

Vorhersage von Präferenzen: Experimentelle Untersuchung zu Antwortskalen und Framing

Prediction of Preferences: Experimental Investigation of Response Scales and Framing

Mauricia Port

Fachbereich 06 Psychologie und Sportwissenschaften, Justus-Liebig-Universität Gießen

PSY-BA-WPAM Abschlussmodul Psychologische Methodenlehre

Erstgutachterin: M.Sc. Jaqueline Nahler

Zweitgutachter: Prof. Dr. Christof Schuster

Gießen, 18. Juni 2025

Abstract

Da Menschen häufig vor Entscheidungen unter unsicheren Bedingungen stehen, kann es hilfreich sein, Vorhersagen über Präferenzen von Menschen treffen zu können. Für bestmögliche Vorhersagen ist es notwendig, Einflussfaktoren zu verstehen. Frühere Befunde deuten darauf hin, dass verbale Antwortskalen im Vergleich zu numerischen bessere Präferenzvorhersagen ermöglichen. Zudem legen bisherige Befunde nahe, dass der Einfluss von Framing bei vielen Arten von Entscheidungen auftritt. Daraus resultierend ist das Ziel der vorliegenden Arbeit, den Einfluss von Antwortskalen und Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit zu untersuchen. Zusätzlich wurde der Einfluss von Risikokompetenz berücksichtigt. Eine empirische Untersuchung in Form eines zweifaktoriellen Experiments mit Zwischensubjekt-Design wurde als Online-Studie durchgeführt. Dabei wurde der Einfluss der Antwortskala (numerisch, verbal) und des Framings (positiv, negativ) auf die Korrelation zwischen der Selbsteinschätzung bei zwölf Fragen und der anschließenden Präferenzeinordnung der Fragen in eine Rangliste untersucht. Die Kovarianzanalyse mit der Kovariate Risikokompetenz zeigte keinen statistisch signifikanten Einfluss der Antwortskala, des Framings sowie deren Interaktion auf die Präferenzvorhersage. Auch die Risikokompetenz hatte keinen statistisch signifikanten Einfluss. Mögliche Gründe für die Ergebnisse könnten methodische Faktoren, aber auch zeitliche und kulturelle Veränderungen sein. Die vorliegende Arbeit leistet einen Beitrag zum Verständnis der Einflussfaktoren auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit und bietet einen Ausgangspunkt für zukünftige Forschung.

Inhaltsverzeichnis

| | |
|--|----|
| 1. Theorie | 1 |
| 1.1 Einleitung..... | 1 |
| 1.2 Das Experiment 2 von Windschitl und Wells (1996) | 3 |
| 1.3 Der Einfluss von Antwortskalen..... | 4 |
| 1.4 Das Konzept des Framings | 6 |
| 1.5 Der Einfluss von Risikokompetenz | 8 |
| 1.6 Herleitung der Hypothesen..... | 9 |
| 2. Methode | 10 |
| 2.1 Versuchsdesign | 10 |
| 2.2 Beschreibung des Fragebogens | 12 |
| 2.3 Berlin Numeracy Test..... | 13 |
| 2.4 Versuchsablauf..... | 14 |
| 2.5 Stichprobenbeschreibung..... | 15 |
| 3. Ergebnisse | 17 |
| 3.1 Datenaufbereitung | 17 |
| 3.2 Vorläufige Analysen..... | 21 |
| 3.3 Deskriptive Statistik | 22 |
| 3.4 Inferenzstatistik | 24 |
| 4. Diskussion..... | 26 |
| 4.1 Ergebnisdiskussion | 26 |
| 4.2 Diskussion der explorativen Datenanalyse..... | 28 |
| 4.3 Limitationen..... | 31 |
| 4.4 Fazit und Ausblick | 33 |
| Literaturverzeichnis..... | V |
| Anhang | IX |

Abbildungsverzeichnis

| | |
|---|-------|
| Abbildung 1 Histogramm der Bearbeitungsdauer..... | 19 |
| Abbildung 2 Histogramm der Z-Werte vor Ausschluss extremer Ausreißer | 20 |
| Abbildung 3 Bereinigte Z-Mittelwerte für die experimentellen Bedingungen | 23 |
| Abbildung B1 Adaptives Testformat des BNT inklusive Punktevergabe | XIII |
| Abbildung C1 Versuchsablauf der experimentellen Untersuchung | XIV |
| Abbildung F1 Bereinigte Z-Mittelwerte zwischen Basisrate und Selbsteinschätzung | XIX |
| Abbildung G1 Standardabweichungen der Versuchspersonen bei den zwölf Fragen..... | XXI |
| Abbildung G2 Mittelwerte der Versuchspersonen bei den zwölf Fragen | XXIII |

Tabellenverzeichnis

| | |
|---|-------|
| Tabelle 1 Experimentelle Bedingungen | 11 |
| Tabelle 2 Deskriptive Statistik der Z-Werte | 24 |
| Tabelle 3 Deskriptive Statistik der Punktzahl im BNT je experimenteller Bedingung..... | 24 |
| Tabelle A1 Themen und dazugehörige Basisraten der zwölf Fragen des Fragebogens | X |
| Tabelle A2 Beschriftung der 11-stufigen numerischen und verbalen Antwortskala | XI |
| Tabelle B1 Fragen des BNT mit der jeweils korrekten Antwort | XII |
| Tabelle D1 Absolute und relative Häufigkeiten der Stichprobencharakteristika | XV |
| Tabelle F1 Deskriptive Statistik der Z-Werte zwischen Basisrate und Selbsteinschätzung | XVIII |
| Tabelle F2 Inferenzstatistik der Z-Werte zwischen Basisrate und Selbsteinschätzung | XVIII |
| Tabelle F3 Deskriptive Statistik der Z-Werte zwischen Basisrate und Präferenzrangliste | XX |
| Tabelle F4 Inferenzstatistik der Z-Werte zwischen Basisrate und Präferenzrangliste | XX |

1. Theorie

Das erste Kapitel umfasst die Einleitung und den theoretischen Hintergrund. Zu Beginn wird die Relevanz der vorliegenden Arbeit erläutert. Die Forschungsfragen werden abgeleitet und die Zielsetzung wird dargestellt. Es werden grundlegende Begriffe definiert, die für das Verständnis der Arbeit notwendig sind. Da das Experiment 2 von Windschitl und Wells (1996) repliziert wird, wird dieses näher beschrieben. Darauffolgend wird der Einfluss von Antwortskalen dargelegt und das Konzept des Framings erklärt. Zudem wird erläutert, inwiefern es in der vorliegenden Arbeit sinnvoll ist, die Risikokompetenz zu berücksichtigen. Zum Abschluss des Kapitels werden anhand der Forschungsfragen und des angeführten theoretischen Hintergrundes die Hypothesen hergeleitet.

1.1 Einleitung

Menschen sind in vielen Lebenssituationen mit Entscheidungen unter unsicheren Bedingungen konfrontiert (Platt & Huettel, 2008). Das wird beispielsweise bei der Betrachtung des medizinischen, wirtschaftlichen oder versicherungstechnischen Bereichs unmittelbar deutlich. Wird sich eine erkrankte Person für die Einnahme eines Medikaments entscheiden oder natürliche Heilmethoden verwenden? Wird eine Person, die über die nötigen finanziellen Mittel verfügt, in eine Aktie investieren oder das Geld auf einem Bankkonto sparen? Wird eine arbeitstätige Person eine Berufsunfähigkeitsversicherung abschließen oder das Risiko im Falle selbst tragen?

Daraus ergibt sich unmittelbar die praktische Relevanz der vorliegenden Arbeit, sowohl für den Fachbereich der Psychologie als auch für andere Fachbereiche. Anhand der oben genannten Beispiele wird deutlich, dass Entscheidungen unter unsicheren Bedingungen erhebliche Konsequenzen mit sich bringen können. Aufgrund dessen wird ersichtlich, wie relevant es ist, Entscheidungen von Menschen nicht nur retrospektiv zu erklären, sondern auch Vorhersagen über die Präferenzen von Menschen unter psychologischer Unsicherheit treffen zu können. Um dabei eine bestmögliche Vorhersage von Präferenzen ermöglichen zu können, ist es essenziell, potenzielle Einflussfaktoren wie beispielsweise das Fragen- und das Antwortformat zu untersuchen. Die Untersuchung der Einflussfaktoren schafft die Grundlage, um besonders geeignete Prädiktoren für die Präferenzen von Menschen unter psychologischer Unsicherheit herausarbeiten zu können.

Ein wichtiger Bestandteil bei der Vorhersage von Präferenzen unter psychologischer Unsicherheit ist die Auswahl von Messinstrumenten, da unterschiedliche Antwortskalen zu unterschiedlichen Einschätzungen führen können (Lermer et al., 2018). Auch das

Fragenformat kann einen Einfluss auf die Einschätzungen haben (Levin et al., 1998). Folglich stellt sich die Frage, wie Informationen präsentiert und abgefragt werden sollten, um eine bestmögliche Vorhersage von Präferenzen unter psychologischer Unsicherheit zu ermöglichen.

Einer ähnlichen Frage sind Windschitl und Wells (1996) in ihrem Experiment 2 nachgegangen. Sie haben den Einfluss der Antwortskala auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit untersucht. Das Experiment weist darauf hin, dass verbale Unsicherheitsantworten im Vergleich zu numerischen bessere Prädiktoren für Präferenzen abbilden (Windschitl & Wells, 1996). Demzufolge lieferte das Experiment 2 von Windschitl und Wells (1996) wichtige Erkenntnisse über den Einfluss des Antwortformats auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit. Der Einfluss des Fragenformats hingegen wurde in dem Experiment 2 nicht betrachtet. Aufgrund dessen untersucht die vorliegende Arbeit im Hinblick auf das Fragenformat das Framing als einen zusätzlichen Faktor. Dabei wird der Frage nachgegangen, welchen Einfluss die unterschiedliche Formulierung von objektiv äquivalenten Informationen auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit hat.

Um herauszufinden, welchen Einfluss Antwortskalen und Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit haben, untersucht die vorliegende Arbeit die im Folgenden dargestellten Forschungsfragen. Die erste Forschungsfrage thematisiert, analog zu dem Experiment 2 von Windschitl und Wells (1996), den Einfluss verschiedener Antwortskalen auf die Vorhersage von Präferenzen. Die zweite Forschungsfrage überprüft den Einfluss von Framing auf die Vorhersage von Präferenzen. Konkretisiert lauten die Forschungsfragen folgendermaßen:

- 1 Ermöglicht die verbale Antwortskala im Vergleich zu der numerischen Antwortskala eine bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit?
- 2 Gibt es einen Einfluss des Framings auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit?

Zur Beantwortung der Forschungsfragen wurde eine experimentelle Untersuchung konzipiert, die auf dem Experiment 2 von Windschitl und Wells (1996) aufbaut. Die Kernidee der experimentellen Untersuchung ist die systematische Untersuchung des Einflusses von Antwortskalen und Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit. Das Ziel besteht darin, herauszufinden, welche Antwortskala, welche Framingversion und welche Kombination von Framing und Antwortskala bestmögliche Prädiktoren für Präferenzen unter psychologischer Unsicherheit abbilden. Dabei baut die

vorliegende Arbeit auf früheren Arbeiten auf, geht aber auch über bisher bestehende Erkenntnisse hinaus. Die Ergebnisse von Windschitl und Wells (1996) werden nach über 25 Jahren im deutschen Sprachraum überprüft und zusätzlich wird der Einfluss von Framing auf die Vorhersage von Präferenzen, die Interaktion von Framing und Antwortskala sowie der Einfluss von Risikokompetenz als ein individuelles Merkmal betrachtet.

1.2 Das Experiment 2 von Windschitl und Wells (1996)

Windschitl und Wells (1996) untersuchten in ihrem Experiment 2, ob eine verbale Antwortskala im Vergleich zu einer numerischen Antwortskala eine bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit ermöglicht.

Demzufolge ist die Thematik der Unsicherheit ein zentraler Aspekt. Die American Psychological Association (2018) definiert Unsicherheit als einen Zustand, in dem etwas, beispielsweise die Wahrscheinlichkeit eines bestimmten Ergebnisses, nicht genau oder präzise bekannt ist. Unsicherheit wird häufig als psychologisches Konstrukt betrachtet, da keine Unsicherheit vorliegen würde, wenn das Wissen einer Person vollständig wäre (Windschitl & Wells, 1996). Deshalb wird von psychologischer Unsicherheit gesprochen.

Zur Beantwortung der Fragestellung wurden die Versuchspersonen in dem Experiment 2 von Windschitl und Wells (1996) randomisiert entweder zu der Bedingung mit der numerischen oder zu der Bedingung mit der verbalen Antwortskala zugeteilt. Durchgeführt wurde das Experiment mit $N = 209$ Studierenden der Iowa State University. Dabei wurden den Versuchspersonen zunächst zwölf verschiedene Themen präsentiert. Zu jedem Thema wurde eine fiktive Prozentzahl angegeben, durch die mitgeteilt wurde, wie viele der zuvor befragten Versuchspersonen die Frage zu dem jeweiligen Thema richtig beantwortet haben. Als Unsicherheitsinformationen verwendeten Windschitl und Wells (1996) in dem Experiment 2 demnach Basisraten, also die Häufigkeit des Vorkommens eines Merkmals in der Grundgesamtheit (Wirtz, 2019). Zu den zwölf verschiedenen Themen sollten die Versuchspersonen entweder auf der numerischen oder auf der verbalen Antwortskala eine Einschätzung darüber abgeben, wie wahrscheinlich es ist, dass sie eine Frage zu dem jeweiligen Thema richtig beantworten würden (Windschitl & Wells, 1996). Im Anschluss daran sollten die Versuchspersonen die zwölf Themen in eine Präferenzrangliste bringen. Dabei sollten sie angeben, zu welchem Thema sie am liebsten eine Frage gestellt bekommen würden, wenn sie für eine richtige Beantwortung der Frage einen Geldpreis gewinnen würden. Die Annahme dabei ist, dass Menschen eher dazu neigen, eine Option mit einer sichereren Belohnung zu bevorzugen als eine Option mit einer weniger sicheren Belohnung von gleichem Wert (Windschitl & Wells, 1996). Da die Versuchspersonen lediglich über die

zwölf verschiedenen Themen sowie die dazugehörige Basisrate informiert wurden, basierte ihre Selbsteinschätzung und ihre Präferenzrangliste allein auf diesen Informationen und nicht auf der eigentlichen Frage selbst (Windschitl & Wells, 1996).

Als Ergebnis zeigte sich, dass die verbale Antwortskala im Vergleich zu der numerischen Antwortskala einen besseren Prädiktor für Präferenzen unter psychologischer Unsicherheit darstellt (Windschitl & Wells, 1996). Das heißt, die verbale Antwortskala ermöglichte eine statistisch signifikant bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit als die numerische Antwortskala (Windschitl & Wells, 1996).

Die vorliegende Arbeit hat das Experiment 2 von Windschitl und Wells (1996) repliziert. Sie stellt somit einen Ansatz dar, um zu überprüfen, ob die Ergebnisse aus dem Experiment 2 von Windschitl und Wells (1996) nach über 25 Jahren im deutschen Sprachraum und in Form einer Online-Studie reproduziert werden können. Replikationsstudien sind sowohl für die Generalisierung existierender wissenschaftlicher Theorien über bereits bekannte Gültigkeitsbereiche hinaus als auch für die Überprüfung und Rechtfertigung wissenschaftlicher Theorien unverzichtbar (Erdfelder & Ulrich, 2018).

1.3 Der Einfluss von Antwortskalen

Das Experiment 2 von Windschitl und Wells (1996) weist darauf hin, dass die Vorhersage von Präferenzen unter psychologischer Unsicherheit durch die Antwortskala beeinflusst wird. Demnach können unterschiedliche Antwortskalen zu unterschiedlichen Einschätzungen und somit zu unterschiedlichen Ergebnissen führen (Lerner et al., 2018). Allerdings scheint der Einfluss von unterschiedlichen Antwortskalen in der psychologischen Forschung häufig vernachlässigt und die Verwendung bestimmter Antwortskalen nicht ausreichend hinterfragt zu werden (Rosenbaum & Valsiner, 2011; Windschitl & Wells, 1996).

Unsicherheitseinschätzungen von Menschen sind für die Forschung oftmals in numerischer und somit quantitativer Form einfacher zu verarbeiten als in verbaler und somit qualitativer Form (Lerner et al., 2018). Das liegt nicht zuletzt daran, dass numerische Unsicherheitseinschätzungen für viele statistische Analysen einfacher zu nutzen sind und überdies direkt mit objektiven Wahrscheinlichkeiten verglichen werden können (Windschitl & Wells, 1996). Darüber hinaus können verbale Quantifizierungen von Wahrscheinlichkeiten subjektiv interpretiert werden und somit eine gewisse Ungenauigkeit mit sich bringen, wodurch die Vergleichbarkeit zwischen Unsicherheitseinschätzungen eingeschränkt sein kann (Lerner et al., 2018; Schwarz, 1999). Demgegenüber bevorzugen die Versuchspersonen oftmals qualitative Antworten und weisen Schwierigkeiten mit quantitativen Antworten auf (Rohrman, 1978).

Insgesamt scheint es derzeit keinen Standard und dadurch zahlreiche Antwortskalen zur Messung von Unsicherheitseinschätzungen zu geben (Haase et al., 2013; Lermer et al., 2018; Menold, 2019). Aufgrund dieser Vielfalt fokussiert die vorliegende Arbeit, analog zu dem Experiment 2 von Windschitl und Wells (1996), die grundlegende Unterscheidung zwischen numerischen und verbalen Antwortskalen.

Die Ergebnisse aus dem Experiment 2 wurden von Windschitl und Wells (1996) dadurch erklärt, dass durch die numerische und die verbale Antwortskala jeweils zwei unterschiedliche Arten von Denkprozessen angeregt werden: Das deliberative und das intuitive Denken.

Nach Windschitl und Wells (1996) führen numerische Antwortskalen zu deliberativem Denken. Das heißt, es werden logikgeleitete und regelbasierte Prozesse ausgelöst, die Menschen für Unsicherheitsinformationen wie die Basisraten sensibilisieren (Windschitl & Wells, 1996). Dieses deliberative Denken kann Menschen dazu leiten, eine möglichst präzise Antwort zu geben, die jedoch nicht zwangsläufig mit ihrer subjektiv empfundenen Einschätzung übereinstimmen muss (Lermer et al., 2013). Windschitl und Wells (1996) gehen davon aus, dass numerische Antwortskalen nicht unbedingt widerspiegeln, wie Menschen natürlicherweise denken. Stattdessen neigen Menschen in vielen Alltagssituationen eher zu intuitivem Denken (Windschitl & Wells, 1996). Intuitives Denken ist schnell, assoziativ, automatisch und wird tendenziell mit verbalen Antwortskalen assoziiert (Windschitl & Wells, 1996). Laut Windschitl und Wells (1996) verwenden die meisten Menschen auch im Alltag eher verbale anstatt numerische Ausdrücke, wenn sie über Unsicherheiten und Wahrscheinlichkeiten sprechen. Zudem werden Grundregeln der Wahrscheinlichkeit durch verbale Antwortskalen weniger salient als durch numerische Antwortskalen (Windschitl & Wells, 1996). Aufgrund des intuitiven Denkens, das durch verbale Antwortskalen angeregt wird, können diese im Vergleich zu numerischen Antwortskalen hinsichtlich der Denkweise von Menschen in Bezug auf Unsicherheiten informativer sein (Windschitl & Wells, 1996). Hinzukommt, dass verbale Antwortskalen nicht nur bessere Prädiktoren für die Präferenzen, sondern auch bessere Prädiktoren für die Verhaltensintentionen sowie für das Verhalten von Menschen darzustellen scheinen (Weinstein et al., 2007; Windschitl & Wells, 1996).

