

# Die

# POLIZEI

FACHZEITSCHRIFT FÜR DIE ÖFFENTLICHE SICHERHEIT MIT BEITRÄGEN AUS DER DEUTSCHEN HOCHSCHULE DER POLIZEI

## Schwerpunkt Kriminalistik

### HERAUSGEBER

Hans-Jürgen Lange  
Holger Münch

### REDAKTION

Dieter Müller (Schriftleitung)  
Ralph Berthel  
Sabrina Schönrock  
Sandra Schmidt

### AUS DEM INHALT

#### Aufsätze

##### Plank

Hat die (Kriminal-) Polizei ihr »Herz für die Gefahrenabwehr« entdeckt bzw. neu definiert? Eine erste kriminalwissenschaftliche Skizze S. 333

##### Kepura

Wer kann Kriminalprävention? S. 338

##### Brünsing/Biedermann

Experimentelle Untersuchung von Einflussfaktoren auf die Genauigkeit von Altersschätzungen durch Zeugen S. 343

##### Schaeff

Nutzung dienstlicher Smartphones und Applikationen in der Kriminalitätsbekämpfung S. 351

##### Goertz

Aktuelle Trends und Akteure im deutschen Extremismus – Eine vergleichende Analyse S. 357

#### Aus der Wissenschaft

Kurzportraits bemerkenswerter Masterarbeiten S. 366

Heft 10  
Oktober 2023  
Seiten 333–368  
114. Jahrgang  
Art.-Nr. 56244310  
PVSt 5624

# 10

Carl Heymanns Verlag

ventionsarbeit balanciert das Spannungsfeld zwischen der *Gewährleistung staatlicher Ordnung* und der *Befähigung zur gesellschaftlichen Teilhabe*. Polizistinnen und Polizisten sind in ihrer Präventionsarbeit mit jungen Menschen pädagogisch orientiert und das obwohl sie sich nicht selbst als Pädagoginnen und Pädagogen sehen. Allein diese zentralen Erkenntnisse bieten eine Fülle von Anregungen für die Konzeptionierung kriminalpräventiver Arbeit. So ist in der Polizei in einem viel stärkeren Maße als bisher die Erschlie-

ßung pädagogischen Wissens als Handlungsgrundlage und Reflexionswissen wünschenswert.<sup>49</sup> Vor allem bedarf es aber einer Klärung, ob die kriminalpräventive Arbeit in instruierender Weise Normen verdeutlichen oder den Sinn von Normen vermitteln soll.<sup>50</sup>

<sup>49</sup> Vgl. *Kepura* (Fn. 1), S. 301.

<sup>50</sup> Vgl. *Kepura* (Fn. 1), S. 304 f.

## Experimentelle Untersuchung von Einflussfaktoren auf die Genauigkeit von Altersschätzungen durch Zeugen

von Ulf Brünsing, Potsdam\* und Prof. Dr. Jürgen Biedermann, Oranienburg\*\*

Akkurate Altersschätzungen zu tatverdächtigen Personen sind für die polizeiliche Ermittlungsarbeit von großer Bedeutung. Gleichzeitig lässt sich der bisherige empirische Forschungsstand diesbezüglich als unzureichend bezeichnen. Ziel der vorliegenden Untersuchung war die Identifizierung von Einflussfaktoren auf die Genauigkeit der Altersschätzung bei fremden Personen, um diese in polizeilichen Anwendungsfällen nutzen zu können. Hierzu wurde ein Experiment durchgeführt, bei welchem 163 Versuchspersonen das Alter von 30 auf Lichtbildern abgebildeten Stimulipersonen schätzen sollten. Es wurden in der Forschungsliteratur bereits in anderen Kontexten festgestellte Einflussfaktoren auf die Qualität von Zeugenaussagen auf ihre Übertragbarkeit bei Altersschätzungen überprüft (Own-Age-Effekt, Own-Sex-Effekt, Grad der subjektiven Sicherheit der Entscheidung). Es konnte ein statistisch signifikanter Einfluss der Zugehörigkeit von Versuchspersonen und Stimuliperson zur gleichen Altersgruppe (Own-Age-Effekt) und des subjektiven Sicherheitsempfindens auf die Genauigkeit von Altersschätzungen nachgewiesen werden. Die Übereinstimmung des Geschlechts zwischen Versuchsperson und Stimuliperson (Own-Sex-Effekt) zeigte hingegen keinen Einfluss auf die Güte der Altersschätzung. Als weitere wichtige Einflussgröße auf die Güte der Altersschätzungen konnten Merkmalsausprägungen der dargestellten Stimuligesichter identifiziert werden. Die Ergebnisse werden auch im Hinblick auf Schlussfolgerungen für die polizeiliche Praxis diskutiert.

