind, die für die Beobachtung oder Inhaltsanalyse unzugänglich wären (Möhring & Schlütz, 2010). Ein vollstandardisierter Fragebogen erlaubt es hier schnell und ökonomisch ein großes Sample an Menschen zu befragen (Scheufele & Engelmann, 2009).

Für das Rezeptionsexperiment wurde ein vollstandardisierter Fragebogen erstellt, durchstrukturiert und für jeden Stimulusbeitrag angepasst.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Um FF4 und FF5 zu beantworten, sollten die Überzeugungen der Rezipienten vor und nach der Stimulipräsentation gemessen werden. Zur Überprüfung der H4.4 sollte die zugeschriebene Glaubwürdigkeit der Rezipienten zu den präsentierten TV-Wissenschaftsbeiträgen erfasst werden und zur Überprüfung der H4.5 die Motivation und kognitive Verarbeitungsfähigkeit der Rezipienten sowie weitere Rezipientenvariablen, die auf eine zentrale/systematische oder periphere/heuristische Verarbeitung hinweisen.

Ein Problem vieler Medienwirkungsstudien ist, dass nur Single-Message-Designs verwendet werden (O'Keefe, 2002). Hier besteht die Gefahr einer uneindeutigen Kausalattribution. Ob eine Überzeugungsänderung wirklich von den verschiedenen Frames der Evidenzdarstellung beeinflusst ist oder eher von anderen inhaltlichen Variablen, die sich bei Real-Stimulus-Material zwangsläufig mit verändern, könnte so also nicht gesichert werden. Unklar wäre, ob andere Beiträge des gleichen Frames gleiche Wirkungen erzielen würden, wie der verwendete Stimulusbeitrag. Deswegen wurden für jeden Frame zwei Beiträge herangezogen. Da mittels Inhaltsanalyse drei Frames identifiziert wurden (vgl. Kapitel 8.1.3), sind also insgesamt sechs Stimulusbeiträge ausgewählt worden. Folglich wurden sechs Experimentalgruppen (Experimentalgruppe = EG) vor und nach der Stimuluspräsentation schriftlich nach ihren Überzeugungen befragt.

Zur Kontrolle der Wirkeffekte wurden zwei Kontrollgruppen (Kontrollgruppe = KG) eingeführt. KG1 ist die *klassische* KG, die nur Pretest- und Posttest-Messungen erhielt, aber keinem Stimulus ausgesetzt wurde. So kann die Veränderung der Überzeugungsurteile ohne experimentelles Treatment gemessen werden, welche sich von den Überzeugungsurteilen mit experimentellem Treatment unterscheiden sollte (Scholl, 2009). KG2 erhielt nur ein Treatment und eine Posttest-Messung. Die Überzeugungsurteile der KG2 und der entsprechenden EG mit Nullmessung sollten optimalerweise gleich sein.

Das Untersuchungsdesign des Rezeptionsexperiments ist dementsprechend, wie in Tabelle 4 dargestellt, aufgebaut.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Tabelle 4: Untersuchungsdesign des Rezeptionsexperiments

Nullmessung	Treatment	Zweitmessung	Gruppe
O_A	1_A	O_A	EG1
O_B	1_B	O_B	EG2
O_C	2_C	O_C	EG3
O_D	2_D	O_D	EG4
O_E	3_E	O_E	EG5
O_F	3_F	O_F	EG6
O_C		O_C	KG1
	2_C	O_C	KG2

O_X: Überzeugungsmessung angepasst an das jeweilige Treatment; $1_{A/B}$: Treatments für Frame 1; $2_{C/D}$: Treatments für Frame 2; $3_{E/F}$: Treatments für Frame 3

Folgend wird auf die Anwendung der Methode der Inhaltsanalyse in dieser Untersuchung detailliert eingegangen.

7.2 Inhaltsanalyse

Die genauen Kriterien der Stichprobenauswahl für die Inhaltsanalyse werden im folgenden Kapitel 7.2.1 beschrieben. Mit Hilfe eines erstellten Codebuchs für die Inhaltsanalyse sollte untersucht werden, wie evident Sachverhalte in TV-Wissenschaftsbeiträgen dargestellt werden. In Kapitel 7.2.2 wird folgend auf die Operationalisierung der dargestellten Evidenz eingegangen; dabei werden alle Kriterien des Codebuchs erläutert und in Bezug auf das gestellte Forschungsproblem sowie auf das Untersuchungsmaterial begründet. Alle Entscheidungen und Bestimmungen bezüglich der Selektions- sowie Klassifikationskriterien, Indikatoren, Variablen, Ausprägungen, Messung und der Codierregeln wurden vor der Durchführung der Inhaltsanalyse festgelegt. Die Auswertungsstrategie zur Inhaltsanalyse, um formal-abstrakte Evidenzframes mittels berechneter Evidenzmaße zu erfassen, wird in Kapitel 7.2.3 ausführlich anhand von Beispielen beschrieben. Die Güte der Inhaltsanalyse und der Auswertungsstrategie werden in Kapitel 7.2.4 diskutiert.

7.2.1 Stichprobenziehung

Die Stichprobe für die Inhaltsanalyse sollte optimalerweise ein verkleinertes, strukturgleiches Abbild der Grundgesamtheit darstellen (Brosius et al.,

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

2012). Forschungsgegenstand dieser Untersuchung sind TV-Wissenschaftsmagazine, die im deutschen Fernsehen ausgestrahlt werden. Dieses Format zielt darauf ab, wissenschaftliche Themen, deren Zusammenhänge und Hintergründe nachvollziehbar und verständlich an Laien zu vermitteln (Milde, 2009). TV-Wissenschaftsmagazine können in Anlehnung an Meier (2006) unterteilt werden u. a. in Magazine, die sich primär mit der Wissenschaftsvermittlung im engeren Sinn befassen, wie bspw. Nano auf 3sat, und in Magazine, die Wissen als Spaß und Unterhaltung vermitteln, wie bspw. Galileo auf Pro7. Letztere werden auch als Wissensmagazine bezeichnet; entsprechend der Tendenz, dass diese den Kriterien eines Wissenschaftsmagazins nicht genügen (Göpfert, 2006a). In dieser Untersuchung wurde sich auf die TV-Wissenschaftsmagazine konzentriert, die sich mit der Wissenschaftsvermittlung im engeren Sinne befassen und auch ihr Grundverständnis dahin ausrichten, neue Erkenntnisse aus der Wissenschaft zu vermitteln. Nach einer vollständigen Analyse des aktuellen deutschen Fernsehprogrammes[66] im Juli 2011 flossen in die Stichprobe dann ausschließlich Beiträge aus TV-Wissenschaftsmagazinen von öffentlich-rechtlichen Sendern.[67] Beiträge aus den folgenden TV-Wissenschaftsmagazinen wurden erfasst und gingen in die Stichprobe ein: W wie Wissen (ARD), Nano (3sat), Quarks & Co (WDR), Planet Wissen (SWR), Odysso (SWR), Faszination Wissen (BR), X:enius (ARTE), Scobel (3sat), Alles Wissen (hr).

Ein abgeschlossener Beitrag eines TV-Wissenschaftsmagazins zu einem medizinischen Sachverhalt war die festgelegte Auswahleinheit dieser Untersuchung (die Festlegung auf medizinische Sachverhalte wurde bereits in Kapitel 5 begründet). Medizin wird dabei in dieser Untersuchung nach Ptok (2000) definiert als wissenschaftliche Heilkunde und Lehre von der Vorbeugung, Erkennung und Behandlung von Krankheiten und Verlet-

66 Die Fernsehprogramme der folgenden analog zu empfangenden Sender wurden anhand ihrer Homepages nach Wissenschaftssendungen durchsucht: 3sat, ARD, ARTE, BR, hr, kabel eins, ProSieben, RTL, RTL II, SAT.1, VOX, ZDF, MDR, rbb, SR , SWR, WDR, n-tv, N24, PHOENIX, sixx, Eurosport, KiKA, SUPER RTL, ZDF.

67 Die Wissenschaftsbeiträge aus den Wissensmagazinen Galileo (ProSieben), Welt der Wunder (RTL II) und Abenteuer Leben (kabel eins) der privaten Fernsehsendeanstalten wurden trotzdem aufgezeichnet und kontrolliert, aber boten zu wenige Beiträge zu medizinischen Sachverhalten ($n < 5$). Eine quantitative, getrennte Auswertung und ein Vergleich zwischen öffentlich-rechtlichen und privaten Sendungen wäre so nicht valide möglich gewesen. In dieser Untersuchung wurden, auch deswegen, ausschließlich die Beiträge der TV-Wissenschaftsmagazine öffentlich-rechtlicher Fernsehsender untersucht.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

zungen bei Menschen und Tieren. Anhand dieser Definition wurden wöchentlich die Agenden der Einzelsendungen der ausgewählten TV-Wissenschaftsmagazine durchsucht.

Es handelt sich bei TV-Wissenschaftsmagazinen um ein Format, das aus Einzelbeiträgen besteht, die von einem Moderator zu einer Gesamtsendung verbunden werden. Die Mehrheit der Wissenschaftsmagazine ist multithematisch aufgebaut, das heißt die Beiträge einer Gesamtsendung präsentieren häufig unterschiedliche Themen (Milde, 2009). Online sind diese Beiträge auch einzeln rezipierbar und nicht mehr in eine Gesamtsendung integriert. Die Untersuchung orientiert sich somit an aktuellen Darstellungs- und Rezeptionsbedingungen, indem sich auf Einzelbeiträge konzentriert wurde.

Zur Ermittlung der erforderlichen Stichprobengröße wurde als Stärke der hypothesenkritischen Effekte ein eta²-Wert von .2 angenommen, was einem schwachen Effekt entspricht (Faul, Erdfelder, Lang & Buchner, 2007). Der α-Fehler wurde streng auf 0.10 gesetzt und der β-Fehler auf 0.10. Unter diesen Bedingungen errechnete sich, unter der Anforderung eine Varianzanalyse mit fünf Gruppen zu berechnen (für die verschiedenen Evidenzquellenarten: Review, Studie, Fallbeispiel, Expertenmeinung und Off-Sprecher), eine optimale Gesamtstichprobe von 330 (Faul et al., 2007). Hiervon ausgehend wurde der Start der Stichprobenaufzeichnung auf den 1. August 2011 und das Ende auf den 31. Mai 2012 gesetzt; bis zu diesem Zeitpunkt war eine Stichprobe in der optimalen Größe abzusehen.

Die in dieser Untersuchung durchgeführte Inhaltsanalyse hat den Anspruch, aufgrund der Fülle der erfassten Beiträge aus öffentlich-rechtlichen TV-Wissenschaftsmagazinen valide, verallgemeinerbare Aussagen zu der dort dargestellten Evidenz machen zu können. Nur mit Hilfe einer ausreichend großen Stichprobe können valide konstante Darstellungsmuster in der Berichterstattung identifiziert werden.

7.2.2 Codebuch: Operationalisierung der dargestellten Evidenz

Das mit Hilfe der Inhaltsanalyse zu messende theoretische Konstrukt in dieser Untersuchung ist die *dargestellte Evidenz*. Dieses hat einen indirekten empirischen Bezug und wurde daher über mehrere Indikatoren näher bestimmt und definiert. Durch die operationale Definition wurden die Dimensionen in Indikatoren überführt. Die Variablenbildung in dieser Untersuchung lief sowohl deduktiv (theoriegeleitet aus der Literatur zum Forschungsstand) als auch induktiv (empiriegeleitet aus dem Material) ab. So

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

war gewährleistet, dass der Gegenstandbereich vollständig erfasst werden konnte (Brosius et al., 2012; Neuendorf, 2002). Die Variablen und ihre Definitionen wurden sowohl nach aktuellen Forschungserkenntnissen und nach einer ersten Sichtung des Untersuchungsmaterials entwickelt.

Codiert wurde auf Beitragsebene (erste Analyseeinheit)[68] und auf Evidenzquellenebene (zweite Analyseeinheit). Als Kontexteinheit für die Codierung wurde der gesamte TV-Wissenschaftsbeitrag herangezogen.

Beitragsebene

Als erstes sollten die formalen Codiereinheiten auf Beitragsebene codiert werden.[69] Sie können in dieser Untersuchung vorwiegend als Differenzierungskriterium für die weitergehende Analyse angesehen werden. Bei Variable V1 wurde zu allererst der Name des Codierers nominal erfasst. Bei Variable V2 wurde nominal erhoben, in welchem TV-Wissenschaftsmagazin der Beitrag gesendet, und bei Variable V3 metrisch, wann der Beitrag ausgestrahlt wurde. Untersucht wurden Beiträge aus den folgenden TV-Wissenschaftsmagazinen: W wie Wissen (ARD), Nano (3sat), Quarks & Co (WDR), Planet Wissen (SWR), Odysso (SWR), Faszination Wissen (BR), X:enius (ARTE), Scobel (3sat) oder Alles Wissen (hr), die vom 01.08.2011 bis zum 31.05.2012 ausgestrahlt wurden. Die Gesamtlänge des TV-Wissenschaftsbeitrags wurde bei Variable V4 in Sekunden metrisch erhoben. Der Moderator des TV-Wissenschaftsmagazins kennzeichnete dabei den Anfang und das Ende eines Beitrags, gehörte selbst aber nicht mit zum Beitrag.

Die nächsten drei inhaltlichen Codiereinheiten dienten in dieser Untersuchung hauptsächlich als Differenzierungskriterien für die weitere Analyse. Bei Variable V5 wurde das spezielle Hauptthema und bei Variable V6 das allgemeine Hauptthema nominal erhoben. Die Codierung sollte sich hier nach den in den Codeplänen aufgeführten Themenlisten richten (siehe Online-Anhang I: Codebuch der Inhaltsanalyse). Diese unterscheiden sich in ihrem Auflösungsgrad. Die systematische Erfassung der möglichen Themen erfolgte im Vorfeld zum einen über eine Titelrecherche

68 Ein Beitrag ist als journalistisches Gesamtwerk des Medienangebotes anzusehen, daher muss dieses auch als Ganzes als frametragende Analyseeinheit gewählt werden (Potthoff, 2012).

69 Formale Codiereinheiten sind physisch manifeste Sachverhalte, die keine Inferenz der Codierer benötigen und sich meist durch Messen, Zählen oder Transkribieren erheben lassen. Im Gegensatz dazu sind inhaltliche Codiereinheiten generell vom Erkenntnisinteresse abhängige Bedeutungsdimensionen, bei denen die Codierer Inferenzen ziehen müssen, um diese zu klassifizieren. (Rössler, 2010)

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

online. Die Titel der meisten Beiträge in TV-Wissenschaftsmagazinen werden auf den Homepages der TV-Sender bereitgestellt. Zum anderen erfolgte die Erfassung über die Themenliste des Codebuchs von Ruhrmann, Guenther, Kessler und Milde (2015), welches für die Analyse von TV-Wissenschaftssendungen zum Thema Molekulare Medizin erstellt wurde.

Die Hauptthese des TV-Wissenschaftsbeitrags wurde bei Variable V7 als String codiert. Diese inhaltliche Variable erleichterte es, den Beitrag wiederzuerkennen. Des Weiteren war diese Variable wichtig als Hilfestellung für die Codierer. Die Codierer konnten und sollten sich die Hauptthese immer wieder vor Augen führen. Die Hauptthese eines TV-Wissenschaftsbeitrags sollte in einem Satz formuliert werden und musste im Mittelpunkt des Betrages stehen und hier diskutiert oder erläutert werden. Der Großteil der Evidenzquellen im Beitrag musste sich auf diese Hauptthese beziehen; sie bewerten, diskutieren oder erläutern. In einigen Beiträgen wird die Hauptthese schon in der Anmoderation explizit genannt, wurde sie das nicht, musste sie von den Codierern aus dem Beitrag erschlossen werden. Die Hauptthese ist, nach Toulmins Argumentationsanalysen (1996), als Schlussregel anzusehen, die implizit oder explizit sein kann, auf diese sollten sich hier definiert die Argumentationen der Evidenzquellen beziehen. Gerichtet wird die Hauptthese nach Titel der Sendung, Anmoderation bzw. Haupttonus der Sendung (Anzahl der Evidenzquellen für eine Position). Die Hauptthese ist die zentrale, hierarchie-höchste Konklusion der Argumentationen in der Sendung (vgl. Bayer, 2007). Die Codierer sollten versuchen bei jeder Sendung erst einmal kognitiv einen Schritt zurück zu gehen und sich zu fragen: Worum geht es dem Wissenschaftsjournalisten bei diesem Beitrag? Welche These möchte er herüberbringen? Welche Hauptaussage möchte er begründen? Die Mehrheit der Evidenzquellen und der vorgetragenen Argumente sollten sich auf diese Hauptaussage beziehen.

Mit Hilfe der Variable V8 sollte im Anschluss die Einschätzung der Codierer erhoben werden. Gefragt war hier die subjektive Meinung der Codierer, ob diese den zu untersuchenden TV-Wissenschaftsbeitrag als insgesamt eher pro oder kontra oder ausgeglichen ausgerichtet in Bezug auf die Hauptthese ansahen. Diese Variable sollte bewirken, dass sich die Codierer im ersten Schritt den gesamten Beitrag im Ganzen anschauen und weiter die Gerichtetheit der aufgestellten Hauptthese überprüfen.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Bei der inhaltlichen Codiereinheit Variable V9 wurde die Anzahl relevanter Evidenzquellen erfasst. Der Begriff *Evidenzquelle* in einem TV-Wissenschaftsbeitrag bezeichnet die spezifische (Bezugs-)Quellen von Belegen, mit denen medizinische Botschaften argumentativ belegt oder widerlegt werden. Dargestellte Evidenzquellen in TV-Wissenschaftsbeiträgen konnten sein (vgl. Kapitel 3.2.1): Reviews, Studien, Fallbeispiele, Expertenmeinungen, Off-Sprecher oder andere Evidenzquellen. Den Codierern wurde für jede der aufgeführten Evidenzquellen eine eindeutige Definition gegeben (siehe Online-Anhang I: Codebuch der Inhaltsanalyse).[70] Von der relevanten Evidenzquelle mussten Aussagen, Argumente oder Belege zum Sachverhalt, für oder gegen die Hauptthese gerichtet, dargestellt werden. Nahm eine Evidenzquelle keine Stellung, bzw. präsentierte keine Argumente, Aussagen oder Belege mit Bezug auf die Hauptthese, so war diese Evidenzquelle für die Diskussion der Hauptthese irrelevant und sollte nicht codiert werden. Ein Off-Sprecher, eine Expertenmeinung oder ein Fallbeispiel sollten von den Codierern nur codiert werden, wenn sie Argumente, Aussagen oder Belege für oder gegen die Hauptthese des Beitrags generierten. Studien werden bspw. häufig von präsentierten Akteuren (z. B. einem Wissenschaftler) oder vom Off-Sprecher im TV-Wissenschaftsbeitrag stellvertretend beschrieben. Hier sollte dann ausschließlich die Evidenzquelle *Studie* codiert werden, nicht der beschreibende Akteur oder Off-Sprecher.[71]

70 In dieser Untersuchung wurde jede Form von Review berücksichtigt. Sobald quantitativ oder qualitativ zusammenfassende Aussagen über mehrere Studien getätigt wurden, wie bspw. *Viele Studien, die diesen Effekt untersuchen, kommen zu dem Ergebnis, dass...* oder *Im Vergleich verschiedener Studien zu diesem Aspekt, zeigt sich…* zählten diese als Meta-Aussagen und wurden als Evidenzquelle *Review* codiert. Synonym zur Evidenzquelle *Studie* konnten in TV-Wissenschaftsbeiträgen auch Begriffe wie *Experiment, Testreihe* und *Versuche* benutzt werden.

71 Sagt bspw. der Off-Sprecher eines Beitrags: „Das ist die neue Studie xy…" und ein Professor sagt: „Die Probanden in dieser Studie mussten diese Medikamente schlucken...", so sollte nur die Evidenzquelle *Studie* codiert werden (ohne den Professor oder den Off-Sprecher zu codieren). In solchen Fällen war als Evidenzquelle immer die ranghöhere Art der Evidenzquelle (siehe Variable V10) zu codieren. Dies galt allerdings nur, wenn Akteure ausschließlich eine weitere Evidenzquelle beschrieben. Sobald Akteure eigene Meinungs- oder Erfahrungsäußerungen machten und neue Argumente generierten, wurden diese extra codiert. Wenn bspw. ein Experte im Beitrag als erstes seine Meinung zu einem Thema kundtat und dann über eine Studie sprach, so sollten hier zwei Evidenzquellen codiert werden, Studie und Expertenmeinung. Wurde eine Evidenzquellenargumentation durch eine andere Evidenzquelle unterbrochen und später im Beitrag dann weitergeführt, so war die weitergeführte Argumentation der Evidenzquelle nicht als neue Evidenzquellenargumentation zu codieren. Die Argumentation einer Evidenzquelle im Beitrag wurde folglich als Ganzes codiert, auch wenn sie unterbrochen wurde von anderen Evidenzquellen. Beispielsweise wurden zwei Experten in einer wechselseitigen Diskussion immer auch

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Wurden alle Variablen von V1 bis V9 auf Beitragsebene codiert, wechselte die Analyseebene nun auf Evidenzquellenebene.

Evidenzquellenebene

Für jede Evidenzquelle wurden die folgenden Variablen V10 bis V26 gesondert codiert. Das Skalenniveau jeder Variable ist konstant ordinal. Die Analyseeinheit auf Evidenzquellenebene war die (Sprech-)Handlungssequenz einer Evidenzquelle. Die (Sprech-)Handlungssequenz wurde als semantische Einheit dadurch definiert, dass ein Sprechakt respektive eine Sprechhandlung (argumentieren, berichten, erklären, beschreiben, kritisieren, diskutieren usw.) mit Bezug zur Hauptthese des TV-Wissenschaftsbeitrags vorhanden war. Die Analyseeinheit *Evidenzquelle* kann sehr kurz und auch sehr lang sein: So kann es sein, dass eine Evidenzquelle (z. B. Experte) nur in einem Satz seine Meinung kurz äußert oder dass eine Evidenzquelle (z. B. Studie) über fünf Minuten genau erläutert wird.