Insgesamt können numerische und verbale Antwortskalen für unterschiedliche Fragestellungen unterschiedlich informativ und demnach unterschiedlich geeignet sein (Windschitl & Wells, 1996). Das verdeutlicht, wie relevant es ist, das Ziel einer Untersuchung vor der Messung festzulegen (Lermer et al., 2018). So kann die Antwortskala ausgewählt werden, die den für das jeweilige Ziel relevanten Denkprozess anregt (Lermer et al., 2018).

Einerseits können numerische Antwortskalen im Vergleich zu verbalen Antwortskalen besonders geeignet sein, um die Genauigkeit von menschlichen Unsicherheitseinschätzungen zu untersuchen (Windschitl & Wells, 1996). Das kann beispielsweise sinnvoll sein, wenn untersucht werden soll, wie gut Menschen darin sind, objektive Wahrscheinlichkeiten möglichst präzise einzuschätzen. Andererseits können mit verbalen Antwortskalen die subjektiv empfundenen Einschätzungen von Menschen über Unsicherheiten besser abgebildet werden (Windschitl & Wells, 1996).

Darüber hinaus scheinen Antwortskalen, die intuitives Denken anregen, wie die verbale Antwortskala, sensitiver in Bezug auf Kontextmanipulationen und Formulierungen der Frage zu sein (Lermer et al., 2013; Windschitl & Wells, 1996). Insbesondere dieser Aspekt kann für den Einfluss von Framing relevant sein, da Framing eine Kontextmanipulation darstellt und in der vorliegenden Arbeit die Formulierung der Frage betrifft.

1.4 Das Konzept des Framings

Tversky und Kahneman haben im Jahr 1981 dem Konzept des Framings seinen Ursprung gegeben. Im Kern geht es beim Framing um die Art der Formulierung von Informationen. Menschen reagieren oft unterschiedlich auf verschieden formulierte, aber objektiv äquivalente Informationen (Levin et al., 1998). Wenn dieselbe Information unterschiedlich formuliert wird, führen psychologische Prinzipien, die die Wahrnehmung und Einschätzung von Wahrscheinlichkeiten beeinflussen, zu vorhersehbaren Veränderungen von Präferenzen (Tversky & Kahneman, 1981).

Um den Einfluss von Framing bei Entscheidungen unter Unsicherheit zu untersuchen, haben Tversky und Kahneman (1981) das Problem der asiatischen Krankheit entwickelt. Dabei ging es um ein fiktives Szenario, in dem eine asiatische Krankheit ausbrechen wird, die voraussichtlich 600 Menschen töten wird. Die Personen wurden in zwei Gruppen eingeteilt, in denen ihnen jeweils zwei Programme zum Umgang mit der Krankheit zur Auswahl gestellt wurden. In der Gruppe mit positivem Framing lautete das erste Programm, dass 200 Menschen gerettet werden. Das zweite Programm lautete, dass mit einer Wahrscheinlichkeit von einem Drittel alle 600 Menschen gerettet werden und mit einer Wahrscheinlichkeit von zwei Dritteln niemand gerettet wird. In der Gruppe mit negativem Framing lautete das erste Programm, dass 400 Menschen sterben werden. Das zweite Programm lautete, dass mit einer Wahrscheinlichkeit von einem Drittel niemand sterben wird und mit einer Wahrscheinlichkeit von zwei Dritteln alle 600 Menschen sterben werden. Beiden Gruppen wurden objektiv identische Programme präsentiert. Der Unterschied liegt in der Formulierung, indem bei dem positiven Framing die Rettung und bei dem negativen Framing das Sterben von Menschen

betont wurde (Tversky & Kahneman, 1981). Die Mehrheit der Personen in der Gruppe mit positivem Framing entschied sich für das erste Programm, also die risikoärmere Variante (Tversky & Kahneman, 1981). Die Mehrheit der Personen in der Gruppe mit negativem Framing entschied sich für das zweite Programm, also die risikoreichere Variante (Tversky & Kahneman, 1981). Somit konnten Tversky und Kahneman (1981) zeigen, dass Präferenzen von der Formulierung des Entscheidungsproblems abhängen und demnach durch das Framing beeinflusst werden. Obwohl die Basisraten bei den Entscheidungen jeweils identisch waren, zeigte sich, dass sich die Mehrheit der Personen bei Entscheidungen, die Gewinne beinhalten, risikoavers verhält, wohingegen sich die Mehrheit der Personen bei Entscheidungen, die Verluste beinhalten, risikofreudig verhält (Tversky & Kahneman, 1981).

Diese zu beobachtende Abweichung von der Rationalität bei dem Entscheidungsprozess wurde mit Hilfe der *Prospect Theory* von Kahneman und Tversky (1979) erklärt. Der Prospect Theory zufolge werden Ergebnisse als Gewinn oder Verlust bewertet, wobei das Framing eine zentrale Rolle spielt. Die subjektive Wertfunktion für Gewinne und Verluste ist dabei S-förmig, also konkav für Gewinne und konvex für Verluste. Zudem ist die Wertfunktion im Bereich der Gewinne flacher als im Bereich der Verluste. Das bedeutet, dass Verluste stärker gewichtet werden als Gewinne und somit das Unbehagen, das durch Verlust entsteht, im Allgemeinen stärker ist als die Freude über einen gleich hohen Gewinn (Kahneman & Tversky, 1979). Aufgrund dessen scheinen Menschen eher bereit zu sein, ein höheres Risiko einzugehen, um Verluste zu vermeiden. Anhand dessen kann das oben beschriebene Muster von Risikoaversion bei dem positiven Framing sowie Risikofreudigkeit bei dem negativen Framing erklärt werden.

Tversky und Kahneman (1981) haben mit ihrer Arbeit die Basis für das Konzept des Framings gelegt. Dennoch gibt es mittlerweile eine Vielzahl an aufbauenden Studien, die darauf hinweisen, dass der Einfluss von Framing nicht stets allgemeingültig und in gleicher Art und Weise auftritt. Vielmehr müssen weitere Unterscheidungen im Hinblick auf methodische und inhaltliche Aspekte in Bezug auf Framing getroffen werden (Kühberger, 1998). Einige solcher Aspekte werden von Kühberger (1998) in einer Meta-Analyse zum Einfluss von Framing beschrieben. So unter anderem die Art der Risikomanipulation. Bei dem Problem der asiatischen Krankheit erfolgte die Risikomanipulation durch Bezugnahme auf ein riskantes Ereignis, das mit einer bestimmten Wahrscheinlichkeit in der Zukunft eintritt. Demgegenüber kann sich die Risikomanipulation allerdings auch auf die Kennzeichnung des aktuellen Zustandes beziehen (Kühberger, 1998). Ein Beispiel hierfür ist, wie in der vorliegenden Arbeit verwendet, die Angabe, dass 67% der zuvor befragten

Personen die Frage richtig beantwortet haben (positives Framing), oder alternativ, dass 33% der zuvor befragten Personen die Frage falsch beantwortet haben (negatives Framing). In diesem Fall wird von der sogenannten Risikomanipulation durch Ergebnissalienz gesprochen (Kühberger, 1998).

Insgesamt konnte der Einfluss von Framing bei vielen Arten von Entscheidungen nachgewiesen werden und es zeigt sich ein breites Anwendungsfeld von Framing, wie unter anderem bei medizinischen und klinischen Entscheidungen, Wahrnehmungsurteilen sowie Verbraucherentscheidungen (Levin et al., 1998). Auch die Vorhersage von Präferenzen unter psychologischer Unsicherheit weist, wie in Kapitel 1.1 beschrieben, ein breites Anwendungsfeld auf. Deshalb soll der Einfluss von Framing auch in der vorliegenden Arbeit untersucht werden. Dies geht über das Experiment 2 von Windschitl und Wells (1996) hinaus, da die Basisrate dort stets als Gewinn, also positiv, präsentiert wurde. Es drängt sich jedoch die Frage auf, ob die Darstellung der Basisrate als Gewinn oder Verlust einen Einfluss auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit hat.

1.5 Der Einfluss von Risikokompetenz

Zusätzlich zu den Einflüssen des Fragen- und Antwortformats, die in Kapitel 1.3 und Kapitel 1.4 beschrieben wurden, können auch andere Faktoren die Vorhersage von Präferenzen unter psychologischer Unsicherheit beeinflussen. Ein solcher Faktor ist die Risikokompetenz. Die Risikokompetenz stellt ein individuelles Merkmal der Versuchspersonen dar und ist weder dem Fragen- noch dem Antwortformat zuzuordnen.

Ein zentraler Aspekt dieser Arbeit sind Risikodaten, wie beispielsweise die Darstellung der Unsicherheitsinformationen in Form von Basisraten. Für das Verständnis von Risikodaten spielt die Risikokompetenz eine Rolle. Die Risikokompetenz und das statistische Verständnis beeinflussen, ob Personen Situationen der psychologischen Unsicherheit anhand von Risikodaten, in Form von Wahrscheinlichkeiten, Proportionen und Prozentsätzen, richtig verstehen (Cokely et al., 2012). Dabei können geringe numerische Fähigkeiten die Risikowahrnehmung verzerren und das Verständnis von Risikoinformationen beeinträchtigen (Reyna et al., 2009). Auch die Anfälligkeit für die Art der Informationspräsentation kann durch geringe numerische Fähigkeiten verstärkt werden (Reyna et al., 2009). Ebenso weisen Peters et al. (2006) darauf hin, dass Personen mit geringen numerischen Fähigkeiten anfälliger für den Einfluss von Framing sind. Demzufolge kann eine gering ausgeprägte Risikokompetenz den Einfluss von Framing erhöhen.

Da in der vorliegenden Arbeit im Wesentlichen mit Risikodaten gearbeitet wird, ist es sinnvoll die Risikokompetenz als Kovariate zu erheben. So besteht die Möglichkeit den

Einfluss der Risikokompetenz auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit zu kontrollieren.

1.6 Herleitung der Hypothesen

Aufgrund des dargestellten theoretischen Hintergrundes und der formulierten Forschungsfragen leiten sich die experimentellen Hypothesen ab.

Hypothese 1 (H1): Die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste ist bei der verbalen Antwortskala höher als bei der numerischen Antwortskala.

Die H1 bezieht sich auf die erste Forschungsfrage. Es handelt sich um die Hypothese zum Haupteffekt der Antwortskala. Sie resultiert aus den in Kapitel 1.2 dargestellten Ergebnissen des Experiments 2 von Windschitl und Wells (1996). Basierend auf dem in Kapitel 1.3 erläuterten Einfluss von Antwortskalen im Hinblick auf das deliberative und das intuitive Denken postuliert die H1, dass die verbale Antwortskala im Vergleich zu der numerischen Antwortskala eine bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit ermöglicht.

Hypothese 2 (H2): Das Framing der Basisrate hat einen Einfluss auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste.

Die H2 bezieht sich auf die zweite Forschungsfrage. Es handelt sich um die Hypothese zum Haupteffekt des Framings. Basierend auf den in Kapitel 1.4 erläuterten Erkenntnissen zum Konzept des Framings postuliert die H2, dass das Framing einen Einfluss auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit hat.

Hypothese 3 (H3): Der Einfluss des Framings auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste variiert zwischen der numerischen und der verbalen Antwortskala.

Die H3 bezieht sich sowohl auf die erste als auch auf die zweite Forschungsfrage. Es handelt sich um die Hypothese zum Interaktionseffekt der Faktoren Framing \times Antwortskala. Basierend auf den in Kapitel 1.3 und Kapitel 1.4 erläuterten Erkenntnissen zum Einfluss von Antwortskalen sowie zum Konzept des Framings postuliert die H3, dass sich der Einfluss des Framings zwischen der numerischen und der verbalen Antwortskala unterscheidet.

2. Methode

Das zweite Kapitel beschäftigt sich mit der Methode. Dafür wird anfangs erläutert, weshalb sich für ein Experiment als Methode entschieden wurde. Zudem wird das Versuchsdesign dargelegt. Daraufhin folgt eine Beschreibung des Fragebogens und des *Berlin Numeracy Tests*, der in der vorliegenden Arbeit zur Messung von Risikokompetenz hinzugezogen wurde. Im Anschluss wird der Versuchsablauf erklärt. Abgerundet wird das Kapitel durch die Beschreibung der Stichprobe.

2.1 Versuchsdesign

Um die in Kapitel 1.1 angeführten Forschungsfragen beantworten und die in Kapitel 1.6 aufgestellten Hypothesen testen zu können, wurde eine empirische Untersuchung in Form eines zweifaktoriellen Experiments mit Zwischensubjekt-Design als Online-Studie realisiert, welches im Folgenden näher erläutert wird.

Bei einem Experiment handelt es sich um eine randomisierte kontrollierte Studie (Schreier et al., 2023). Andere Methoden, wie zum Beispiel qualitative Literaturanalysen oder Befragungen, wären für die Untersuchung der Forschungsfragen sowie das Testen der Hypothesen in der vorliegenden Arbeit nicht sinnvoll gewesen. Ausschließlich ein Experiment ermöglicht es, Ursache-Wirkungs-Beziehungen zu untersuchen und ist demnach geeignet, um Kausalität zu prüfen (Schreier et al., 2023). Dabei wird der kausale Einfluss der unabhängigen Variablen auf die abhängigen Variablen untersucht. Hierzu werden in einem Experiment die verschiedenen Ausprägungen der unabhängigen Variablen systematisch variiert, um deren Einfluss auf die abhängigen Variablen prüfen zu können (Schreier et al., 2023). In der vorliegenden Arbeit sind die unabhängigen Variablen des Experiments zum einen das Format der Antwortskala und zum anderen das Framing der Basisrate. Die abhängige Variable ist die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste.

Jede Versuchsperson wurde randomisiert zu einer der vier experimentellen Bedingungen des 2 (Antwortskala: numerisch, verbal) \times 2 (Framing: positiv, negativ) Designs des Experiments zugeteilt. Die vier experimentellen Bedingungen sind in Tabelle 1 dargestellt.

Tabelle 1*Experimentelle Bedingungen*

| | Antwortskala | |
|---------|-------------------------|----------------------|
| | Numerisch | Verbal |
| Positiv | Numerisch, positiv (NP) | Verbal, positiv (VP) |
| Negativ | Numerisch, negativ (NN) | Verbal, negativ (VN) |

Anmerkung. Die vier experimentellen Bedingungen ergeben sich aus der Kombination der Stufen des Faktors Antwortskala und der Stufen des Faktors Framing.

Aufgrund der randomisierten Zuteilung der Versuchspersonen zu einer der vier Bedingungen des Experiments wird dazu beigetragen, dass personenbedingte Störvariablen nahezu gleich verteilt in jeder der vier experimentellen Bedingungen vorliegen (Bortz & Schuster, 2010). Dadurch wird mittels Randomisierung eine höhere interne Validität als bei Untersuchungen ohne Randomisierung erreicht (Bortz & Schuster, 2010). Interne Validität ist als Stärke anzusehen, da eine hohe interne Validität damit einhergeht, dass die Ergebnisse weitgehend frei von störenden Einflüssen und somit eindeutig interpretierbar sind (Bortz & Schuster, 2010).

Da jede Versuchsperson jeweils nur eine der vier experimentellen Bedingungen durchläuft, handelt es sich um ein Zwischensubjekt-Design (Greenwald, 1976). Im Vergleich zum Innersubjekt-Design, bei dem jede Versuchsperson mehrere oder alle Bedingungen durchläuft, kann das Zwischensubjekt-Design unter anderem den Vorteil haben, dass Übungs-, Sensibilisierungs- und Übertragungseffekte vermieden werden können (Greenwald, 1976). Ein Nachteil beim Zwischensubjekt-Design im Vergleich zum Innersubjekt-Design ergibt sich im Hinblick auf die statistische Power. Da die Versuchspersonen im Zwischensubjekt-Design auf die verschiedenen Bedingungen aufgeteilt werden, steigt die benötigte Stichprobengröße mit der Anzahl der Bedingungen an (Greenwald, 1976). Demzufolge ist im Zwischensubjekt-Design eine größere Anzahl an Versuchspersonen notwendig, um eine vergleichbare statistische Power wie im Innersubjekt-Design zu erreichen (Greenwald, 1976). Da die Untersuchung online durchgeführt wurde, ist der Bedarf nach einer größeren Anzahl an Versuchspersonen ein geringeres Hindernis im Vergleich zu Untersuchungen, die in Präsenz durchgeführt werden. Online-Studien bieten die Möglichkeit, einfach und kostengünstig sowie zeitlich und räumlich flexibel viele Versuchspersonen in einem kurzen Zeitraum zu akquirieren (Mayer, 2013; Sue & Ritter, 2007).

Die mindestens notwendige Stichprobengröße für die vorliegende Arbeit wurde mittels *G*Power 3.1* ermittelt (Faul et al., 2009). Aus den Ergebnissen von Windschitl und Wells (1996) ergibt sich eine Effektstärke von Cohen's $d = 0.385$ für den Effekt der Antwortskala. Nach Cohen (1988/2013) entspricht das einem kleinen Effekt. Da bei Windschitl und Wells (1996) nicht hervorgeht, wie viele Versuchspersonen jeweils in den beiden experimentellen Gruppen waren, wurde von gleichgroßen Gruppen bei der Berechnung von Cohen's d ausgegangen. Die Effektstärke Cohen's $d = 0.385$ entspricht einer Effektstärke von Cohen's $f = 0.193$ (Cohen, 1988/2013). Durch eine a priori Power-Analyse für eine Kovarianzanalyse mit einer Kovariate, vier Bedingungen, einem Freiheitsgrad, einer Power von 95% ($\alpha = 5\%$) und einem kleinen Effekt von $f = 0.193$ wurde die mindestens notwendige Stichprobengröße als $N = 351$ bestimmt.