### I. Einleitung

Nach einem versuchten Tötungsdelikt werden drei Zeugen, welche den Täter bei der Flucht beobachtet haben, durch die Polizei zum Täteralter befragt. Die Zeugen machen hierbei unterschiedliche Altersangaben. Dem Ermittler stellt sich nun die Frage, welcher Zeuge das Alter am genauesten geschätzt hat bzw. bei welcher Person die Aussagegüte am höchsten ist, um möglichst effektive und adäquate polizeiliche Maßnahmen einleiten zu können. Die Güte der Altersschätzung ergibt sich aus dem Abstand zwischen dem Schätzalter und Realalter der einzuschätzenden Person, wobei es hierbei zu Altersüberschätzungen und Altersunterschätzungen kommen kann.

Aufgrund der polizeilichen Relevanz von Altersschätzungen, nicht nur bei Verfahren der Schwerestrafbarkeit, besteht ein großes Interesse daran, auf empirischem Weg Einflussfaktoren zu ermitteln, um die Zuverlässigkeit von Altersschätzungen bewerten zu können. Erkenntnisse zu derartigen Einflussfaktoren sollen in vorliegender Untersuchung im Rahmen eines Experimentes geschaffen bzw. vertieft werden. Es soll also ein Wissensstand generiert werden, der es uns erlaubt, Entscheidungen nicht vorrangig aufgrund eines bestimmten »Bauchgefühls« zu treffen, sondern aufgrund einer wissenschaftlichen Faktenlage.

Das oben genannte Beispiel nimmt Bezug auf die Zuverlässigkeit visueller Wahrnehmungen der drei Zeugen. Diese haben Informationen registriert, verarbeitet und wiedergegeben. Die daraus resultierenden Zeugenaussagen sind ein komplexer kognitiver Leistungsprozess, dessen Verlauf von dem Wahrnehmen einer Situation, der Enkodierung der Informationen (Überführung der Informationen ins Gedächtnis) und dem Abrufen bis hin zur verbalen Rekonstruktion des damaligen Geschehens reicht (*Greuel*, 2016). In allen Phasen dieses Prozesses können Faktoren auftreten, welche das Rekonstruktionsergebnis hinsichtlich der Genauigkeit, der Zuverlässigkeit und somit der Verwertbarkeit beeinflussen (ebd.).

Wissenschaftlichen Studien zufolge wirkt sich unter bestimmten Bedingungen die Zugehörigkeit des Zeugen zur gleichen Altersgruppe wie die einzuschätzende Person der Wahrnehmungssituation bedeutsam auf die Güte der Zeugenaussagen aus (*Anastasi & Rhodes*, 2005; Landeskriminalamt Nordrhein-Westfalen, 2011; *Meissner et al.*, 2007). Der *Own-Age-Effekt* besagt, dass die Wahrnehmung, Erinnerung und Wiedererkennung von Personen und Gesichtern bei der eigenen Altersgruppe leichter fällt als bei Angehörigen anderer Altersgruppen (*Greuel*, 2016). Dieser Effekt zeigte sich auch bei einer wissenschaftlichen Auswertung von

\* Der Verfasser ist Dezernent bei der Mordkommission Potsdam in Brandenburg. Der vorliegende Artikel beruht auf einer experimentellen Untersuchung im Kontext der Masterarbeit des Erstautors im Masterstudiengang »Kriminalistik« an der Hochschule der Polizei des Landes Brandenburg.