Das theoretische Konstrukt *dargestellte Evidenz* setzt sich aus den Komponenten *externe* und *interne Evidenz* zusammen. Die Komponente *externe Evidenz* wurde durch die Dimensionen *Art der einzelnen Evidenzquellen* und *Qualität der einzelnen Evidenzquellen* operationalisiert. Die Komponente *interne Evidenz* wurde durch die Dimensionen *Argumentationsweise der einzelnen Evidenzquelle* und *Verwendung von evidenzstiftendem Bildmaterial der einzelnen Evidenzquelle* messbar gemacht. Alle folgenden Variablen wurden als Indikatoren für die einzelnen Dimensionen erhoben. Bevor nun diese einzeln definiert werden, folgt in Tabelle 5 eine Übersicht.

als genau zwei Expertenmeinungen als Evidenzquellen codiert, egal wie oft sie sich gegenseitig unterbrachen.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Tabelle 5: Überblick über die Variablenzuordnung

Komponente	Dimension	Indikator (Variable)	Variablenname
Externe Evidenz	Art der Evidenzquelle	Art der Evidenzquelle	V10
	Qualität der Evidenzquelle	Validität der Evidenzquelle	V11
Interne Evidenz	Argumentationsweise der Evidenzquelle	Anzahl der Argumente	V12+V13+V14
		Gewichtung der Argumente	V15
		Neuigkeit der Argumente	V16
		Dargestellte Unsicherheit -explizit	V17
		Dargestellte Unsicherheit -implizit	V18
		Homogenität der Argumentation	V19
		Detaillierung & Hintergrund	V20
		Konstanz der Argumentation	V21
		Sekundäre Bewertung	V22
	Bildmaterial der Evidenzquelle	Bilder	V23

Die *externe Evidenz* wird durch die Art und die Qualität der dargestellten Evidenzquelle generiert. Die Art der Evidenzquelle wurde bei Variable V10 erfasst. Die Ausprägungen dieser Variable wurden in Anlehnung an die Evidenzgrade des SIGN (2000) aus der Medizin (vgl. Kapitel 2.1) eingeteilt.[72] Die Qualität wurde bei den verschiedenen Evidenzquellen durch jeweils unterschiedliche dargestellte interne Validität erfasst, daher wurde

72 Die zahlenmäßig höchste Ausprägung haben entsprechend die beim höchsten Evidenzlevel angesiedelten Reviews. Die nächstkleinere Ausprägung 4 wurde codiert, wenn die Evidenzquelle eine einfache Studie war, die Ausprägung 3, wenn die Evidenzquelle ein Fallbeispiel oder eine Fallbeispielserie war und die Ausprägung 2, wenn es sich um eine Expertenmeinung handelte. War die Evidenzquelle ein Off-Sprecher war hier die Ausprägung 1 zu codieren. Um die Vollständigkeit der Kategorie zu gewährleisten, wurden bei der Ausprägung 0 alle möglichen anderen Evidenzquellenarten codiert.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

die Variable _V11 evidenzquellenspezifisch geteilt in _V11a Validität Review[73], _V11b Validität Studie[74], _V11c Validität Fallbeispiel[75] und _V11d Validität Expertenmeinung[76]. Für den Off-Sprecher liegen in der bisherigen Forschungsliteratur noch keine validen evidenzspezifischen Qualitäts- bzw. Validitätsmerkmale vor, die hier kontrolliert werden könnten bzw. sollten (vgl. Kapitel 3.2.1).

Das theoretische Konstrukt *dargestellte Evidenz* wird nicht nur durch die Komponente *externe Evidenz* gebildet, sondern auch durch die Komponente *interne Evidenz*. Die interne Evidenz wird generiert durch die Argumentationsweise der einzelnen Evidenzquelle (V12 bis V22) und Verwendung von evidenzstiftendem Bildmaterial bei der Präsentation der einzelnen Evidenzquelle (V23).

Bei Variable V12 wurde erhoben, wie viele Argumente pro, und bei Variable V13 folgend, wie viele Argumente von der Evidenzquelle kontra die

73 Die Ausprägungen wurden hier in Anlehnung an die Evidenzlevel des SIGN (2000) aus der Medizin erstellt. Zwischen hoher, mittlerer und geringer interner Validität wurde unterschieden. Die Ausprägung 0 wurde codiert, wenn Informationen zur internen Validität der Evidenzquelle im Beitrag nicht benannt wurden. Eine hohe interne Validität war im Gegensatz zur geringeren internen Validität gegeben, wenn nur ein geringes Risiko für Verzerrungen oder Fehler bestand und bspw. aufgezeigt wird, dass systematisch eine hohe Anzahl an Studien miteinander verglichen wurden oder die verglichenen Studien selbst qualitativ sehr gut waren (bspw. eine Meta-Analyse von randomisierten, kontrollierten Studien). Wurden Aussagen, die für eine hohe Validität sprachen und Aussagen, die für eine geringe Validität sprachen, geäußert, sollte hier die mittlere Validität codiert werden. Wenn bspw. im Beitrag an der Systematik der Auswertung, an der Generalisierbarkeit, Genauigkeit oder Studienauswahl gezweifelt wird, sollte hier nur eine geringe Validität codiert werden. Die Validität der Evidenzquelle konnte im Beitrag explizit benannt, aber auch implizit dargestellt sein.

74 Die Ausprägungen wurden hier ebenfalls in Anlehnung an die Evidenzlevel des SIGN (2000) aus der Medizin erstellt und folglich zwischen hoher, mittlerer und geringer Validität unterschieden. Die Ausprägung 0 wurde codiert, wenn die Validität oder Qualität einer Studie im Beitrag nicht benannt wurde. Eine hohe Validität ist im Gegensatz zur geringen internen Validität gegeben, wenn nur ein geringes Risiko für Verzerrungen und/oder Fehler in der Studie bestand (bspw. bei einer randomisierten, kontrollierten Studie). Wurden Aussagen, die für eine hohe Qualität sprechen und Aussagen, die für eine geringe Qualität sprechen, geäußert, sollte hier die mittlere Validität codiert werden. Die Qualität der Studie konnte explizit benannt werden, indem bspw. gesagt wurde, dass die Qualität der Messung hoch war und/oder die Risiken für systematische Verzerrungen kontrolliert wurden. Die Qualität der Studie konnte aber auch implizit dargestellt werden, indem Qualitätsmerkmale aufgeführt wurden.

75 Die Erkenntnisse, die in einem Fallbericht dargestellt sind, der sehr viele andere Fälle repräsentieren soll, gelten als intern valider als die Erkenntnisse aus einem Einzelfallbericht (vgl. Kapitel 3.1.1). Als am validesten gelten hier die Erkenntnisse aus einer Fallberichtserie.

76 Aussagen eines Experten mit höherer Reputation (Professor, Nobelpreisträger, u. a.) sind als intern valider anzusehen als Aussagen eines Experten ohne Titel, Preise oder expliziter Benennung (vgl. Kapitel 3.2.1).

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Hauptthese gerichtet gegeben wurden. Als Argument wird eine Äußerung bezeichnet, die funktional als eine Begründungshandlung erscheint (Grundler, 2011). Oder einfacher ausgedrückt: Argumente sind die Gründe, die für oder gegen eine (strittige) These gerichtet sind (Kienpointner, 2001; vgl. Kapitel 3.2.2).[77] Diese (strittige) These ist in den Beiträgen schon definiert als Hauptthese. Diese Hauptthese ist nach Toulmins Argumentationsanalysen (1996) als Schlussregel anzusehen, auf die sich die begründete Behauptung hier beziehen muss. Um als Argument zu gelten, müssen Aussagen die Ebene der Beteuerung, des Bekenntnisses oder Appels deutlich überschreiten (Schultz, 2006). Argumentieren heißt letztendlich Behauptungen belegen (Geißner, 2004). Um eine Behauptung, Annahme oder These zu stützen, braucht es Argumente. Jedes Argument wurde danach eingestuft, welche Position es in Bezug auf die Hauptthese hat, ob es diese stützt oder widerlegt (vgl. auch *argumentative Polarisierung* bei Weiß, 1989). In einem Satz konnten mehrere Argumente enthalten sein. Ein Argument konnte aber auch mehrere Sätze umfassen. Aus der jeweiligen Ausprägung bzw. Anzahl der Pro- und Kontraargumente bei den Variablen V12 und V13 sollten die Codierer bei Variable V14 die Polarität der Evidenzquelle bestimmen. Je nachdem, wo die Mehrheit der Argumente gegeben wurde, wird die Polarität der Argumentation von der Evidenzquelle eher codiert als pro, kontra oder ausgeglichen bezüglich der Hauptthese des TV-Wissenschaftsbeitrags. Sollte eine Evidenzquelle genauso viele Argumente pro- oder kontra die Gegenthese gerichtet haben, wurde sie als ausgeglichen codiert.[78]

77 Bis heute wird häufig Toulmins Argumentationsschema zur Analyse von Argumentationen eingesetzt (Grundler, 2011). In Toulmins (1996) Argumentationsschema besteht eine Argumentation aus einem Datum (Behauptung) und einer Schlussregel, die dazu führt, dass die dritte Komponente, wieder eine Behauptung, zur Konklusion erklärt werden kann (wobei die Schlussregel implizit bleiben darf). Ein Beispiel wäre: *Es gab keine Heilung* (Datum), *folglich funtionierte die neue Therapie nicht* (Konklusion). Als implizite Schlussregel kann angenommen werden: *Wenn es keine Heilung gibt, funktioniert eine Therapie nicht.* Letzendlich ist in dieser Untersuchung als Argument eine Behauptung definiert worden, die in Bezug auf die Hauptthese im Beitrag gerichtet geäußert und begründet wird. Mit den Codierern wurde das Erkennen eines Argumentes eingehend geübt. Bestimmte Wörter und Wendungen helfen, diese aufzufinden, bspw. Worte wie „weil", „daher", „also", „somit", „folglich" und „deswegen" (Bayer, 2007). Die Beziehung zwischen Hauptthese und Argument sollte mit Hilfe der Grundformel „p, weil q" abgebildet werden können (Angabe von Geltungsgründen). Aber wie bei Schultz (2006) war es nicht gefordert, dass eine argumentative Kette chronologisch oder formal vollständig sein musste. Schlichte Behauptungen, Meinungsäußerungen, Polemik, Drohungen, bloße Information, Narration oder Selbstoffenbarungen zählten hier ausdrücklich nicht als Argument (vgl. Schultz, 2006).

78 In dieser Codierung wurde nicht erfasst, welchen Wert ein Argument hat, bspw. wie gut oder schlecht es ist. Dies kann von den Codierern nicht ohne unverhältnismäßig großen

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Doch nicht nur die Quantität spielt, wie in Kapitel 3.2.2 herausgestellt, eine große Rolle bei der Argumentationsweise. Mit Hilfe der Variable V15 sollte die Gewichtung bzw. die Relevanz der Argumente erfasst werden.[79] Relevant ist ein Argument, wenn es die Konklusion/Hauptthese stützt, d. h. wenn die Wahrheit der Prämisse ein guter Grund ist, auch die Konklusion für wahr zu halten (Bayer, 2007). Bei Variable V16 wurde codiert, ob mindestens ein präsentiertes Argument der Evidenzquelle inhaltlich neu war; wiederholte und tautologische Argumente bieten nichts inhaltlich Neues. Variable V17 erfasste, ob bei der Darstellung der einzelnen Evidenzquelle explizit die Unsicherheit und/oder Sicherheit ihrer Erkenntnisse oder Aussagen hervorgehoben wird. Es konnte im TV-Wissenschaftsbeitrag bspw. explizit gesagt werden, dass die Belege/Argumente einer Evidenzquelle unsicher sind und somit die Belegkraft einer Argumentation gering ist (vgl. Kapitel 3.2.2).[80] Bei Variable V18 wurde erfasst,

Aufwand geleistet werden. Auch war es, wie bei Analysen zur Logik und Argumentationstheorie üblich, nicht relevant, ob ein Argument wahr oder falsch ist (Bayer, 2007). Es lief bei der Codierung auch nicht darauf hinaus, die exakte Anzahl an Argumenten zu eruieren. Dies könnte aufgrund von überlappenden, unvollständigen oder teilweise wiederholten Argumenten (Bayer, 2007; Grundler, 2011; Rieke, Sillars & Peterson, 2013; Toulmin, 1996; Weiß, 1989) von den Codierern ebenfalls nicht ohne unverhältnismäßig großen Aufwand geleistet werden. Oft sind die Argumentationen in der Berichterstattung jedoch einfach strukturiert, hier ist die Argumentationsanalyse relativ problemlos möglich (Weiß, 1989). Codiert wurden hier jeweils drei Ausprägungen (1 = von der Evidenzquelle werden keine Argumente gegeben, die für die Hauptthese sprechen; 2 = von der Evidenzquelle werden 1 bis 3 Argumente gegeben; 3 = von der Evidenzquelle werden mehr als drei Argumente gegeben, die für die Hauptthese sprechen). In den Codiererschulungen zeigte sich, dass diese Ausprägungsvarianz von den Codierern relativ schnell und reliabel codiert werden konnte.

79 Wurde mindestens ein präsentiertes Argument der Evidenzquelle ausführlich dargestellt und dieses ist zentral/wichtig für die Diskussion der Hauptthese, sollte hier Ausprägung 1 codiert werden. Wurden die aufgezeigten Argumente einer Evidenzquelle nur kurz/beiläufig erwähnt oder sind randständig und unwichtig für die Diskussion der Hauptthese, so sollte hier Ausprägung 0 codiert werden.

80 Als weitere Anzeichen der expliziten Unsicherheitsbekundung galten z. B. Aussagen über mangelnde Repräsentativität und Transparenz, Vorläufigkeit, Wissenslücken, Messfehler, Unklarheit, zu wenig Daten oder Praxisanwendungen, mangelnde Übertragbarkeit, mangelnde Objektivität, Reliabilität, Validität oder Qualität. Umgekehrt konnte im TV-Wissenschaftsbeitrag auch von einem Sprecher explizit gesagt werden, dass die Belege/Argumente einer Evidenzquelle sehr sicher sind und somit die Belegkraft einer Argumentation sehr stark ist. Als weitere Anzeichen der expliziten Sicherheitsbekundung galten bspw. Aussagen über die hohe Messgenauigkeit, große Stichprobe oder Qualität des Erkenntnisprozesses.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

ob bei der Darstellung der einzelnen Evidenzquelle implizit die Unsicherheit und/oder Sicherheit ihrer Erkenntnisse oder Aussagen hervorgehoben wird.[81]

Wie in Kapitel 3.2.2 erläutert, sollte eine Argumentation, um eine gewisse Qualität für sich zu beanspruchen, homogen und konstant sein und jedes Argument sollte detailliert, mit Erläuterungen und Hintergrundinformationen versehen sein. Bei Variable V19 wurde die Homogenität der Argumentation erfasst. Die Argumente/Aussagen der Evidenzquelle sind intern homogen, wenn keine inhaltlichen oder logischen Widersprüche in der Argumentation zu finden sind (Arntzen, 2011; Lewandowski et al., 2012). Ein Widerspruch ist die gleichzeitige Behauptung einer Aussage und ihrer Negation. Ein argumentativer Widerspruch entsteht, wenn ein Grund gleichzeitig für und gegen eine These angeführt wird (Bayer, 2007). Bei Variable V20 wurde erfasst, ob die Argumente der Evidenzquelle mit Erläuterungen und Hintergrundinformationen versehen wurden, also einen hohen Detaillierungsgrad aufwiesen.[82] Fehlte es einer Aussage an Homogenität oder quantitativen Detailreichtum, sprach dies nicht automatisch gegen die Belegkraft dieser. Wie wichtig auch die Konstanz der Argumentationstendenz für die Qualität ist, wurde bereits in Kapitel 3.2.2 aufgezeigt. Änderte sich bei einer Evidenzquelle im Laufe des Beitrags die Meinung, Argumentationsseite oder Ergebnisdeutung, so war die Argumentation nicht konstant. Bei Variable V21 wurde die Konstanz der Argumentationstendenz in der Argumentation der einzelnen Evidenzquelle erfasst, indem codiert wurde, ob die Gegenthese laut der Argumentation der Evidenzquelle existieren kann und annehmbar ist oder nicht. Da die Evidenzkraft einer Evidenzquelle auch durch die Argumentation einer anderen Evidenzquelle sekundär erhärtet oder wertgemindert werden kann,

81 Implizite Anzeichen für Unsicherheit konnten sein: Konjunktivformen (Lehmkuhl & Peters, 2016), Schätzungen, Satzanfänge wie *noch nicht erforscht ist*, *eine vorläufige Annahme ist* oder Worte wie *vielleicht, möglicherweise* oder *wahrscheinlich*. Implizite Anzeichen für die Sicherheit einer Argumentation sind bspw. Satzanfänge wie *Sicher ist, Klar ist, Fakt ist, Bewiesen ist, Fest steht* und Worte wie *evident, logisch, offensichtlich, natürlich, wirklich, einleuchtend* oder *gesichert*, die in der Argumentation verwendet werden. (Peters, 2014; Maurer, 2011; vgl. Kapitel 3.2.2)

82 Als relevante Hintergrundinformationen oder Details konnten bspw. Aussagen über den Forschungsprozess an sich, über Ursachen und Folgen einer Problematik, über Ablauf, Zeitpunkt, Art und Anordnung eines Versuches und theoretische Annahmen sowie Aussagen über Messinstrumente, Stichprobe, Auswertungsverfahren, Auftraggeber und Forschungsinstitut gelten. Wiederholungen und überflüssige Details oder randständige Hintergrundinformationen (der Zusammenhang zur Hauptthese ist nicht erkennbar) bieten kein neues, relevantes Moment in der Argumentation.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

wurde bei Variable V22 die sekundäre Bewertung einer Evidenzquelle erfasst.[83]

Die Komponente *interne Evidenz* wird, wie schon mehrfach aufgezeigt, generiert durch die Argumentationsweise der einzelnen Evidenzquelle und durch die Verwendung von evidenzstiftendem Bildmaterial bei der Präsentation der einzelnen Evidenzquelle. Bei den Variablen V23a, V23b und V23c wurden nun die bei der Darstellung einer Evidenzquelle verwendeten Bilder codiert. Die drei Bildformen mit der höchsten Evidenzkraft wurden hier codiert. Da diese Variable später für die Berechnung eines prozentuierten Indexes genutzt wird (vgl. Kapitel 7.2.3.1), ist es wichtig, einen zu erreichenden Maximalwert zu setzen. Deshalb wurde sich hier auf maximal drei Bilder festgelegt. Bildern, die die Erkenntnisse mehrerer Fälle visuell kumulieren, wie bspw. Diagramme, kann aus wissenschaftlicher Perspektive, analog zu den Evidenzlevel der Medizin (Grade Working Group, 2004), die höchste Evidenzkraft zugeschrieben werden. Apparativ erstellte Bilder, wie Fotographien oder Mikroskopbilder, haben wissenschaftlich gesehen weniger Evidenzkraft, da hier nur die Erkenntnisse eines einzelnen Falles dargestellt werden. Noch weniger Evidenzkraft ist einem Bild mit ausschließlich einer Person (bspw. Experten) zuzuschreiben, da dies kein wissenschaftliches Ergebnis darstellt. In Kapitel 3.2.2 wurde des Weiteren ausführlich dargelegt, dass Bilder als inhaltliche Codiereinheit auch aufgrund ihrer Herstellungsprozesse unterschiedlich starke Evidenzkraft innehaben. Die Ausprägungen der Variablen wurden dementsprechend in vier Stufen eingeteilt.[84]

83 Als Anzeichen für die (teilweise) Wertminderung der Belegkraft einer Evidenzquelle durch eine andere Evidenzquelle galt bspw., wenn auf die Unehrlichkeit der Evidenzquelle von einer anderen Evidenzquelle verwiesen wurde, die Lauterkeit von Motiven der Evidenzquelle in Frage gestellt wurde oder Hinweise auf Widersprüchlichkeit von Ergebnissen oder Ausführungen der Evidenzquelle gegeben wurden. Anzeichen für die (teilweise) Erhärtung der Belege und/oder die Belegkraft einer Evidenzquelle durch eine andere Evidenzquelle waren dementsprechend bspw. gegeben, wenn auf die Rechtschaffenheit einer Evidenzquelle verwiesen wurde, die Lauterkeit bestätigt wurde oder Hinweise auf die Schlüssigkeit gegeben wurden. Die sekundäre Bewertung konnte sich auf den Inhalt der Argumentation einer Evidenzquelle und/oder auf die Evidenzquelle an sich beziehen. Beispiele wären hier: Ein Wissenschaftler, der sagt, dass die durchgeführte und im Beitrag präsentierte Metastudie völlig unbrauchbar ist, weil die verwendeten Studien nicht miteinander zu vergleichen sind; oder: Ein Betroffener, der sagt, dass der Wissenschaftler, der eben seine Meinung als Evidenzquelle im Beitrag präsentierte, ein Lügner sei, der nur auf Profit aus ist.

84 Bei der höchsten Ausprägung waren die verwendeten Bilder mit der höchsten Evidenzkraft zu codieren. Wurde ein digital erstelltes, computervermitteltes Bild, wie eine Statistik (Diagramm, Kurve, Tabelle) dargestellt, so sollten die Codierer für dieses Bild die Ausprägung 3 codieren. Die Ausprägung 2 war bei einem Bild zu codieren, welches durch ein Aufschreibesystem erstellt wurde, wie bspw. eine Mikrofotografie, ein Mikroskopbild,

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Es wurde im gesamten Codebuch nicht auf einen speziellen Themenbereich fokussiert; alle Variablen wurden folglich themenunabhängig erhoben und ausgewertet.

7.2.3 Auswertungsstrategie

Um die dargestellte Evidenz im Gesamtbeitrag erfassen zu können, wurde zuerst die dargestellte externe und interne Evidenz der einzelnen Evidenzquellen in einem Beitrag erfasst. Dann wurde ein prozentuierter Index generiert, welcher die dargestellte externe und interne Evidenz pro Evidenzquelle im Beitrag repräsentiert. Dieser ist die Voraussetzung für die weitere Berechnung der dargestellten Evidenzmaße im Gesamtbeitrag. In Kapitel 7.2.3.1 wird die Bildung des Indexes für die dargestellte Evidenz jeder Evidenzquelle erläutert. Zur Berechnung der dargestellten Evidenzmaße im Gesamtbeitrag wurde die Evidenztheorie von Dempster & Shafer (ETDS) genutzt. Auf diese Berechnungsmethode wird in Kapitel 7.2.3.2 eingegangen, indem die Anwendung der ETDS an einem Beispielbeitrag demonstriert wird. Um die FF3 zu beantworten, werden die mit Hilfe der ETDS berechneten Evidenzmaße im Anschluss als clusterbildende Variablen dazu verwendet, Muster der dargestellten Evidenz, also formal-abstrakte Evidenzframes, reliabel zu identifizieren. Auf die verwendete Clusteranalyse zur Bildung der Evidenzframes wird in Kapitel 7.2.3.3 eingegangen.

7.2.3.1 Indexbildung

Da die Komponenten der dargestellten Evidenz einer einzelnen Evidenzquelle nicht mit einer Messung abgedeckt, sondern durch mehrere Indikatoren erfasst wurden, wurde in dieser Untersuchung zunächst ein Index für die externe und interne Evidenz jeder Evidenzquelle gebildet. Indizes repräsentieren einen einzelnen Merkmalsraum, in dem mehrere Indikatoren, welche operational definiert wurden, rechnerisch zusammengefasst werden (Brosius et al., 2012). Meyer (2011) unterstreicht den Vorteil, dass mit Hilfe eines Indexes durch einen einzigen Wert Entwicklungen und

eine Röntgenaufnahme oder eine Aufnahme aus einem anderen Messgerät. Noch weniger Evidenzkraft wurde einer Comiczeichnung, einer Computeranimation, einem präsentierten Modell oder einem Bild, auf dem ausschließlich eine gefilmte Person, eine gefilmte Operation oder ein gefilmter Forschungsprozess zu sehen ist, zugeschrieben. Deshalb war hier die Ausprägung 1 zu codieren. Am wenigsten Evidenzkraft wurde dem Füllbild (ohne Kontext zur Evidenzquelle, zur Hauptthese und zum gesprochenen Text) zugeschrieben; hier war von den Codierern die Ausprägung 0 zu codieren.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Zusammenhänge einfach und übersichtlich dargestellt werden können und Bewertungen ermöglicht werden, welche ohne diesen nur schwierig vermittelbar oder bewertbar wären. Er zeigt aber auch die Nachteile auf, dass Indizes sehr abstrakt sind, die eventuell interessanten Einzelergebnisse der Indikatoren unsichtbar werden und es schwierig ist, die einzelnen Indikatoren für einen Index begründet einheitlich zu skalieren sowie deren Gewichtung theoretisch zu begründen.