2.2 Beschreibung des Fragebogens

Bei der Teilnahme an der empirischen Untersuchung wurden den Versuchspersonen, analog zu Windschitl und Wells (1996), zwölf Fragen dargeboten. Die Fragen hatten dabei den gleichen Wortlaut, variierten aber jeweils in Thema und Basisrate. Im Hinblick auf die Themen unterschieden sich die Fragen der vorliegenden Arbeit zum Teil von den Fragen in dem Experiment 2 von Windschitl und Wells (1996). Da das Experiment 2 vor über 25 Jahren und in den USA durchgeführt wurde, wurden aktuellere und auf den deutschen Sprachraum angepasste Themen gewählt. Zu jedem Thema war eine Basisrate für richtige (positives Framing) beziehungsweise falsche Antworten (negatives Framing) zugehörig. Die Themen und dazugehörigen Basisraten sind in Tabelle A1 (siehe Anhang A) dargestellt. Der Wortlaut der Fragen lautete wie folgt: „Wenn wir Ihnen eine Frage über [Thema] stellen würden, die von [Basisrate in %] der zuvor befragten Personen [richtig/falsch] beantwortet wurde, wie wahrscheinlich ist es, dass Sie die Frage richtig beantworten würden?“.

Unmittelbar nach jeder der zwölf Fragen folgte eine numerische oder verbale Antwortskala, auf der die Versuchspersonen ihre Selbsteinschätzung abgeben sollten. Bei den Antwortskalen handelte es sich, analog zu Windschitl und Wells (1996), um durchgängig beschriftete 11-stufige Rating-Skalen (siehe Anhang A, Tabelle A2). Insbesondere eine Rating-Skala mit elf Skalenpunkten ist in diesem Fall konzeptionell sinnvoll, da die Skalenpunkte bei der numerischen Antwortskala sinnig die Antwortoptionen 0% bis 100% in Zehnerschritten repräsentieren. Auch Leung (2011) spricht eine Empfehlung für die Verwendung einer 11-stufigen Skala aus, da diese die Skalensensitivität zu erhöhen, der Normalverteilung am nächsten zu kommen und leicht verständlich zu sein scheint. Darüber hinaus wird erwartet, dass eine durchgängige Beschriftung der Skalen das Verständnis über

die Bedeutung jeder Antwortoption erhöhen und eine uneindeutige Interpretation der Skalen vermindern kann. Dementsprechend können die Antworten der Versuchspersonen bei durchgängig beschrifteten Skalen konsistenter und stabiler sein, was sich in einer höheren Test-Retest-Reliabilität zeigt (Weng, 2004).

Um eine höhere Übersichtlichkeit sicherzustellen, wurde jede Frage auf einer neuen Seite des Online-Fragebogens präsentiert. Die Reihenfolge, in der die Fragen präsentiert wurden, variierte nicht. Das war in diesem Fall sinnvoll, da so die Möglichkeit bestand, bei der nachfolgenden Präferenzrangliste die zwölf Themen inklusive zugehöriger Basisrate in der gleichen Reihenfolge zu präsentieren, in der sie im ersten Teil des Fragebogens abgefragt wurden. Dies sollte die Übersichtlichkeit erhöhen und dazu beitragen, Verwirrungen, die durch wechselnde Anordnungen bei den Versuchspersonen entstehen können, zu reduzieren.

Im Anschluss an die zwölf Fragen wurden die Versuchspersonen auf einer neuen Seite des Online-Fragebogens aufgefordert eine Präferenzrangliste der zuvor präsentierten Fragen zu erstellen. Die Anweisungen lauteten dabei wie folgt: „Stellen Sie sich vor, dass Ihnen eine der zuvor genannten Fragen gestellt wird und dass Sie einen Geldpreis gewinnen würden, wenn Sie sie richtig beantworten würden. Bitte ordnen Sie dementsprechend die zwölf möglichen Fragen von der Frage, die Sie am liebsten gestellt bekommen würden (1), bis zu der Frage, die Sie sich am wenigsten wünschen würden (12). Mit anderen Worten: Setzen Sie die Frage, von der Sie glauben, dass Sie sie am ehesten richtig beantworten könnten auf die 1 und die für Sie zweitleichste Frage auf die 2 und so weiter.“. Diese Anweisung ist vergleichbar mit der Anweisung in dem Experiment 2 von Windschitl und Wells (1996). Unter der Anweisung wurden die zwölf Themen inklusive zugehöriger Basisrate aufgeführt.

2.3 Berlin Numeracy Test

Wie in Kapitel 1.5 bereits erläutert, ist es in der vorliegenden Arbeit sinnvoll, die Risikokompetenz zu erheben. Dazu liefert der Berlin Numeracy Test (BNT) ein geeignetes Messinstrument. Mit Hilfe des BNT können schnelle und valide Einschätzungen bezüglich Risikokompetenz und statistischem Verständnis getroffen werden (Cokely et al., 2012). In mehreren Studien konnte gezeigt werden, dass der BNT in verschiedenen Ländern und in unterschiedlichen Bevölkerungsgruppen eine verlässliche psychometrische Unterscheidungskraft aufweist (Cokely et al., 2012). Zudem konnten Cokely et al. (2012) nachweisen, dass der BNT über konvergente und diskriminante Validität verfügt. Das heißt, der BNT weist eine hohe Korrelation mit verwandten Konstrukten, wie beispielsweise anderen numerischen Tests und Maßen kognitiver Fähigkeiten, und eine niedrige Korrelation mit nicht verwandten Konstrukten, wie beispielsweise Motivation und Persönlichkeit, auf.

Dabei ist der BNT ein starker Indikator für das Verständnis alltäglicher Risiken und statistischer Wahrscheinlichkeiten (Cokely et al., 2012). Aus diesem Grund wurde zusätzlich zu dem in Kapitel 2.3 beschriebenen Fragebogen der BNT eingesetzt.

Es wurde das adaptive Testformat des BNT für hochgebildete Stichproben verwendet. Diese Version des BNT war für die vorliegende Arbeit am sinnvollsten, da aufgrund der Rekrutierung der Versuchspersonen vorwiegend im universitären Kontext eine hochgebildete Stichprobe erwartet werden konnte. Eine grafische Darstellung des adaptiven Testformats inklusive der Punktevergabe ist in Abbildung B1 (siehe Anhang B) zu finden. Da den Versuchspersonen bei dem adaptiven Testformat des BNT lediglich zwei bis drei Fragen gestellt werden, benötigt die Bearbeitung in der Regel nur wenige Minuten und ist deshalb gut geeignet, um an den Fragebogen angeschlossen zu werden.

In Tabelle B1 (siehe Anhang B) sind die vier Fragen des adaptiven Testformats des BNT mit den jeweils korrekten Antworten dargestellt. Das Antwortformat der Fragen war offen.

2.4 Versuchsablauf

Der Online-Fragebogen wurde mit Hilfe der Webanwendung *SoSci Survey* erstellt. Zur Rekrutierung der Versuchspersonen wurde der Fragebogen über den Rundmailverteiler der Justus-Liebig-Universität Gießen und über das Versuchspersonen-Verwaltungssystem SONA verbreitet. Zudem wurde der Fragebogen über die sozialen Medien WhatsApp und Instagram an Bekannte zur Bearbeitung weitergegeben. Bei der Teilnahme über das Versuchspersonen-Verwaltungssystem SONA erhielten die Versuchspersonen eine halbe Versuchspersonenstunde als Vergütung. Wenn nicht über das Versuchspersonen-Verwaltungssystem SONA teilgenommen wurde, wurden die Versuchspersonen nicht für ihre Teilnahme entlohnt.

Die Erhebung wurde vom 19. Dezember 2024 bis zum 13. Januar 2025 durchgeführt. Der Erhebungszeitraum umfasst demnach 26 Tage.

Die Versuchspersonen führten die empirische Untersuchung online über eigene Endgeräte in ihrem natürlichen Umfeld durch. Zu Beginn wurden allgemeine Informationen zum Experiment angezeigt und die Einwilligung der Versuchspersonen zur Teilnahme am Experiment eingeholt. Anschließend wurden die Versuchspersonen randomisiert einer der vier experimentellen Bedingungen zugeteilt. Das Experiment begann mit der Präsentation der zwölf Fragen inklusive der zugehörigen Selbsteinschätzung auf der numerischen oder verbalen Antwortskala. Daraufhin wurde ein Kontrollitem eingefügt. Dies kann bei der Identifizierung von Versuchspersonen, die den Fragebogen unaufmerksam, nicht sorgfältig oder nicht ernsthaft beantwortet haben, helfen (Gummer et al., 2021; Meade & Craig, 2012).

Das Kontrollitem unterschied sich zwischen der numerischen und der verbalen Bedingung. Die Anweisung bei dem Kontrollitem lautete wie folgt: „Um zu zeigen, dass Sie bei der Beantwortung aufmerksam sind, kreuzen Sie hier bitte die Antwortoption „[40% / Etwas unwahrscheinlich]“ an.“. Unter der Anweisung war die 11-stufige numerische oder verbale Antwortskala, je nach Bedingung, abgebildet (siehe Anhang A, Tabelle A2). Hierauf folgten die Präferenzrangliste und der BNT. Anschließend wurden als soziodemografische Daten das Alter, das Geschlecht, der höchste Bildungsabschluss und die Beschäftigung der Versuchspersonen abgefragt. Wenn die Versuchspersonen als Beschäftigung angaben, dass sie Studierende sind, wurden sie im Anschluss zu der Frage weitergeleitet, welchen Studiengang sie belegen. Das Abfragen der soziodemografischen Daten ist am Ende des Fragebogens sinnvoll, weil das Interesse der Versuchspersonen während der Bearbeitung des Fragebogens abnehmen kann und sie deshalb am Ende eher gewillt sind, Fragen zur eigenen Person zu beantworten (Mayer, 2013). Zum Abschluss des Experiments wurde den Versuchspersonen für ihre Teilnahme gedankt. Der gesamte Versuchsablauf ist als Flussdiagramm in Abbildung C1 (siehe Anhang C) abgebildet.

Während des gesamten Experiments wurde ein Fortschrittsbalken oben rechts auf jeder Seite des Online-Fragebogens dargestellt, der den Versuchspersonen prozentual ihren Bearbeitungsfortschritt anzeigte. Alle Fragen, ausgenommen der soziodemografischen Daten, mussten bei der Bearbeitung zwingend beantwortet werden. Es war kein Zurückblättern möglich. Die Bearbeitung konnte jederzeit von den Versuchspersonen unter- und abgebrochen werden.

2.5 Stichprobenbeschreibung

Die Stichprobenbeschreibung bezieht sich auf die Stichprobe nach dem Ausschluss von Versuchspersonen im Zuge der Datenaufbereitung (siehe Kapitel 3.1).

Insgesamt umfasst die Stichprobe $N = 289$ Versuchspersonen. Dabei nahmen $n = 193$ Versuchspersonen über das Versuchspersonen-Verwaltungssystem SONA teil (66.8%) und $n = 96$ Versuchspersonen nicht (33.2%). Das Durchschnittsalter in der Stichprobe beträgt $M = 25.08$ Jahre ($SD = 9.44$). Das Alter der Versuchspersonen liegt dabei zwischen 18 und 67 Jahren. In der Stichprobe identifizieren sich $n = 212$ Versuchspersonen als weiblich (73.4%), $n = 71$ Versuchspersonen als männlich (24.6%) und $n = 1$ Versuchsperson als divers (0.3%). Zudem machte $n = 1$ Versuchsperson keine Angabe zu ihrem Geschlecht (0.3%). Demzufolge besteht der Großteil der Stichprobe aus weiblichen Versuchspersonen. Im Hinblick auf den höchsten Bildungsabschluss verfügt der Großteil der Stichprobe über ein Abitur ($n = 224$, 77.5%). Außerdem weisen $n = 35$ Versuchspersonen einen Hochschulabschluss als höchsten

Bildungsabschluss auf (12.1%). Bezüglich der Beschäftigung besteht der Großteil der Stichprobe aus Studierenden ($n = 242$, 83.7%). Überdies sind $n = 33$ Versuchspersonen arbeitstätig (11.4%). Dazu zählen Angestellte, Beamte und Selbstständige. Die Studierenden der Stichprobe sind verschiedenen Fachrichtungen zugehörig. Dabei sind die am häufigsten vertretenen Studiengänge Lehramt ($n = 78$, 27.0%) und Psychologie ($n = 62$, 21.5%). Darüber hinaus studieren weitere $n = 51$ Versuchspersonen Medizin und Gesundheitswissenschaften (17.6%).

Jeweils $n = 4$ Versuchspersonen haben die Fragen zu den soziodemografischen Daten Alter, Geschlecht, höchster Bildungsabschluss und Beschäftigung nicht beantwortet (1.4%).

Aus Gründen der Übersichtlichkeit sind lediglich Charakteristika der Stichprobe mit Angaben über zehn Prozent bei den soziodemografischen Daten höchster Bildungsabschluss, Beschäftigung und Studiengang im Text angegeben. Um die Vollständigkeit der Stichprobenbeschreibung zu gewährleisten, sind die weiteren Angaben zu den soziodemografischen Charakteristika der Stichprobe in Tabelle D1 (siehe Anhang D) abgetragen.

3. Ergebnisse

Im dritten Kapitel werden die Ergebnisse der experimentellen Untersuchung präsentiert. Dazu wird zunächst ein Überblick über die Datenaufbereitung und die vorläufigen Analysen gegeben. Im Anschluss wird die deskriptive Statistik beschrieben. Abgeschlossen wird das Kapitel mit dem Bericht der inferenzstatistischen Ergebnisse.

Für die Datenaufbereitung, die vorläufigen Analysen und die Datenauswertung wurde *IBM SPSS Statistics 29.0* verwendet.

3.1 Datenaufbereitung

Als Vorbereitung für die anschließenden Analysen wurde für jede Versuchsperson ein Spearman-Rangkorrelationskoeffizient zwischen der Selbsteinschätzung bei den zwölf Fragen und der Präferenz in der Rangliste für die entsprechende Frage berechnet. Im Zuge dessen wurden die Angaben der Versuchspersonen bei der Präferenzrangliste invertiert, sodass ein hoher Wert einer starken Präferenz bezüglich der Beantwortung einer Frage zu dem jeweiligen Thema entsprach. Das erleichterte im Nachfolgenden die Interpretation der Korrelationen, da diese so intuitiv interpretiert werden konnten. Für die Berechnung der Spearman-Rangkorrelationskoeffizienten wurden die Daten umstrukturiert, sodass die Spalten für die zwölf Fragen und die Spalten für die zwölf Ränge der Präferenzrangliste in Zeilen umgewandelt wurden, die die Selbsteinschätzung und die Präferenz in der Rangliste jeder Versuchsperson pro Frage enthielten. Bei den Spearman-Rangkorrelationskoeffizienten wurde eine Fisher Z -Transformation durchgeführt. Dies ist aus zwei Gründen sinnvoll. Zum einen werden Z -Werte benötigt, wenn mehrere Korrelationskoeffizienten gemittelt werden sollen (Bortz & Schuster, 2010). Das ist in der vorliegenden Arbeit notwendig, um Mittelwerte der Korrelationskoeffizienten für die experimentellen Bedingungen berechnen und diese so miteinander vergleichen zu können. Zum anderen sind Z -Werte approximativ normalverteilt, was für die inferenzstatistische Absicherung der Stichprobenkorrelation relevant ist (Bortz & Schuster, 2010).

Zur besseren Lesbarkeit wird die Fisher Z -Transformation der Spearman-Rangkorrelationskoeffizienten zwischen der Selbsteinschätzung und der Präferenzrangliste im Folgenden verkürzt als Z -Wert bezeichnet.

Des Weiteren erfolgte der Ausschluss von Versuchspersonen im Zuge der Datenaufbereitung. Dazu wurden verschiedene Kriterien herangezogen, welche im Folgenden thematisiert werden. Vor dem Ausschluss umfasste die Stichprobe $N = 340$ Versuchspersonen.

Zunächst erfolgte der Ausschluss durch das Kontrollitem. Dabei war das Ausschlusskriterium, dass das Kontrollitem nicht oder falsch beantwortet wurde. Dadurch wurden $n = 23$ Versuchspersonen (6.8%) ausgeschlossen. Nach dem Ausschluss durch das Kontrollitem verblieben $N = 317$ Versuchspersonen.

Daraufhin erfolgte der Ausschluss durch die Summe der Bearbeitungsdauer. Dazu wurde die Variable *TIME_SUM* von SoSci Survey verwendet. Diese gibt die Summe der Bearbeitungsdauer nach der Bereinigung von Unterbrechungen an (SoSci Survey). Als Unterbrechungen werden dabei Bearbeitungsdauern länger als zwei Stunden pro Fragebogenseite klassifiziert. Zudem gilt ein Median der Bearbeitungsdauer höher als der dreifache Interquartilabstand dividiert durch 1.34 ebenfalls als Unterbrechung, da dies mehr als drei Standardabweichungen in einer normalverteilten Stichprobe entspricht (SoSci Survey). Zur Beurteilung, welche Summe der Bearbeitungsdauer noch als realistisch zu betrachten ist, wurde ein Histogramm mit den Versuchspersonen nach dem Ausschluss durch das Kontrollitem angefertigt (siehe Abbildung 1). Die Bearbeitungsdauer in der Stichprobe betrug im Mittel $M = 7.84$ Minuten ($SD = 3.22$). Mit Hilfe des Histogramms sowie durch logische Begründung kann argumentiert werden, dass eine Summe der Bearbeitungsdauer kleiner als vier Minuten als unrealistisch zu bewerten ist. In unter vier Minuten ist eine sinnvolle und ernsthafte Bearbeitung des Fragebogens nicht möglich. Demzufolge war das Ausschlusskriterium hierbei, dass die Summe der Bearbeitungsdauer weniger als vier Minuten betrug. Dadurch wurden $n = 19$ Versuchspersonen (5.6%) ausgeschlossen. Nach dem Ausschluss durch die Summe der Bearbeitungsdauer verblieben $N = 298$ Versuchspersonen.

Da die Risikokompetenz in Form der Punktzahl im BNT als Kovariate in die Analyse eingeschlossen wurde, erfolgte anschließend der Ausschluss durch den BNT. Dabei war das Ausschlusskriterium, dass der BNT nicht bearbeitet wurde. Dadurch wurden $n = 7$ Versuchspersonen (2.1%) ausgeschlossen. Nach dem Ausschluss durch fehlende Werte bei dem BNT verblieben $N = 291$ Versuchspersonen.