\*\* Der Verfasser ist Professor für Psychologie an der Hochschule der Polizei des Landes Brandenburg.

229 Ermittlungsakten zu Raubüberfällen auf Geldinstitute des Landeskriminalamtes Nordrhein-Westfalen. Bei dieser Untersuchung konnte festgestellt werden, dass Zeugen das Täteralter genauer schätzen, wenn die Differenz zwischen Täter- und Zeugenalter weniger als 10 Jahre betrug. Bei Zeugen, die eine große Altersdifferenz zum Täter aufwiesen, wurden hingegen größere Unsicherheiten bei der Altersschätzung beobachtet (Landeskriminalamt Nordrhein-Westfalen, 2011).

Als weiterer Einflussfaktor auf die Zuverlässigkeit von Zeugenaussagen wird in der wissenschaftlichen Literatur ein Geschlechtsunterschied zwischen Zeugen und den einzuschätzenden Personen der Wahrnehmungssituation beschrieben (Ellis et al., 1973; Meissner et al., 2007; Shepherd & Deregowski, 1981; Sporer, 1992). Der *Own-Sex-Effekt* besagt, dass die Wiedererkennung von Personen des gleichen Geschlechts einfacher gelingt als die Wiedererkennung von Personen des anderen Geschlechts (Greuel, 2016). Es existieren allerdings vergleichsweise wenige Untersuchungen, die sich außerhalb von Identifizierungsleistungen speziell mit dem Einfluss des Geschlechtes auf die spezifische Wahrnehmung, Erinnerung und Beschreibung von Gesichtern befassen.

Die subjektive Sicherheit in Bezug auf getätigte Zeugenaussagen gilt in der Praxis ebenfalls als Gradmesser für die Beurteilung der Validität von Zeugenaussagen und wird bei der Beweiswürdigung von Identifizierungsleistungen herangezogen. Der Ausspruch »Ich bin mir absolut sicher« bedeutet wissenschaftlichen Erkenntnis zufolge allerdings nicht, dass die Identifizierungs- oder Beschreibungsleistung, die der Zeuge oder die Zeugin erbracht hat, quasi zwangsläufig richtig ist (Greuel, 2016). Stattdessen ergibt sich trotz vorhandener Zusammenhänge zwischen der subjektiven Sicherheit und Akkuratheit von Zeugeneinschätzungen ein nicht zu vernachlässigendes Risiko von Falschidentifizierungen, welche dennoch mit hoher subjektiver Sicherheit vertreten werden (Martschuk & Sporer, 2021; Sporer & Sauerland, 2008).

Darüber hinaus haben wissenschaftliche Studien eine Vielzahl weiterer Faktoren identifiziert, die sich auf die Güte von Zeugenaussagen auswirken können. Diese betreffen unter anderem situative Rahmenbedingungen während der Wahrnehmungsphase (Lichtverhältnisse, Beobachtungsdistanz/-dauer, Maskierung des Täters, Komplexität des Geschehens, ethnische Zugehörigkeit des Täters) sowie personengebundene Faktoren beim Zeugen (Stresslevel und Aufmerksamkeit während des Ereignisses, Alter, Intelligenz). Im Weiteren spielt das zeitliche Intervall zwischen Wahrnehmung und Gedächtnisabruf eine Rolle, ebenso wie Informationen, die während dieses Intervalls auf die Person eingewirkt haben (bspw. Aussagen von anderen Zeugen; Übersicht bei Buchner & Brandt, 2017; Greuel, 2016). Auch ist die Art und Weise der Gestaltung späterer Befragungen, insbesondere im Hinblick auf die Verwendung suggestiver Befragungsmethoden zu berücksichtigen (Milne & Bull, 2003; Volbert & Steller, 2014; Wells et al., 2006).