Im Anschluss an die Codierung wurden zunächst zwei Indizes für jede Evidenzquelle aus den einzelnen Ausprägungen der codierten Variablen für die externe und interne Evidenzdarstellung gebildet. Alle relevanten Variablen wurden einheitlich ordinal skaliert erhoben, um diese Voraussetzungen für die Indexbildung zu erfüllen (Brosius et al., 2012; Meyer, 2011). Des Weiteren ist jede relevante Variable so skaliert, dass sie einen Extremwert hat, der theoretisch nicht überschritten werden kann, aber bei einer Messung real, gleichwahrscheinlich auftreten könnte. Dies sind die Voraussetzungen, um später Indizes generieren zu können, in denen die Skalen durch Prozentuierung standardisiert werden (Meyer, 2011). Die Variablen, die in die Indizes einflossen, hatten jeweils unterschiedliche Ausprägungsvarianzen. In dieser Untersuchung wurde die Anzahl der einzelnen Skalenpunkte pro Variable inhaltlich begründet und schon mit Bezug auf die Indexbildung sinnvoll vergeben, wie in Kapitel 7.2.2 aufgezeigt. Die Ausprägungen und Ausprägungsanzahl haben für die inhaltlichen Variablen auf Evidenzquellenebene auch eine inhaltliche Bedeutung:

- Die Codes wurden im positiven Bereich für jede dieser Variablen so vergeben, dass sie immer einer Steigerung der dargestellten Evidenz der jeweiligen Evidenzquelle entsprechen.
- Die Ausprägungscodes wurden im negativen Bereich für jede dieser Variablen so vergeben, dass sie immer einer Minderung der dargestellten Evidenz der jeweiligen Evidenzquelle entsprechen.
- Der Ausprägungscode 0 wurde für jede dieser Variablen immer so vergeben, dass er weder für eine Steigerung noch für eine Minderung der dargestellten Evidenz der einzelnen Evidenzquelle spricht.

Die Komponenten *externe* und *interne Evidenz* jeder Evidenzquelle wurden, um eine gemeinsame, standardisierte und damit unmittelbar vergleichbare Basis zu generieren, getrennt voneinander durch Indizes erfasst und prozentuiert, wie im Folgenden genauer beschrieben wird. Die Komponenten

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

konnten so in den Gesamtindex des theoretischen Konstrukts, die dargestellte Evidenz der einzelnen Evidenzquelle, zu gleich starken Teilen eingehen. Da die externe und interne Evidenz in dieser Untersuchung das erste Mal getrennt voneinander untersucht wurden, konnten im Vorhinein keine validen Aussagen zum Zusammenhang zwischen externer und interner Evidenz gemacht werden und eine Gleichgewichtung scheint und schien am besten geeignet.

In der folgenden Tabelle 6 sind nun die Komponenten *externe* und *interne Evidenz* mit ihren Dimensionen und mit allen dazugehörigen Indikatoren, die das theoretische Konstrukt *dargestellte Evidenz* messen, zusammengetragen. In Klammern stehen die jeweilig möglichen Ausprägungen für jede Variable.

Tabelle 6: Operationalisierte Komponenten, ihre Dimensionen und Indikatoren des Konstrukts dargestellte Evidenz

Komponente	Dimension	Indikator
Externe Evidenz	Art der Evidenzquelle	Art der Evidenzquelle (0/1/2/3/4/5)
	Qualität der Evidenzquelle	Validität der Evidenzquelle (0/1/2/3)
Interne Evidenz	Argumentationsweise der Evidenzquelle	Anzahl/Differenzmenge der Argumente (0/1/2) Gewichtung der Argumente (0/1) Dargestellte Unsicherheit -explizit (-1/0/1) Dargestellte Unsicherheit -implizit (-1/0/1) Homogenität (-1/1) Neuigkeit (0/1) Detaillierung & Hintergrundinformationen (0/1/2) Konstanz der Argumentationstendenz (-1/0/1) Sekundäre Bewertung (-1/0/1)
	Bildmaterial der Evidenzquelle	Bild 1 (0/1/2/3) Bild 2 (0/1/2/3) Bild 3 (0/1/2/3)

Für die Komponente *externe Evidenz* wurde Index A gebildet. Index A umfasst die beiden Dimensionen aus denen die externe Evidenz einer Evi-

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

denzquelle im TV-Wissenschaftsbeitrag entstehen kann. Die beiden Dimensionen sind, wie in Kapitel 3.2 erläutert, die *Art der einzelnen Evidenzquelle* und die *Qualität der einzelnen Evidenzquelle*. Die Ausprägungscodes der Variablen V10 *Art der Evidenzquelle* und V11 *Qualität der Evidenzquelle* bilden addiert den Index A. Dieser hat einen Ergebnisraum von 0 bis 8. Maximal konnte folglich von einer Evidenzquelle bei Index A der Wert 8 erreicht werden und minimal der Wert 0.[85]

Diese Skala des Indexes A von 0 bis 8 wurde nun prozentuiert, denn zur gleichwertigen Verknüpfung der Komponenten *interne* und *externe Evidenz* zu einem gemeinsamen Gesamtindex mussten deren Skalen durch Prozentuierung vereinheitlicht werden. Meyer (2011) beschreibt, dass die Transformation von Skalen durch Prozentuierung sehr gut geeignet ist, um eine standardisierte Intervallskala für inhaltlich unterschiedliche Indikatoren herzustellen.

Die Ausprägungscodes der Variablen von V15 bis V23 bildeten die Komponente *interne Evidenz* ab (vgl. Kap. 7.2.2).[86] Die Komponente *interne Evidenz* wird gebildet, wie in Kapitel 3.2 aufgezeigt, aus den beiden Dimensionen *Argumentationsweise der einzelnen Evidenzquelle* und *Verwendung von*

85 Hat eine Evidenzquelle beim Index A den Wert 8, entspricht dieser einer hundertprozentigen Erfüllung der Messkategorien für die Komponente *externe Evidenz*. Der Wert 8 ist hier der maximal zu erreichende Indexwert und dieser wird gleichgesetzt mit 100 Prozent. Mit diesen zugeordneten 100 Prozent ist hier nicht gemeint, dass die Evidenzquelle zu 100 Prozent extern evident ist, explizit ist hier ausschließlich der Bezug auf die gemessenen Kategorien. Eine hundertprozentige respektive größtmögliche Erfüllung der Messkategorien zur dargestellten externen Evidenz ist für die Evidenzquelle bei dem Indexwert 8 erreicht. Die geringstmögliche Erfüllung der Messkategorien zur dargestellten externen Evidenz ist bei dem Indexwert 0 gegeben. Das kennzeichnet wiederum jedoch nicht, dass die Evidenzquelle gar nicht extern evident dargestellt ist. Dieser Indexwert bedeutet ausschließlich, dass diese Evidenzquelle die geringstmögliche Erfüllung der Messkategorien zur dargestellten externen Evidenz erreicht hat. Somit wird einer Evidenzquelle, die einen Indexwert von 0 hat, die geringstmögliche Erfüllung der Messkategorien von 0 Prozent zugeordnet.

86 Für diesen Indikator wurden im ersten Schritt die Ausprägungscodes der Variablen für die interne Evidenz der einzelnen Evidenzquelle addiert. Die Variablen V12, V13 und V14 wurden zusammengefasst in eine neue Kategorie *Differenzmenge der Argumente;* diese wurde im zweiten Schritt dazu addiert. Wurde bei der Variable V14 (Polarität der Evidenzquelle) die Ausprägung 1 codiert (die Evidenzquelle argumentiert für die Hauptthese des Beitrags), soll der Ausprägungscode von Variable V12 (Anzahl der Argumente pro) mit dem Ausprägungscode von Variable V13 (Anzahl der Argumente kontra) subtrahiert werden. Die Differenz wurde zur Ausprägung einer neuen Kategorie *Differenzmenge der Argumente* und diese wurde dann mit in den Index B addiert. Wurde bei der Variable V14 die Ausprägung 2 codiert (die Evidenzquelle argumentiert gegen die Hauptthese des Beitrags), soll der Ausprägungscode von Variable V13 (Anzahl der Argumente kontra) mit dem Ausprägungscode von Variable V12 (Anzahl der Argumente pro) subtrahiert werden. Diese Differenz wurde zur Ausprägung der Kategorie *Differenzmenge der Argumente* und wurde dann in den Index B addiert.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

evidenzstiftendem Bildmaterial der einzelnen Evidenzquelle. Die Ausprägungscodes der Variablen V23a, V23b und V23c (erfassen das verwendete evidenzstiftende Bildmaterial)[87] wurden für den zweiten Index B mit den Variablen für den Indikator *Argumentationsweise der einzelnen Evidenzquelle* addiert. Die Ausprägungscodes der Variablen des Indikators *Argumentationsweise der einzelnen Evidenzquelle* und des Indikators *Verwendung von evidenzstiftendem Bildmaterial der einzelnen Evidenzquelle* bildeten also addiert den zweiten Index B. Dieser hat einen Ergebnisraum von -5 bis 20. Maximal konnte folglich von einer Evidenzquelle hier der zweite Indexwert 20 erreicht werden und minimal der zweite Indexwert -5. Diese Skala des Indexes B von -5 bis 20 wurde nun, entsprechend der Transformation von Index A, prozentuiert.[88]

Der Gesamtindex, der die dargestellte Evidenz der Evidenzquelle repräsentiert, kombiniert die zwei Komponenten *externe* und *interne Evidenz* und verknüpft dabei die verschiedenen Dimensionen des theoretischen Konstrukts *dargestellte Evidenz* in eine einzige Skala. Gebildet wurde der Gesamtindex durch den Mittelwert von Index A und Index B, indem der Indexwert des Indexes A mit dem Indexwert des Indexes B addiert und dann durch zwei geteilt wird:

(Indexwert A + Indexwert B) : 2 = Gesamtindex

Beide Indizes der Komponenten des theoretischen Konstrukts bildeten so zu gleichen Teilen den Gesamtindex, der die dargestellte Evidenz der einzelnen Evidenzquelle repräsentiert.

Die für den jeweiligen Indikator gemessenen Merkmalsausprägungen wurden zu einem Zahlenwert reduziert, zu einem Index addiert und dieser wurde dann durch Prozentuierung standardisiert. Durch die Prozentuie-

87 Durch die Codierung von bis zu drei Bildern bekam die Kategorie *Bild* einen großen Einfluss im später gebildeten Index der internen Evidenz der jeweilig dargestellten Evidenzquelle. In Kapitel 3.2.2 wurde bereits aufgezeigt, warum es wichtig ist dem Bild eine entsprechend hohe Aussagekraft bei der Evidenzstiftung zuzubilligen.

88 Bei Index B kann von einer Evidenzquelle der höchstmögliche Indexwert von 20 erreicht werden. Dieser entspricht einer hundertprozentigen Erfüllung der Messkategorien für die Komponente *interne Evidenz*. Der Wert 20 ist hier der maximal zu erreichende Indexwert und dieser wird gleichgesetzt mit 100 Prozent. Mit diesen zugeordneten 100 Prozent ist hier nicht gemeint, dass die Evidenzquelle zu 100 Prozent intern evident ist, explizit ist hier ausschließlich der Bezug auf die gemessenen Kategorien. Eine hundertprozentige respektive größtmögliche Erfüllung der Messkategorien zur dargestellten internen Evidenz ist für die Evidenzquelle bei dem Indexwert 20 erreicht.

https://doi.org/10.5771/9783748918168-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

rung des Gesamtindexes bildet sich eine kompakte Intervallskala für inhaltlich unterschiedliche Indikatoren bzw. Komponenten (Meyer, 2011). Ziel der Indexbildung war es, die *dargestellte Evidenz* der einzelnen Evidenzquelle einfach und übersichtlich darzustellen. Die dargestellte Evidenz der einzelnen Evidenzquelle wird durch den Gesamtindexwert ausgedrückt.[89] Der prozentuale Gesamtindexwert ist des Weiteren die Voraussetzung für die weitere Berechnung der Evidenzmuster jedes Beitrags mit Hilfe der ETDS. Die numerische Evidenztheorie verlangt, dass die Evidenz der einzelnen Evidenzquelle durch Gewichte (Werte oder Prozente) ausgedrückt wird, die dann bei der Kombination der dargestellten Evidenzen benutzt werden (Spies, 2008). Die Voraussetzungen, um mit Hilfe der ETDS die einzelnen Evidenzquellen zusammenzuführen, sind nun erfüllt, da eine inhaltsanalytische Indexbildung der dargestellten Evidenz der einzelnen Evidenzquelle und deren Überführung in Prozente erfolgte.

7.2.3.2 Anwendung der Dempster & Shafer Evidenztheorie

Entscheidend für die Wahl der ETDS zur Berechnung der dargestellten Evidenz eines TV-Wissenschaftsbeitrags war die Möglichkeit Evidenzmaße der Hauptthese im TV-Wissenschaftsbeitrag ermitteln zu können. Evidenzen aus unterschiedlichen dargestellten Evidenzquellen konnten so verknüpft werden, dass Aussagen zur dargestellten Gesamtevidenz im TV-Wissenschaftsbeitrag möglich sind und auch die Konflikte zwischen gegebenen Evidenzen modelliert werden konnten.

89 Eine sehr niedrige Punktzahl würde, um ein praktisches Beispiel zu geben, folgende Evidenzquellendarstellung erhalten: Ein Experte, der nicht explizit benannt wird, gibt in seinen Ausführungen explizit an, dass er unsicher ist und dass die Gegenthese auch richtig sein kann. Ohne Hintergrundinformationen oder Erläuterungen zu geben, gibt er nur ein Argument für die Hauptthese des Beitrags an. Dieses Argument berührt inhaltlich nur peripher die Hauptdiskussion im Beitrag und ist im Konjunktiv verfasst. In seiner Argumentation verstrickt er sich in Widersprüche, und seine Aussagen sind weder neu noch konstant. Im Bild wird ausschließlich der Experte selbst gezeigt. Diese Evidenzquelle wird nun von der folgenden Evidenzquelle auch noch als sehr unglaubwürdig deklariert.
Eine sehr hohe Punktzahl würde, um auch das andere Extrem als praktisches Beispiel zu geben, folgende Evidenzquellendarstellung erhalten: Eine Meta-Analyse von randomisierten, kontrollierten Studien wird dargestellt. Es wird explizit und implizit gesagt, dass die Ergebnisse sicher sind und die Gegenthese falsch sein muss. Die Argumente sind zentral für die Diskussion der Hauptthese und werden mit Details, Hintergrundinformationen und Bildern, wie Diagrammen und Tabellen, erläutert. Die Argumentation dieser Evidenzquelle ist intern homogen, konstant und bietet inhaltlich viele neue Erkenntnisse, die für die Hauptthese des Beitrags sprechen. Alle folgenden Evidenzquellen im Beitrag unterstützen die Ausführungen dieser Evidenzquelle.

https://doi.org/10.5771/9783748275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Die Anwendung der ETDS unterliegt verschiedenen Grundannahmen und Bedingungen. Die Regel zur Kombination der Evidenzquellen ist sowohl assoziativ als auch kommutativ (Wu, 2003). Das heißt, dass das Ergebnis der Kombination unabhängig von der Reihenfolge ist, in der die Verknüpfung vorgenommen wird und unabhängig ist von der Reihenfolge der einzelnen Evidenzquellen (Dempster, 1967).[90]

Um die intersubjektive Nachvollziehbarkeit zu gewährleisten und dem Leser die Rechenregeln der ETDS näher zu bringen, werden diese im Folgenden aufgezeigt und parallel dazu wird eine Beispielberechnung durchgeführt.[91] Der Beispielbeitrag ist ein Element aus der Stichprobe dieser Untersuchung. Am 17.11.2011 wurde der Beitrag mit dem Thema *Gesundheit* und der Hauptthese *Kaffee macht den Menschen geistig fitter* bei Odysso im SWR gesendet. Drei Evidenzquellen werden in diesem Beitrag präsentiert. Auf das dargestellte Fallbeispiel, dessen Argumente für die Hauptthese sprechen, folgt im Beitrag eine Studie, deren Aussagen ebenfalls die Hauptthese stützen. Die dritte Evidenzquelle ist ebenfalls eine Studie, aber diese ist kontra die Hauptthese gerichtet und spricht dafür, dass Kaffee den Menschen nicht geistig fitter macht. Ein Beitrag, in dem mindestens eine Evidenzquelle für die Hauptthese spricht und mindestens eine Evidenzquelle gegen die Hauptthese gerichtet ist, wird folgend als *kontrovers* definiert. In dem gewählten Beispielbeitrag werden sowohl Evidenzquellen (e_i), die pro (e_i^+) als auch Evidenzquellen, die kontra (e_i^-) die Hauptthese gerichtet sind, präsentiert. Die dargestellte externe (Index A) und interne Evidenz (Index B) jeder Evidenzquelle im Beitrag wurde mit Hilfe der Inhaltsanalyse erfasst und durch einen prozentuierten Gesamtindexwert für jede Evidenzquelle repräsentiert, wie in Kapitel 7.2.3.1 beschrieben.

Die Evidenzquelle 1 (e_1^+) argumentiert für die Hauptthese und hat den Indexwert A von 5 und den Indexwert B von 7. Diese Indizes werden, wie in Kapitel 5.2 beschrieben, prozentuiert. Der Wert von Index A entspricht 62.5 Prozent und der Wert von Index B entspricht 35 Prozent des jeweils

90 Jede Evidenzquelle ist im Kontext der zu bewertenden bedingten Evidenz unabhängig (Spies, 2000). Keine Evidenzquelle darf eine andere Evidenzquelle bedingen (Jones, Lowe & Harrison, 2002). Allein unter diesen Voraussetzungen ist gewährleistet, dass jede neu eingeführte Evidenz eine These unabhängig bewerten kann und inkrementell kombiniert zu einer neuen Verteilung aller Evidenzen beiträgt (Berndt, 2009; Sentz & Ferson, 2002).

91 Die Regeln zur Berechnung wurden entnommen aus den Publikationen von Beierle und Kern-Isberner (2008), Berndt (2009), Boersch, Heinsohn und Socher (2007), Heinsohn und Socher-Ambrosius (1999), Parikh, Pont und Jones (2001), Salicone (2007), Sentz und Ferson (2002) und Spies (1993, 2008).

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

maximal zu erreichenden Indexwertes. Der Gesamtindex für die dargestellte Evidenz wird nun aus dem Mittelwert von Index A und Index B gebildet und beträgt für diese Evidenzquelle $e_1^+(H) = 48.8\% = .488$.

Die Evidenzquelle 2 (e_2^+) argumentiert gegen die Hauptthese und hat den Indexwert A von 5 und den Indexwert B von 13. Diese Indizes werden prozentuiert. Der Wert von Index A entspricht 62.5 Prozent und der Wert von Index B entspricht 65 Prozent des jeweils maximal zu erreichenden Indexwertes. Der Gesamtindex für die dargestellte Evidenz beträgt für diese Evidenzquelle $e_2^+(H) = 63.8\% = .638$.

Die Evidenzquelle 3 (e_1^-) ist kontra die Hauptthese gerichtet und hat den Indexwert A von 7 und den Indexwert B von 12. Diese Indizes werden prozentuiert. Der Wert von Index A entspricht 87.5 Prozent und der Wert von Index B entspricht 60 Prozent des jeweils maximal zu erreichenden Indexwertes. Der Gesamtindex für die dargestellte Evidenz beträgt für diese Evidenzquelle $e_1^-(H) = 73.8\% = .738$.

Die Gesamtindizes der unterschiedlich evident dargestellten Evidenzquellen können nun zusammengeführt werden, um das dargestellte Beliefmaß, das Doubtmaß, die unterstützende Plausibilität für das Zutreffen der Hauptthese, die konträre Plausibilität gegen das Zutreffen der Hauptthese und die dargestellte Ungewissheit im Gesamtbeitrag zu bestimmen. Als erstes wird nun ein Prowert (pro) und ein Kontrawert (kon) für die Hauptthese mit Hilfe dieser Formeln ermittelt:

$$\mathbf{pro(H) = 1 - (1-e_1^+(H))*(1-e_2^+(H))...(1-e_n^+(H))}$$

$$\mathbf{kon(H) = 1 - (1-e_1^-(H))*(1-e_2^-(H))...(1-e_n^-(H))}$$

Im Beispielbeitrag beträgt der Prowert:

$$pro(H) = 1 - ((1-0.488)*(1-0.638)) \approx .815$$

und der Kontrawert:

$$kon(H) = 1 - (1-0.738) \approx .738.$$

Diese zwei Werte können nun dazu genutzt werden die dargestellten Evidenzmaße zu berechnen. Das *Beliefmaß* kann mit Hilfe dieser Formel berechnet werden:

$$\mathbf{bel(H) = ((pro(H)*(1-con(H))/(1-pro(H)*con(H))}$$

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

und beträgt im Beispielbeitrag:

$$bel(H) = (0.815*(1-0.738))/(1-(0.815*0.738)) \approx .53 \rightarrow 53\%.$$

Die Belief-Funktion repräsentiert die Gesamtsumme der dargestellten Belege für das sichere Zutreffen der Hauptthese. Der Wert 0 bedeutet, dass kein Beleg für das Zutreffen der Hauptthese im Beitrag präsentiert wird. Der Wert 1 steht dafür, dass absolute Belegkraft für das Zutreffen einer Hauptthese präsentiert wird. Ein Beliefwert von .53 würde hier heißen, dass die Hauptthese, die im TV-Wissenschaftsbeitrag präsentiert wurde, so dargestellt ist, dass sie zu 53 Prozent belegt zutrifft.

Das *Doubtmaß* kann mit Hilfe dieser Formel berechnet werden:

$$\mathbf{bel^c(H) = ((con(H)*(1-pro(H))/(1-pro(H)*con(H))}$$

und beträgt im Beispielbeitrag:

$$bel^c(H) = (0.738*(1-0.815)/(1-(0.815*0.738)) \approx .34 \rightarrow 34\%.$$

Die Doubt-Funktion repräsentiert die Gesamtsumme der dargestellten Belege für das Nicht-Zutreffen der Hauptthese. Der Wert 0 bedeutet, dass kein Beleg für das Nicht-Zutreffen der Hauptthese im Beitrag präsentiert wird. Der Wert 1 steht dafür, dass absolute Belegkraft für das Nicht-Zutreffen der Hauptthese präsentiert wird. Ein Doubtwert von .34 würde heißen, dass die Hauptthese des TV-Wissenschaftsbeitrags so dargestellt ist, dass sie zu 34 Prozent belegt nicht zutrifft.

Das Belief- und das Doubtmaß haben jeweils die Grenzen 1 und 0. Wird die Hauptthese in einem Beitrag so präsentiert, dass sie sicher zutrifft, so ist der Beliefwert der Hauptthese 1 und der Doubtwert 0. Wird die Hauptthese in einem Beitrag so präsentiert, dass sie definitiv nicht zutrifft, so ist der Beliefwert der Hauptthese 0 und der Doubtwert 1. Der Doubtwert und der Beliefwert ergeben zusammengezogen eine Summe, die kleiner oder gleich 1 ist: $bel^c(H)+bel(H) \leq 1$.

In wahrscheinlichkeitstheoretischen Berechnungen ergibt das Beliefmaß zusammen mit dem Doubtmaß immer genau 1. Es wird also davon ausgegangen, dass die Hauptthese entweder eintritt oder nicht, bzw. so dargestellt ist, dass sie entweder belegt oder widerlegt wird. Existieren aber Evidenzen, also nur vages oder unvollständiges Wissen, und Gegenevidenzen, ist es sinnvoll mit Plausibilitäts- und Ungewissheitsmaßen zu

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

rechnen (Spreckelsen & Spitzer, 2008). Die dargestellte *Ungewissheit* un(H) lässt sich mittels des Belief- und Doubtmaßes berechnen:

$$\mathbf{un(H) = 1- bel^c(H)-bel(H)}$$

Das Ungewissheitsmaß der Hauptthese im Beispielbeitrag beträgt:

$$un(H) = 1\text{-}0.34\text{-}0.53 = .13 \rightarrow 13\%.$$

Je kleiner das Ungewissheitsintervall ist, desto spezifischer ist die Evidenz dargestellt. Ein Ungewissheitswert von 1 bedeutet, dass völlige Ungewissheit dargestellt wird, indem keine Belege für oder gegen einen Sachverhalt präsentiert werden. Ein Ungewissheitswert von .13 würde hier heißen, dass die Hauptthese, die im TV-Wissenschaftsbeitrag präsentiert wurde, so dargestellt ist, dass es zu 13 Prozent ungewiss ist, ob sie zutrifft und es in einem Intervall von 13 bis 100 Prozent wahrscheinlich ist, dass sie entweder sicher zutrifft oder sicher nicht zutrifft.