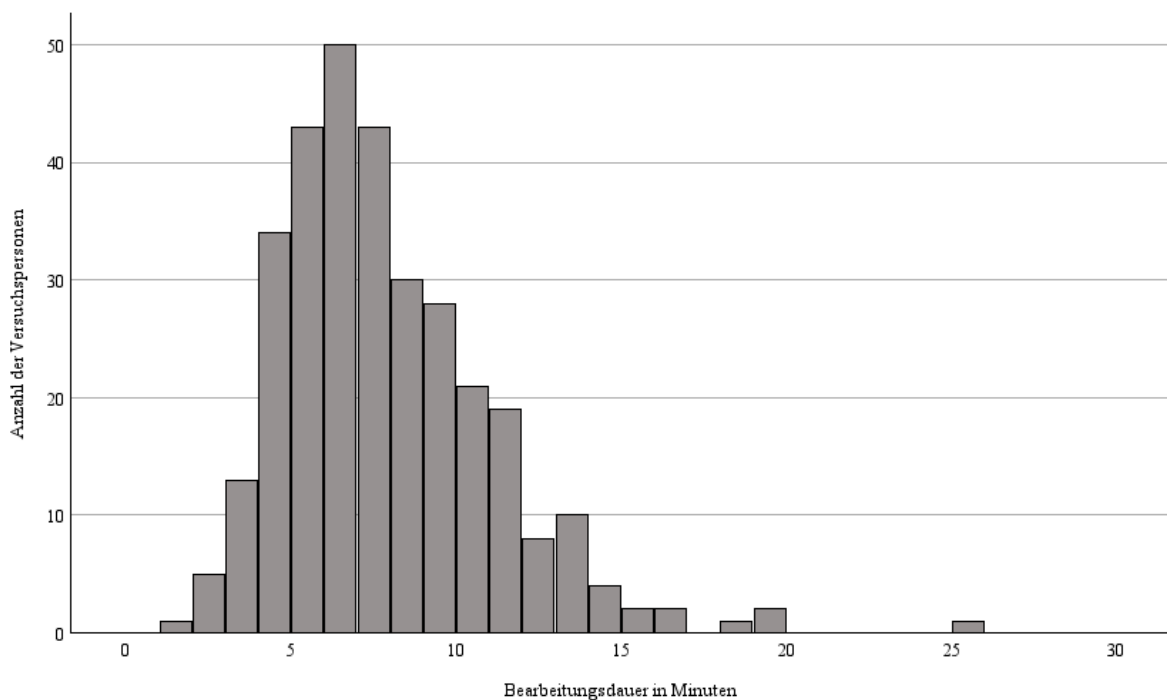
Die Bearbeitungsdauer des BNT wurde nicht zum Ausschluss von Versuchspersonen herangezogen. Durch die Verwendung von künstlicher Intelligenz zum Lösen der Fragen des BNT können Bearbeitungsdauern erreicht werden, die nicht von Bearbeitungsdauern durch menschliche Rechnungen unterschieden werden können. Somit stellt die Bearbeitungsdauer des BNT kein geeignetes Ausschlusskriterium dar, um Versuchspersonen, die den BNT nicht ernsthaft bearbeitet haben, ausschließen zu können.

Da die Versuchspersonen vor der Bearbeitung des BNT zwangsläufig die Selbsteinschätzung bei den zwölf Fragen sowie die Präferenzrangliste beantworten mussten,

konnte auf den Ausschluss von Versuchspersonen durch die letzte bearbeitete Seite des Fragebogens verzichtet werden. Im Zuge des Ausschlusses von Versuchspersonen, die den BNT nicht bearbeitet haben, wurden Versuchspersonen, die den Fragebogen nicht bis zum BNT bearbeitet haben, ebenfalls ausgeschlossen. Dadurch wurde sichergestellt, dass die Daten, die zur Beantwortung der aufgestellten Hypothesen zwingend notwendig sind, vorhanden sind.

Abbildung 1

Histogramm der Bearbeitungsdauer

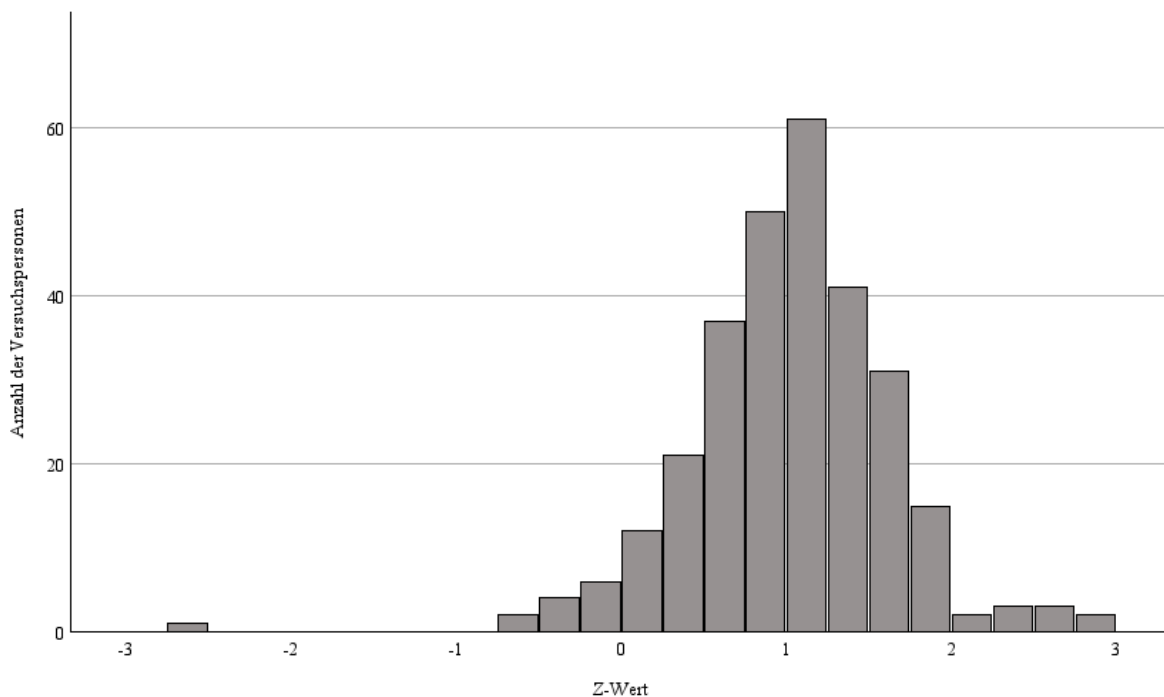


Anmerkung. Summe der Bearbeitungsdauer von $N = 317$ Versuchspersonen. Auf der x -Achse ist die Summe der Bearbeitungsdauer in Minuten und auf der y -Achse ist die Anzahl der Versuchspersonen abgebildet. $M = 7.84$ Minuten ($SD = 3.22$).

Anhand der Verteilung der abhängigen Variable mittels des in Abbildung 2 dargestellten Histogramms wird deutlich, dass es einen extremen Ausreißer bei der abhängigen Variable gab. Aufgrund dessen erfolgte der Ausschluss dieser Versuchsperson. Der Vergleich des Mittelwerts der abhängigen Variable ($M = 1.04$) mit dem um 5% getrimmten Mittelwert ($M_{\text{trim}} = 1.04$) gibt Grund zu der Annahme, dass weitere Ausreißer bei der abhängigen Variable keinen zu berücksichtigenden Einfluss haben. Demzufolge wurde lediglich der extreme Ausreißer, also $n = 1$ Versuchsperson (0.3%), ausgeschlossen. Nach dem Ausschluss extremer Ausreißer bei der abhängigen Variable verblieben $N = 290$ Versuchspersonen.

Abbildung 2

Histogramm der Z-Werte vor Ausschluss extremer Ausreißer



Anmerkung. Z-Werte von $N = 291$ Versuchspersonen. Auf der x -Achse ist der Z-Wert und auf der y -Achse ist die Anzahl der Versuchspersonen abgebildet. $M = 1.03$ ($SD = 0.61$).

Im Rahmen der explorativen Datenanalyse wurde eine Versuchsperson mit auffälligem Antwortmuster festgestellt (siehe Anhang G). Antwortmuster stellen die Tendenz von Versuchspersonen dar, unabhängig vom Inhalt der Fragen auf bestimmte Art und Weise zu antworten, wodurch die Ergebnisse verzerrt werden können (van Vaerenbergh & Thomas, 2013). Deshalb wurde diese $n = 1$ Versuchsperson (0.3%) ausgeschlossen. Nach dem Ausschluss durch Antwortmuster verblieben $N = 289$ Versuchspersonen.

Insgesamt wurden $n = 51$ Versuchspersonen ausgeschlossen. Das entspricht 15.0% ausgeschlossener Versuchspersonen der ursprünglichen Stichprobe. Somit verblieben nach Abschluss des Ausschlusses von Versuchspersonen im Zuge der Datenaufbereitung $N = 289$ Versuchspersonen.

Nach dem Ausschluss von Versuchspersonen existierten $n = 12$ Versuchspersonen mit negativen Z -Werten. Diese negativen Z -Werte könnten darauf hinweisen, dass die Anweisungen bei der Präferenzrangliste von den Versuchspersonen missverstanden wurden. Jedoch scheint es bei eingehender Betrachtung der Daten dieser $n = 12$ Versuchspersonen keine systematischen Hinweise auf ein Missverständnis der Anweisungen bei der Präferenzrangliste zu geben. Das heißt, Fragen mit einer hohen Selbsteinschätzung auf der numerischen oder verbalen Antwortskala wurden bei der Präferenzrangliste nicht systematisch auf die unteren Ränge verteilt und umgekehrt, was auf ein Missverständnis der Anweisungen bei der Präferenzrangliste hinweisen würde. Deshalb wurden diese Versuchspersonen nicht aus der Analyse ausgeschlossen.

3.2 Vorläufige Analysen

Um herauszufinden, welche Variablen als Kovariate mit in die Analyse einfließen sollten, wurden die potenziell in Frage kommenden Variablen vor der Analyse dahingehend überprüft. Jeweils eine einfaktorielle Varianzanalyse ohne Messwiederholung wurde durchgeführt, um den Einfluss der soziodemografischen Daten Geschlecht, höchster Bildungsabschluss, Beschäftigung und Studiengang auf die Z -Werte zu untersuchen. Die einfaktorielle Varianzanalyse ohne Messwiederholung zeigte keinen statistisch signifikanten Unterschied der Z -Werte für die soziodemografischen Daten Geschlecht, $F(3, 281) = 0.12, p = .950$, höchster Bildungsabschluss, $F(5, 279) = 0.47, p = .798$, Beschäftigung, $F(8, 276) = 0.67, p = .723$, und Studiengang, $F(11, 227) = 0.41, p = .949$. Außerdem wurde eine einfaktorielle Varianzanalyse ohne Messwiederholung durchgeführt, um den Einfluss der Teilnahme über das Versuchspersonen-Verwaltungssystem SONA auf die Z -Werte zu untersuchen. Auch hier zeigte die einfaktorielle Varianzanalyse ohne Messwiederholung keinen statistisch signifikanten Unterschied der Z -Werte für die Teilnahme über das

Versuchspersonen-Verwaltungssystem SONA, $F(1, 287) = 2.71, p = .101$. Die Annahmen zur Berechnung von Varianzanalysen wurden im Vorhinein überprüft. Die Beziehung zwischen dem Alter und den Z -Werten wurde mit Hilfe eines Pearson Produkt-Moment-Korrelationskoeffizienten untersucht. Es zeigte sich keine statistisch signifikante Korrelation zwischen dem Alter und den Z -Werten, $r = 0.06, n = 285, p = .356$. Auch die Beziehung zwischen der Punktzahl im BNT und den Z -Werten wurde mit Hilfe eines Pearson Produkt-Moment-Korrelationskoeffizienten untersucht. Auch hier zeigte sich keine statistisch signifikante Korrelation zwischen der Punktzahl im BNT und den Z -Werten, $r = 0.06, n = 289, p = .299$. Die Annahmen zur Berechnung von Pearson Produkt-Moment-Korrelationskoeffizienten wurden ebenfalls im Vorhinein überprüft.

Da sich kein statistisch signifikanter Einfluss der oben thematisierten Variablen auf die Z -Werte zeigte, müssen diese nicht in der Analyse berücksichtigt werden. In Kapitel 1.5 und Kapitel 2.3 wurde jedoch eine theoriegestützte Begründung für die Verwendung der Risikokompetenz in Form der Punktzahl im BNT als Kovariate in der vorliegenden Arbeit dargestellt. Aufgrund dessen wird diese, obwohl sich keine statistisch signifikante Korrelation zwischen der Punktzahl im BNT und den Z -Werten zeigte, als Kovariate in die Analyse eingeschlossen.

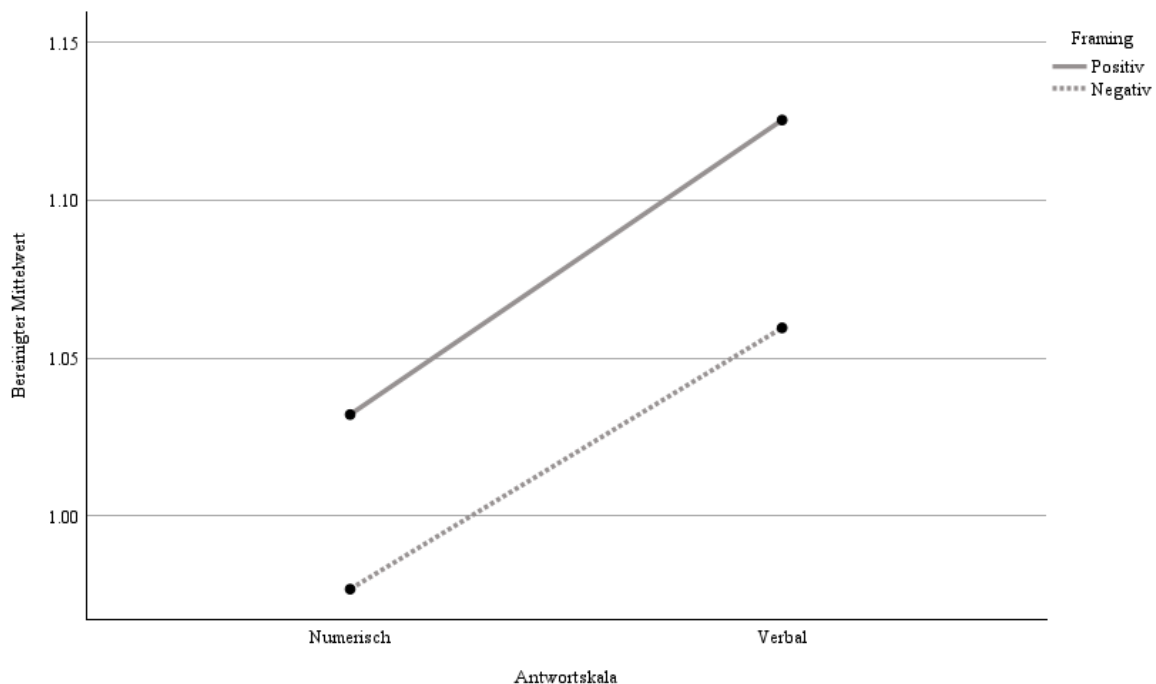
3.3 Deskriptive Statistik

Der Z -Wert lag im Mittel bei $M = 1.05$ ($SD = 0.57$). Bei der separaten Betrachtung der Z -Werte für jede der vier experimentellen Bedingungen zeigte sich, dass der Z -Wert in der experimentellen Bedingung VP im Mittel am höchsten war ($M = 1.11, SD = 0.58$). Darauf folgten die Bedingung VN ($M = 1.07, SD = 0.60$) und die Bedingung NP ($M = 1.03, SD = 0.60$). Der im Mittel niedrigste Z -Wert zeigte sich bei der Bedingung NN ($M = 0.98, SD = 0.52$). Bei der separaten Betrachtung der Z -Werte für die zwei unabhängigen Variablen, Antwortskala und Framing, zeigten sich die im Folgenden beschriebenen Ergebnisse. Der Z -Wert lag bei der verbalen Antwortskala im Mittel bei $M = 1.09$ ($SD = 0.59$) und war damit höher als bei der numerischen Antwortskala ($M = 1.01, SD = 0.56$). Bei dem positiven Framing lag der Z -Wert im Mittel bei $M = 1.07$ ($SD = 0.59$) und war damit höher als bei dem negativen Framing ($M = 1.02, SD = 0.56$). Diese Ergebnisse sind in Abbildung 3 mittels einer grafischen Darstellung der bereinigten Mittelwerte der Z -Werte für die experimentellen Bedingungen veranschaulicht. Insgesamt waren die mittleren Z -Werte positiv, was auf einen positiven Zusammenhang zwischen der Selbsteinschätzung und der Präferenzrangliste hinweist. Die oben beschriebenen deskriptiven Statistiken sowie die bereinigten Mittelwerte inklusive der Standardfehler sind in Tabelle 2 abgetragen.

Die Punktzahl im BNT lag im Mittel bei $M = 2.52$ Punkten ($SD = 1.16$). Dabei deckten die erreichten Punktzahlen im BNT den gesamten Bereich möglicher Punktzahlen ab (Min = 1, Max = 4). Diese Werte stimmen mit den Angaben von Cokely et al. (2012) zu dem verwendeten adaptiven Testformat des BNT überein. Die Angaben der deskriptiven Statistiken des BNT je experimenteller Bedingung sind in Tabelle 3 abgebildet.

Abbildung 3

Bereinigte Z-Mittelwerte für die experimentellen Bedingungen



Anmerkung. Bereinigte Mittelwerte der Z-Werte von $N = 289$ Versuchspersonen für die zwei Framingversionen (positiv, negativ) in Abhängigkeit von der Antwortskala (numerisch, verbal). Auf der x -Achse sind die Stufen des Faktors Antwortskala und auf der y -Achse ist der bereinigte Mittelwert der Z-Werte abgetragen. Zu beachten ist, dass die y -Achse nicht bei dem Wert Null beginnt. Die durchgezogene Linie repräsentiert die Bedingung mit positivem Framing und die gestrichelte Linie repräsentiert die Bedingung mit negativem Framing. Die im Modell enthaltene Kovariate Risikokompetenz wurde mit dem Wert BNT = 2.52 (mittlere Punktzahl im BNT in der Stichprobe der vorliegenden Arbeit) bewertet.

Tabelle 2*Deskriptive Statistik der Z-Werte*

| | <i>n</i> | <i>M</i> | <i>SD</i> | <i>M</i> _{bereinigt} | <i>SEM</i> |
|-------------------------|----------|----------|-----------|-------------------------------|------------|
| NP | 72 | 1.03 | 0.60 | 1.03 | 0.07 |
| NN | 76 | 0.98 | 0.52 | 0.98 | 0.07 |
| VP | 73 | 1.11 | 0.58 | 1.13 | 0.07 |
| VN | 68 | 1.07 | 0.60 | 1.06 | 0.07 |
| Numerische Antwortskala | 148 | 1.01 | 0.56 | 1.00 | 0.05 |
| Verbale Antwortskala | 141 | 1.09 | 0.59 | 1.09 | 0.05 |
| Positives Framing | 145 | 1.07 | 0.59 | 1.08 | 0.05 |
| Negatives Framing | 144 | 1.02 | 0.56 | 1.02 | 0.05 |

Anmerkung. $N = 289$. Die Bereinigung bezieht sich auf die Tatsache, dass der Effekt der Kovariate statistisch entfernt wurde. Die im Modell enthaltene Kovariate Risikokompetenz wurde mit dem Wert $BNT = 2.52$ (mittlere Punktzahl im BNT in der Stichprobe der vorliegenden Arbeit) bewertet.