In Bezug auf die Beurteilung der Qualität von Zeugenaussagen sind grundsätzlich die Prozesse der *Personenbeschreibung* und *Personenidentifizierung* zu unterschei-

den. Bisherige Forschungsergebnisse verweisen auf die Beteiligung unterschiedlicher kognitiver Prozesse bei Beschreibungs- und Identifizierungsleistungen, wobei der Zusammenhang zwischen beiden Leistungsbereichen eher schwach ausfällt, d.h. hohe oder niedrige Leistungen in einem Bereich gehen nicht unbedingt mit hohen oder niedrigen Leistungen in dem jeweils anderen Bereich einher (Meissner et al., 2007).

### 1. Zielstellung der Untersuchung

Wie dargestellt, existieren bereits wissenschaftliche Erkenntnisse hinsichtlich relevanter Einflussfaktoren auf die Güte von Zeugenaussagen. Diese betreffen allerdings vorrangig Einflüsse auf die Fähigkeit zur korrekten Identifizierung von Personen, die zuvor im Rahmen eines bestimmten Settings beobachtet wurden (bspw. im Kontext einer polizeilichen Gegenüberstellung). Inwieweit bisherige Erkenntnisse auf den spezifischen Bereich von Zeugenaussagen übertragen werden können, die sich auf die Schätzung des Alters einer wahrgenommenen Person beziehen, ist hingegen unzureichend beforscht. Für den deutschsprachigen Kontext lassen sich diesbezüglich erste Erkenntnisse aus der Studie des LKA NRW (2011) ableiten, deren empirische Bestätigung/Erweiterung allerdings weiterer Studien bedarf. Gleichzeitig sind empirisch fundierte Erkenntnisse zur Frage der Güte von Altersschätzungen für die polizeiliche Aufgabenerfüllung im Bereich der Strafverfolgung und Gefahrenabwehr von großer Relevanz, u.a. um geeignete Maßnahmen zur Tatabklärung abzuleiten und Fahndungsmaßnahmen nach unbekanntem Tatverdächtigen zielgerichtet zu gestalten.

Mittels eines Experiments soll in vorliegender Studie untersucht werden, welchen Einflussfaktoren die Güte der Schätzung des Alters einer erwachsenen Person unterliegt, wenn hierfür auf die Wahrnehmung einer fotografischen Aufnahme des Gesichts zurückgegriffen wird und die Altersschätzungen der Versuchspersonen unmittelbar im Anschluss an die Gesichtswahrnehmung erfolgen. Als Maß für die Güte der Altersschätzung diente die Präzision der Altersschätzung, wobei zwischen dem Treffen und Nicht-Treffen eines Altersbereichs von +/- 5 Jahren um das tatsächliche Alter der einzuschätzenden Person unterschieden wurde. Auf Basis bisheriger Untersuchungen zur Zuverlässigkeit von Zeugenaussagen unterlagen die erwarteten Ergebnisse des Experiments den folgenden drei Hypothesen:

- (1) Versuchspersonen treffen die Altersgruppe anderer Personen mit einer höheren Wahrscheinlichkeit, wenn die Personen der Altersgruppe der Versuchsperson entsprechen (*Own-Age-Effekt*).
- (2) Versuchspersonen treffen die Altersgruppe anderer Personen mit einer höheren Wahrscheinlichkeit, wenn die Personen dem gleichen Geschlecht wie die Versuchsperson angehören (*Own-Sex-Effekt*).
- (3) Die seitens der Versuchspersonen berichtete subjektive Sicherheit der Altersschätzung hat keinen Einfluss auf die Wahrscheinlichkeit, die Altersgruppe anderer Personen zu treffen.