Werden Belief-, Ungewissheits- und Doubtwert einer Hauptthese miteinander addiert, muss dies immer 1 ergeben → $bel(H)+un(H)+bel^c(H) = 1$. Sobald eine neue Evidenz hinzukommt, müssen nach Dempster und Shafer alle verfügbaren Evidenzen neu skaliert werden, so dass die Summe der Evidenzmaße wieder 1 ergibt.

Die unterstützende Plausibilitätsfunktion ermittelt, bis zu welchem Maß die dargestellten Evidenzen dafür sprechen, dass die Hauptthese im günstigsten Fall zutreffen kann. Es betrachtet alle Evidenzen, die gegen die dargestellte Hauptthese sprechen. Der Wahrscheinlichkeitsbereich für das Zutreffen der Hauptthese ist begrenzt durch den Doubtwert. Das unterstützende Plausibilitätsmaß ist die Wahrscheinlichkeitsmasse, die nicht dem Doubtmaß zugeteilt werden kann. Die unterstützende Plausibilität pl(H) wird berechnet, indem der Doubtwert, also der Zweifel an das Zutreffen der Hauptthese, von 1 subtrahiert wird:

$$\mathbf{pl(H) = 1\text{-}bel^c(H)}$$

Im Beispielbeitrag beträgt der unterstützende Plausibilitätswert:

$$pl(H) = 1\text{-}0.34 = .66 \rightarrow 66\%.$$

Das heißt, dass die gegebenen Belege hier so dargestellt sind, dass es zu 66 Prozent plausibel ist, dass die Hauptthese zutrifft.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Das konträre Plausibilitätsmaß ist die Wahrscheinlichkeitsmasse, die nicht dem Beliefmaß zugeteilt werden kann. Folglich begrenzt das Beliefmaß der Hauptthese das Plausibilitätsmaß. Dieses wird mit folgender Formel berechnet:

$$\mathbf{pl^c(H) = 1\text{-}bel(H)}$$

Im Beispielbeitrag beträgt der konträre Plausibilitätswert:

$$pl^c(H) = 1\text{-}0.53 = .47 \rightarrow 47\%.$$

Das heißt, dass die gegebenen Belege hier so dargestellt sind, dass es zu 47 Prozent plausibel ist, dass die Gegenthese zutrifft.

Einige Zusammenhänge der Evidenzmaße werden nun zum besseren Verständnis für den Leser aufgeführt:

- Das unterstützende Plausibilitätsmaß und das konträre Plausibilitätsmaß bilden zusammen eine Summe, die größer oder gleich 1 ist. → $pl(H)+pl^c(H) \geq 1$
- Am besten gestützt ist eine Hauptthese bzw. am stärksten evident dargestellt ist eine Hauptthese, wenn der Beliefwert und der unterstützende Plausibilitätswert 1 sind. → $bel(\Omega) = pl(\Omega) = 1$
- Am schlechtesten gestützt ist eine Hauptthese, wenn der Beliefwert und der unterstützende Plausibilitätswert 0 sind. → $bel(\Omega) = pl(\Omega) = 0$
- Das Belief- und unterstützende Plausibilitätsmaß sind duale Kapazitäten, wobei der Beliefwert immer kleiner oder maximal gleich dem unterstützenden Plausibilitätswert ist. → $bel(H) \leq pl(H)$

In der Abbildung 7 sind die Evidenzmaße des Beispielbeitrags modelliert.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

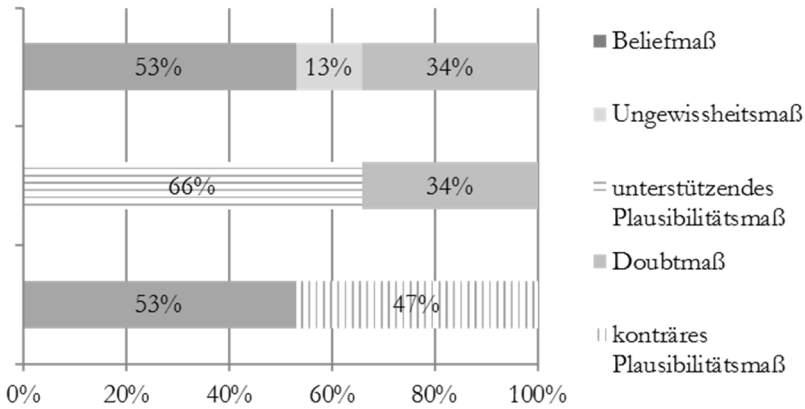

Abbildung 7: Evidenzmaße des Beispielbeitrags.

Ein Beitrag, in dem entweder ausschließlich Evidenzquellen präsentiert werden, die für die Hauptthese argumentieren oder ausschließlich Evidenzquellen, die gegen die Hauptthese argumentieren, wird folgend als *monodirektional* definiert. Werden in einem monodirektionalen TV-Wissenschaftsbeitrag ausschließlich Evidenzquellen (e_i) präsentiert, die die Hauptthese unterstützen, so ist der Beliefwert des Beitrags wie folgt zu berechnen:

$$\mathbf{bel(H) = 1 - (1 - e_1(H))*(1 - e_2(H))...(1 - en(H))}$$

Ein Beliefwert bspw. von .9 würde hier heißen, dass die Hauptthese, die im TV-Wissenschaftsbeitrag präsentiert wurde, so dargestellt ist, dass sie zu 90 Prozent sicher zutrifft und es in einem Intervall von 90 bis 100 Prozent wahrscheinlich ist, dass sie zutrifft. Der Doubtwert beträgt 0 Prozent, wenn keine Evidenz dafür präsentiert wird, dass die Hauptthese nicht zutreffen könnte. Es bleibt allerdings mitunter eine gegebene Ungewissheit. Die dargestellte Plausibilität für eine Hauptthese in einem Beitrag ohne Doubtmaß beträgt 100 Prozent und die konträre Plausibilität 0 Prozent.

Werden in einem monodirektionalen TV-Wissenschaftsbeitrag ausschließlich Evidenzquellen (e_i) dargestellt, die die Hauptthese anzweifeln, so wird der Doubtwert des Beitrags wie folgt, berechnet:

$$\mathbf{bel^c(H) = 1 - (1 - e_1(H))*(1 - e_2(H))...(1 - en(H))}$$

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Der Beliefwert beträgt 0 Prozent, denn es wird keine Evidenz dafür geliefert, dass die Hauptthese zutreffen könnte.

Für jeden Beitrag wurde ein Muster der dargestellten Evidenz mit Hilfe der ETDS berechnet. Für jeden Beitrag der Stichprobe liegt also ein Belief-, Doubt-, und Ungewissheitsmaß sowie unterstützendes und konträres Plausibilitätsmaß für die jeweilige Hauptthese vor. Diese Maße sind nur in bestimmter Kombination mathematisch logisch möglich. Das Evidenzmaßmuster eines jeden Beitrages ist also ein empirischer Befund innerhalb eines mathematisch engeschränkten Möglichkeitsraumes. Genau dies ist aber die Voraussetzung, um eben einen vergleichbaren Maßstab zwischen den Beiträgen zu entwickeln. Es kommt dann auf die genauen Werte der Maße an, die aussagekräftig sind für die Identifizierung der Evidenzmaßmuster. Um die Evidenzdarstellungsmuster über alle TV-Wissenschaftsbeiträge hinweg zu identifizieren, flossen diese Evidenzmaße jeden Beitrags als clusterbildende Variablen in eine Two-Step-Clusteranalyse ein.

7.2.3.3 Frame-Identifizierung durch Clusteranalyse

Wenn sich die Ausprägungen der frame-bildenden Elemente, die Evidenzmaße, in einer charakteristischen Weise in jedem Beitrag gruppieren, so verschiedene Muster formen, und wenn ein Muster dann über mehrere Beiträge hinweg identifiziert werden kann, dann kann von einem Frame gesprochen werden (Entman, 1993; Matthes & Kohring, 2004; Scheufele, 2010). Ist dieser Frame generalisiert, themenunabhängig und bildet ausschließlich die Struktur der Evidenzdarstellung ab, dann kann von einem formal-abstrakten Frame ausgegangen werden (vgl. Kapitel 4.3).

Am reliabelsten lassen sich, nach Matthes und Kohring (2004), Frames mit Hilfe einer Clusteranalyse erfassen (vgl. auch Dahinden, 2006; Potthoff, 2012). Die Clusteranalyse gehört zu den multivariaten Verfahren der explorativen Datenanalyse (Marcinkowski et al., 2008). Der Grundgedanke der Clusteranalyse liegt in der Reduktion umfangreicher Daten. Clusteranalysen können Daten so gruppieren, dass die Daten innerhalb eines Clusters möglichst ähnliche Variablenausprägungen aufweisen (Intracluster-Homogenität) und die Daten verschiedener Cluster möglichst unähnlich sind (Intercluster-Heterogenität; Backhaus, Erichson, Plinke & Weiber, 2006; Schendera, 2010). Um homogene Gruppen der dargestellten Evidenz in TV-Wissenschaftsbeiträgen aufzufinden, ist es also zielführend eine Clusteranalyse zu nutzen.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Die Struktur der dargestellten Evidenz in den einzelnen Beiträgen konnte mit Hilfe der ETDS berechnet werden. Werden ähnliche Evidenzmaß-Strukturen signifikant oft präsentiert, so können diese in Gruppen zusammengefasst werden, welche die formal-abstrakten Evidenzframes repräsentieren. Um diese Frames zu selektieren, werden ausschließlich die berechneten Evidenzmaße als framebildende Beitragsvariablen zur Musterbildung mit Hilfe der Clusteranalyse herangezogen. Die Evidenzmaße selber sind, wie in Kapitel 7.2.3.2 beschrieben, entstanden aus Indizes, die aus einer Vielzahl von Variablen gewonnen wurden. Shafer (1976) hat die fünf Evidenzmaße als notwendig und hinreichend für die Evidenzbestimmung eines Sachverhalts definiert.[92]

Ziel der Clusteranalyse ist es hier keineswegs, jeden Beitrag eindeutig einem Frame zuzuordnen, sondern jeder Beitrag hat je nach Ausprägung der Evidenzmaße eine bestimmte Wahrscheinlichkeit, einer latenten Gruppe von Beiträgen anzugehören (Matthes, 2008). Als Charakteristik von Clusterverfahren gilt es, dass alle vorliegenden Eigenschaften der Untersuchungsobjekte gleichzeitig und gleichgewichtet zur Gruppenbildung herangezogen werden (Backhaus et al., 2006; Schendera, 2010). Matthes und Kohring (2004) empfehlen zur Frame-Analyse eine hierarchische Clusteranalyse durchzuführen. Als Clusteranalyseverfahren wurde hier der Two-Step-Clusteranalyse Vorrang gegeben. Die Two-Step-Clusteranalyse ist eine Kombination der Verfahren der hierarchischen Clusteranalyse und der Clusterzentrenanalyse. Janssen und Laarz (2007) beschreiben, dass in der ersten Stufe der Clusterung alle Fälle sequenziell abgearbeitet und Subcluster mit sehr ähnlichen Fällen gebildet werden. In der zweiten Stufe werden diese dann mittels eines agglomerativen hierarchischen Verfahrens zu den eigentlichen Clustern fusioniert. Die Two-Step-Clusteranalyse ist insbesondere für Untersuchungen mit mehr als 250 Fällen geeignet (Schendera, 2010). Die clusterbildenden Variablen/Evidenzmaße weisen, wie von der Two-Step-Clusteranalyse vorausgesetzt, kein gemischtes Skalenniveau auf und wurden zuverlässig gemessen.

Die Clusteranalyse ist zu den Interdependenzanalyseverfahren zu zählen. Diese erfordern keine Vorspezifikation in abhängige und unabhängige Variablen. Gefordert werden allerdings möglichst voneinander unabhängige framebildende Elemente. Die Evidenzmaße korrelieren aber mitunter

92 Schendera (2010) zeigt auf, dass die Anzahl der Variablen Einfluss auf die Ergebnisse der Clusteranalyse hat. Die Variablenanzahl jedoch künstlich durch irrelevante Variablen hoch zu setzen, sei nicht sinnvoll und nicht notwendig. Die Variablen, welche zur Clusterbildung benutzt wurden, sind inhaltlich begründet worden und dementsprechend festgesetzt.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

aufgrund der gemeinsamen Basis zu ihrer Berechnung signifikant miteinander.[93] Die fünf Maße sind unterschiedliche Funktionen auf der Basis von zwei empirischen Indizes pro Beitrag. Die Maße sind deswegen mathematisch nicht unabhängig voneinander, das heißt, sie können nicht jede beliebige Kombination von Werten annehmen. Da jedes Evidenzmaß in dieser Untersuchung als wichtig für die Clusteranalyse definiert wurde, konnte keine Variable als überflüssig (nicht notwendig oder nicht hilfreich) charakterisiert und weggelassen werden. Viele Autoren weisen in diesem Zusammenhang darauf hin, dass die Two-Step-Clusteranalyse ein sehr robustes Verfahren ist, dass auch bei Verletzung der Unabhängigkeitsannahme zu brauchbaren Ergebnissen führt (bspw. Janssen & Laatz, 2005; Schendera, 2010).

Ein Vorteil der Two-Step-Clusteranalyse ist es, dass keine Clusterzahl vorgegeben werden muss, sondern diese wird automatisch bestimmt (Marcinkowski et al., 2008). Da sich aus der theoretischen Analyse in dieser Untersuchung keine Anhaltspunkte für eine Clusterzahl ergaben, wurde diese automatisch ermittelt. Es kann jedoch aufgrund des eingeschränkten mathematischen Möglichkeitsraumes, in dem sich die Werte der Evidenzmaße bewegen, plausibel auf mögliche Muster geschlossen werden. Möglich ist bspw. ein Evidenzmaßmuster mit starker, eines mit mittlerer und eines mit schwacher einseitiger Evidenz sowie ein Evidenzmaßmuster mit starker konfligierender, eines mit mittlerer und eines mit schwacher konfligierender Evidenz. Es kommt auf die Mittelwerte der Clusterlösungen an und auf die Anzahl der Beiträge, die eben ähnliche Evidenzmaße haben und gruppiert werden.

Im Anschluss an die Clusterung wurde zur Überprüfung der Clusterlösung eine Diskriminanzanalyse durchgeführt. Im folgenden Kapitel wird auf die Diskriminanzanalyse und die weitere Sicherung der Güte der Inhaltsanalyse und der Auswertungsstrategie explizit Bezug genommen.

93 Für metrische Variablen diente zur Prüfung der Unabhängigkeit der Signifikanztest für Korrelationskoeffizienten nach Pearson und ergab mitunter signifikante Zusammenhänge. Überlegt wurde eine Faktorenanalyse vorzuschalten, um dem Problem der Korrelationen entgegen zu wirken, diese erbrachte aber keine sinnvollen Ergebnisse bzw. war die Faktorenanalyse mit einem erheblichen Informationsverlust verbunden.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

7.2.4 Reliabilität und Validität

Folgend werden die Güte der Inhaltsanalyse, der Indexbildung, der Anwendung der ETDS und der Clusteranalyse getrennt voneinander diskutiert.

Inhaltsanalyse

Zunächst wird die Güte der Stichprobe für die Inhaltsanalyse diskutiert. Nachfolgend werden die Inferenzvalidität, die Konstruktvalidität und die Reliabilität der Inhaltsanalyse reflektiert.

Die Inhaltsanalyse wurde anhand der Berichterstattung über medizinische Sachverhalte durchgeführt, da nur in der Medizin valide Evidenzlevel existieren (vgl. Kapitel 3.1.1), welche für die Erstellung des Codebuchs wichtig waren (vgl. Kapitel 7.2.2). Da es um die Identifizierung formalabstrakter, also themenunabhängiger, Darstellungsmuster ging, wurden alle Variablen im Codebuch themenunspezifisch definiert. Das heißt, dass das Codebuch auch auf die Berichterstattung über andere wissenschaftliche Sachverhalte angewendet werden kann. Die errechnete erforderliche Stichprobe, um formal-abstrakte Evidenzdarstellungsmuster in TV-Wissenschaftsbeiträgen der öffentlich-rechtlichen Fernsehsender identifizieren zu können und um Unterschiede in den Darstellungen der einzelnen Evidenzquellen hypothesenkritisch zu untersuchen, wurde in der Untersuchung erreicht (vgl. Kapitel 7.2.1).

Die Inferenzvalidität definiert das Ausmaß, in dem die in der Inhaltsanalyse gezogenen Schlussfolgerungen gültig sind. Inferenzen werden in der Inhaltsanalyse nur insofern gezogen, dass von der Stichprobe auf die Grundgesamtheit geschlossen wird. Inferenzen zur Wirkung des Medieninhalts auf den Rezipienten werden erst mit Hinzunahme des zweiten Untersuchungsteils, in dem die Überzeugungsurteile der Rezipienten erfasst werden, getätigt. Obwohl die Ergebnisse aus der Inhaltsanalyse als gut generalisierbar gelten können, werden die Erkenntnisse, wie alle Erkenntnisse in der Kommunikationswissenschaft, ausschließlich als probabilistisch angesehen (Brosius et al., 2012; Stocking, 2010). Es werden also keine Gesetzmäßigkeiten mit Hilfe der Inhaltsanalyse herausgestellt, nur bestimmte Wahrscheinlichkeiten, unter denen die Hypothesen und Erkenntnisse gelten, aufgezeigt. Durch die sehr große Stichprobe (zeitlich begrenzte Vollerhebung) ist dieser Untersuchung eine hohe Inferenzvalidität bei der Übertragung der Ergebnisse auf die Grundgesamtheit zu attestieren. Aufgrund der Fülle der erfassten Beiträge aus öffentlich-rechtlichen

https://doi.org/10.5771/9783845232968-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

TV-Wissenschaftsmagazinen können valide verallgemeinerbare Aussagen zur dort dargestellten Evidenz über medizinische Sachverhalte gemacht werden.

Das Konstrukt *dargestellte Evidenz* bzw. die Konstrukte *interne* und *externe Evidenz* wurden in dieser Untersuchung erstmals definiert. Die Konstruktvalidität der dargestellten Evidenz kann also in dieser Untersuchung nicht durch vergleichbare Studien gesichert werden. Um die Vollständigkeit des Erhebungsinstrumentes zu sichern, wurde versucht, das Konstrukt als auch das Kategorienschema anhand eines umfassenden Forschungsüberblicks, anhand praktischer Vorerfahrungen mit dem Untersuchungsmaterial und anhand theoretischer Vorüberlegungen mit Verweis auf empirische Studien zu validieren. Denkbar wären mitunter aber noch weitere Variablen gewesen, welche bei der Belegkraft von Argumentationen eine Rolle spielen könnten. Für diese fehlten aber empirische Studien, die diese Rolle begründen. Beispielsweise bei der bildlichen Kategorie könnten noch Faktoren, wie die Perspektive, die ästhetische Darstellungsstrategie oder die Bild-Text-Beziehung, relevant sein.

Es wurde im Codebuch darauf geachtet, dass die Definitionen der Kategorien sowie die Definitionen der einzelnen Ausprägungen trennscharf sind. Die Güte der Inhaltsvalidität in dieser Untersuchung, als Teilaspekt der Konstruktvalidität, wird auch dadurch gestützt, dass das Codebuch auf alle Beiträge der Stichprobe anwendbar war. Des Weiteren mussten Auffangwerte nur sehr selten codiert werden.

Zwei Codierer sind für die Codierung der TV-Wissenschaftsbeiträge eingestellt worden.[94] Durch die Codierung von mehreren Codierern konnten Erwartungseffekte der Forschungsleiterin sowie Bias kontrolliert werden. Um die Reliabilität zu sichern, wurden möglichst eindeutige Codierregeln, Kategorienschemata und Begriffsdefinitionen im Codebuch festgelegt, den Codierern Beispielcodierungen gegeben und drei intensive Codiererschulungen durchgeführt, so dass eine verlässliche Messung möglich war. Insgesamt wurden in den Codiererschulungen 20 Beispielbeiträge gemeinsam codiert und diskutiert.

Jede Reliabilitätsmessung beruht auf der Idee der Messwiederholung anhand desselben Materials und sagt sowohl etwas über die Güte des Codebuchs als auch über die Sorgfalt der Codierer aus (Früh, 2011; Rössler, 2010). Bei wiederholter Anwendung des Codebuchs müssen intersubjektiv dieselben Ergebnisse geliefert werden. Die Prüfung der Reliabilität

94 Hier sei nochmal ein großes Dankeschön an meine beiden Codierer C. Knigge und M. Bayer ausgesprochen. Das war sehr gute Arbeit!

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

fand in dieser Untersuchung sowohl während der Pretestphase nach der zweiten Codiererschulung als auch nach der Codierung statt. Die für die Reliabilitätstests ausgewählten Beiträge sind Bestandteil des insgesamt codierten Materials und wurden zufällig daraus gezogen. In den Reliabilitätstest, der nach Abschluss der Codierungen gerechnet wurde, flossen 30 Beiträge ein. Da einige Variablen je nach Anzahl der vorkommenden Evidenzquellen (V9) pro Beitrag öfter codiert wurden, flossen einige Variable sogar mit 52 bzw. 57 Codierungen in den Reliabilitätstest ein.[95]

Untersucht wurde die Intercoderreliabilität mit Hilfe des Reliabilitätstests nach Holsti (1969) und mit Hilfe von Cohens Kappa (Cohen, 1960). Durch Cohens Kappa kann eine Korrektur für die Anzahl zufälliger Übereinstimmungen erfolgen (Wirtz & Caspar, 2002). Die Reliabilitäten sind in Tabelle 7 für die formalen, wertenden und inhaltlichen Variablen getrennt aufgeführt, da für die formalen Variablen eine höhere Reliabilität gefordert werden muss als für die wertenden und inhaltlichen Variablen. Es wird bei den einzelnen untersuchten Variablen auch dokumentiert, wie viele Ausprägungen überhaupt möglich waren. Bei Variablen mit nur zwei Ausprägungen liegt die Zufallswahrscheinlichkeit für einen Treffer schon bei einem Reliabilitätskoeffizienten von .50. Nach Holsti wird somit bei diesen Variablen ein höherer Reliabilitätskoeffizient gefordert. Bei Variablen mit vielen Ausprägungen (V2, V3, V4, V5, V6, V7 und V9) können Zufallstreffer quasi ausgeschlossen werden. Die Beurteilung von Reliabilitätskoeffizienten erfolgte deshalb in dieser Untersuchung anhand des Schwierigkeitsgrades und Art der jeweiligen Variable. Tendenziell sind für inhaltliche Variablen Werte nach Holsti ab .80 und nach Cohens Kappa ab .40 und für formale Variablen Werte nach Holsti und nach Cohens Kappa nahe 1 zu fordern (Neuendorf, 2002; Rössler, 2010; Wirtz & Caspar, 2002).