Tabelle 3*Deskriptive Statistik der Punktzahl im BNT je experimenteller Bedingung*

| Experimentelle Bedingung | <i>n</i> | <i>M</i> | <i>SD</i> | <i>SEM</i> | Min | Max |
|--------------------------|----------|----------|-----------|------------|-----|-----|
| NP | 72 | 2.50 | 1.05 | 0.12 | 1 | 4 |
| NN | 76 | 2.61 | 1.13 | 0.13 | 1 | 4 |
| VP | 73 | 2.16 | 1.17 | 0.14 | 1 | 4 |
| VN | 68 | 2.82 | 1.21 | 0.15 | 1 | 4 |

Anmerkung. $N = 289$.

3.4 Inferenzstatistik

Um die in Kapitel 1.1 angeführten Forschungsfragen beantworten und die in Kapitel 1.6 aufgestellten Hypothesen testen zu können, wurde eine 2×2 Kovarianzanalyse ohne Messwiederholung zwischen den experimentellen Bedingungen durchgeführt. Diese untersuchte den Einfluss der Antwortskala und des Framings auf den Z-Wert. Die Punktzahl im BNT wurde als Kovariate verwendet, um für die individuelle Risikokompetenz kontrollieren zu können. Es wurden vorläufige Analysen durchgeführt, um sicherzustellen, dass die Annahmen für die Kovarianzanalyse erfüllt sind (siehe Anhang E).

Der Interaktionseffekt der Faktoren Framing \times Antwortskala wurde statistisch nicht signifikant, $F(1, 284) = 0.01, p = .939, \eta_p^2 = .000$. Das weist darauf hin, dass der Einfluss des

Framings auf den Z -Wert nicht systematisch in Abhängigkeit von der Version der Antwortskala variiert und umgekehrt.

Auch der Haupteffekt des Faktors Antwortskala wurde statistisch nicht signifikant, $F(1, 284) = 1.68, p = .098$ (einseitig), $\eta_p^2 = .006$. Das weist darauf hin, dass der Z -Wert nicht signifikant davon abhängt, ob die numerische oder die verbale Antwortskala verwendet wird.

Ebenso wurde der Haupteffekt des Faktors Framing statistisch nicht signifikant, $F(1, 284) = 0.78, p = .379, \eta_p^2 = .003$. Das weist darauf hin, dass der Z -Wert nicht signifikant davon abhängt, ob positives oder negatives Framing verwendet wird.

4. Diskussion

Das vierte Kapitel diskutiert die Ergebnisse und zieht theoretische sowie praktische Schlüsse. Hier werden die gewonnenen Erkenntnisse über den Einfluss von Antwortskalen und Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit zusammengefasst. Zu Beginn werden die Ergebnisse interpretiert und hinsichtlich der Forschungsfragen und Hypothesen bewertet. Es werden mögliche Erklärungen für die Ergebnisse und Fehlerquellen thematisiert. Zusätzlich werden die Ergebnisse der explorativen Datenanalyse diskutiert. Zum Schluss werden die Limitationen der Arbeit aufgezeigt sowie ein Ausblick für zukünftige Forschung gegeben.

4.1 Ergebnisdiskussion

Es wurde einerseits die Forschungsfrage untersucht, ob die verbale Antwortskala im Vergleich zu der numerischen Antwortskala eine bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit ermöglicht. Demzufolge postuliert die entsprechende Hypothese, dass die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste bei der verbalen Antwortskala höher ist als bei der numerischen Antwortskala (siehe H1 in Kapitel 1.6). Es zeigte sich kein statistisch signifikanter Einfluss der Antwortskala auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste (siehe Kapitel 3.4). Dementsprechend konnte die in Kapitel 1.6 aufgestellte H1 nicht bestätigt werden. Infolgedessen lautet die Antwort auf die erste Forschungsfrage in der vorliegenden Arbeit, dass die verbale Antwortskala im Vergleich zu der numerischen Antwortskala keine bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit ermöglicht. Die Befunde von Windschitl und Wells (1996) konnten nicht repliziert werden. Dennoch entspricht der Effekt der hypothesenkonformen Richtung. In Abbildung 3 ist zu erkennen, dass die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste bei der verbalen Antwortskala höher war als bei der numerischen Antwortskala (siehe Kapitel 3.3). Werden die Ergebnisse mit den Ergebnissen von Windschitl und Wells (1996) verglichen, fällt auf, dass der mittlere Z-Wert bei der numerischen Antwortskala in der vorliegenden Arbeit höher war. Bei Windschitl und Wells (1996) ergab sich bei der numerischen Antwortskala ein mittlerer Z-Wert von $M = 0.84$ ($SD = 0.51$). Demzufolge war die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste bei der numerischen Antwortskala in der vorliegenden Arbeit im Mittel höher als bei dem Experiment 2 von Windschitl und Wells (1996). Dahingegen unterschied sich der mittlere Z-Wert bei der verbalen Antwortskala in der vorliegenden Arbeit lediglich geringfügig von dem mittleren Z-Wert bei Windschitl und Wells

(1996). Dort ergab sich bei der verbalen Antwortskala ein mittlerer Z -Wert von $M = 1.02$ ($SD = 0.42$). In der vorliegenden Arbeit waren die mittleren Z -Werte sowohl bei der numerischen als auch bei der verbalen Antwortskala höher als bei dem Experiment 2 von Windschitl und Wells (1996). Zudem lagen die mittleren Z -Werte bei der numerischen und der verbalen Antwortskala näher beieinander, weshalb sich kein statistisch signifikanter Einfluss der Antwortskala zeigte. Für die beobachteten Unterschiede kann es verschiedene Ursachen geben. Zunächst können die Unterschiede auf die unterschiedlichen Stichproben der beiden Untersuchungen zurückzuführen sein. Obwohl beide Stichproben vorwiegend aus Studierenden bestanden, gibt es dennoch zwei augenscheinliche Unterschiede. Zum einen liegen über 25 Jahre zwischen den beiden Untersuchungen. Zum anderen wurde das Experiment 2 von Windschitl und Wells (1996) in den USA und die vorliegende Arbeit in Deutschland durchgeführt. Aus der zeitlichen Veränderung und dem kulturellen Unterschied könnten die unterschiedlichen mittleren Z -Werte resultieren. Auch methodisch weisen das Experiment 2 von Windschitl und Wells (1996) und die vorliegende Arbeit Unterschiede auf. Beispielsweise wurde die vorliegende Arbeit online durchgeführt, während es sich bei dem Experiment 2 von Windschitl und Wells (1996) um einen Fragebogen im traditionellen Papierformat handelte. So ist es denkbar, dass die Unterschiede in den mittleren Z -Werten bei der numerischen Antwortskala aus Versuchsleitereffekten im traditionellen Papierformat bei Windschitl und Wells (1996) resultieren. Versuchsleitereffekte beziehen sich auf Merkmale und Verhaltensweisen der Versuchsleitenden, die die Untersuchungsergebnisse beeinflussen können (Döring, 2023). Sie treten beispielsweise auf, wenn Versuchspersonen die Erwartungen der Versuchsleitenden errahnen und sich entsprechend verhalten (Döring, 2023). Da die Versuchspersonen bei Online-Studien nicht mit den Versuchsleitenden in Kontakt treten, können Versuchsleitereffekte vermieden werden. Als weiterer methodischer Unterschied kommt hinzu, dass in der vorliegenden Arbeit eine zusätzliche Variation durch das Framing vorgenommen wurde. Das geht über das Experiment 2 von Windschitl und Wells (1996) hinaus. Auch aus dieser zusätzlichen experimentellen Manipulation könnten die Unterschiede in den mittleren Z -Werten resultieren.

Andererseits wurde die Forschungsfrage untersucht, ob es einen Einfluss des Framings auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit gibt. Die entsprechende Hypothese postuliert, dass das Framing der Basisrate einen Einfluss auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste hat (siehe H2 in Kapitel 1.6). Es zeigte sich kein statistisch signifikanter Einfluss des Framings auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste (siehe Kapitel 3.4).

Dementsprechend konnte die in Kapitel 1.6 aufgestellte H2 nicht bestätigt werden. Infolgedessen lautet die Antwort auf die zweite Forschungsfrage in der vorliegenden Arbeit, dass es keinen Einfluss des Framings auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit gibt. In einer Meta-Analyse zum Einfluss von Framing wies Kühberger (1998) darauf hin, dass die Art der Risikomanipulation einen Einfluss auf den Effekt von Framing hat. Wurde als Risikomanipulation die Bezugnahme auf ein riskantes Ereignis verwendet, zeigte sich ein Effekt des Framings (Kühberger, 1998). Dahingegen zeigte die Risikomanipulation durch Ergebnissalienz Nulleffekte (Kühberger, 1998). In der vorliegenden Arbeit wurde jedoch die Risikomanipulation durch Ergebnissalienz verwendet. Das kann erklären, warum der Haupteffekt des Framings keine statistische Signifikanz erreicht hat. Nichtsdestotrotz ist in Abbildung 3 zu erkennen, dass die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste bei dem positiven Framing sowohl in der numerischen als auch in der verbalen Bedingung höher war als bei dem negativen Framing (siehe Kapitel 3.3).

Die Interaktionshypothese postuliert, dass der Einfluss des Framings auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste zwischen der numerischen und der verbalen Antwortskala variiert (siehe H3 in Kapitel 1.6). Der Interaktionseffekt der Faktoren Framing \times Antwortskala erreichte keine statistische Signifikanz (siehe Kapitel 3.4). Das weist darauf hin, dass sich der Einfluss des Framings nicht systematisch zwischen der numerischen und der verbalen Antwortskala unterscheidet. Ebenso deutet das darauf hin, dass sich der Einfluss der Antwortskala nicht systematisch zwischen dem positiven und dem negativen Framing unterscheidet. Folglich scheint es keine Interaktion zwischen dem Fragenformat in Form des Framings und dem Antwortformat in Form der Antwortskala zu geben. Dementsprechend konnte auch die in Kapitel 1.6 aufgestellte H3 nicht bestätigt werden.

4.2 Diskussion der explorativen Datenanalyse

Aufgrund des in Kapitel 1.3 beschriebenen deliberativen Denkens, das durch die numerische Antwortskala angeregt werden kann, kann angenommen werden, dass die Versuchspersonen bei der numerischen Antwortskala sensitiver auf die angegebene Basisrate reagieren (Windschitl & Wells, 1996). Demzufolge kann ihre Selbsteinschätzung stärker durch die Basisrate beeinflusst werden. Um diese Idee zu untersuchen, wurde eine 2×2 Kovarianzanalyse ohne Messwiederholung zwischen den experimentellen Bedingungen durchgeführt. Diese überprüfte den Einfluss der Antwortskala und des Framings auf die Korrelation zwischen der jeweils angegebenen Basisrate bei den zwölf Fragen und der

Selbsteinschätzung. Die Punktzahl im BNT wurde als Kovariate verwendet, um für die individuelle Risikokompetenz kontrollieren zu können. Zudem zeigte sich in vorläufigen Analysen ein signifikanter Einfluss des Alters auf die Z-Werte der Korrelation zwischen der Basisrate und der Selbsteinschätzung. Deshalb wurde das Alter ebenfalls als Kovariate verwendet, um dafür kontrollieren zu können. Die Annahmen für die Kovarianzanalyse wurden analog zu dem in Anhang E beschriebenen Vorgehen überprüft und gelten als erfüllt. Die Pearson Produkt-Moment-Korrelationskoeffizienten wurden vergleichbar mit dem in Kapitel 3.1 beschriebenen Vorgehen berechnet. Auch bei den Pearson Produkt-Moment-Korrelationskoeffizienten wurde eine Fisher Z-Transformation durchgeführt.

Die Interaktion der Faktoren Framing \times Antwortskala erreichte keine statistische Signifikanz (siehe Anhang F, Tabelle F2). Das weist darauf hin, dass der Einfluss des Framings auf die Z-Werte der Korrelation zwischen der Basisrate und der Selbsteinschätzung nicht systematisch in Abhängigkeit von der Version der Antwortskala variiert und umgekehrt.

Des Weiteren zeigte sich in der explorativen Datenanalyse kein statistisch signifikanter Einfluss der Antwortskala auf die Z-Werte der Korrelation zwischen der Basisrate und der Selbsteinschätzung (siehe Anhang F, Tabelle F2). Demzufolge konnte die im Rahmen der explorativen Datenanalyse aufgestellte Annahme, dass Versuchspersonen bei der numerischen Antwortskala aufgrund des angeregten deliberativen Denkens sensitiver auf die Basisrate reagieren und demnach ihre Selbsteinschätzung stärker durch die Basisrate beeinflusst wird, nicht bestätigt werden. Die Befunde von Windschitl und Wells (1996) konnten auch hier nicht repliziert werden.

Indessen zeigte sich ein statistisch signifikanter Einfluss des Framings auf die Z-Werte der Korrelation zwischen der Basisrate und der Selbsteinschätzung (siehe Anhang F, Tabelle F2). Bei positivem Framing zeigte sich eine statistisch signifikant höhere Korrelation zwischen der Basisrate und der Selbsteinschätzung (siehe Anhang F, Tabelle F1 und Abbildung F1). Positives Framing scheint die Sensitivität der Versuchspersonen für die Basisrate zu erhöhen. Das bedeutet, dass die Selbsteinschätzung der Versuchspersonen stärker durch die Basisrate beeinflusst wurde, wenn diese in der positiven Framingversion präsentiert wurde. Eine mögliche Erklärung dafür liefern Tversky und Kahneman (1981). Sie wiesen darauf hin, dass sich Personen bei positivem Framing häufig risikoavers verhalten, wohingegen sich Personen bei negativem Framing häufig risikofreudig verhalten (siehe Kapitel 1.4). Daraus ergibt sich eine mögliche Erklärung für die erhöhte Sensitivität der Versuchspersonen auf die Basisrate bei positivem Framing. Die Risikoaversion bei dem positiven Framing könnte dazu führen, dass die Versuchspersonen sensitiver auf die Basisrate

reagieren. Die Basisrate kann von den Versuchspersonen dazu genutzt werden, um einen Anhaltspunkt für den Schwierigkeitsgrad der Frage zu erhalten. Somit könnte die stärkere Orientierung der Selbsteinschätzung an der Basisrate zur Reduktion psychologischer Unsicherheit genutzt werden und so vermeintlich das Risiko reduzieren.

Weiterführend stellt sich die Frage, ob nicht nur die Selbsteinschätzung, sondern auch die Präferenzrangliste bei positivem Framing stärker durch die Basisrate beeinflusst wird. Um den Einfluss der Antwortskala und des Framings auf die Z -Werte der Korrelation zwischen der Basisrate und der Präferenzrangliste zu untersuchen, wurde eine 2×2 Kovarianzanalyse ohne Messwiederholung zwischen den experimentellen Bedingungen durchgeführt. Die Punktzahl im BNT wurde als Kovariate verwendet, um für die individuelle Risikokompetenz kontrollieren zu können. Zudem zeigte sich in vorläufigen Analysen ein signifikanter Einfluss des höchsten Bildungsabschlusses sowie der Beschäftigung auf die Z -Werte der Korrelation zwischen der Basisrate und der Präferenzrangliste. Deshalb wurden der höchste Bildungsabschluss und die Beschäftigung ebenfalls als Kovariate verwendet, um dafür kontrollieren zu können. Die Annahmen für die Kovarianzanalyse wurden analog zu dem in Anhang E beschriebenen Vorgehen überprüft und gelten als erfüllt. Die Korrelationen wurden analog zu dem in Kapitel 3.1 beschriebenen Vorgehen berechnet. Auch hierbei wurde eine Fisher Z -Transformation durchgeführt.

Die Interaktion der Faktoren Framing \times Antwortskala erreichte keine statistische Signifikanz (siehe Anhang F, Tabelle F4). Das weist darauf hin, dass der Einfluss des Framings auf die Z -Werte der Korrelation zwischen der Basisrate und der Präferenzrangliste nicht systematisch in Abhängigkeit von der Version der Antwortskala variiert und umgekehrt. Ebenso zeigte sich in der explorativen Datenanalyse weder ein statistisch signifikanter Einfluss der Antwortskala noch ein statistisch signifikanter Einfluss des Framings auf die Z -Werte der Korrelation zwischen der Basisrate und der Präferenzrangliste (siehe Anhang F, Tabelle F4). Das weist darauf hin, dass die Z -Werte der Korrelation zwischen der Basisrate und der Präferenzrangliste nicht signifikant davon abhängen, ob die numerische oder die verbale Antwortskala beziehungsweise ob positives oder negatives Framing verwendet wird.

Demnach scheint das positive Framing die Sensitivität der Versuchspersonen gegenüber der Basisrate zu erhöhen, aber nur im Hinblick auf die Selbsteinschätzung. Bei der Präferenzrangliste hingegen scheint das positive Framing keinen Einfluss auf die Sensitivität der Versuchspersonen gegenüber der Basisrate zu haben. Das liegt möglicherweise darin begründet, dass das Framing bei der Selbsteinschätzung durch die Wörter *richtig* beziehungsweise *falsch* in dem Wortlaut der zwölf Fragen salient auftritt (siehe Wortlaut der

zwölf Fragen in Kapitel 2.2). Bei der Präferenzrangliste hingegen wurden lediglich die Themen inklusive der Basisrate angegeben, ohne dass explizit positives oder negatives Framing eingesetzt wurde. Dies könnte die Ergebnisse erklären.

Im Allgemeinen zeigte sich die im Mittel höchste Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste (siehe Kapitel 3.3). Danach folgte die Korrelation zwischen der Basisrate und der Selbsteinschätzung (siehe Anhang F, Tabelle F1). Die im Mittel niedrigste Korrelation zeigte sich zwischen der Basisrate und der Präferenzrangliste (siehe Anhang F, Tabelle F3).

Das weist darauf hin, dass Menschen sich im Hinblick auf ihre Präferenzen stärker an ihrer Selbsteinschätzung als an Unsicherheitsinformationen wie der Basisrate orientieren. Positives Framing scheint zwar die Sensitivität gegenüber der Basisrate bei der Selbsteinschätzung zu erhöhen, allerdings scheint diese erhöhte Sensitivität gegenüber der Basisrate nicht nachhaltig zu sein, da die Präferenzen der Menschen lediglich in geringem Maß durch die Basisrate beeinflusst wurden.

4.3 Limitationen

Trotz der Stärken der vorliegenden Arbeit sind auch die Limitationen zu berücksichtigen. Diese sollten sowohl bei der Interpretation der Ergebnisse als auch bei zukünftiger Forschung in dem Themengebiet einbezogen werden.