Über die Untersuchung dieser zentralen Hypothesen hinaus wurden noch weitere Einflussfaktoren auf die Güte der

Altersschätzung betrachtet, die sich einerseits Eigenschaften der Versuchspersonen und andererseits Eigenschaften der einzuschätzenden Personen zuordnen ließen. Für den erstgenannten Bereich wurde (unabhängig vom Alter der einzuschätzenden Person) der Einfluss der Altersgruppe der Versuchspersonen untersucht, wobei kindliche, jugendliche und Versuchspersonen ab 60 Jahre gegenüber erwachsenen Personen unter 60 Jahren kontrastiert wurden. Bisherige Untersuchungen legen nahe, dass jüngere Menschen gegenüber älteren Personen/Senioren Vorteile bei der Beschreibung und Identifikation von anderen Personen aufweisen (Martschuk & Sporer, 2021; Schwöbel, 2014; Yarmey et al., 1994), allerdings sehr junge, kindliche Versuchspersonen bei einzuschätzenden erwachsenen Personen schlechter als erwachsene Versuchspersonen abschneiden (Hills & Lewis, 2011; Lindholm, 2005)

Neben der Möglichkeit eines Own-Sex-Effektes stellte sich auf Basis früherer Untersuchungen (Herlitz & Lovén, 2013; Megreya et al., 2011; Sporer & Sauerland, 2008) die Frage, ob Frauen gegebenenfalls generell Vorteile in der Wahrnehmung und Verarbeitung von Gesichtern aufweisen, welche sich auch auf die Güte der Altersschätzung auswirken. Zudem sollte der Einfluss eines polizeilichen beruflichen Hintergrunds auf die Altersschätzung überprüft werden. Diesbezüglich existieren bereits Untersuchungen, wobei sich der polizeiliche berufliche Hintergrund im Hinblick auf unterschiedliche Gütekriterien von Zeugenaussagen sowohl vor- als auch nachteilhaft auswirken kann (Kaminski & Sporer, 2016), spezifische Untersuchungen zur Güte der Altersschätzung, auf die das vorliegende Experiment abzielt, allerdings nicht vorliegen.

Im Weiteren standen Faktoren der äußerlichen Erscheinung der hinsichtlich des Alters einzuschätzenden Personen im Blickpunkt der Datenanalysen. Diese betrafen das Tragen von Make-Up, eines Barts, eine graue Haarfarbe sowie das Vorliegen einer Glatze/Halbglatte. Diese Faktoren sind bis dato im Hinblick auf die Güte von Altersschätzungen ebenfalls unzureichend beforscht, wengleich deren Einfluss auf die Güte von Altersschätzungen naheliegt (vgl. Guski, 1996).

## II. Methode

### 1. Stichprobe

Die Versuchspersonen wurden seitens des Erstautors und einer weiteren Person akquiriert, welche gleichzeitig als Versuchsleiter für die Durchführung des Experiments zuständig waren. Die Kontaktaufnahme zu den Versuchspersonen erfolgte über das berufliche und persönliche Umfeld. Insgesamt nahmen an dem Experiment 163 Versuchspersonen teil. Das Alter variierte zwischen 7 und 84 Jahren, bei einem durchschnittlichen Alter von 41,7 Jahren (Median = 42 Jahre, Sd = 18,2 Jahre). Die Versuchspersonen waren mit einem Anteil von 53 % weiblichen und von 47 % männlichen Geschlechts. Eine nähere Analyse der Altersverteilung verwies auf einen Anteil von 9 % an kindlichen Versuchspersonen, einen Anteil von 2 % an jugendlichen Versuchspersonen, einen Anteil von 70 % an erwachsenen Personen zwischen 18 und 59 Jahren und einen Anteil von 19 % an Personen ab 60 Jahren. 13 % der Versuchspersonen waren beruflich bei der Polizei tätig.