95 Im Rahmen der ersten Codiererschulung wurde offensichtlich, dass Variable V9 *Anzahl der Evidenzquellen* eine kritische Variable ist. Daher wurde sich in der zweiten Codiererschulung intensiv und sehr ausführlich mit dieser befasst. Sollten zwei Codierer zu einem Beitrag bei Variable V10 *Evidenzquellenart* eine ganz unterschiedliche Evidenzquelle codiert haben, so fielen alle folgenden Variablen, die sich auf diese Evidenzquelle bezogen, aus dem Reliabilitätstest. Da die Codierer dann auf Basis differenter Evidenzquellenarten weitercodierten. Die Folgevariablen sind somit nicht mehr miteinander vergleichbar. Insgesamt fielen von den 57 Fällen fünf Fälle aus dem Reliabilitätstests der Variablen V11 bis V23.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Tabelle 7: Intercoderreliabilitätswerte getrennt nach Variablen

Variable	Art der Variable	Anzahl möglicher Ausprägungen	Anzahl der Codierungen	Holsti (Reihenfolge berücksichtigt)	Cohens Kappa
V1 *Codierer*	formal	3	30	0	0
V2 *Wissenschaftsmagazin*	formal	9	30	1	1
V3 *Ausstrahlungsdatum*	formal	1.08.11 - 31.05.12	30	1	1
V4[a] *Beitragsdauer*	formal	0 bis x	30	1	1
V5 *Thema speziell*	inhaltlich	1 bis x	30	.93	.92
V6 *Thema allgemein*	inhaltlich	1 bis x	30	.91	.90
V7[b] *Hauptthese*	inhaltlich	x	30	.93	.92
V9 *Anzahl der Evidenzquellen*	inhaltlich	1 bis x	30	.98	.82
V10 *Evidenzquellenart*	inhaltlich	6	57	.93	.91
V11 *Validität der Evidenzquelle*	inhaltlich	4	52	.90	.85
V12 *Argumente pro*	inhaltlich	3	52	.81	.63
V13 *Argumente kontra*	inhaltlich	3	52	.92	.73
V14 *Polarität*	inhaltlich	3	52	.99	.92
V15 *Gewichtung*	inhaltlich	2	52	.96	.58
V16 *Neuigkeit*	inhaltlich	2	52	.95	.54
V17 *Unsicherheit – explizit*	inhaltlich	3	52	.86	.39
V18 *Unsicherheit – implizit*	inhaltlich	3	52	.79	.45
V19 *Homogenität*	inhaltlich	2	52	.95	.51
V20 *Detaillierung*	inhaltlich	2	52	.92	.41
V21 *Konstanz*	inhaltlich	3	52	.93	.71
V22 *Sekundäre Bewertung*	inhaltlich	3	52	.67	.42
V23a *Bild1*	inhaltlich	4	52	.90	.63
V23b *Bild 2*	inhaltlich	4	52	.80	.58
V23c *Bild 3*	inhaltlich	4	52	.84	.54

[a] Bei der Messung der Beitragsdauer wurde ein Toleranzintervall von 5 Sekunden eingeräumt, da der Übergang von Moderation und Beitrag mitunter 2-3 Sekunden dauern kann.

[b] Hier wurde nicht die genaue, wörtliche Übereinstimmung gemessen, sondern die Übereinstimmung im Sinn bzw. Grundgedanken der Hauptthese.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Manche Variablen haben mitunter schlechtere Reliabilitätswerte nach Cohen, da ein unzureichender Variationsgrad der Merkmalsausprägungen einzelner Variablen bei den ausgewählten Beiträgen ausgewiesen ist. Beispielsweise V17 *Unsicherheit – explizit* und V20 *Detailreichtum* weisen sehr gute Reliabilitätswerte nach Holsti, aber nur mittelmäßige Reliabilitätswerte nach Cohen auf. Das liegt an den sehr liberalen statistischen Eigenschaften des Holsti-Koeffizienten bzw. an den sehr konservativen statistischen Eigenschaften von Cohens Kappa (Potthoff, 2012). Bei V17 wurde sehr selten die Ausprägung -1 codiert und bei der Variable V20 kein Mal die Ausprägung 2. Da die Stichprobe für den Reliabilitätstest zufällig gezogen wurde, kann es passieren, dass bspw. überdurchschnittlich viele Beiträge herangezogen werden, bei denen die Evidenzquellen detailreiche (V20) Argumente ohne explizite Anzeichen für Unsicherheit (V17) aufzeigen. Aber auch wenn in allen Beiträgen der Untersuchung nur Evidenzquellen wären, die ausschließlich detailreiche Argumente ohne explizite Anzeichen für Unsicherheit aufzeigen, ist dies ein wertvolles Ergebnis für diese Untersuchung. Alle Variablen sind in ihren Ausprägungen und in ihrer Existenz begründet und so fließen auch die Variablen V17 und V20 mit in die Untersuchung ein. Der traditionell berechnete Reliabilitätskoeffizient nach Holsti ist hier oft aussagekräftiger und Cohens Kappa eher unbrauchbar (Potthoff, 2012). Demnach urteilt Cohens Kappa insbesondere bei dichotomen Variablen zu streng, da zu große Teile der Übereinstimmung dem Zufall zugeschrieben werden. Ein mittelmäßiger Wert nach Cohens Kappa spricht hier also nicht grundsätzlich gegen die Reliabilität oder Validität der Variablen. Ähnlich ist es bei der Variable V22 *sekundäre Bewertung*; hier wurde von den drei Codierern in den Beiträgen des Reliabilitätstests nur einmal die Ausprägung -1 codiert. Ein Wert nach Holsti von rund .70 ist aufgrund der Komplexität dieser Variable noch in einem akzeptablen Bereich. Der schlechte Reliabilitätswert nach Cohen bei V18 *Unsicherheit – implizit* ist ebenfalls zurückzuführen auf die hohe Komplexität der Variable und eine zu geringe Variation der Merkmalsausprägungen in den Beiträgen, die in den Reliabilitätstest einflossen.

Die Intercoderreliabilität beträgt für die formalen und wertenden Variablen in dieser Untersuchung insgesamt nach Holsti und nach Cohens Kappa 1 und ist somit sehr gut. Für die inhaltlichen Variablen beträgt die Intercoderreliabilität insgesamt nach Holsti .89 und nach Cohen .67. Die Reliabilität der Inhaltsanalyse ist somit auch für die inhaltlichen Variablen zufriedenstellend. Die Analysevalidität insgesamt kann als gut angesehen werden.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Indexbildung

Voraussetzung für die Bildung eines additiven Indexes ist ein einheitliches Skalenniveau (Brosius et al., 2012; Meyer, 2011). Alle relevanten Variablen wurden einheitlich ordinal skaliert erhoben, um die Voraussetzungen für die Indexbildung zu erfüllen. Des Weiteren ist jede dieser Variablen so skaliert, dass sie einen Extremwert hat, der theoretisch nicht überschritten werden kann, aber bei einer Messung real gleichwahrscheinlich auftreten könnte. Auch diese Voraussetzungen wurden erfüllt, um die Indizes für die externe und interne Evidenz zu generieren, deren Skalen dann durch Prozentuierung standardisiert werden konnten (Meyer, 2011). Weitere Voraussetzung der Indexbildung ist es, dass die Skalen der Indikatoren durch Standardisierung auch in ihrer Ausprägungsvarianz vereinheitlicht werden (Meyer, 2011). In dieser Untersuchung wurde die Anzahl der einzelnen Skalenpunkte pro Variable schon mit Bezug auf die Indexbildung inhaltlich begründet vergeben (vgl. Kapitel 7.2.3.1). Folglich ist die Ausprägungsvarianz nicht einheitlich gehalten. Dies ist durchaus als kritisch zu reflektieren, aber auch eine z-Transformation der Skalen wäre hier aufgrund des Skalenniveaus keine Lösung gewesen, denn über ordinale Daten lässt sich (meist) keine z-Standardisierung sinnvoll berechnen (Meyer, 2011).

Auch die Wertigkeit der einzelnen Variablen, welche in die Indizes einflossen, ist unterschiedlich, da in den Index für die externe Evidenz weniger Variablen einflossen als in den Index für die interne Evidenz. Wären weitere oder andere Variablen erfasst worden, welche mit in die Indizes einfließen würden, so wäre eventuell auch die Clusteranalyse zu anderen Ergebnissen gekommen. Alle Variablen flossen theoretisch begründet in die Indizes ein. Weil der bisherige Forschungsstand keine Lösungen für eine Gewichtung aufzeigte, wurden bei der internen Evidenz die Dimensionen der Argumentationsweise und des evidenzstiftenden Bildmaterials nahezu gleich gewertet, wie in Kapitel 7.2.3.1 beschrieben.

Die externe und interne Evidenz in dieser Untersuchung wurden, wie schon erwähnt, in dieser Untersuchung das erste Mal getrennt voneinander untersucht. Deswegen konnten keine validen Aussagen zum Zusammenhang zwischen externer und interner Evidenz gemacht werden und die Gewichtung der Indizes wurde ebenfalls pragmatisch entschieden. Die interne und die externe Evidenz flossen zu genau gleichen Teilen in einen Index der dargestellten Evidenz pro Evidenzquelle ein. Zur gleichwertigen Verknüpfung der Komponenten interne und externe Evidenz zu einem

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

gemeinsamen Gesamtindex wurden deren Skalen durch Prozentuierung vereinheitlicht, wie von Meyer (2011) empfohlen. So war es möglich durch die Transformation der Skalen eine standardisierte Intervallskala für inhaltlich unterschiedliche Indikatoren herzustellen.

Die pragmatischen Lösungen zur Indexbildung waren notwendig für die weitere Berechnung mit Hilfe der ETDS. Die Anwendung der ETDS kann, aufgrund der Einschränkungen und Vorläufigkeiten in ihrer Anwendung, auch in Bezug auf die Indexbildung, nur als explorativer Vorschlag zur systematischen Erfassung der dargestellten Evidenzmaße in journalistischen Beiträgen gelten.

Anwendung der ETDS

Die Anwendung der ETDS in dieser Untersuchung bot viele Möglichkeiten und Vorteile, aber auch einige Nachteile. Ein Nachteil der Anwendung der ETDS war, neben den bereits erläuterten Schwierigkeiten bei der Indexbildung und dem zusätzlichen Berechnungsaufwand, die allgemeine Unkenntnis der Theorie in der Kommunikationswissenschaft. Diese Nachteile werden allerdings kompensiert durch den Vorteil, dass die ETDS zu aussagekräftigen Ergebnissen bei der Beschreibung der dargestellten Evidenz kam und dabei die dargestellte Ungewissheit mit einbezog. Die ETDS konnte bestätigt werden als effektive Methode, um gegebene Evidenzen und Ungewissheit zu modellieren (Cortes-Rello & Golshani, 1999; Sentz & Ferson, 2002). Die Informationen, welche TV-Wissenschaftsbeiträge vermitteln wollen, werden von unterschiedlichen, dargestellten Evidenzen gestützt oder widerlegt und werden somit auch unterschiedlich ungewiss dargestellt. Ein Berechnungssystem, welches die dargestellte Evidenz in TV-Wissenschaftsbeiträgen modellieren soll, muss fähig sein auch dargestellte Ungewissheit aufzuzeigen. Ungewissheit und Evidenz sind untrennbar miteinander verbunden (vgl. Kapitel 3.1). Die Anwendung der ETDS in dieser Untersuchung bot auch den Vorteil, dass eine Evidenz, die nur teilweise oder schwach für eine Hauptthese spricht, nicht gleichzeitig auch als Evidenz berechnet wurde, die dafür spricht, dass das Gegenteil der Hauptthese zutrifft – wie es bei anderen Berechnungsmethoden zur Modellierung von Evidenzen der Fall wäre (Bao et al., 2012; Spreckelsen & Spitzer, 2008). Mit Hilfe der ETDS konnten in den Beiträgen Evidenzmaße für das Zutreffen der Hauptthese (Belief), für die Evidenz gegen das Zutreffen der Hauptthese (Doubt), für die Plausibilität des Zutreffens der Hauptthese, für die Plausibilität des Zutreffens der Gegenthese und für die Ungewissheit, ob die Hauptthese zutrifft, modelliert

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

werden. Die Evidenzmaße bilden eine einheitliche, systematische Vergleichsgrundlage in Bezug auf die dargestellte Evidenz.

Clusteranalyse

Die Clusteranalyse erlaubt es reliabel konstante Frames zu identifizieren (Matthes & Kohring, 2008; Scheufele & Scheufele, 2010). Scheufele und Scheufele (2010) kritisieren das Verfahren der Clusteranalyse bezüglich dessen Validität zur Erfassung von inhaltlichen Medienframes, weil hier ausschließlich vorher festgelegte Frameelemente untersucht werden. Sie halten es für unwahrscheinlich, dass durch die Clusteranalyse auch die latenten Inhalte von Medienframes adäquat und vollständig abgebildet werden können. In dieser Untersuchung sollte es nicht darum gehen, einzelne thematische und inhaltsabhängige Medienframes in ihrer einmaligen, spezifischen Form adäquat und vollständig abzubilden. Es ging darum, typische formal-abstrakte Evidenzframes zu identifizieren, welche konstant in der Medienberichterstattung zu finden sind. Deswegen ist die Clusteranalyse hier ein geeignetes, exploratives Werkzeug (vgl. Scheufele & Scheufele, 2010). Es interessieren theoretisch begründet ausschließlich die Evidenzmaße als framebildende Elemente.

Am reliabelsten lassen sich, nach Matthes und Kohring (2004), Frames mit Hilfe einer Clusteranalyse erfassen (vgl. auch Dahinden, 2006; Potthoff, 2012). Sie empfehlen für die Frame-Analyse eine hierarchische Clusteranalyse durchzuführen. Als Clusteranalyseverfahren wurde in dieser Untersuchung der Two-Step-Clusteranalyse Vorrang gegeben. Die Two-Step-Clusteranalyse ist eine Kombination der Verfahren der hierarchischen Clusteranalyse und der Clusterzentrenanalyse. Der Vorteil der Clusterzentrenanalyse ist, dass die Beiträge im Laufe des Verfahrens auch die Gruppen wechseln können und somit eine bessere Zuordnung als bei den ausschließlich hierarchischen Methoden möglich ist. Weitere Vorteile der Two-Step-Clusteranalyse waren, dass sie bei einer großen Anzahl an Fällen durchgeführt werden kann und keine Clusterzahl vorgegeben werden musste, sondern diese automatisch bestimmt wurde. Dass die Analyse sehr robust gegenüber Verletzungen der Verteilungs- und Unabhängigkeitsregel ist, ist ein weiteres Plus (Janssen & Laatz, 2005; Schendera, 2010). Dies war in dieser Untersuchung insbesondere wichtig, weil die Evidenzmaße zum Teil miteinander korrelierten und somit trotzdem eine zuverlässige Clusterlösung erreicht werden konnte.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Im Anschluss an die Clusterung wurde zur Überprüfung der Güte der Clusterlösung eine Diskriminanzanalyse durchgeführt. Die Diskriminanzanalyse ist ein strukturprüfendes Verfahren (Backhaus et al., 2006) und eignet sich daher als statistischer Plausibilitätstest unter Kontrolle der *F*- und *t*-Werte zur Einschätzung der Güte einer Clusterlösung (Schendera, 2010). Sie kehrt die Vorgehensweise der Clusteranalyse um und versucht aus den clusterbildenden Variablen (die Evidenzmaße) auf die ermittelte Clusterzugehörigkeit zu schließen (Schendera, 2010). Die Diskriminanzanalyse zeigt für die Drei-Clusterlösung mit 99 Prozent korrekt klassifizierter Fälle sehr gute Modellanpassung, ein sehr gutes Trennvermögen und eine hohe Leistungsfähigkeit des Modells (siehe Janssen & Laatz 2007).[96] Der Gruppenmittelwertevergleich lässt darauf schließen, dass die

96 Die Diskriminanzanalyse zeigte, dass die Funktionswerte der Diskriminanzfunktion für die unterschiedlichen Fallgruppen sehr verschieden sind. Je weiter diese Werte auseinander liegen, umso einfacher ist es, einen Fall anhand seines Funktionswertes zuzuordnen. Bei der Betrachtung der Kovarianzmatrizen fällt auf, dass die Kovarianzwerte der Gruppen zum Teil unterschiedliche Vorzeichen haben, dies resultiert aus der fehlenden Unabhängigkeit der Variablen. Der Box M-Test bestätigt mit einem signifikanten Wert ($p < .001$), dass statistische bedeutsame Unterschiede zwischen den Kovarianzmatrizen vorhanden sind. In der Diskriminanzanalyse wurden jeweils immer die ersten zwei kanonischen Diskriminanzfunktionen, welche in der Analyse verwendet wurden, berechnet. Die Mittelwerte (Gruppen-Zentroide) der Diskriminanzwerte in den Clustern zeigen große Unterschiede zwischen den Clustern durch die Funktion 1 und noch stärker dann durch die Funktion 2. Wilks-Lambda wurde berechnet und liegt bei allen Variablen unter .25. Es ist ein inverses Gütemaß; kleinere Werte bedeuten höhere Unterschiedlichkeiten der Gruppen (Schendera, 2010). Die Unterschiedlichkeit ist univariat signifikant ($p < .001$) und zeigt für sich zumindest auf, dass das Clustermodell nicht vollkommen ungeeignet zur Erklärung der abhängigen Variablen ist und signifikante Gruppenunterschiede bestehen. Alle clusterbildenden Variablen (Evidenzmaße) sind folglich potentielle Diskriminantoren. Das multivariate Wilks-Lambda deutet schon nach Schritt 1 auf große Gruppenunterschiede hin, ebenfalls die exakte *F*-Statistik (Schritt 1: $\lambda = .11$, exakte *F*-Signifikanz = .001; Schritt 2: $\lambda = .027$, exakte *F*-Signifikanz = .001). Der paarweise Gruppenvergleich zeigt, dass sich die Cluster signifikant ($p < .001$) unterscheiden. Der Eigenwert der Funktion 1 ist 8.2 und der Funktion 2 ist 3.1. Ein Eigenwert von 8.2 erklärt ca. drei Viertel der Varianz (73%). Der Eigenwert von Funktion 2 erklärt ca. ein Viertel der Varianz (27%). Zusammen erklären beide Funktionen die gesamte Varianz. Die erste Funktion leistet offenbar einen größeren Beitrag zur Unterscheidung zwischen den Gruppen. Die beiden hohen Eigenwerte ergeben sich, da die Streuung zwischen den Gruppen im Verhältnis zur Streuung innerhalb der Gruppen sehr groß ist (Schendera, 2010). Dies ist die von einer Diskriminanzanalyse angestrebte Situation. Es ist folglich gewährleistet, dass sich die Funktionswerte der einzelnen Gruppen deutlich voneinander unterscheiden, während die Werte innerhalb einer Gruppe sehr ähnlich sind. Die Kanonischen Korrelationskoeffizienten beider Funktionen sind bei .944 und .869. Beide Werte liegen sehr nah an 1 und weisen somit auf eine sehr hohe Trennkraft der Diskriminanzfunktionen und folglich der Diskriminanz an sich hin. Umso größer der Korrelationskoeffizient, desto größer ist die angestrebte Streuung zwischen den Gruppen im Verhältnis zur Streuung innerhalb der Gruppen (Schendera, 2010). Die ausgesprochen hohen Eigenwerte und kanonischen Korrelationskoeffizienten der Funktionen weisen darauf hin, dass die Streuung zwischen den

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

unabhängige Variable *Clusterzugehörigkeit* durchaus zur Erklärung der abhängigen Variablen (Evidenzmaße) geeignet ist.[97] Zur Validitätsanalyse der Clusterlösung wurden des Weiteren die BIC-Werte und AIC-Werte (vgl. Tabelle 8)[98] und die *F*- und *t*-Werte für die clusterbildenden Variablen in den Clustern kontrolliert (vgl. Tabelle 9)[99].

Clustern im Vergleich zur Streuung in den Clustern groß ist und somit eine gute Trennung zwischen den Gruppen vorliegt (Schendera, 2001). Auch die Streudiagramme der kanonischen Diskriminanzfunktionen zeigen, dass die Zentroide relativ zentral in der jeweiligen Punktewolke liegen und voneinander deutlich getrennt sind; Überlappungen sind nicht zu erkennen.

97 Deutlich bzw. groß sind die Unterschiede beim durchschnittlichen Doubtwert von Cluster 2 gegenüber den anderen Clustern, beim durchschnittlichen Beliefwert von Cluster 1 gegenüber den anderen Clustern und beim durchschnittlichen Ungewissheitswert von Cluster 3 gegenüber den anderen Clustern; die Unterschiede zwischen den Standardabweichungen sind bei den benannten Fällen optimalerweise relativ gering (vgl. Tabelle 17). Alle erklärenden Variablen weisen des Weiteren signifikante Mittelwertunterschiede auf ($p < .001$).

98 Beide Auswahlmaße zum Bestimmen der Clusteranzahl, Schwarz's Bayesian Criterion (BIC) und Akaike Information Criterion (AIC), erzielten das gleiche Clusterergebnis. Der BIC-Wert ist bei einer Clusterlösung von 3 sehr klein und das Verhältnis der Distanzmaße ist sowohl nach BIC als auch nach AIC bei der Drei-Cluster-Lösung sehr groß. In der Analyse wurde das Distanzmaß Log-Likelihood verwendet. Das euklidische Distanzmaß kam bei der Analyse mit drei Clustern zu fast identischen Ergebnissen (lediglich vier Beiträge wurden in ein anderes Cluster eingeordnet) und hätte ebenfalls verwendet werden können, da die Variablen, welche bei der Clusteranalyse herangezogen wurden, metrisch sind. Die Mittelwerte der Clustervariablen in den Clustern sind gleich, egal welches Distanzmaß verwendet wird.

99 Der *F*-Wert ist für alle Variablen innerhalb aller Cluster unter 1, das heißt, dass im Cluster keine höhere Streuung der Variable vorliegt, als in der Grundgesamtheit und das Cluster ist damit im Hinblick auf die clusterbildenden Variablen als homogen zu betrachten. Die *t*-Werte unterstützen die Annahme der Clusterbeschreibung/-charakterisierung und zeigen deutlich, dass die Variablen der *unterstützenden Plausibilität* und *Beliefmaß* in Cluster 1 stärker besetzt sind als in der Grundgesamtheit, bzw. die Variablen *Doubtmaß, kontroverse Plausibilität* und *Unsicherheitsmaß* schwächer. In Cluster 2 sind die Variablen *Doubt* und *kontroverse Plausibilität* etwas stärker besetzt als in der Grundgesamtheit, bzw. die Variablen *unterstützenden Plausibilität, Ungewissheitsmaß* und *Beliefmaß* etwas schwächer. In Cluster 3 ist wieder deutlich zu sehen, dass hier die Variablen *Beliefmaß* und *Doubtmaß* schwächer besetzt sind als in der Grundgesamtheit bzw. die Variablen *Unsicherheitsmaß, unterstützende* und *konträre Plausibilität* stärker.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Tabelle 8: BIC- und AIC-Kennzahlen der Clusterbildung

Anzahl der Cluster	BIC	BIC-Änderung[a]	Verhältnis der BIC-Änderungen[b]	Verhältnis der Distanzmaße[c]	AIC	AIC-Änderung[a]	Verhältnis der AIC-Änderungen[b]	Verhältnis der Distanzmaße[c]
1	1167.71				1130.00			
2	677.62	-49.09	1.00	1.81	602.19	-527.81	1.00	1.81
3	431.83	-245.79	.50	3.93	318.68	-283.51	.54	3.93
4	412.37	-19.45	.04	1.45	261.52	-57.17	.11	1.45
5	416.80	4.42	-.01	1.74	228.22	-33.29	.06	1.74
6	443.95	27.15	-.06	1.68	217.66	-1.57	.02	1.68
7	483.44	39.50	-.08	1.07	219.44	1.78	-.00	1.07
8	524.08	4.64	-.08	1.82	222.36	2.92	-.01	1.82
9	572.38	48.31	-.10	1.29	232.95	1.59	-.02	1.29

[a] Die Änderungen wurden von der vorherigen Anzahl an Clustern in der Tabelle übernommen.
[b] Die Änderungsquoten sind relativ zu der Änderung an den beiden Cluster-Lösungen.
[c] Die Quoten für die Distanzmaße beruhen auf der aktuellen Anzahl der Cluster im Vergleich zur vorherigen Anzahl der Cluster.