Die Untersuchung wurde online durchgeführt. Auf der einen Seite bringt dies beispielsweise den Vorteil einer einfachen, kostengünstigen, zeit- und ortsunabhängigen Akquirierung vieler Versuchspersonen in einem kurzen Zeitraum mit sich (Mayer, 2013; Sue & Ritter, 2007). Dennoch konnte die mittels einer a priori Power-Analyse berechnete mindestens notwendige Stichprobengröße in der vorliegenden Arbeit nicht erreicht werden (siehe Kapitel 2.1 und Kapitel 2.5). Aufgrund des begrenzten Erhebungszeitraums und des Ausschlusses von Versuchspersonen im Zuge der Datenaufbereitung war es nicht möglich, diese Stichprobengröße zu realisieren. Allerdings ist die bestimmte Stichprobengröße erforderlich, um den Effekt bei einem festgelegten Signifikanzniveau mit der entsprechenden statistischen Power entdecken zu können (Cohen, 1988/2013). Demzufolge kann die zu geringe Stichprobengröße ein Grund dafür sein, dass der Effekt der Antwortskala keine statistische Signifikanz erreicht hat. Die mindestens notwendige Stichprobengröße wurde für die Effektstärke des Effekts der Antwortskala anhand der Ergebnisse des Experiments 2 von Windschitl und Wells (1996) ermittelt (siehe Kapitel 2.1). Deshalb können an dieser Stelle keine Aussagen bezüglich der Stichprobengröße im Hinblick auf den Effekt des Framings getroffen werden.

Auf der anderen Seite kann die Durchführung online auch Nachteile haben. Unter anderem ist die Kontrolle von Einflussfaktoren eingeschränkt, da die Durchführungsbedingungen nicht kontrolliert werden können (Thielsch & Weltzin, 2009). So ist die Durchführungsobjektivität beeinträchtigt, da nicht gewährleistet werden kann, dass alle Versuchspersonen die Untersuchung unter standardisierten und damit vergleichbaren Untersuchungsbedingungen bearbeiten (Moosbrugger & Kelava, 2008). Bei eingeschränkter Objektivität sind die Ergebnisse nicht mehr unabhängig von Einflüssen außerhalb der Versuchsperson, sondern werden beispielsweise durch situationale Faktoren beeinflusst und hängen von diesen ab, was einen Einfluss auf die Ergebnisse haben kann (Moosbrugger & Kelava, 2008). Allerdings weisen Gosling et al. (2004) darauf hin, dass die Ergebnisse bei Online-Studien konsistent zu Ergebnissen bei Verwendung von Fragebögen im traditionellen Papierformat sind. Es kann auch argumentiert werden, dass die Durchführungsobjektivität online durch die vermiedenen Versuchsleitereffekte erhöht wird (Thielsch & Weltzin, 2009).

Aus einer mangelnden Durchführungsobjektivität resultiert ebenfalls die nicht auszuschließende Nutzung von Hilfsmitteln bei der Bearbeitung des BNT. Trotz der Anweisung zu Beginn des BNT, dass keine Hilfsmittel verwendet und bei der Beantwortung der Fragen gewissenhaft gearbeitet werden soll, ist nicht auszuschließen, dass Versuchspersonen die Fragen nicht selbstständig gelöst haben. Beispielsweise durch die Verwendung von künstlicher Intelligenz oder dem Suchen der Lösungen im Internet, ist es nur eingeschränkt möglich, valide Einschätzungen bezüglich der Risikokompetenz durch den BNT treffen zu können. Dies könnte ein Grund dafür sein, warum die Punktzahl im BNT in der vorliegenden Arbeit keinen statistisch signifikanten Einfluss auf die Korrelation zwischen der Selbsteinschätzung und der Präferenzrangliste aufweist.

Eine weitere methodische Limitation liegt im Hinblick auf die nicht vorgenommene Randomisierung der Fragenreihenfolge vor. Obwohl es, wie in Kapitel 2.2 beschrieben, in der vorliegenden Arbeit durchaus sinnvoll war, die zwölf Fragen nicht in ihrer Reihenfolge zu randomisieren, kann dies dennoch Nachteile mit sich bringen. So können beispielsweise Reihenfolgeeffekte auftreten, die durch systematisches Variieren der Fragenreihenfolge zwischen den Versuchspersonen hätten vermieden werden können (Döring, 2023). Solche Reihenfolgeeffekte sind der *Primacy*- und der *Recency-Effekt*. Der *Primacy*- und der *Recency-Effekt* beschreiben, dass Inhalte und in diesem Fall die Fragen, die am Anfang und am Ende präsentiert werden, besser erinnert werden als die Fragen in der Mitte (Felser, 2023). Das heißt, die Anfangs- und Endpositionen haben einen Einfluss (Döring, 2023). Diese Reihenfolgeeffekte können das Urteil der Versuchspersonen verzerren (Döring, 2023).

Des Weiteren weist die in der vorliegenden Arbeit verwendete Art der Risikomanipulation Limitationen auf. Die Risikomanipulation durch Ergebnissalienz zeigte in der Meta-Analyse von Kühberger (1998) Nulleffekte im Hinblick auf den Effekt von Framing (siehe Kapitel 4.1). Demzufolge kann die in der vorliegenden Arbeit verwendete Art der Risikomanipulation die Untersuchung des Einflusses von Framing einschränken.

Auch die empirische Untersuchung in Form eines Experiments kann trotz der zahlreichen Vorteile Limitationen mit sich bringen. So stellt sich im Hinblick auf die externe Validität die Frage, inwiefern die Ergebnisse des Experiments auf natürliche, im Alltag auftretende Situationen übertragbar sind (Bortz & Schuster, 2010).

Überdies wird die externe Validität durch die homogene Stichprobe eingeschränkt (Bortz & Schuster, 2010). Wie in Kapitel 2.5 beschrieben, besteht der Großteil der Versuchspersonen aus weiblichen Studierenden im Alter von etwa 25 Jahren. Somit ist die Stichprobe nicht repräsentativ für die Allgemeinbevölkerung. Aufgrund der homogenen und nicht repräsentativen Stichprobe ist die Generalisierbarkeit der Ergebnisse eingeschränkt, wodurch die Aussagekraft der Ergebnisse begrenzt sein kann (Schmidt-Atzert et al., 2021).

4.4 Fazit und Ausblick

Die vorliegende Arbeit setzt sich mit dem Einfluss von Antwortskalen und Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit auseinander. Die in Kapitel 1.1 aufgestellten Forschungsfragen mussten dabei verneinend beantwortet werden und die in Kapitel 1.6 aufgestellten Hypothesen konnten nicht bestätigt werden. Die Ergebnisse weisen darauf hin, dass frühere Befunde nicht ohne Weiteres repliziert werden können. Es ließ sich nicht feststellen, dass die verbale Antwortskala im Vergleich zu der numerischen Antwortskala eine statistisch signifikant bessere Vorhersage von Präferenzen unter psychologischer Unsicherheit ermöglicht. Auch zeigte sich kein statistisch signifikanter Einfluss des Framings auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit. Ebenso variierte der Einfluss des Framings auf die Vorhersage von Präferenzen nicht statistisch signifikant zwischen der numerischen und der verbalen Antwortskala und umgekehrt. Dennoch leistet die vorliegende Arbeit einen Beitrag zum Verständnis der Einflussfaktoren auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit. Aufbauend auf den Ergebnissen der Arbeit ergeben sich die im Folgenden beschriebenen Perspektiven für zukünftige Forschung.

Im Gegensatz zu dem Experiment 2 von Windschitl und Wells (1996) erreichte der Haupteffekt der Antwortskala in der vorliegenden Arbeit keine statistische Signifikanz. Auch der Befund von Windschitl und Wells (1996), dass die numerische Antwortskala die

Sensitivität gegenüber der Basisrate bei der Selbsteinschätzung erhöht, konnte in der vorliegenden Arbeit nicht repliziert werden. Es ist zu beachten, dass die Befunde von Windschitl und Wells (1996) lediglich auf einen kleinen Effekt der Antwortskala hinwiesen. Dabei sollte berücksichtigt werden, dass ein statistisch signifikanter Effekt nicht automatisch auch von praktischer Bedeutung ist (Pallant, 2020). Demzufolge ist weitere Forschung notwendig, um den Einfluss der Antwortskala auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit sowie die Größe des Effekts zu untersuchen.

Sowohl bei Windschitl und Wells (1996) als auch in der vorliegenden Arbeit war die Stichprobe homogen. Wie in Kapitel 4.3 erläutert, kann dies die Generalisierbarkeit und die Aussagekraft der Ergebnisse einschränken (Schmidt-Atzert et al., 2021). Hinzukommt, dass die Stichprobengröße der vorliegenden Arbeit für den zu erwartenden kleinen Effekt der Antwortskala zu gering war. Deshalb sollten in zukünftiger Forschung Einflussfaktoren auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit mittels einer repräsentativen Stichprobe sowie einer angemessenen Stichprobengröße untersucht werden.

Darüber hinaus sollte in zukünftiger Forschung betrachtet werden, ob andere Antwortskalen als die 11-stufige Rating-Skala bessere Prädiktoren für Präferenzen unter psychologischer Unsicherheit abbilden können. Beispielsweise können anstatt der numerischen und der verbalen 11-stufigen Rating-Skala, wie sie bei Windschitl und Wells (1996) und in der vorliegenden Arbeit verwendet wurden, die quantitative und die qualitative visuelle Analogskala eingesetzt werden. Bei der visuellen Analogskala sind lediglich die Enden der Skala beschriftet (Schmidt-Atzert et al., 2021). Visuelle Analogskalen weisen keine Kategorien auf und den Versuchspersonen wird zur Beantwortung ermöglicht, jede Stelle des Kontinuums zu markieren (Schmidt-Atzert et al., 2021). Aufgrund dessen könnten diese Skalen, insbesondere die qualitative visuelle Analogskala, möglicherweise noch stärker den intuitiven Denkprozess anregen und somit einen besseren Prädiktor für Präferenzen unter psychologischer Unsicherheit abbilden. Demnach sollten diese und auch andere in Frage kommende Antwortskalen in zukünftiger Forschung betrachtet werden.

Im Hinblick auf den Effekt des Framings stellt die in der vorliegenden Arbeit verwendete Art der Risikomanipulation eine Limitation dar (siehe Kapitel 4.1 und Kapitel 4.3). Aufgrund dessen sollte in zukünftiger Forschung die Risikomanipulation durch Bezugnahme auf ein riskantes Ereignis verwendet werden, um den Einfluss von Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit zu untersuchen. So kann der Frage nachgegangen werden, ob es tatsächlich keinen Einfluss von Framing auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit gibt oder ob der statistisch

nicht signifikante Einfluss von Framing durch die verwendete Art der Risikomanipulation zu erklären ist.

Des Weiteren sollte in zukünftiger Forschung in Betracht gezogen werden, die experimentelle Untersuchung in Präsenz durchzuführen, um die Nutzung von Hilfsmitteln bei der Bearbeitung des BNT kontrollieren zu können. So kann herausgefunden werden, ob es tatsächlich keinen Einfluss der Risikokompetenz auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit gibt oder ob der statistisch nicht signifikante Einfluss der Risikokompetenz durch die Verwendung von Hilfsmitteln im Rahmen der Online-Studie zu erklären ist.

Im Allgemeinen zeigte sich in der Arbeit kein statistisch signifikanter Einfluss möglicher Kovariate auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit (siehe Kapitel 3.2). In zukünftiger Forschung ist es sinnvoll, weitere mögliche Kovariate, wie beispielsweise die individuelle Risikoeinstellung der Versuchspersonen, zu untersuchen. Menschen unterscheiden sich in ihrer Risikoeinstellung, also der Art und Weise, wie sie Entscheidungen, die mit Risiko und Unsicherheit verbunden sind, treffen (Blais & Weber, 2006). Demnach kann es in zukünftiger Forschung sinnvoll sein, zu untersuchen, ob die Risikoeinstellung einen Einfluss auf die Vorhersage von Präferenzen unter psychologischer Unsicherheit hat.

Zusammenfassend bietet das Forschungsfeld zahlreiche Perspektiven für zukünftige Forschung, wobei sich die vorliegende Arbeit als Ausgangspunkt eignet. Durch zukünftige Forschung können Einflussfaktoren besser verstanden und bestmögliche Vorhersagen über die Präferenzen von Menschen unter psychologischer Unsicherheit ermöglicht werden. Wie zu Beginn bereits dargelegt, sind Menschen in vielen Lebenssituationen mit Entscheidungen unter unsicheren Bedingungen konfrontiert (Platt & Huettel, 2008). Aufgrund dessen sind die aus dem Forschungsfeld resultierenden Befunde sowohl von theoretischer als auch von praktischer Relevanz und können in vielen Fachbereichen Anwendung finden. Über die psychologische Methodenlehre hinaus können die Erkenntnisse zu den Einflussfaktoren und der bestmöglichen Vorhersage von Präferenzen unter psychologischer Unsicherheit insbesondere in der Markt- und Meinungsforschung wertvoll sein. So können unter anderem im medizinischen, wirtschaftlichen oder versicherungstechnischen Bereich, aber auch in den Bereichen der Werbung sowie der Politik die Präferenzen von Menschen unter psychologischer Unsicherheit bestmöglich vorhergesagt werden. Das kann einen weitreichenden Nutzen mit sich bringen.

Literaturverzeichnis

- American Psychological Association (Hrsg.). (2018). *APA Dictionary of Psychology: uncertainty*. <https://dictionary.apa.org/uncertainty>
- Blais, A.-R. & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1(1), 33–47. <https://doi.org/10.1017/S1930297500000334>
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-12770-0>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Routledge. <https://doi.org/10.4324/9780203771587> (Erstveröffentlichung 1988)
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S. & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7(1), 25–47. <https://doi.org/10.1017/S1930297500001819>
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-64762-2>
- Erdfelder, E. & Ulrich, R. (2018). Zur Methodologie von Replikationsstudien. *Psychologische Rundschau*, 69(1), 3–21. <https://doi.org/10.1026/0033-3042/a000387>
- Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Felser, G. (2023). *Werbe- und Konsumentenpsychologie* (5. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-65471-2>
- Gosling, S. D., Vazire, S., Srivastava, S. & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological bulletin*, 83(2), 314–320. <https://doi.org/10.1037/0033-2909.83.2.314>

- Gummer, T., Roßmann, J. & Silber, H. (2021). Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation. *Sociological Methods & Research*, 50(1), 238–264. <https://doi.org/10.1177/0049124118769083>
- Haase, N., Renkewitz, F. & Betsch, C. (2013). The Measurement of Subjective Probability: Evaluating the Sensitivity and Accuracy of Various Scales. *Risk Analysis*, 33(10), 1812–1828. <https://doi.org/10.1111/risa.12025>
- Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–292. <https://doi.org/10.2307/1914185>
- Kühberger, A. (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational behavior and human decision processes*, 75(1), 23–55. <https://doi.org/10.1006/obhd.1998.2781>
- Lerner, E., Streicher, B. & Raue, M. (2018). Measuring Subjective Risk Estimates. In M. Raue, E. Lerner & B. Streicher (Hrsg.), *Psychological Perspectives on Risk and Risk Analysis* (S. 313–327). Springer. https://doi.org/10.1007/978-3-319-92478-6_14
- Lerner, E., Streicher, B., Sachs, R. & Frey, D. (2013). How risky? The impact of target person and answer format on risk assessment. *Journal of Risk Research*, 16(7), 903–919. <https://doi.org/10.1080/13669877.2012.761267>
- Leung, S.-O. (2011). A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Service Research*, 37(4), 412–421. <https://doi.org/10.1080/01488376.2011.580697>
- Levin, I. P., Schneider, S. L. & Gaeth, G. J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational behavior and human decision processes*, 76(2), 149–188. <https://doi.org/10.1006/obhd.1998.2804>
- Mayer, H. O. (2013). *Interview und schriftliche Befragung: Grundlagen und Methoden empirischer Sozialforschung* (6. Aufl.). Oldenbourg Wissenschaftsverlag. <https://doi.org/10.1524/9783486717624>
- Meade, A. W. & Craig, S. B. (2012). Identifying Careless Responses in Survey Data. *Psychological methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Menold, N. (2019). Effekte der Verbalisierung von Ratingskalen auf die Messqualität. In N. Menold & T. Wolbring (Hrsg.), *Schriftenreihe der ASI - Arbeitsgemeinschaft*

- Sozialwissenschaftlicher Institute. Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente* (S. 103–131). Springer. https://doi.org/10.1007/978-3-658-24517-7_4
- Moosbrugger, H. & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In *Testtheorie und Fragebogenkonstruktion* (S. 7–26). Springer. https://doi.org/10.1007/978-3-540-71635-8_2
- Pallant, J. (2020). *SPSS Survival Manual: A step by step guide to data analysis using IBM SPSS* (7. Aufl.). Routledge. <https://doi.org/10.4324/9781003117452>
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K. & Dickert, S. (2006). Numeracy and Decision Making. *Psychological science*, 17(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2006.01720.x>
- Platt, M. L. & Huettel, S. A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nature neuroscience*, 11(4), 398–403. <https://doi.org/10.1038/nn2062>
- Razali, N. M. & Wah, Y. B. (2011). Power comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Reyna, V. F., Nelson, W. L., Han, P. K. & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological bulletin*, 135(6), 943–973. <https://doi.org/10.1037/a0017327>
- Rohrman, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozial-Psychologie Frankfurt/Main*, 9(3), 222–245.
- Rosenbaum, P. J. & Valsiner, J. (2011). The un-making of a method: From rating scales to the study of psychological processes. *Theory & Psychology*, 21(1), 47–65. <https://doi.org/10.1177/0959354309352913>
- Schmidt-Atzert, L., Krumm, S. & Amelang, M. (2021). *Psychologische Diagnostik* (6. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-61643-7>

- Schreier, M., Echterhoff, G., Bauer, J. F., Weydmann, N. & Hussy, W. (2023). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (3. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-66673-9>
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>
- SoSci Survey (Hrsg.). *Zusätzliche Variablen in der Datenausgabe*. <https://www.sosicisurvey.de/help/doku.php/de:results:variables>
- Sue, V. M. & Ritter, L. A. (2007). *Conducting Online Surveys*. SAGE.
- Thielsch, M. T. & Weltzin, S. (2009). Online-Befragung in der Praxis. In T. Brandenburg & M. T. Thielsch (Hrsg.), *Praxis der Wirtschaftspsychologie: Themen und Fallbeispiele für Studium und Anwendung* (S. 69–85). MV-Verlag.
- Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science (New York, N.Y.)*, 211(4481), 453–458. <https://doi.org/10.1126/science.7455683>
- van Vaerenbergh, Y. & Thomas, T. D. (2013). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*, 25(2), 195–217. <https://doi.org/10.1093/ijpor/eds021>
- Weinstein, N. D., Kwitel, A., McCaul, K. D., Magnan, R. E., Gerrard, M. & Gibbons, F. X. (2007). Risk perceptions: Assessment and relationship to influenza vaccination. *Health Psychology*, 26(2), 146–151. <https://doi.org/10.1037/0278-6133.26.2.146>
- Weng, L.-J. (2004). Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability. *Educational and Psychological Measurement*, 64(6), 956–972. <https://doi.org/10.1177/0013164404268674>
- Windschitl, P. D. & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, 2(4), 343–364. <https://doi.org/10.1037/1076-898X.2.4.343>
- Wirtz, M. A. (2019). *Basisrate*. Dorsch - Lexikon der Psychologie. <https://dorsch.hogrefe.com/stichwort/basisrate>

Anhang

| | |
|---|-------|
| Anhang A: Fragebogenitems und Antwortskalen..... | X |
| Anhang B: Berlin Numeracy Test | XII |
| Anhang C: Versuchsablauf | XIV |
| Anhang D: Stichprobencharakteristika | XV |
| Anhang E: Überprüfung der Annahmen für die Kovarianzanalyse | XVI |
| Anhang F: Ergebnisse der explorativen Datenanalyse..... | XVIII |
| Anhang G: Explorative Datenanalyse von Antwortmustern | XXI |

Anhang A

Tabelle A1

Themen und dazugehörige Basisraten der zwölf Fragen des Fragebogens

| Thema | Basisrate richtig | Basisrate falsch |
|------------------------|-------------------|------------------|
| Künstliche Intelligenz | 57 | 43 |
| Fußball | 70 | 30 |
| Deutsche Bundesländer | 24 | 76 |
| Berühmte Künstler | 44 | 56 |
| Umweltschutz | 32 | 68 |
| Literatur | 80 | 20 |
| Fahrräder | 67 | 33 |
| Deutsche Geschichte | 92 | 8 |
| Berühmte Frauen | 58 | 42 |
| Fitness | 19 | 81 |
| Aktien | 13 | 87 |
| Videospiele | 77 | 23 |

Anmerkung. Angabe der Basisraten in %. Die Basisraten sind identisch zu dem Experiment 2 von Windschitl und Wells (1996). Jeder Versuchsperson werden entweder die Basisraten für richtige Antworten oder die Basisraten für falsche Antworten (Gegenwahrscheinlichkeit) präsentiert.