### 2. Versuchsmaterialien und Versuchsdurchführung

Vor der eigentlichen Durchführung des Experiments wurden die Versuchspersonen über die übergeordnete Zielstellung und den Ablauf des Experiments informiert. Es erfolgte eine Aufklärung über die Anonymität der Datenerhebung und -weiterverarbeitung sowie die Freiwilligkeit der Teilnahme an dem Experiment. Anschließend wurde das Prozedere der Versuchsdurchführung und die Aufgabenstellung an die Versuchspersonen nochmals präzisiert.

Im Rahmen der Versuchsdurchführung wurden den Versuchspersonen in der Form einer PowerPoint-Präsentation jeweils 30 Farblichtbilder von Personen (Stimulibilder mit Stimulipersonen) präsentiert. Die Stimulibilder wurde mittels der Festlegung automatischer Übergangszeiten für jeweils fünf Sekunden präsentiert. Anschließend sollten die Versuchspersonen eine Schätzung des Alters der Stimuliperson abgeben und zusätzlich ihren Grad an Sicherheit hinsichtlich dieser Altersschätzung mitteilen, auf einer Skala von 1 (»absolut unsicher«) bis 6 (»absolut sicher«). Darüber hinaus sollten die Versuchspersonen darlegen, ob Sie mit der Stimuliperson verwandt sind, oder Ihnen diese Person persönlich näher bekannt, flüchtig bekannt oder unbekannt ist, wobei diese Bekanntheitskategorien im Vorfeld näher definiert wurden. Lediglich in den Kategorien »verwandt« und »persönlich näher bekannt« durften der Versuchsperson nähere Informationen zum Alter Stimuliperson vorliegen. Für die späteren Analysen wurden Versuchsperson-Stimuliperson-Paare mit solchen Bekanntschaftsverhältnissen ausgeschlossen. Die Antworten der Versuchspersonen wurden auf einem Antwortbogen protokolliert. Anschließend erfolgte ein manueller Wechsel zum nächsten Stimulibild. Vor der Präsentation der Stimulibilder zur Gewinnung der Analysedaten wurde das Versuchsprozedere mit drei Beispielbildern eingeübt. Als Motivationselement für die Versuchspersonen, möglichst valide Alterseinschätzungen zu tätigen, diente ein zuvor in Aussicht gestellter Abgleich der Altersschätzungen mit dem tatsächlichen Alter der Stimulipersonen im Nachgang an das eigentliche Experiment. Abbildung 1 illustriert den Ablauf der Untersuchung in schematischer Form.

Die Versuchsdurchführung fand überwiegend im Einzelsetting statt, in welchem der Versuchsleiter der Versuchs-



Abbildung 1. Schematische Veranschaulichung des Versuchsablaufs.

person die Stimulibilder auf einem Laptop präsentierte und die Antworten auf dem Antwortbogen protokollierte. In rund einem Viertel der Fälle erfolgte die Versuchsdurchführung im Gruppensetting (Unterrichtsraum) unter Nutzung eines Beamers. Die Stimulipersonen wurden unter standardisierten Bedingungen abgelichtet und zeigten jeweils ihr Gesicht in Frontalansicht. Die Oberbekleidung der Stimulipersonen wurde für die Aufnahmen abgedeckt und die Personen sollten sich hinsichtlich ihrer Mimik möglichst neutral präsentieren. Die äußeren Rahmenbedingungen der Aufnahmen sollten somit für alle Fotos möglichst konstant gehalten werden.