Tabelle 9: *F*- und *t*-Werte für die clusterbildenden Variablen

Frame (Nr.)	Variable	*F*	*t*
Wissenschaftlich gesicherte Evidenz (1)	Beliefmaß	.153	0.99
	Doubtmaß	.018	-1.38
	Konträre Plausibilität	.018	1.38
	Unterstützende Plausibilität	.153	-0.99
	Ungewissheitsmaß	.157	-0.81
Konfligierende Evidenz (2)	Beliefmaß	.356	-0.35
	Doubtmaß	.942	0.23
	Konträre Plausibilität	.942	-0.23
	Unterstützende Plausibilität	.356	0.35
	Ungewissheitsmaß	.266	-0.34
Fragile Evidenz (3)	Beliefmaß	.395	-0.69
	Doubtmaß	.017	-1.54
	Konträre Plausibilität	.017	1.54
	Unterstützende Plausibilität	.395	0.69
	Ungewissheitsmaß	.409	0.73

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Nachdem nun ausführlich die Güte der Inhaltsanalyse und der Auswertungsstrategie zur Inhaltsanalyse reflektiert wurde, steht nun das Rezeptionsexperiment im Fokus.

7.3 Rezeptionsexperiment

Im folgenden Kapitel 7.3.1 werden die Kriterien der Stichprobenziehung und der Stimuliauswahl für das Rezeptionsexperiment detailliert aufgezeigt. In Kapitel 7.3.2 wird dann die Operationalisierung der abhängigen Variablen, die Überzeugungsurteile und der potentiellen Einflussvariablen dargelegt. Anschließend wird in Kapitel 7.3.3 die Güte der Stichprobenziehung und Stimuliauswahl sowie der Befragung, des Fragebogens und der experimentellen Variation explizit reflektiert.

7.3.1 Stichprobenziehung und Stimuliauswahl

Zuerst werden in diesem Kapitel die Kriterien der Stichprobenziehung und dann die Stimuliauswahl für das Rezeptionsexperiment erläutert.

Stichprobe

Die Stichprobe des Experiments soll es ermöglichen, Kausalschlüsse reliabel ziehen zu können. Zur Ermittlung der erforderlichen Stichprobengröße wurde als Stärke der hypothesenkritischen Effekte ein eta²-Wert von .2 angenommen, was einem schwachen Effekt entspricht (Faul, et al., 2007). Der α-Fehler wurde streng auf 0.01 gesetzt und der β-Fehler auf 0.05. Unter diesen Bedingungen errechnete sich, unter der Anforderung eine Varianzanalyse mit zwei Messwiederholungen zu berechnen, eine erforderliche Gesamtstichprobe von 501 (Faul et al., 2007). Folglich wurde in dieser Untersuchung eine Stichprobe von 82 Personen pro Experimentalgruppe angestrebt.

Für die Stichprobe des Rezeptionsexperimentes wurden aus finanziellen und pragmatischen Gründen Studierende der Friedrich-Schiller-Universität Jena herangezogen. Studierende sind als homogene Gruppe mit ähnlichen soziodemographischen Merkmalen und kognitiven Fähigkeiten zu betrachten, daher lassen sich Gruppen von Studierenden gut miteinander vergleichen. Damit die Studierenden aber möglichst noch keine Erfahrungen mit der Wissenschaft und medizinische Evidenzeinordnung hatten, wurden Studierende rekrutiert, die sich im ersten Semester in ihrer ersten universitären, nicht-medizinischen Lehrveranstaltung befanden.

https://doi.org/10.5771/9783845254685-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Um diese zu akquirieren, wurde das Experiment in Bachelor-Einführungsvorlesungen durchgeführt.[100]

Stimuliauswahl

Im Rezeptionsexperiment sollte die Wirkung von Real-Stimulus-Material untersucht werden. Sinn war es keine weitere Studie mit künstlich generiertem Material durchzuführen (vgl. Kapitel 4.2), deren externe Validität und Generalisierbarkeit eingeschränkt ist. Real-Stimuli repräsentieren die Medienberichterstattung valider als selbst geschaffene Stimuli (Lecheler & De Vreese, 2011; Scheufele, 2004b).

Zur Beantwortung der Forschungsfragen sollten möglichst TV-Wissenschaftsbeiträge ausgewählt werden, welche die drei gefundenen Evidenzframes optimal abbilden (vgl. Kapitel 8.1.3). Die drei Frames wurden mit Hilfe einer Clusteranalyse identifiziert, in die die fünf Evidenzmaße eingingen, die für jeden Beitrag berechnet wurden: Beliefmaß, Doubtmaß, Ungewissheitsmaß, unterstützendes und konträres Plausibilitätsmaß. Im Ergebniskapitel der Inhaltsanalyse 8.1 werden die typischen Evidenzmaße der Frames detailliert aufgezeigt. Die Beiträge mit der höchsten Wahrscheinlichkeit, in jeweils eines der Cluster eingeordnet zu werden, beinhalten am ehesten die durchschnittlichen Evidenzmaße der Cluster. Diese Beiträge repräsentieren die Cluster also jeweils am besten. Um möglichst cluster-prototypische Beiträge der drei erfassten Evidenzframes zu selektieren, wurden alle Beiträge nach ihrer Prototypizität, bezogen auf die durchschnittlichen Evidenzmaße der Frames, sortiert. Des Weiteren sollten für FF5 einerseits Beiträge über Sachverhalte ausgewählt werden, zu denen bei den Versuchspersonen möglichst Voreinstellungen bestanden und andererseits Beiträge zu Sachverhalten ausgewählt werden, zu denen möglichst noch keine Einstellungen vorhanden waren, damit sie diese eventuell erst nach der Stimuluspräsentation ausbilden. Je Evidenzframe wurden zwei Beiträge ausgewählt:

100 Um den Professoren nicht zu viel Zeit ihrer Vorlesung zu nehmen, war die Dauer des Experiments auf ca. 15 Minuten angesetzt. Folgende Vorlesungen wurden ausgewählt: „Einführung in die Fachdidaktik“ bei Prof. Dr. Dickel, „Einführung in die angewandte Ethik“ bei Prof. Dr. Knöpfler, „Einführung in die Kommunikationswissenschaft“ bei Prof. Dr. Ruhrmann, „Wozu Soziologie“ bei Prof. Dr. Reitz, „Einführung in die Forschungsmethoden der Erziehungswissenschaft“ bei Prof. Dr. Frey, „Einführung in die Statistik“ bei Herrn Jost, „Romantik“ bei Prof. Dr. Matuschek, „Die Weltreligionen als Einführung in die Religionsgeschichte“ bei Prof. Dr. Schmitz und „Einführung in die Betriebswirtschaftslehre“ bei Prof. Dr. Lukas.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

- Als prototypische Beiträge für Frame 1 *Wissenschaftlich gesicherte Evidenz* mussten Beiträge mit einem Beliefmaß von ca. 84 Prozent, einem Doubtmaß von 0 Prozent und einem Ungewissheitsmaß von ca. 16 Prozent, einem unterstützenden Plausibilitätsmaß von 100 Prozent und einem durchschnittlichen konträren Plausibilitätsmaß von 16 Prozent gefunden werden. Als ein geeigneter prototypischer Beitrag des Frames 1 für das Rezeptionsexperiment wurde der Beitrag *Kältetherapie* herausgefiltert, welcher am 24.02.2011 bei Faszination Wissen im BR ausgestrahlt wurde. Die Hauptthese dieses Beitrags ist: *Die Kältetherapie hilft bei schwersten Schlaganfällen*. Das Beliefmaß des Beitrags ist 88 Prozent, das Doubtmaß 0 Prozent, das Ungewissheitsmaß 12 Prozent, das unterstützende Plausibilitätsmaß 100 Prozent und das konträre Plausibilitätsmaß beträgt 12 Prozent. Der Beitrag ist 308 Sekunden lang und beinhaltet zwei Evidenzquellen: ein *Fallbeispiel* und eine *Expertenmeinung*. Als weiterer geeigneter Beitrag wurde ein Beitrag mit der Hauptthese *Körperliche Bewegung macht auch geistig fit* ausgewählt, welcher am 01.12.2012 bei Planet Wissen (SWR) ausgestrahlt wurde. Dieser Beitrag hat ein Beliefmaß von 91 Prozent, das Doubtmaß beträgt 0 Prozent, das Ungewissheitsmaß 9 Prozent, das unterstützende Plausibilitätsmaß 100 Prozent und das konträre Plausibilitätsmaß 9 Prozent. Der Beitrag ist 286 Sekunden lang und beinhaltet vier Evidenzquellen: zwei *Studien*, eine *Expertenmeinung* und ein *Off-Sprecher*. Zu dem Sachverhalt *Bewegung* sollten Studierende im Allgemeinen eine Einstellung besitzen im Gegensatz zum Sachverhalt *Kältetherapie*.
- Für Frame 2 *Konfligierende Evidenz* sollten zwei Beiträge mit möglichst jeweils einem Beliefmaß von ca. 56 Prozent, einem durchschnittlichen Doubtmaß von 24 Prozent, einem durchschnittlichen Ungewissheitsmaß von 20 Prozent, einem unterstützenden Plausibilitätsmaß von ca. 76 Prozent und einem konträren Plausibilitätsmaß von ca. 44 Prozent gefunden werden. Hier wurde zum einen der Beitrag *Darmspiegelung*, der am 17.11.2011 bei X:enius auf ARTE ausgestrahlt wurde, ausgewählt. Die Hauptthese des Beitrags ist: *Vorsorgliche Darmspiegelungen sind anzuraten*. Der Beitrag hat ein Beliefmaß von 55 Prozent und ein Doubtmaß von 29 Prozent, ein Ungewissheitsmaß von 16 Prozent, das konträre Plausibilitätsmaß liegt bei 45 Prozent und das unterstützende Plausibilitätsmaß bei 71 Prozent. Des Weiteren ist der Beitrag 199 Sekunden lang und beinhaltet vier Evidenzquellen; ein *Off-Sprecher*, zwei *Fallbeispiele* und eine *Expertenmeinung*. Für Frame 2 wurde zum anderen ein Beitrag über die Wirkung von Kaffee auf den Körper ausgewählt. Bei diesem Sachverhalt ist zu erwarten, dass die Studierenden bereits eine Voreinstellung besitzen. Der Beitrag *Kaffee* wurde am 17.11.2011 bei

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Odysso im SWR ausgestrahlt. Die Hauptthese des Beitrags ist: *Kaffee macht geistig nicht fitter*. Der Beitrag hat ein Beliefmaß von 53 Prozent und ein Doubtmaß von 34 Prozent, das Ungewissheitsmaß beträgt 13 Prozent, das konträre Plausibilitätsmaß beträgt 47 Prozent und das unterstützende Plausibilitätsmaß 66 Prozent. Des Weiteren ist der Beitrag 328 Sekunden lang und im Beitrag sind drei Evidenzquellen dargestellt: ein *Fallbeispiel* und zwei *Studien*.

- Für Frame 3 *Fragile Evidenz* musste möglichst ein Beitrag mit einem Beliefmaß von ca. 41 Prozent, einem Doubtmaß von 0 Prozent, einem durchschnittlichen Ungewissheitsmaß von ungefähr 59 Prozent, einem unterstützenden Plausibilitätsmaß von 100 Prozent und einem konträren Plausibilitätsmaß von ca. 59 Prozent gefunden werden. Der Beitrag *Alzheimer*, welcher am 10.01.2011 bei X:enius auf ARTE ausgestrahlt wurde, konnte als geeignet herausgefiltert werden; hier wird nur eine Evidenzquelle gezeigt, welche der *Off-Sprecher* ist. Die Hauptthese des Beitrags ist: *Bei Alzheimer sterben Nervenzellen im Gehirn der Betroffenen*. Das Beliefmaß des Beitrags beträgt 34 Prozent, das Doubtmaß 0 Prozent, das Ungewissheitsmaß 66 Prozent, das unterstützende Plausibilitätsmaß 100 Prozent und das konträre Plausibilitätsmaß 66 Prozent. Der Beitrag ist 207 Sekunden lang und behandelt als Sachverhalt eine Krankheit von der Studierende nur in seltenen Fällen selbst betroffen sind. Als weiterer geeigneter Beitrag wurde ein Beitrag mit der Hauptthese *Zu viel Hygiene ist ungesund* ausgewählt. Dieser Sachverhalt wurde als alltagsnaher Sachverhalt für Studierende angesehen, zu dem sie wahrscheinlich bereits eine Einstellung besitzen. Der Beitrag *Hygiene* wurde am 14.11.2011 bei X:enius (ARTE) ausgestrahlt, ist 129 Sekunden lang und beinhaltet ebenfalls nur die Evidenzquelle *Off-Sprecher*. Das Beliefmaß des Beitrags beträgt 29 Prozent, das Doubtmaß 0 Prozent, das Ungewissheitsmaß 71 Prozent, das unterstützende Plausibilitätsmaß 100 Prozent und das konträre Plausibilitätsmaß 71 Prozent.

In der folgenden Tabelle sind die Evidenzmaße der ausgewählten Stimulusbeiträge gegenübergestellt.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Tabelle 10: Evidenzmuster der ausgewählten Stimulusbeiträge

Beitrag	Belief-maß	Doubt-maß	Ungewiss-heitsmaß	Unterstützendes Plausibilitätsmaß	Konträres Plausibilitätsmaß
		Frame 1 (*Wissenschaftlich gesicherte Evidenz*)			
Kältetherapie	.88	0	.12	1	.12
Bewegung	.91	0	.09	1	.09
		Frame 2 (*Konfligierende Evidenz*)			
Darmspiegelung	.55	.29	.16	.71	.45
Kaffee	.53	.34	.13	.66	.47
		Frame 3 (*Fragile Evidenz*)			
Alzheimer	.34	0	.66	1	.66
Hygiene	.29	0	.71	1	.71

Reynolds und Reynolds (2002) und auch Peter (2013) weisen darauf hin, wie wichtig es ist, die Qualität des Nachrichteninhaltes der einzelnen Stimulusbeiträge im Rezeptionsexperiment zu kontrollieren. Daher wurden alle Beiträge auf ihre Eignung als Stimulusmaterial hin gepretestet. Jeweils 18 bis 38 Versuchspersonen (Studierende der Friedrich-Schiller-Universität Jena) haben dafür die Beiträge auf einer fünfstufigen Ratingskala (von 1 *sehr wenig* bis 5 *sehr stark*) nach ausgewählten Eigenschaften zur Qualität des Beitrags bewertet und ausgedrückt, wie qualitativ hochwertig, fachmännisch, objektiv, verständlich, vertrauenswürdig, realistisch und informativ sie den jeweiligen Beitrag fanden. Schon in Kapitel 4.2 wurde herausgestellt, wie einflussreich die Lebendigkeit und Emotionalität bei der Wirkung von Evidenzen sein können; daher wurde ebenfalls mittels einer fünfstufigen Ratingskala gepretestet wie lebendig, emotional, interessant und langweilig die Beiträge empfunden wurden. Zur Überprüfung, ob ein Merkmal in den drei Stichproben identisch verteilt ist, wurden dann Chi²-Tests berechnet.

Der Anteil der Merkmalsausprägungen fast aller Variablen ist in den sechs Stichproben (zwei Beiträge pro Frame) gleich; es wurden hier keine signifikanten Unterschiede entdeckt (vgl. Tabelle 11). Nur bei der Variable *emotional* zeigten sich signifikante Unterschiede ($\chi^2 = 72.70$; $df = 20$; $p < .001$; $\varphi_c = .42$). Im direkten Vergleich der Ausprägungen in den Beiträgen zeigt sich allerdings, dass sich die Variablen zwischen den Beiträgen meist nicht um mehr als einen Punkt im Durchschnitt unterscheiden (vgl. Tabelle 12). Der Beitrag *Darmspiegelung* wird generell als am emotionalsten

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

bewertet; der Beitrag *Hygiene* als am wenigsten emotional. Ob sich diese Eigenschaft auf die Wirkung der Beiträge auswirkt, wurde in Bezug auf die Ergebnisse (vgl. Kapitel 8.2.2) überprüft.

Tabelle 11: Ergebnisse des Pretests der Stimuliauswahl

Variablen	*M*	*SD*	χ^2	*df*
Beitrag x emotional	2.4[a]	1.3	72.70***	20
Beitrag x glaubwürdig	4.4	0.6	23.53	15
Beitrag x verständlich	4.6	0.6	28.59	20
Beitrag x informativ	4.4	0.7	31.23	20
Beitrag x langweilig	2.3	1.0	21.12	20
Beitrag x qualitativ hochwertig	3.6	0.9	35.97	25
Beitrag x lebendig	3.1	1.0	18.71	20
Beitrag x interessant	4.0	0.7	18.58	15
Beitrag x realistisch	4.1	0.7	23.02	15
Beitrag x objektiv	3.7	0.9	30.59	20
Beitrag x vertrauenswürdig	4.0	0.8	18.74	15
Beitrag x fachmännisch	3.9	0.9	27.82	20

$n = 149$; ***$p < .001$; [a] fünfstufige Skala von 1 *sehr wenig* bis 5 *sehr stark*

Tabelle 12: Zugeschriebene Emotionalität zu den Beiträgen

Beitrag x emotional	*M*	*SD*	*n*
Kaffee	2.3[a]	1.2	21
Bewegung	2.0	1.0	38
Darmspiegelung	3.8	1.0	18
Kältetherapie	3.1	1.1	19
Alzheimer	3.4	1.2	18
Hygiene	1.6	1.0	35
Insgesamt	2.5	1.3	149

[a] fünfstufige Skala von 1 *sehr wenig* bis 5 *sehr stark*

Nachdem nun die Untersuchungseinheiten für das Rezeptionsexperiment definiert wurden, werden folgend das Forschungsinstrument und die Operationalisierungen der relevanten Konstrukte aufgezeigt.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

7.3.2 Fragebogen: Operationalisierung relevanter Konstrukte

Zunächst wird in diesem Kapitel auf die Operationalisierung der abhängigen Variablen, die Überzeugungsurteile, eingegangen, folgend dann auf die Operationalisierung der potentiellen Einflussvariablen, der zugeschriebenen Glaubwürdigkeit und der Informationsverarbeitung sowie auf die Operationalisierung weiterer möglicher Dritt- bzw. Einflussvariablen.

7.3.2.1 Überzeugungsurteile

Unter einer Einstellung wird in dieser Untersuchung die individuelle, mentale und bilanzierende Bewertung eines gedanklichen Objektes verstanden (vgl. Kapitel 4.1.1). Überzeugungen sind dabei wesentliche kognitive Komponenten von Einstellungen und subjektive Schätzungen der Wahrscheinlichkeit, dass bestimmtes Wissen gültig ist. Der traditionellen Messung von expliziten Einstellungen und Überzeugungen in Form von Befragungen ist generell eine hohe Zuverlässigkeit zu attestieren bspw. wegen der vergleichsweise hohen internen Konsistenz (Cunningham, Preacher & Banaji, 2001). Kämpfe (2005) zeigt auf, dass allein von einem Fragebogenitem ausgehend, ein umfangreiches assoziatives Netzwerk aktiviert wird und sowohl kognitive, als auch emotionale und motivationale Anteile reflektiert würden. Über die Fragen zur Messung von expliziten Überzeugungen werden somit unterschiedliche Aspekte eines Überzeugungsobjektes erfasst. Ob der Rezipient der Überzeugung ist, dass etwas zutrifft, ungewiss ist oder er etwas anzweifelt, ist beeinträchtigt von seinen Emotionen, Kognitionen und seinem Verhalten in Bezug auf das Bewertungsobjekt (vgl. Kapitel 4.1.1).

In Anlehnung an die Evidenzmaße der Frames werden in dieser Untersuchung fünf verschiedene Überzeugungsurteile differenziert:

(1) Als Belief ist das Maß an subjektiver Überzeugung definiert, in dem Rezipienten einer Hauptthese Glauben schenken.
(2) Als Doubt ist das Maß an subjektiver Überzeugung definiert, in dem Rezipienten eine Hauptthese anzweifeln.
(3) Als Ungewissheit ist das Maß an subjektiver Überzeugung definiert, in dem Rezipienten sich selbst ungewiss sind, ob sie einer Hauptthese glauben oder sie anzweifeln.
(4) Als Plausibilität wird das Maß an subjektiver Überzeugung definiert, in dem Rezipienten eine Hauptthese plausibel finden.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

(5) Als Gegenplausibilität wird das Maß an subjektiver Überzeugung definiert, in dem Rezipienten die Gegenthese plausibel finden.

Überzeugungen variieren, so wie Einstellungen, auch in ihrer Stärke (vgl. Kapitel 4.1.1); diese kann ausgedrückt werden in Einheiten subjektiver Sicherheit oder Wahrscheinlichkeit von 0 bis 1 (Wyer & Albarracín, 2005). Diese Eigenschaft spielt bei der Erfassung der Überzeugungen in dieser Untersuchung eine wichtige Rolle. Zur Erfassung der expliziten Überzeugungsurteile wurden die Rezipienten gebeten in Prozenten anzugeben, (1) wie sehr sie glauben, dass die Hauptthese des jeweiligen Beitrags gültig ist, (2) wie sehr sie daran zweifeln, dass die Hauptthese im jeweiligen Beitrag zutrifft und (3) wie sehr sie ungewiss sind, ob die Hauptthese im jeweiligen Beitrag gültig ist. Des Weiteren wurde die empfundene Plausibilität der Hauptthese und Gegenthese abgefragt. Auch hier sollten die Rezipienten in Prozent angeben, (4) wie plausibel sie die präsentierte Hauptthese im TV-Wissenschaftsbeitrag finden und (5) wie plausibel sie die entsprechende Gegenthese finden.

Jede der Überzeugungsfragen war, mit Hilfe einer numerischen Prozentsatz-Skala, als Prozentzahl, für das jeweilige Zutreffen der Überzeugungen, zu beantworten. Als optische Hilfestellung wurde ein Maßband von 1 bis 100 dargestellt, die genaue Prozentzahl war in einem unterstrichenden Feld einzutragen (siehe Online-Anhang III: Nachher-Fragebogenversion des Rezeptionsexperiments). Die Einteilung von 0 bis 100 ist damit begründet, dass, wie schon in Kapitel 4.1.1 beschrieben, Überzeugungen als kognitive Komponente in einem Wahrscheinlichkeitsraum von 0 bis 1 bestehen (Wyer & Albarracín, 2005). Des Weiteren ist die intervallskalierte Einteilung von 0 bis 100 optimal für die Vergleichbarkeit der einzelnen Überzeugungsurteile mit den Evidenzmaßen der Evidenzframes. Die Überzeugungsurteile (Belief-, Doubt-, Ungewissheits- und Plausibilitäts-Urteile) wurden auf ihre interne Konsistenz hin überprüft. Mit einem Cronbachs α von .85 wurde hier ein guter Reliabilitätswert erreicht.