Tabelle A2*Beschriftung der 11-stufigen numerischen und verbalen Antwortskala*

| Numerisch | Verbal |
|-----------|--|
| 0 | Unmöglich |
| 10 | Extrem unwahrscheinlich |
| 20 | Ziemlich unwahrscheinlich |
| 30 | Eher unwahrscheinlich |
| 40 | Etwas unwahrscheinlich |
| 50 | So wahrscheinlich wie unwahrscheinlich |
| 60 | Etwas wahrscheinlich |
| 70 | Eher wahrscheinlich |
| 80 | Ziemlich wahrscheinlich |
| 90 | Extrem wahrscheinlich |
| 100 | Sicher |

Anmerkung. Angabe der Beschriftung der numerischen Antwortskala in %. Übersetzung der Beschriftung der verbalen Antwortskala aus dem Englischen von dem Experiment 2 von Windschitl und Wells (1996). Jeder Versuchsperson wird entweder die numerische oder die verbale Antwortskala präsentiert.

Anhang B

Tabelle B1

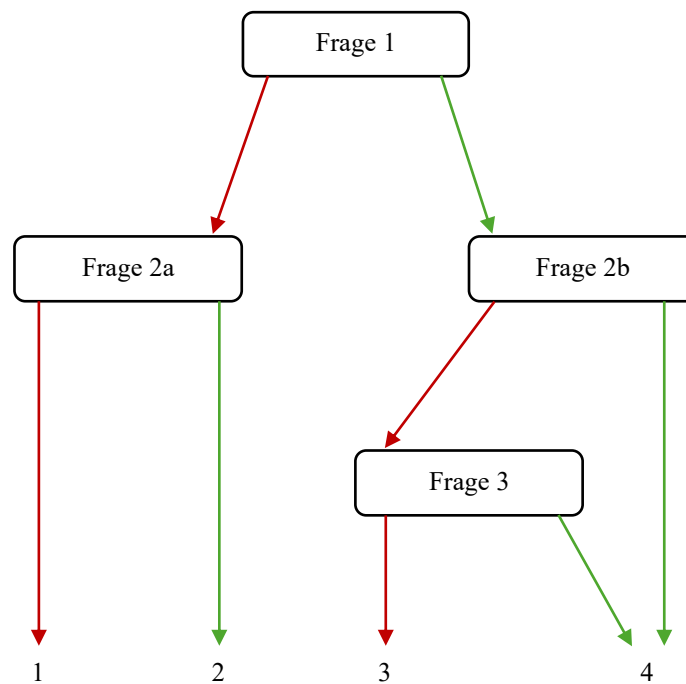
Fragen des BNT mit der jeweils korrekten Antwort

| Nummer der Frage | Frage | Korrekte Antwort |
|---------------------|---|---------------------|
| 1 | <p>Von 1.000 Leuten in einer Kleinstadt sind 500 Mitglied im Gesangsverein. Von diesen 500 Mitgliedern im Gesangsverein sind 100 Männer. Von den 500 Einwohnern, die nicht im Gesangsverein sind, sind 300 Männer.</p> <p>Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Mann ein Mitglied des Gesangsvereins ist? Bitte geben Sie die Wahrscheinlichkeit in % an.</p> | 25 |
| 2a | <p>Stellen Sie sich vor, wir werfen einen fünfseitigen Würfel 50 mal.</p> <p>Bei wie vielen dieser 50 Würfe würde dieser fünfseitige Würfel erwartungsgemäß eine ungerade Zahl zeigen (1, 3 oder 5)?</p> | 30 |
| 2b | <p>Stellen Sie sich vor, wir werfen einen gezinkten Würfel (6 Seiten). Die Wahrscheinlichkeit, dass der Würfel eine 6 zeigt, ist doppelt so hoch wie die Wahrscheinlichkeit jeder der anderen Zahlen.</p> <p>Von 70 Würfeln, bei wie vielen dieser 70 Würfe würde dieser Würfel erwartungsgemäß eine 6 zeigen?</p> | 20 |
| 3 | <p>In einem Wald sind 20% der Pilze rot, 50% braun und 30% weiß. Ein roter Pilz ist mit einer Wahrscheinlichkeit von 20% giftig. Ein Pilz, der nicht rot ist, ist mit einer Wahrscheinlichkeit von 5% giftig.</p> <p>Wie hoch ist die Wahrscheinlichkeit, dass ein giftiger Pilz im Wald rot ist?</p> | 50 |

Anmerkung. Verwendet wird die deutsche Version des adaptiven Testformats des BNT von Cokely et al. (2012).

Abbildung B1

Adaptives Testformat des BNT inklusive der Punktevergabe

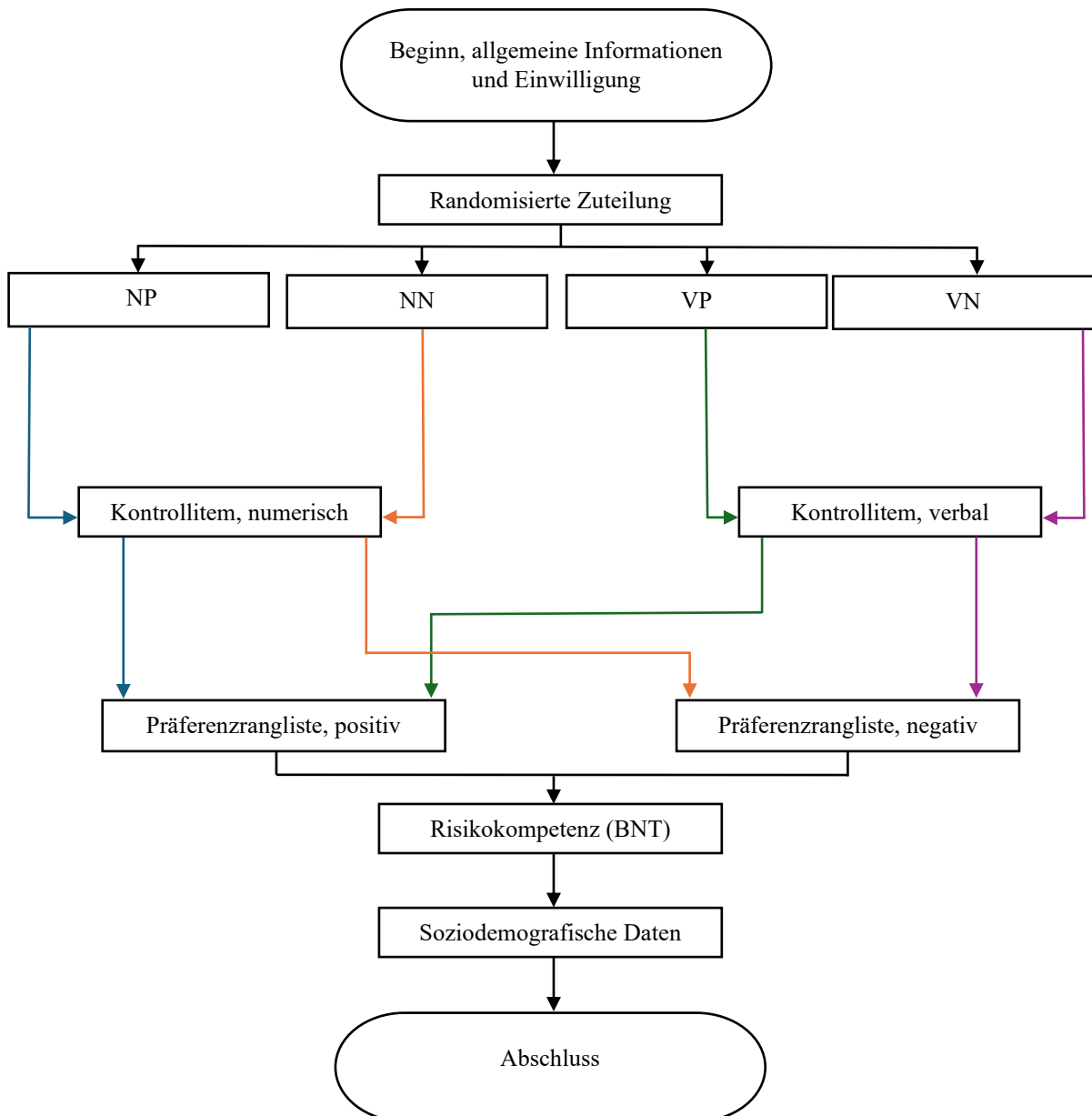


Anmerkung. Nach Cokely et al. (2012). Vor der Bearbeitung des BNT wurde folgende Anweisung angezeigt: „Bitte beantworten Sie die folgenden Fragen. Verwenden Sie keinen Taschenrechner, aber nutzen Sie gerne Schmierpapier für Notizen. Es ist wichtig, dass Sie dabei gewissenhaft arbeiten.“. Die grünen Pfeile geben korrekte Antworten an und die roten Pfeile geben inkorrekte Antworten an. Jede Frage hatte eine Wahrscheinlichkeit von 50% korrekt beantwortet zu werden. Wenn eine Frage korrekt beantwortet wurde, wurde eine schwerere Frage präsentiert. Wenn eine Frage inkorrekt beantwortet wurde, wurde eine leichtere Frage präsentiert. Um die Motivation zur Bearbeitung der Fragen zu erhöhen und die Abbruchquote zu verringern, wurde vor der Frage 2a sowie vor der Frage 3 folgender Satz angezeigt: „Nur noch diese Aufgabe muss gerechnet werden :)“. Vor der Frage 2b wurde folgender Satz angezeigt: „Bleiben Sie dran, es ist gleich geschafft!“.

Anhang C

Abbildung C1

Versuchsablauf der experimentellen Untersuchung



Anmerkung. Die bunten Pfeile stellen den Pfad von den zwölf Fragen mit der zugehörigen Selbsteinschätzung bis zu der Präferenzrangliste in den vier verschiedenen experimentellen Bedingungen dar. Dabei repräsentieren die blauen Pfeile den Pfad in der Bedingung NP. Die orangenen Pfeile repräsentieren den Pfad in der Bedingung NN. Die grünen Pfeile repräsentieren den Pfad in der Bedingung VP. Die lila Pfeile repräsentieren den Pfad in der Bedingung VN.

Anhang D

Tabelle D1

Absolute und relative Häufigkeiten der Stichprobencharakteristika

| Stichprobencharakteristika | <i>n</i> | Relative Häufigkeit |
|--|----------|---------------------|
| Höchster Bildungsabschluss | 285 | 98.6 |
| Noch Schüler oder Schülerinnen | 2 | 0.7 |
| Realschulabschluss | 8 | 2.8 |
| Fachhochschulreife | 14 | 4.8 |
| Abitur | 224 | 77.5 |
| Hochschulabschluss | 35 | 12.1 |
| Promotion | 2 | 0.7 |
| Beschäftigung | 285 | 98.6 |
| Schüler oder Schülerinnen | 2 | 0.7 |
| Ausbildung | 1 | 0.3 |
| Studierende | 242 | 83.7 |
| Angestellte | 27 | 9.3 |
| Beamte | 4 | 1.4 |
| Selbstständige | 2 | 0.7 |
| Arbeitslos / Arbeit suchend | 1 | 0.3 |
| Hausfrau oder Hausmann | 1 | 0.3 |
| Rentner oder Rentnerinnen | 5 | 1.7 |
| Studiengänge | 239 | 82.7 |
| Lehramt | 78 | 27.0 |
| Psychologie | 62 | 21.5 |
| Medizin und Gesundheitswissenschaften | 51 | 17.6 |
| Interdisziplinärer Studiengang | 12 | 4.2 |
| Naturwissenschaften | 10 | 3.5 |
| Sozial- und Gesellschaftswissenschaften | 9 | 3.1 |
| Wirtschaftswissenschaften | 5 | 1.7 |
| Sprach- und Kulturwissenschaften | 3 | 1.0 |
| Umwelt-, Agrar-, Forst- und Ernährungswissenschaften | 3 | 1.0 |
| Ingenieurwissenschaften | 2 | 0.7 |
| Kunst, Musik, Design und Medien | 2 | 0.7 |
| Rechtswissenschaften | 2 | 0.7 |

Anmerkung. Angaben der relativen Häufigkeiten in %.

Anhang E

Für die Datenauswertung wurde eine 2×2 Kovarianzanalyse ohne Messwiederholung zwischen den experimentellen Bedingungen durchgeführt. Die Punktzahl im BNT wurde als Kovariate verwendet, um für die individuelle Risikokompetenz kontrollieren zu können. Die Kovarianzanalyse zählt zu den parametrischen statistischen Verfahren, weshalb einige Annahmen erfüllt sein müssen (Pallant, 2020). Da die Kovarianzanalyse eine Erweiterung der Varianzanalyse ist, müssen einerseits die Annahmen der Varianzanalyse erfüllt sein (Pallant, 2020). Eine der Annahmen für die Varianzanalyse betrifft das Niveau der Messung. Diese Annahme besagt, dass die abhängige Variable auf Intervall- oder Verhältnisskalenniveau gemessen wird (Pallant, 2020). Aufgrund der Konzeption der vorliegenden Arbeit, bei der Z-Werte als abhängige Variable verwendet werden, gilt diese Annahme als erfüllt. Zwei weitere Annahmen besagen zum einen, dass eine Zufallsstichprobe aus der Grundgesamtheit verwendet wird und zum anderen, dass die Unabhängigkeit der Beobachtungen gegeben ist (Pallant, 2020). Diese Annahmen können aufgrund des Versuchsdesigns als erfüllt gelten (siehe Kapitel 2.1). Eine weitere Annahme ist die Normalverteilung der Werte für die kontinuierlichen Variablen (Pallant, 2020). Die Annahme der Normalverteilung wurde mit Hilfe des *Shapiro-Wilk-Tests* überprüft. Die Untersuchung von Razali und Wah (2011) weist darauf hin, dass der Shapiro-Wilk-Test im Vergleich zu anderen häufig verwendeten Normalverteilungstests über die höchste statistische Power verfügt. Deshalb wurde dieser Normalverteilungstest betrachtet. Der Shapiro-Wilk-Test zeigte, dass die Z-Werte in der Stichprobe normalverteilt sind, $p = .051$. Überdies sollte das zentrale Grenzwerttheorem berücksichtigt werden. Das zentrale Grenzwerttheorem besagt: „Die Verteilung von Mittelwerten aus Stichproben des Umfangs n , die derselben Grundgesamtheit entnommen wurden, geht mit wachsendem Stichprobenumfang in eine Normalverteilung über.“ (Bortz & Schuster, 2010). Bei einem Stichprobenumfang von $n > 30$ pro Bedingung kann bei beliebigen Verteilungsformen von einer normalverteilten Mittelwertverteilung ausgegangen werden, sodass eine Verletzung der Normalverteilungsannahme keine größeren Probleme verursachen sollte (Bortz & Schuster, 2010; Pallant, 2020). Dies war in der vorliegenden Arbeit mit $n = 72$ Versuchspersonen in der Bedingung NP, $n = 76$ Versuchspersonen in der Bedingung NN, $n = 73$ Versuchspersonen in der Bedingung VP und $n = 68$ Versuchspersonen in der Bedingung VN der Fall. Eine weitere Annahme ist die Homogenität der Varianzen in den experimentellen Bedingungen (Pallant, 2020). Die Annahme der Varianzhomogenität wurde mit Hilfe des *Levene-Tests*

überprüft. Es zeigte sich, dass die Varianzen der experimentellen Bedingungen homogen sind, $F(3, 285) = 0.40, p = .753$.

Andererseits müssen bei der Kovarianzanalyse, über die Annahmen der Varianzanalyse hinaus, weitere Annahmen erfüllt sein, welche im Folgenden thematisiert werden. Eine Annahme für die Kovarianzanalyse ist die Reliabilität der Kovariate (Pallant, 2020). Diese Annahme kann durch die Verwendung des BNT, einem gut validierten Test zur Erfassung der Risikokompetenz, als erfüllt gelten (siehe Kapitel 2.3). Des Weiteren stellt die lineare Beziehung zwischen der abhängigen Variable und der Kovariate eine Annahme für die Kovarianzanalyse dar (Pallant, 2020). Die Annahme der Linearität wurde mit Hilfe von Streudiagrammen, die die Beziehung zwischen den Z-Werten und der Punktzahl im BNT für jede der vier experimentellen Bedingungen abbildeten, untersucht und gilt als erfüllt. Eine weitere Annahme für die Kovarianzanalyse ist die Homogenität der Regressionssteigung (Pallant, 2020). Die Interaktion zwischen der unabhängigen Variable Antwortskala und der Kovariate Risikokompetenz in Form der Punktzahl im BNT war statistisch nicht signifikant, $F(1, 283) = 0.60, p = .440, \eta_p^2 = .002$. Auch die Interaktion zwischen der unabhängigen Variable Framing und der Kovariate Risikokompetenz in Form der Punktzahl im BNT war statistisch nicht signifikant, $F(1, 283) = 0.89, p = .346, \eta_p^2 = .003$. Aufgrund dessen gilt die Annahme der Homogenität der Regressionssteigung als erfüllt.