Für FF5 wurde bei der Stimuliauswahl darauf geachtet, dass ein im Beitrag präsentierter Sachverhalt pro Frame für die Versuchspersonen alltagsfern ist, also möglichst keine Überzeugung durch eigene Erfahrungen zum Sachverhalt bestand, und dass ein Beitrag ausgewählt wird, welcher einen Sachverhalt präsentiert, der alltagsnäher ist, zu dem Überzeugungen hoch wahrscheinlich schon vor der Stimuluspräsentation vorhanden waren. Ob Rezipienten schon eine Einstellung zu einem Sachverhalt hatten oder

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

nicht, wurde zum einen anhand einer Trichterfrage kontrolliert. Die Versuchspersonen wurden hier danach gefragt, ob sie sich unter dem spezifischen Sachverhalt im Beitrag bspw. vorsorgliche Darmspiegelung etwas vorstellen können. Antworteten sie auf diese Frage mit *Nein*, sahen sie als nächstes den Stimuli. Antworteten sie auf diese Frage mit *Ja*, wurden sie weiter nach ihren vorherigen Überzeugungen zum jeweiligen Sachverhalt befragt, damit später ihre Überzeugungen, die sie vor dem Treatment hatten mit denen, die sie nach dem Treatment haben, verglichen werden konnten. Ob bei den Rezipienten eine Überzeugung zu einem Sachverhalt bestand, wurde zum anderen dadurch überprüft, dass ihnen immer die Möglichkeit gegeben wurde, bei der genauen Erfassung der vorherigen Überzeugungen, mit *weiß nicht* zu antworten. So wurden die Überzeugungen bei den Rezipienten nie erzwungen. Auch wenn sie bspw. wissen, was die vorsorgliche Darmspiegelung ist, heißt dies nicht, dass sie auch eine Überzeugung zu diesem Sachverhalt aufgebaut haben müssen. Weiter war es auch nach der Stimuluspräsentation möglich, dass sich die Rezipienten noch keine Überzeugung über einen Sachverhalt gebildet hatten; sie konnten dann ebenfalls bei allen Überzeugungsfragen mit *weiß nicht* antworten. Die vorherigen Überzeugungen wurden mit den gleichen Fragen erfasst, die auch für die Überzeugungserfassung nach der Rezeption eingesetzt wurden (siehe Online-Anhang II: Vorher-Fragebogenversion des Rezeptionsexperiments). Die erfassten vorherigen und die resultierenden Überzeugungsurteile waren somit sehr gut miteinander vergleichbar.

Da die Überzeugungsurteile in dieser Untersuchung als kognitive Komponente der Einstellung gelten, wurden die konative und affektive Einstellungskomponente zusätzlich einzeln operationalisiert. Dies geschah nicht zuletzt aber auch, weil der Forschungsstand zu Überzeugungen deutlich macht, dass diese ebenfalls affektive und konative Komponenten beinhalten (vgl. Kapitel 4.11). Die Versuchspersonen mit einer Voreinstellung wurden vor der Stimuluspräsentation nach ihrem derzeitigen Verhalten und nach ihren Emotionen in Bezug auf den jeweiligen Sachverhalt der Hauptthese befragt. Nach der Stimuluspräsentation wurden alle Versuchspersonen nach ihrem zukünftigen Verhalten und nach ihren Emotionen in Bezug auf den jeweiligen Sachverhalt der Hauptthese befragt. Dies wurde nicht in Prozent erhoben, da diese Komponenten nicht im direkten Vergleich mit den dargestellten Evidenzmustern gestellt, sondern zusätzlich erfasst wurden. Die Frage nach der konativen Komponente wurde mit Hilfe einer fünfstufigen Ratingskala erfasst. Für die Frage nach der emoti-

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

onalen Komponente wurde, wie bei Möhring und Schlütz (2010) vorgeschlagen, die visualisierte Smiley-Skala nach Jäger (2004) verwendet. Diese ist anschaulich und insbesondere für emotionale Bewertungen geeignet. Die dargestellten Smileys sind experimentell hinsichtlich ihrer Veränderung des emotionalen Ausdrucks geprüft und validiert; sie gelten als eindimensional und äquidistant (somit können Daten auf Intervallniveau erhoben werden).

Die Glaubwürdigkeitszuschreibung kann, wie schon mehrfach betont (vgl. Kapitel 4.1.3), eine wichtige beeinflussende, aber auch abhängige Variable bei der Framingwirkung sein. Im Folgenden wird auf die Operationalisierung dieser Variable eingegangen.

7.3.2.2 Glaubwürdigkeit

Die Glaubwürdigkeit ist ein hypothetisches Konstrukt, welches aus Wechselbeziehungen zwischen Quelle, Nachricht und Rezipient entsteht (vgl. Kapitel 4.1.3). Die Glaubwürdigkeitszuschreibung der Rezipienten findet auf der Basis verschiedenster Stimmungen, Vorbildungen, Situationen und Erfahrungen statt. Glaubwürdigkeit kann nicht inhaltsanalytisch erfasst werden (Bentele, 1988). Die meisten Untersuchungen zur Glaubwürdigkeit untersuchen die zugeschriebene Glaubwürdigkeit nur von einer Entität des journalistischen Beitrags und messen entweder die Glaubwürdigkeit der Nachricht oder der Quelle. Roberts (2010) bilanziert aus der Forschungsliteratur zur Glaubwürdigkeit, dass dies wegen der komplizierten Wechselbeziehungen zwischen Quelle und Nachricht ein Fehler ist.

Instrumente zur Messung von Nachrichten- oder Quellenglaubwürdigkeit sind oft Skalen mit semantischen Differentialen, die in der Regel auf die Operationalisierung einer oder mehrerer allgemeiner Dimensionen konzentriert sind (vgl. Metzger et al., 2003). Die Quellenglaubwürdigkeit gilt schon zu Beginn der kommunikationswissenschaftlichen Forschung als ein mehrdimensionales Konzept, das aus mindestens zwei unterscheidbaren Komponenten besteht. Bei den Studien der Yale-Gruppe um Hovland (Hovland & Janis, 1959; Hovland & Weiß, 1951; Hovland et al., 1953) ist die Glaubwürdigkeit einer Quelle das Produkt aus dessen Kompetenz (*expertiseness* oder *competency*) und dessen Vertrauenswürdigkeit (*trustworthiness*). Vertrauenswürdigkeit wird einer Quelle zugeschrieben, die nach Einschätzung der Rezipienten Wissen nicht verzerrt kommuniziert, sondern richtige und vollständige Aussagen macht. Kompetenz wird einer Quelle zugeschrieben, die nach Einschätzung der Rezipienten aufgrund

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

ihrer Qualifikation über relevantes Wissen verfügt. Wichtig ist nach Hovland und Janis (1959): Wer sagt was, auf welche Art und Weise und in welchem Kontext.[101] Für viele Forscher gelten in der Glaubwürdigkeitsforschung noch heute Kompetenz und Vertrauenswürdigkeit als die notwendigen, konstitutiven Komponenten der Glaubwürdigkeit (bspw. Eisend, 2003; Kohring & Matthes, 2007; Küster-Rohde, 2010; O'Keefe, 2002). Pornpitakpan's Review (2004) zur Glaubwürdigkeitsforschung zeigte auf, dass bis heute nicht klar ist, ob entweder die Kompetenz oder die Vertrauenswürdigkeit wichtiger ist. Teilweise variieren bei der zugeschriebenen Glaubwürdigkeit die Ausprägungen der zugeschriebenen Kompetenz und Vertrauenswürdigkeit zwischen verschiedenen Quellen beträchtlich.[102] Trotz der Vielzahl von Studien über die Glaubwürdigkeit fehlen für die Operationalisierung der beiden Komponenten noch immer klare und systematische Kriterien.

Schon Berlo, Lemert und Mertz (1969) kritisieren die fehlende theoretische Fundierung der Faktoren der Yale-Gruppe. Durch eine Serie von Faktoranalysen von semantischen Differenzialen kamen Berlo et al. (1969) zu dem Ergebnis, dass Quellenglaubwürdigkeit drei Dimensionen hat: Sicherheit, Qualifikation und Dynamik. Dynamik sei dabei aber nicht Wesenskern des Konstrukts *Glaubwürdigkeit*. Es folgen unzählige weitere Untersuchungen mit vielen weiteren Dimensionen (vgl. Studienreview von Eisend, 2003). Für die meisten Forscher haben aber alle zusätzlichen Dimensionen nur eine untergeordnete, intensivierende Funktion (Eisend, 2003; O'Keefe, 2002). Koch und Lindemann (2013) beschreiben, dass bspw. die durch Studien herausgestellte Einflussgrößen auf die Quellenglaubwürdigkeit, wie Sympathie, Ähnlichkeit oder physische Attraktivität, meist nur dann von Bedeutung sind, wenn Informationen über Vertrauenswürdigkeit oder Kompetenz nicht vorliegen.

101 Die Glaubwürdigkeit wird bei Hovland und Janis (1959) jedoch als objektiver Faktor der Quelle definiert, nicht als zugeschriebener Faktor innerhalb der Rezeption. Quellen-Faktoren (Kompetenz und Vertrauenswürdigkeit) und Nachrichten-Faktoren (Reihenfolge der Argumente und Dauer der Darstellung) könnten daher inhaltsanalytisch in einer Botschaft erfasst werden. Nach der in der heutigen Forschung und auch in dieser Untersuchung vertretenen rezipientenorientierten Ansicht kann Glaubwürdigkeit nicht mehr aus der Botschaft inhaltsanalytisch erfasst werden (vgl. schon Bentele, 1988).

102 Beispielsweise zeigte die Studie von Peters (1992) zur Glaubwürdigkeit von Quellen in Westdeutschland zum Thema *Chernobyl-Desaster*, dass sich zwar die zugeschriebene Glaubwürdigkeit zwischen den untersuchten Quellen nur wenig voneinander unterschieden, die einzelnen klassischen Glaubwürdigkeitsfaktoren allerdings beträchtlich variieren können.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Da Glaubwürdigkeit als multidimensionales Konstrukt konzipiert und nicht direkt messbar ist, kann sie nur durch eine Multi-Item-Messung vollständig erfasst werden (Rieh, 2007). Eine Vielzahl von Studien nutzte semantische Differenziale, um Glaubwürdigkeit zu messen, und Itembatterien, die verschiedene Dimensionen der Glaubwürdigkeit abdecken sollen. Diese Studien wendeten oftmals die explorative Faktorenanalyse an, um die Dimensionen von Glaubwürdigkeit zu identifizieren. Eine große Anzahl an Dimensionen der Glaubwürdigkeit konnte so über die Jahre ermittelt werden (Eisend, 2006).[103] Die Einbeziehung weiterer Faktoren/Dimensionen erfolgte dabei uneinheitlich (Hellmüller & Trilling, 2012).[104] Dies deutet auf fehlende Konsistenz hin und lässt sich vor allem auf methodische Probleme zurückführen (vgl. Eisend, 2006). Die Dimensionen der Glaubwürdigkeit wurden meist weder theoretisch abgeleitet, noch in eine Theorie der Glaubwürdigkeit integriert (vgl. Eisend, 2006; Kohring & Matthes, 2007). Den meisten faktoranalytischen Untersuchungen kann in Bezug auf zusätzlich gefundene Dimensionen vorgeworfen werden, dass sie ihre Skalen zufällig auswählen, dass sie Faktoren gleiche Namen geben, diese aber auf unterschiedlichen Skalen basieren und dass bestimmte Glaubwürdigkeitsfaktoren unzulässig auf jegliche Bezugsobjekte generalisiert werden (vgl. Hellmüller & Trilling, 2012; Pornpitakpan, 2004; Seidenglanz, 2007).

Roberts (2010) verwendete in seiner Untersuchung ein überzeugendes Messinstrument zur Erfassung der zugeschriebenen Glaubwürdigkeit. Er untersucht die Glaubwürdigkeit von Inhalt und Quelle gesondert und führt die Befragungsergebnisse dann zusammen. Dazu verwendete er für die Quellenglaubwürdigkeit die Skala von Meyer (1988). Meyer untersuchte 1988 die Glaubwürdigkeit von Zeitungen als Informationsquellen. Für die Glaubwürdigkeit des Inhaltes verwendete Roberts (2010) die Skala

103 Eisend (2006) gibt einen sehr umfassenden Überblick über die wichtigsten bisherigen Glaubwürdigkeitsstudien. Dabei zeigt sich, dass bisher mehr als 30 verschiedene Dimensionen der Glaubwürdigkeit zugeordnet wurden. Die meisten Studien haben zwischen zwei und fünf verschiedene Dimensionen und beziehen sich mitunter nur auf die Quellenglaubwürdigkeit. Die Dimensionen *Kompetenz* und *Vertrauenswürdigkeit* werden in vielen Studien als konstitutiv angesehen und werden oftmals um intensivierende, untergeordnete Dimension wie bspw. Dynamik oder Attraktivität ergänzt.

104 In einem systematischen Review der führenden internationalen Wissenschaftsmagazine von 1951 bis 2011 untersuchten Hellmüller und Trilling (2012) wie Quellen-, Nachrichten- und Medienglaubwürdigkeit konzeptualisiert und operationalisiert wurden. Auch sie fanden viele verschiedene Dimensionen und Skalen, um Glaubwürdigkeit zu messen. Diese seien meist inkonsistent in ihrer theoretischen Herleitung, zudem fehle die methodische/operationale Präzision und die Skalen ließen sich zumeist nicht replizieren oder validieren.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

von Flanagin und Metzger (2000). Flanagin und Metzger untersuchten 2000 die Glaubwürdigkeit des Inhalts von Online-Informationen. Die Analyse von Roberts (2010) zeigte die Zuverlässigkeit der Skala von Meyer (1988) und der Skala von Flanagin und Metzger (2000). Beide Skalen können gut zusammengeführt und zusammen verwendet werden, um die Glaubwürdigkeit von Quelle und Botschaft gleichzeitig zu messen.

Wie von Roberts (2010) geraten, wurde in dieser Studie sowohl die Glaubwürdigkeit der Quelle als auch ihrer Botschaft untersucht. So wurde die Glaubwürdigkeit der TV-Wissenschaftsbeiträge als Informationsquelle mit Hilfe der Items von Meyer (1988) erfasst und die Glaubwürdigkeit des Inhalts der jeweiligen TV-Wissenschaftsbeiträge mit Hilfe der Items von Flanagin und Metzger (2000) untersucht. Wie bei Roberts (2010), Flanagin und Metzger (2000) und bei Meyer (1988) wurde dazu ein fünfstufiges semantisches Differenzial genutzt. Die Items wurden ins Deutsche übersetzt und werden folgend angeführt:

Messenger scale nach Roberts (2010) und Übersetzung
- cannot/can be trusted → nicht vertrauenswürdig/vertrauenswürdig
- inaccurate/accurate → ungenau/genau
- unfair/fair → unfair/fair
- does not/does tell the whole story → vollständig (erzählt die ganze Story)/unvoll-ständig (erzählt nicht die ganze Story)[105]
- biased/not biased → voreingenommen/unvoreingenommen

Message scale nach Roberts (2010) und Übersetzung
- unbelievable/believable → nicht glaubhaft/glaubhaft
- inaccurate/accurate → ungenau/genau
- not trustworthy/trustworthy → nicht vertrauenswürdig/vertrauenswürdig
- biased/not biased → verzerrt/unverzerrt
- incomplete/complete → unvollständig/vollständig

Um einen möglichen Zusammenhang zwischen Überzeugungsurteilen und zugeschriebener Glaubwürdigkeit, wie in H4.4 postuliert, zu untersuchen, muss die dem jeweiligen Beitrag zugeschriebene Glaubwürdigkeit mit der gebildeten Überzeugung verglichen werden. Menschen, die TV-Wissenschaftsbeiträgen im Allgemeinen nichts glauben, würden nach den Konsistenztheorien (Harmon-Jones & Harmon-Jones, 2007) auch dem

105 Dieses Item wirkte im Pretest in der Originalform „tells the hole story" zu personifizierend, daher wurde ein beschreibendes Adjektiv hinzugefügt.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

einzelnen Beitrag eher nicht glauben und ihre Überzeugungen nicht in Richtung der präsentierten Evidenzmuster eines rezipierten Beitrags ändern. Daher wurde kontrolliert, wie viel Glaubwürdigkeit die Rezipienten TV-Wissenschaftsbeiträgen im Allgemeinen und speziell dem gesehenen Beitrag zuschreiben. Die zugeschriebene Glaubwürdigkeit zu TV-Wissenschaftsbeiträgen im Allgemeinen wurde vor der Stimuluspräsentation abgefragt und nach der Stimuluspräsentation wurde die Glaubwürdigkeit abgefragt, die speziell dem jeweiligen Stimulusbeitrag zugeschrieben wurde. So konnten beide miteinander verglichen und deren Einfluss auf die Überzeugungsbildung und -änderung der Rezipienten untersucht werden. Beide Male wurde die interne Konsistenz der Skala von Roberts (2010) über Cronbachs α bestimmt und zeigte sehr gute Reliabilitätswerte von über .86 auf.

Dass auch die Wirkung der Glaubwürdigkeit einer Nachricht von der Informationsverarbeitungsroute der Rezipienten abhängen kann, wurde bereits in Kapitel 4.1.2 erläutert, auch deswegen ist es wichtig diese zu erfassen.

7.3.2.3 Informationsverarbeitung

Die Informationsverarbeitung kann nach den Zweiprozessmodellen auf zwei verschiedenen Routen stattfinden (vgl. Kapitel 4.1.2). Auf welcher Route Rezipienten eine mediale Information verarbeiten, hängt im Wesentlichen davon ab, wie motiviert diese sind und welche kognitiven Verarbeitungsfähigkeiten sie im Moment der Rezeption haben.

Um die Motivation und Verarbeitungsfähigkeit der Rezipienten bei der Beitragsrezeption zu erfahren, wurden die Versuchspersonen zum einen nach der Stimuluspräsentation direkt gefragt, ob sie den Beitrag aufmerksam angesehen haben, ob sie motiviert waren den Inhalt des Beitrags zu verstehen, ob sie den Inhalt des Beitrags verstanden haben, ob sie sich auf den Beitrag konzentrieren konnten und ob sie Lust darauf hatten, an der Befragung teilzunehmen. Auf einer Ratingskala von 1 für *Nein, überhaupt nicht.* bis 5 für *Ja, sehr.* sollten sich die Versuchspersonen dann selbst einordnen. Für die Items der Motivation und kognitiven Verarbeitungsfähigkeit wurde die interne Konsistenz über Cronbachs α bestimmt und erreichte gute Reliabilitätswerte von über .79.

Das Involvement gilt als eng mit der Motivation verknüpft. Wie bei Peter (2013) wurde das Involvement der Rezipienten im Experiment vor der

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Stimuluspräsentation erhoben. Die Versuchspersonen wurden danach gefragt, ob sie selbst von dem präsentierten Sachverhalt des jeweiligen Treatments betroffen sind oder waren (Alternativfrage *Ja./Nein.*). Sie wurden des Weiteren nach ihren bisherigen Erfahrungen mit der Medizin befragt (fünfstufige Ratingskala *Schlecht.* bis *Gut.*) und ob sie sich für neue medizinische Erkenntnisse interessieren (fünfstufige Ratingskala *Ja, sehr.* bis *Nein, generell nicht.*).

Des Weiteren wurde das Need for Cognition der Rezipienten erfragt und so themenunabhängig das Bedürfnis der Rezipienten nach kognitiver Aktivität bzw. nach Engagement und Freude an Denkaufgaben erfasst. Rezipienten mit einem hohen Need for Cognition würden nach den Zweiprozessmodellen eher auf der zentralen/systematischen Route Informationen aufnehmen und aktiv verarbeiten, somit sollten bei diesen mitunter andere und/oder stärkere Medienwirkungseffekte auftreten als bei Rezipienten mit einem niedrigen Need for Cognition (Petty & Cacioppo, 1986b; vgl. Kapitel 4.1.2). Ein hohes Need for Cognition kann eine Einstellungsänderung verstärken (Krämer & Winter, 2014). Zur Messung wurde eine modifizierte und gekürzte Skala von Bless, Fellhauer, Bohner und Schwarz (1991) mit zehn Items verwendet.[106] Diese folgenden Items wurden mittels einer siebenstufigen Skala mit den Endpunkten +3 für *Trifft ganz genau zu.* bis -3 für *Trifft überhaupt nicht zu.* erhoben:

106 Diese Skala stellt eine Übersetzung der Need for Cognition-Skala von Petty und Cacciopo von 1982 dar. Sie erlaubt eine ökonomische Durchführung und weist gute Skalenwerte auf. Die zehn Items aus der Skala von Bless, Fellhauer, Bohner und Schwarz (1991), die die höchsten Faktorladungen > .50 auf dem dominanten Faktor hatten, wurden für die Untersuchung des Need for Cognition ausgewählt.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Need for Cognition-Items (Faktorladung des Items bei Bless et al., 1991)

- Die Aufgabe, neue Lösungen für Probleme zu finden, macht mir wirklich Spaß. (.52)
- Die Vorstellung, mich auf mein Denkvermögen zu verlassen, um es zu etwas zu bringen, spricht mich nicht an. (.55)
- Ich würde lieber etwas tun, das wenig Denken erfordert, als etwas, das mit Sicherheit meine Denkfähigkeit herausfordert. (.57)
- Ich finde wenig Befriedigung darin, angestrengt und stundenlang nachzudenken. (.61)
- In erster Linie denke ich, weil ich muss. (.54)
- Denken entspricht nicht dem, was ich unter Spaß verstehe. (.53)
- Ich versuche Situationen vorauszuahnen und zu vermeiden, in denen die Wahrscheinlichkeit groß ist, dass ich intensiv über etwas nachdenken muss. (.55)
- Ich habe es gern, wenn mein Leben voller kniffliger Aufgaben ist, die ich lösen muss. (.64)
- Ich würde komplizierte Probleme einfachen Problemen vorziehen. (.53)
- Es genügt mir, einfach die Antwort zu kennen, ohne die Gründe für die Antwort eines Problems zu verstehen. (.51)

Die interne Konsistenz der Skala wurde über Cronbachs α bestimmt und zeigte einen guten Reliabilitätswert von .79 auf.

In Kapitel 4.1.3 wurde erläutert, dass die Wirkung von Evidenz und Glaubwürdigkeit davon beeinflusst werden kann, welche Verarbeitungspräferenzen Rezipienten im Allgemeinen haben, also ob sie eher Kopfmenschen (Präferenz für Deliberation) oder Bauchmenschen (Präferenz für Intuition) sind. Hier könnte davon ausgegangen werden, dass eher deliberative Personen Informationen bevorzugt auf der zentralen/systematischen Route nach den Zweiprozessmodellen verarbeiten und dass eher intuitive Personen bevorzugt auf der peripheren/heuristischen Route verarbeiten. Wie in der Studie von Koch und Lindemann (2013) wurden für die Untersuchung Items aus der Skala von Betsch (2004) verwendet, welche die Präferenzen der Rezipienten für Intuition und Deliberation (PID) ermitteln. Die PID-Skala von Betsch (2004) untersucht beide Entscheidungsmodi unabhängig voneinander. Wie bei Koch und Lindemann (2013) werden in dieser Untersuchung nur die vier Items jeder Subskala genutzt, welche die höchsten Faktorladungen bei Betsch (2004) aufwiesen. Mit Hilfe einer fünfstelligen Skala mit den Eckpunkten -2 für *Stimme nicht zu.* und +2 für *Stimme voll zu.* wurden folgende Items untersucht:

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Präferenz für Deliberation
- Bevor ich Entscheidungen treffe, denke ich meistens erst mal gründlich nach.
- Bevor ich Entscheidungen treffe, denke ich meistens erst mal über meine Ziele nach, die ich erreichen will.
- Ich lasse mich bei meinen Schlussfolgerungen von meinen Gefühlen, meiner Menschenkenntnis und Lebenserfahrung leiten.[107]
- Ich denke erst nach, bevor ich handle.