Anhang F

Tabelle F1

Deskriptive Statistik der Z-Werte zwischen Basisrate und Selbsteinschätzung

| | <i>n</i> | <i>M</i> | <i>SD</i> | <i>M</i> _{bereinigt} | <i>SEM</i> |
|-------------------------|----------|----------|-----------|-------------------------------|------------|
| NP | 72 | 0.91 | 0.68 | 0.91 | 0.07 |
| NN | 75 | 0.47 | 0.55 | 0.47 | 0.07 |
| VP | 71 | 0.75 | 0.67 | 0.77 | 0.08 |
| VN | 67 | 0.51 | 0.65 | 0.50 | 0.08 |
| Numerische Antwortskala | 147 | 0.69 | 0.65 | 0.69 | 0.05 |
| Verbale Antwortskala | 138 | 0.64 | 0.67 | 0.64 | 0.05 |
| Positives Framing | 143 | 0.83 | 0.68 | 0.84 | 0.05 |
| Negatives Framing | 142 | 0.49 | 0.60 | 0.48 | 0.05 |

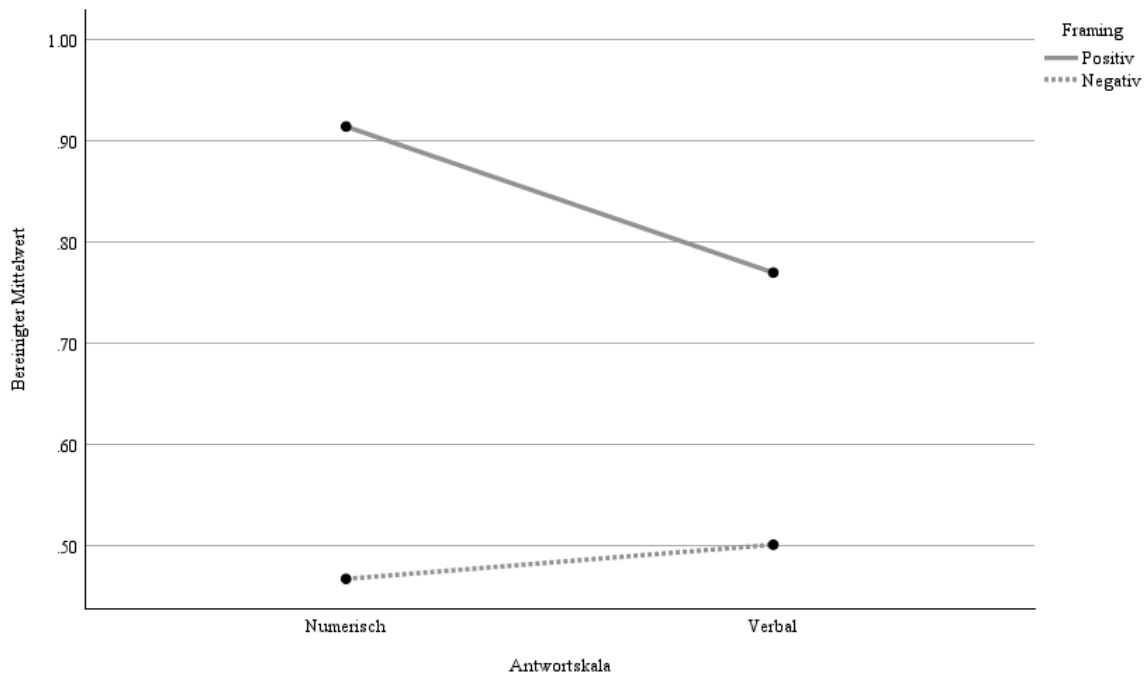
Anmerkung. $N = 285$. Die Fisher Z-Transformation der Pearson Produkt-Moment-Korrelationskoeffizienten zwischen der Basisrate und der Selbsteinschätzung lag im Mittel bei $M = 0.66$ ($SD = 0.66$). Die Bereinigung bezieht sich auf die Tatsache, dass die Effekte der Kovariaten statistisch entfernt wurden. Die im Modell enthaltene Kovariate Risikokompetenz wurde mit dem Wert $BNT = 2.52$ (mittlere Punktzahl im BNT in der Stichprobe der vorliegenden Arbeit) bewertet. Die im Modell enthaltene Kovariate Alter wurde mit dem Wert $Alter = 25.08$ (mittleres Alter in Jahren in der Stichprobe der vorliegenden Arbeit) bewertet.

Tabelle F2

Inferenzstatistik der Z-Werte zwischen Basisrate und Selbsteinschätzung

| | <i>F</i> | <i>p</i> | η_p^2 |
|--------------------------|----------|------------------|------------|
| Interaktionseffekt | 1.41 | .237 | .005 |
| Haupteffekt Antwortskala | 0.55 | .230 (einseitig) | .002 |
| Haupteffekt Framing | 22.28 | < .001 | .074 |

Anmerkung. $N = 285$. Die in der Tabelle abgebildeten inferenzstatistischen Ergebnisse beziehen sich auf die Fisher Z-Transformation der Pearson Produkt-Moment-Korrelationskoeffizienten zwischen der Basisrate und der Selbsteinschätzung. Die Zählerfreiheitsgrade sind $df = 1$ und die Nennerfreiheitsgrade sind $df = 279$. Das partielle Eta-Quadrat von $\eta_p^2 = .074$ zeigt einen mittleren Effekt an (Cohen, 1988/2013). Die Effektstärke zeigt an, dass 7.4% der Varianz in der Korrelation zwischen der Basisrate und der Selbsteinschätzung durch das Framing erklärt werden.

Abbildung F1*Bereinigte Z-Mittelwerte zwischen Basisrate und Selbsteinschätzung*

Anmerkung. Bereinigte Mittelwerte der Fisher Z-Transformation der Pearson Produkt-Moment-Korrelationskoeffizienten zwischen der Basisrate und der Selbsteinschätzung von $N = 285$ Versuchspersonen für die zwei Framingversionen (positiv, negativ) in Abhängigkeit von der Antwortskala (numerisch, verbal). Auf der x -Achse sind die Stufen des Faktors Antwortskala und auf der y -Achse ist der bereinigte Mittelwert der Z-Werte der Korrelation zwischen Basisrate und Selbsteinschätzung abgetragen. Zu beachten ist, dass die y -Achse nicht bei dem Wert Null beginnt. Die durchgezogene Linie repräsentiert die Bedingung mit positivem Framing und die gestrichelte Linie repräsentiert die Bedingung mit negativem Framing. Die im Modell enthaltene Kovariate Risikokompetenz wurde mit dem Wert BNT = 2.52 (mittlere Punktzahl im BNT in der Stichprobe der vorliegenden Arbeit) bewertet. Die im Modell enthaltene Kovariate Alter wurde mit dem Wert Alter = 25.08 (mittleres Alter in Jahren in der Stichprobe der vorliegenden Arbeit) bewertet.

Tabelle F3*Deskriptive Statistik der Z-Werte zwischen Basisrate und Präferenzrangliste*

| | <i>n</i> | <i>M</i> | <i>SD</i> | <i>M</i> _{bereinigt} | <i>SEM</i> |
|-------------------------|----------|----------|-----------|-------------------------------|------------|
| NP | 72 | 0.58 | 1.27 | 0.58 | 0.12 |
| NN | 75 | 0.32 | 0.80 | 0.32 | 0.12 |
| VP | 71 | 0.46 | 0.60 | 0.51 | 0.12 |
| VN | 67 | 0.51 | 1.32 | 0.47 | 0.13 |
| Numerische Antwortskala | 147 | 0.45 | 1.06 | 0.45 | 0.08 |
| Verbale Antwortskala | 138 | 0.49 | 1.01 | 0.49 | 0.09 |
| Positives Framing | 143 | 0.52 | 0.99 | 0.54 | 0.09 |
| Negatives Framing | 142 | 0.41 | 1.08 | 0.39 | 0.09 |

Anmerkung. $N = 285$. Die Fisher Z-Transformation der Spearman-

Rangkorrelationskoeffizienten zwischen der Basisrate und der Präferenzrangliste lag im Mittel bei $M = 0.47$ ($SD = 1.04$). Die Bereinigung bezieht sich auf die Tatsache, dass die Effekte der Kovariaten statistisch entfernt wurden. Die im Modell enthaltene Kovariate Risikokompetenz wurde mit dem Wert $BNT = 2.52$ (mittlere Punktzahl im BNT in der Stichprobe der vorliegenden Arbeit) bewertet. Die im Modell enthaltene Kovariate höchster Bildungsabschluss wurde mit dem Wert $Bildung = 6.96$ (mittlerer höchster Bildungsabschluss in der Stichprobe der vorliegenden Arbeit (entspricht in etwa dem höchsten Bildungsabschluss Abitur)) bewertet. Die im Modell enthaltene Kovariate Beschäftigung wurde mit dem Wert $Beschäftigung = 3.28$ (mittlere Beschäftigung in der Stichprobe der vorliegenden Arbeit (entspricht in etwa der Beschäftigung Studierende)) bewertet.

Tabelle F4*Inferenzstatistik der Z-Werte zwischen Basisrate und Präferenzranglist*

| | <i>F</i> | <i>p</i> | η_p^2 |
|--------------------------|----------|------------------|------------|
| Interaktionseffekt | 0.79 | .377 | .003 |
| Haupteffekt Antwortskala | 0.12 | .728 | .000 |
| Haupteffekt Framing | 1.55 | .108 (einseitig) | .006 |

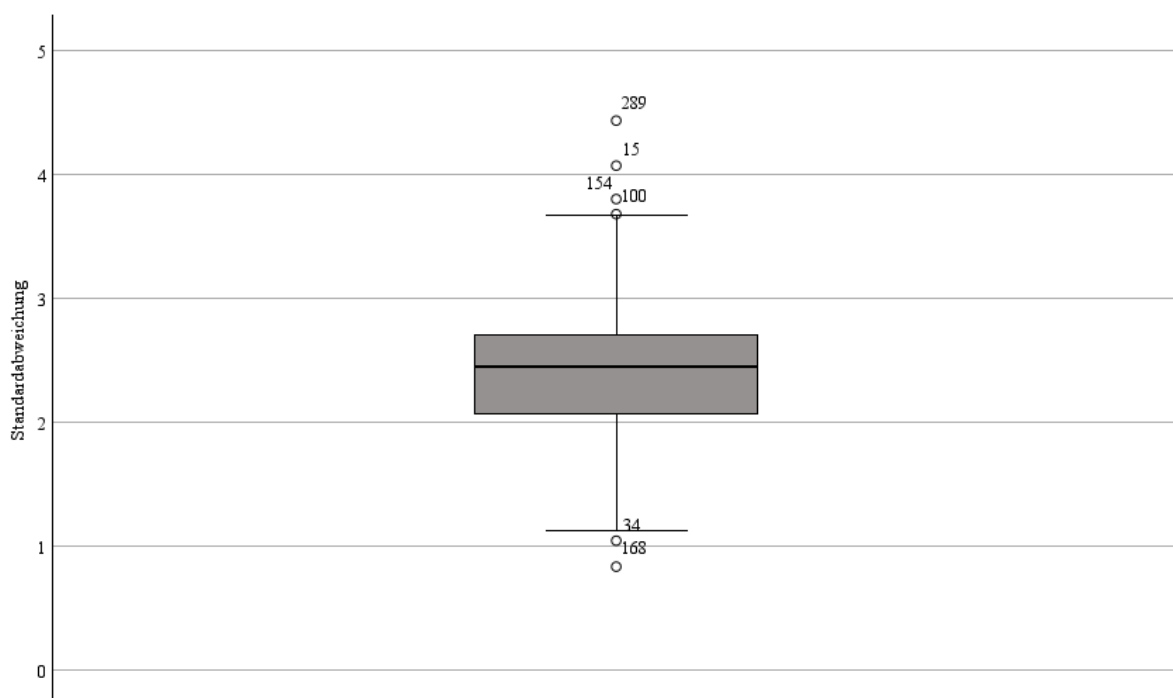
Anmerkung. $N = 285$. Die in der Tabelle abgebildeten inferenzstatistischen Ergebnisse beziehen sich auf die Fisher Z-Transformation der Spearman-Rangkorrelationskoeffizienten zwischen der Basisrate und der Präferenzrangliste. Die Zählerfreiheitsgrade sind $df = 1$ und die Nennerfreiheitsgrade sind $df = 278$.

Anhang G

Im Rahmen der explorativen Datenanalyse wurden zusätzlich Antwortmuster der Versuchspersonen bei den zwölf Fragen untersucht. Antwortmuster, wie beispielsweise die Invarianz der Antworten, die Tendenz zu Extremwerten oder die Tendenz zur Mitte, können die Ergebnisse verzerren (van Vaerenbergh & Thomas, 2013). Zur Untersuchung der Antwortmuster wurde zunächst die Standardabweichung jeder Versuchsperson bei den zwölf Fragen als Maß der Antwortvarianz betrachtet. Je größer die Standardabweichung, desto größer ist die Varianz in den Antworten. Die Ergebnisse weisen nicht auf Versuchspersonen mit starken Invarianzen der Antworten hin (siehe Anhang G, Abbildung G1). Dennoch zeigten sich zwei Ausreißer mit einer geringen und vier Ausreißer mit einer hohen Standardabweichung bei den zwölf Fragen (siehe Anhang G, Abbildung G1).

Abbildung G1

Standardabweichungen der Versuchspersonen bei den zwölf Fragen



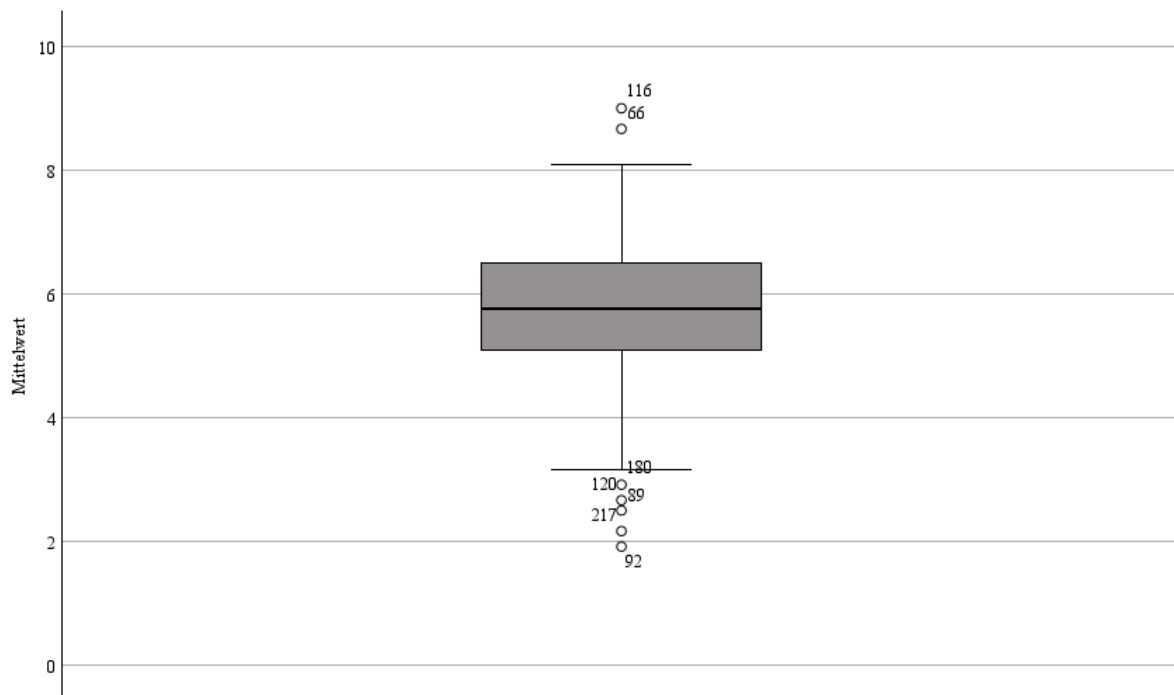
Anmerkung. $N = 290$. Die Länge der Box stellt den Interquartilbereich dar. Die Linie im Inneren der Box stellt den Median dar. Ausreißer sind durch Kreise repräsentiert. Ausreißer am unteren Ende weisen eine Standardabweichung von $SD \leq 1.0$ auf. Ausreißer am oberen Ende weisen eine Standardabweichung von $SD \geq 3.7$ auf. $M = 2.39$ ($SD = 0.50$).

Um diese Fälle im Hinblick auf die Antwortmuster genauer untersuchen zu können, wurde der Mittelwert jeder Versuchsperson bei den zwölf Fragen betrachtet. Es zeigten sich fünf Ausreißer mit einem geringen Mittelwert und zwei Ausreißer mit einem hohen Mittelwert (siehe Anhang G, Abbildung G2). Die Mittelwerte können einen Hinweis auf die Tendenz zu Extremwerten sowie einen Hinweis auf die Tendenz zur Mitte geben. Eine geringe Standardabweichung und ein geringer oder hoher Mittelwert können auf die Tendenz zu Extremwerten hinweisen. Auch eine hohe Standardabweichung und ein mittlerer Mittelwert können auf die Tendenz zu Extremwerten hinweisen. Eine geringe Standardabweichung und ein mittlerer Mittelwert können auf die Tendenz zur Mitte hinweisen. Die Ausreißer wurden mit Blick auf diese Charakteristika manuell untersucht. So wurde eine Versuchsperson mit der Tendenz zur Mitte festgestellt ($M = 6.00$, $SD = 1.04$). Diese Versuchsperson wurde aus der Analyse ausgeschlossen (siehe Kapitel 3.1). Die anderen Ausreißer schienen bei manueller Betrachtung keine auffälligen Antwortmuster aufzuweisen.

Die Untersuchung der Antwortmuster wurde nach dem Ausschluss von Versuchspersonen im Zuge der Datenaufbereitung durchgeführt. Es kann angenommen werden, dass bei dem Ausschluss von Versuchspersonen im Zuge der Datenaufbereitung bereits Versuchspersonen ausgeschlossen wurden, die den Fragebogen nicht ernsthaft beantwortet haben. Dies erfolgte zum Beispiel mit Hilfe des Ausschlusses durch das Kontrollitem sowie mit Hilfe des Ausschlusses durch die Summe der Bearbeitungsdauer (siehe Kapitel 3.1). Aufgrund dessen ist es möglich, dass ein Großteil der Versuchspersonen mit einem auffälligen Antwortmuster bereits vor der Untersuchung von Antwortmustern ausgeschlossen wurde.

Abbildung G2

Mittelwerte der Versuchspersonen bei den zwölf Fragen



Anmerkung. $N = 290$. Die Länge der Box stellt den Interquartilbereich dar. Die Linie im Inneren der Box stellt den Median dar. Ausreißer sind durch Kreise repräsentiert. Ausreißer am unteren Ende weisen einen Mittelwert von $M \leq 2.9$ auf. Ausreißer am oberen Ende weisen einen Mittelwert von $M \geq 8.7$ auf. $M = 5.74$ ($SD = 1.14$).