Präferenz für Intuition
- Ich bin ein sehr intuitiver Mensch. Ich entscheide meist aus dem Bauch heraus.[108]
- Ich mag emotionale Situationen, Diskussionen und Filme.
- Bei meinen Entscheidungen spielen Gefühle eine große Rolle
- Ich nehme bei einem Problem erst mal die harten Fakten und Details auseinander, bevor ich mich entscheide.

Für beide Präferenzskalen wurde die interne Konsistenz über Cronbachs α bestimmt; bei beiden zeigten sich jeweils gute Reliabilitätswerte von über .72. Auch die interne Konsistenz der gesamten Skala wurde über Cronbachs α berechnet und wies einen akzeptablen Reliabilitätswert von .69 auf.

7.3.2.4 Dritt- bzw. Einflussvariablen

Aus der Aufarbeitung des Forschungsstandes zur Wirkung von Evidenz in Kapitel 4 konnten einige mögliche Dritt- bzw. Einflussvariablen eruiert werden. Diese wurden während des Rezeptionsexperiments erhoben und deren Auswirkungen auf die Überzeugungsbildung oder -änderung wurden untersucht. Zu den kontrollierten Variablen zählen die Sehgewohnheiten und soziodemographische Merkmale wie Alter, Geschlecht und Bildungsstand der Rezipienten. Menschen, die viel TV-Wissenschaftssendungen schauen, haben tendenziell weniger Vorbehalte in Bezug auf die Wissenschaft, haben höheres faktuelles und prozedurales wissenschaftli-

107 Vor dem Pretest hieß dieses Item noch, nach Koch und Lindemann (2013), *Ich ziehe Schlussfolgerungen lieber aufgrund meiner Gefühle, Menschenkenntnis und Lebenserfahrung.* Dies war allerdings, so zeigte der Pretest, schwer verständlich und wurde leicht modifiziert.

108 Vor dem Pretest hieß dieses Item noch, nach Koch und Lindemann (2013), *Ich bin ein sehr intuitiver Mensch.* Dies war allerdings, so zeigte der Pretest, schwer verständlich und wurde leicht modifiziert.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

ches Wissen und glauben mitunter eher an die Zusicherungen der Wissenschaft (Nisbet et al., 2002). Die Sehgewohnheiten der Rezipienten wurden erfasst, indem diese gefragt wurden, wie oft sie TV-Wissenschaftsbeiträge bspw. von Nano, Quarks & Co, Planet Wissen, Faszination Wissen, Odysso, X:enius, Scobel oder W wie Wissen schauen (fünfstellige Ratingskala *Nie.* bis *Sehr oft.*). Die soziodemographischen Merkmale wie das Alter, das Geschlecht, das aktuelle Hochschulsemester und der Bildungsstand der Rezipienten wurden, wie von Brosius et al. (2012) und Scheufele und Engelmann (2009) geraten, am Ende des Fragebogens erhoben.

Nachdem die Forschungsinstrumente dieser Untersuchung aufgezeigt wurden, wird im nächsten Kapitel die Güte des Rezeptionsexperiments diskutiert.

7.3.3 Reliabilität und Validität

Begonnen wird folgend mit der Diskussion der Stichprobe und Stimuliauswahl. Im darauffolgenden Unterpunkt wird die Güte der Befragung und des Fragebogens diskutiert. Im letzten Unterpunkt dieses Kapitels wird die Güte der experimentellen Variation reflektiert.

Stichprobe und Stimulusmaterial

Dass die Stichprobe nur aus Studierenden der Universität Jena bestand, könnte die externe Validität des Rezeptionsexperiments negativ beeinflussen (Scholl, 2009). Es gibt allerdings keine plausiblen Hinweise dafür, warum gefundene Effekte artifiziell zu betrachten seien. Schon Daschmann (2001) weist darauf hin, dass Studierende überdurchschnittlich hoch kognitiv strukturiert sind und dass Effekte in der Durchschnittsbevölkerung sogar verstärkt auftreten könnten, daher mindert der Rückgriff auf Studierende die externe Validität nur marginal. Wichtig war es für diese Untersuchung eine homogene Gruppe von Menschen zu untersuchen, damit sich die sechs Experimentalgruppen möglichst nicht in ihrer allgemeinen kognitiven Fähigkeit unterschieden (vgl. Kapitel 7.3.1).

Im Rezeptionsexperiment wurde die Wirkung von Real-Stimuli untersucht. Dies kann die interne Validität dieser Untersuchung unter Umständen negativ beeinflussen. Die Gefahr einer uneindeutigen Kausalattribution besteht, da nicht endgültig gesichert werden kann, ob eine Überzeugungsänderung wirklich von den verschiedenen Frames der Evidenzdarstellung beeinflusst ist oder eher von anderen inhaltlichen Variablen, die sich bei Real-Stimulus-Material zwangsläufig mit verändern. Weiteres

https://doi.org/10.5771/9783845278668-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Problem, das die Verwendung von Real-Stimulusmaterial mit sich brachte, war die inhaltliche Variation zwischen den Beiträgen. In der Stichprobe der Inhaltsanalyse war es nicht möglich frame-prototypische Beiträge zur selben dargestellten Hauptthese zu finden. Eine Möglichkeit, die inhaltliche Variation der Beiträge zu kontrollieren, wäre es gewesen, TV-Wissenschaftsbeiträge selbst zu erstellen. Real-Stimuli-Material zu verwenden, stärkt allerdings die externe Validität der Untersuchung, denn es repräsentiert die Medienberichterstattung valider als selbst geschaffene Stimuli (Lecheler & De Vreese, 2011; Scheufele, 2004b). Auch dass die Wirkung der empirisch gefundenen Frames und nicht die von theoretisch erstellten Frames untersucht wurde, stärkt die externe Validität der Untersuchung, denn die verwendeten Frames kommen real in der Berichterstattung vor. Eine andere Möglichkeit die inhaltliche Variation der Beiträge zu kontrollieren, ist es ein Multi-Message-Design zu verwenden. Dafür wurde sich in dieser Untersuchung entschieden. Um zu erfassen, ob andere Beiträge des gleichen Frames gleiche Wirkungen erzielen würden, wie der verwendete Stimulusbeitrag, wurden für jeden Frame zwei Beiträge herangezogen. Mehr als zwei Stimulusbeiträge pro Frame zu verwenden, hätte die Untersuchungsergebnisse noch weiter sichern können; dies war aber aus forschungsökonomischen Gründen nicht realisierbar.

Das ausgewählte Stimulusmaterial wurde auf seine Eignung gepretestet (vgl. Kapitel 7.3.1). Die Ergebnisse zeigen, dass die ausgewählten Beiträge als Stimulusmaterial gut geeignet waren.

Güte der Befragung und des Fragebogens

Es wurden, um die Güte der Befragung, des Fragebogens und des Rezeptionsexperiments von vornherein zu sichern, zwei Pretests durchgeführt. Der erste Pretest wurde in Form einer Fragebogenkonferenz abgehalten (mit zwei Kollegen des Instituts für Kommunikationswissenschaft der FSU Jena). Beim zweiten Pretest wurde das gesamte Rezeptionsexperiment im Feld getestet (n = 31 Bachelor-Studierende der Kommunikationswissenschaft).

Alle Befragungen liefen, kontrolliert durch die Interviewer, standardisiert nach genau dem gleichen Schema ab, um die Reliabilität zu sichern.[109] Damit ist die Vergleichbarkeit der Befragungssituation gegeben und die

109 Als Maßnahme zur Reduktion und Vermeidung von Interviewereffekten wurden neben der Forschungsleiterin zwei Interviewer ohne Kenntnis der Forschungsfragen eingesetzt. Diese passten des Weiteren darauf auf, dass die Versuchspersonen nicht zurückblätterten und die Befragungssituation, was äußere Kontextgegebenheiten anbelangt, nicht variiert.

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Generalisierbarkeit der Ergebnisse sichergestellt (Brosius et al., 2012; Möhring & Schütz, 2010). Am Anfang des Fragebogens wurden die Rezipienten darauf hingewiesen alle Fragen in Bezug auf ihr Leben im Allgemeinen zu beantworten, so dass die Antworten dem entsprechen, wie die Rezipienten im Allgemeinen entscheiden (analog zu Betsch, 2004). Des Weiteren wurden sie darauf hingewiesen spontan zu antworten, da es keine richtigen oder falschen Antworten gibt. Die Befragten wurden vor Ausfüllen des Fragebogens darüber aufgeklärt, dass ihre Antworten freiwillig sind, absolut vertraulich behandelt und anonym ausgewertet werden.

Der Methode *Befragung* wird aufgrund der künstlichen Befragungssituation mangelnde Standardisierbarkeit attestiert (Möhring & Schütz, 2010; Scholl, 2009). Es wird davon ausgegangen, dass die Interviewsituation und das Messinstrument das Antwortverhalten der Rezipienten verzerren, da eine nicht alltägliche Situation besteht (Reaktivität). Die Versuchspersonen sind sich der Befragungssituation bewusst, dies kann zu *demand effects* (Daschmann, 2001, S. 169) führen. Das heißt, ein bestimmtes auf das Experiment bezogenes nicht natürliches Verhalten zeigen, dass die Befragungsergebnisse beeinflusst, z. B. könnten sie sich zu kooperativ oder besonders unkooperativ verhalten oder wollen meta-analytisch den Sinn hinter einem Experiment verstehen. So kann sich auch die Selbsteinschätzung negativ auf die Reliabilität auswirken, da Rezipienten bspw. sozial erwünscht antworten oder sich selbst ausschließlich im positiven Licht darstellen wollen. Bei der querschnittlich angelegten Befragung der bewussten, verbalisierbaren Kognitionen war die Messintention für die Rezipienten durchschaubar. Implizite Einstellungsmessungen würden dem Abhilfe schaffen, weisen jedoch (noch) vielfach unbefriedigende Reliabilitäten (Hofmann, Gawronski, Gschwendner, Le & Schmitt, 2005) und Validität auf (O'Keefe, 2002). Die Teilnahme an dem Rezeptionsexperiment war für die Studierenden freiwillig und die Daten wurden anonym erfasst. Dies wirkt möglichen *demand effects* wahrscheinlich entgegen.

Damit die Versuchspersonen die Beantwortung des Fragebogens nicht vorzeitig abbrachen oder absichtlich wegen Überforderung falsch beantworteten (response bias), wurde darauf geachtet, dass dieser kognitiv verständlich, kurz und nicht zu monoton war.[110] Alle Fragen wurden eindi-

110 Die Fragebögen von Versuchspersonen, die die Befragung kurzfristig abbrachen und nur den ersten Teil ausfüllten, wurden nicht codiert. Füllten Versuchspersonen der Experimentalgruppen mehr als ein Drittel der Fragen nicht aus, so gingen ihre Fragebögen nicht

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

mensional gestellt. Im Pretest wurden alle Frageformulierungen und Begriffe auf ihre Verständlichkeit hin überprüft und einzelne Items und Formulierungen leicht verändert oder mit zusätzlichen Erläuterungen versehen.[111] Es wurden Überleitungen wie *Den ersten Teil haben Sie fast geschafft!* eingeführt, nie zwei Listenfragen direkt nacheinander abgefragt und bei den semantischen Differenzialen wurde auch darauf geachtet, dass der jeweils positive Pol nicht durchweg auf einer Seite ist.

Zur Sicherung der Inhaltsvalidität wurde das Konstrukt *Überzeugung* durch verschiedene Überzeugungsurteile erfasst. Ob die gemessenen manifesten Überzeugungen mit den latenten realen Überzeugungen übereinstimmen (vgl. Kapitel 4.1.1) bzw., ob nur kurzfristige oder auch langfristige Überzeugungen gemessen wurden, kann in dieser Untersuchung, da sie querschnittlich angelegt ist, nicht eruiert oder vorhergesagt werden. Wie bspw. von Schwarz (2006) befürwortet, sollte sich die Medienwirkungsforschung viel mehr darauf konzentrieren, kontextsensitive Bewertungen zu untersuchen, als nach überdauernden Einstellungen zu fahnden.

Insgesamt gab es bezüglich der Überzeugungen für jeden Beitrag eine andere Fragebogenversion. Besonders im Bereich der Einstellungs- und Überzeugungsfragen können kleine Veränderungen in der Fragenformulierung schon zu anderen Ergebnissen führen (Brosius et al., 2012). Daher wurde besonders darauf geachtet, dass die Frageformen zu den jeweiligen Treatments so identisch wie möglich waren. Untersucht werden ausschließlich die bewussten Überzeugungen zur Hauptthese, da nur diese mittels der Befragung, die letztendlich auf Selbstbeobachtung der Versuchspersonen beruht, erfassbar sind. Welche subjektiven Urteilsheuristiken für die Überzeugungsurteile von Rezipienten genutzt wurden, konnte in dieser Untersuchung bspw. nicht direkt erfasst werden. Dazu wäre ein detaillierterer Fragebogen oder eine andere Methodik, wie bspw. die Methode des lauten Denkens, geeignet gewesen.

Hatten die Rezipienten schon eine Voreinstellung bzw. vorherige Überzeugungen zum präsentierten Sachverhalt im TV-Wissenschaftsbeitrag, waren sie vor und nach der Stimuluspräsentation mit denselben Fragen

mit in die endgültige Stichprobe ein. 14 Fragebögen wurden daher bei der Datenbereinigung entfernt. Dies hat den Vorteil, dass in der endgültigen Stichprobe weniger fehlende Werte sind, die bei der statistischen Berechnung hinderlich sein könnten.

111 Die zugeschriebene Glaubwürdigkeit wurde mit Hilfe der *Message- und Messenger-Scale* nach Roberts (2010) erfasst (vgl. Kapitel 8.3), im Pretest zeigte sich, dass es hier zu Verständnisschwierigkeiten bei den Versuchspersonen kommen kann, da diese Schwierigkeiten hatten Informationsquelle und Inhalt auseinander zu halten. Daher wurden hier präzise Erläuterungen in den Fragebogen eingebaut.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

konfrontiert. Dies konnte zu Erinnerungseffekten führen, welche die Ergebnisse verzerren könnten. Allerdings waren durch die einheitlichen Fragestellungen die Befragungsergebnisse sehr gut miteinander vergleichbar, deshalb wurde sich dafür entschieden, die Fragen nicht umzuformulieren. Um Erinnerungseffekte sowie mögliche Ausstrahlungseffekte (Konsistenz- oder Kontrasteffekte) der einzelnen Überzeugungsurteile aufeinander, möglichst zu minimieren, wurden zwischen die einzelnen Fragen zu den Überzeugungsurteilen andere Fragen (Pufferfragen) gesetzt (wie bspw. Brosius et al. (2012) und Möhring und Schlütz (2010) vorschlagen). Des Weiteren befindet sich ein Hinweis auf dem Fragebogen vor und nach dem Treatment, dass die Versuchspersonen nicht zurückblättern und die Fragen der Reihe nach beantworten sollen.

Es wurden fast ausschließlich geschlossene Fragen mit vorgegebenen Antwortkategorien erstellt, da eine große Fallzahl im Rezeptionsexperiment angestrebt war und nur die Merkmale erfasst werden sollten, die relevant waren.[112] Den Antworten auf geschlossene Fragen kann im Allgemeinen größere Reliabilität zugeschrieben werden als den Antworten offener Fragen (Möhring & Schlütz, 2010). Die geschlossene Frageform macht die Antworten der Versuchspersonen vergleichbar und statistisch bearbeitbar. Der Bezugsrahmen ist vereinheitlicht, das erhöht auch die Validität. Geschlossene Fragen können allerdings auch zu einer Antwortverzerrung führen. Die Versuchspersonen könnten mitunter zu einer Antwortausdifferenzierung gedrängt werden, die real nicht besteht und an anderer Stelle könnten Informationen verloren gehen, die in den Antwortvorgaben der geschlossenen Frageform nicht berücksichtigt sind (Brosius et al., 2012; Möhring & Schlütz, 2010). Bei adjektivischen oder adverbiellen Skalen bleiben individuelle semantische Interpretationen unkontrolliert. Jeder Rezipient könnte unter einem Adjektiv etwas anderes verstehen. Daher wurden im ersten Pretest (Fragebogenkonferenz) die einzelnen Items besprochen und gegebenenfalls durch Erläuterungen ergänzt (vgl. bspw. Kapitel 7.3.2 Need for Cognition-Skala).

112 Bei offenen Fragen besteht die Gefahr der Ergebnisverzerrung durch unterschiedliche Eloquenz der Befragten und es kann zu einer schwer vergleichbaren Zersplitterung der Antworten kommen. Das heißt weiter, dass zu viel Unwesentliches von den Versuchspersonen angeführt werden könnte und das, was wirklich interessiert, käme zu kurz oder darauf könnte überhaupt nicht Bezug genommen werden (Brosius, Haas & Koschel, 2012). Nur die soziodemografischen Daten wurden teilweise im Fragebogen offen erhoben.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Daten aus Ratingskalen sind realistisch gesehen ordinal-skaliert also nur quasimetrisch. In der Untersuchung wurde sich der pragmatischen Sichtweise von Möhring und Schlütz (2010) angeschlossen und davon ausgegangen, dass die Verletzungen der Intervalleigenschaften insgesamt nicht so gravierend sind, als dass auf die Anwendung statistischer Verfahren verzichtet werden müsse. Es wird davon ausgegangen, dass die Abstände zwischen den Antwortvorgaben von den Rezipienten durch die verbale, numerische oder visuelle Unterstützung im Fragebogen als gleich angesehen wurden. Bei den Ratingskalen der Untersuchung wurde überwiegend auf numerische Skalen zurückgegriffen, um eine Gleichabständigkeit zu gewährleisten, da verbale Anker ein Problem für die Validität darstellen könnten (Möhring & Schlütz, 2010).

Alle Fragen im Fragebogen haben eine ungerade Anzahl an Skalenpunkten und folglich einen Mittelpunkt. Keine Versuchsperson sollte künstlich dazu gedrängt werden ihre Einstellung oder Überzeugung in irgendeine Richtung zu verlagern. Dies verbessert auch die Validität der Untersuchung (Möhring & Schlütz, 2010; Scholl, 2009). Die Rezipienten wurden des Weiteren weder vor noch nach der Stimuluspräsentation dazu gedrängt überhaupt eine Einstellung oder Überzeugung zu dem abgefragten Sachverhalt zu haben; jederzeit konnten sie auch mit *weiß nicht* antworten. Bei den Ratingskalen wurde darauf geachtet, dass die Spannweite der Antwortvorgaben bis auf eine Ausnahme bei nur fünf übersichtlichen Stufen liegt. Bei den Überzeugungsurteilen wurde sich im Experiment, u. a. wegen der besseren Vergleichbarkeit mit den dargestellten Evidenzmustern, auf Fragen nach Prozentsätzen (0% bis 100%) festgelegt. Möhring und Schlütz (2010) merken an, dass Ratingskalen auch zu differenziert sein können. Sie argumentieren, dass eine Differenziertheit verlangt werden würde, die so nicht alle Befragten haben können. Doch die Entscheidung der Befragten, sich bspw. nur auf Zehnerschritte zu beschränken, sollte ihnen in diesem Experiment bei den Überzeugungsurteilen selbst überlassen werden. Sollte der Fall eintreten, dass sich die Versuchspersonen nur auf Zehnerschritte beschränken, waren in Bezug auf deren Auswertung der Ergebnisse keine Nachteile vorhersehbar. Bei den Überzeugungsurteilen war es eher als Verlust anzusehen, Differenziertheit aufgrund einer zu geringen Stufenzahl zu verlieren.

Güte der experimentellen Variation

Kontrollgruppen sind ein entscheidendes Kriterium für die Validität der Forschungsergebnisse (Scholl, 2009). Zwei Kontrollgruppen dienen in

Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

dieser Untersuchung als Stütze für die Erkenntnis, dass die unterschiedlichen Überzeugungen der Rezipienten tatsächlich auf der Präsentation des Treatments beruhen (vgl. Kapitel 7). Die Probanden der KG1 (n = 31) bekamen den Vorher- und Nachher-Fragebogen des Beitrags *Darmspiegelung*, sahen diesen Beitrag aber nicht. Es konnte nun im Anschluss kontrolliert werden, ob sich die Überzeugungen der Rezipienten, welche das Treatment rezipierten, von denen unterschieden, welche das Treatment nicht rezipierten. In der folgenden Tabelle 13 sind die Unterschiede der Überzeugungsurteile zwischen den beiden Gruppen anhand der t-Werte aufgelistet.

Tabelle 13: Überzeugungsurteile (t2) von KG1 und EG3

Überzeugungsurteile (t2)	KG1 *M* (*SD*) (*n* = 31)	EG3 *M* (*SD*) (*n* = 174)	*t*	*df*
Belief	.43 (0.21)	.62 (0.24)	2.24**	154
Doubt	.33 (0.21)	.41 (0.80)	0.33	150
Ungewissheit	.66 (0.25)	.46 (0.32)	-2.03**	154
Plausibilität	.46 (0.25)	.65 (0.23)	2.22**	147
Gegenplausibilität	.33 (0.24)	.33 (0.24)	0.93	152

**$p < .05$

Die signifikanten t-Werte bei drei von fünf Überzeugungsurteilen zeigen deutlich, dass es sehr unwahrscheinlich ist, dass die beiden Mittelwerte der Gruppen aus Populationen mit demselben Mittelwert stammen. Es gibt folglich einen systematischen Unterschied zwischen den beiden Gruppen, der die Mittelwertsdifferenz verursacht hat. Dieser kann durch die Gruppenvariable *mit/ohne Treatment* erklärt werden.

Die Probanden der KG2 (n = 19) bekamen das Treatment *Darmspiegelung* mit passendem Nachher-Fragebogen. Die Antworten dieser Gruppe dürften sich, um eine hohe Güte zu attestieren, nicht signifikant unterscheiden von der EG, welche einen Vorher- und Nachher-Fragebogen zum entsprechenden Treatment hatte. Bei den Evidenzvariablen sind alle Varianzen laut dem Levene-Test auf Varianzhomogenität gleich und auch die t-Werte unterscheiden sich nicht signifikant voneinander (vgl. Tabelle 14).

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Tabelle 14: Überzeugungsurteile (t2) von KG2 und EG3

Überzeugungsurteile (t2)	KG2 *M* (*SD*) (*n* = 19)	EG3 *M* (*SD*) (*n* = 174)	*t*	*df*
Belief	.58 (0.23)	.62 (0.24)	0.70	164
Doubt	.35 (0.20)	.41 (0.80)	0.33	158
Ungewissheit	.48 (0.27)	.46 (0.32)	-0.40	161
Plausibilität	.61 (0.23)	.65 (0.23)	0.22	158
Gegenplausibilität	.34 (0.20)	.33 (0.24)	-0.20	161

Das macht deutlich, dass es sehr wahrscheinlich ist, dass die beiden Mittelwerte der Gruppen aus Populationen mit demselben Mittelwert stammen. Es gibt folglich keinen systematischen Unterschied zwischen den beiden Gruppen. Insgesamt weisen also beide Kontrollgruppen auf eine hohe Güte der experimentellen Variation in diesem Rezeptionsexperiment hin. Streng genommen müssten alle Treatments entsprechend der KG1 und KG2 überprüft werden. Aus ökonomischen Gründen wurde sich in dieser Untersuchung auf eines beschränkt.

Im folgenden Kapitel werden nun die Ergebnisse der Inhaltsanalyse und des Rezeptionsexperiments aufgezeigt.

https://doi.org/10.5771/9783845275468-145
Generiert durch IP '3.139.235.164', am 27.09.2024, 08:26:05.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.