Eine gezielte Variation der Stimulibilder erfolgte den Untersuchungsfragen entsprechend in Bezug auf das Alter und Geschlecht der Stimulipersonen. Hierfür wurden fünf verschiedene Altersgruppen gebildet (20–29 Jahre, 30–39 Jahre, 40–49 Jahre, 50–59 Jahre und 60–78 Jahre). Für jede Altersgruppe wurden jeweils drei männliche und drei weibliche Stimulipersonen gewonnen. Die Stimulipersonen unterschieden sich außerhalb des Alters und Geschlechts naturgemäß durch weitere äußerliche Merkmale, die durch das experimentelle Design keiner gezielten Beeinflussung unterlagen, allerdings als mögliche Einflussfaktoren dokumentiert und in die Datenanalysen einbezogen wurden. Diese Merkmale betrafen u.a. eine graue Haarfarbe, das Tragen eines Barts sowie das augenscheinliche Tragen von Make-Up (Geschminkt-Sein). Um möglichen Effekten der Reihenfolge der Stimulibilder zu begegnen, wurden verschiedene Versionen der PowerPoint-Präsentation erstellt und eingesetzt, die sich jeweils durch eine zufällige Reihenfolge der Stimulibilder unterschieden.

**3. Untersuchungsvariablen und Datenanalyse**

Die abhängige Variable der Untersuchung stellte das Treffen der Altersgruppe der einzuschätzenden Stimuliperson dar, wobei die Variable bei einer Differenz zwischen Real- und Schätzalter im Bereich zwischen +/- 5 Jahren mit Ja und in den anderen Fällen mit Nein kodiert wurde. Als Prädiktoren zur Vorhersage der Wahrscheinlichkeit des Treffens der tatsächlichen Altersgruppe (= Trefferwahrscheinlichkeit) wurden mehrere Merkmale untersucht, die sich auf einer übergeordneten Ebene Merkmalen zur unmittelbaren Prüfung der oben genannten Hypothese 1 bis 3 und weiteren Merkmalen der Versuchsperson beziehungsweise Stimulipersonen zuordnen ließen (siehe Tabelle 1).

Für die statistische Datenanalyse kam eine multiple logistische Regressionsanalyse mit sogenannten gemischten Effekten zum Einsatz (siehe Brown, 2021; Eid et al., 2017). Solche Modelle sind insbesondere für die Auswertung von Experimenten geeignet, in denen jeweils die gleiche Versuchsperson mehrere Versuchsdurchgänge durchläuft und dadurch mehrere einzelne Messungen zu einer Versuchsperson vorliegen, wobei die abhängige Variable lediglich zwei Ausprägungen aufweist (in unserem Fall Ja vs. Nein). Gemischte Modelle ermöglichen die Berücksichtigung sogenannter fester Effekte, die den durchschnittlichen Effekt bestimmter Merkmale auf die abhängige Variable in einem experimentellen Setting repräsentieren. In unserem Fall werden die Prädiktoren in Tabelle 1 allesamt als feste Effekte modelliert. Unterschiede zwischen den individuellen Versuchspersonen außerhalb bereits berücksichtigter

<b>Abhängige Variable</b>
Treffen der Altersgruppe der Stimuliperson (Ja = Differenz zwischen Real- und Schätzalter im Bereich zwischen +/- 5 Jahre vs. Nein = absolute Differenz zwischen Real- und Schätzalter größer als 5 Jahre)
<b>Prädiktoren anhand der Untersuchungshypothesen 1 bis 3</b>
Gleiche Altersgruppe von Versuchsperson und Stimuliperson (Nein vs. Ja = Altersdifferenz max. 5 Jahre)
Gleiches Geschlecht von Versuchsperson und Stimuliperson (Nein vs. Ja)
Subjektive Sicherheit der Altersschätzung (von 1 = absolut unsicher bis 6 = absolut sicher)
<b>Weitere Prädiktoren der Versuchspersonen</b>
Kindliche Versuchsperson (Nein vs. Ja = unter 14 Jahre alt)
Jugendliche Versuchsperson (Nein vs. Ja = zwischen 14 und 17 Jahre alt)
Versuchsperson >=60 Jahre (Nein vs. Ja)
Weibliches Geschlecht der Versuchsperson (Nein vs. Ja)
Polizeilicher beruflicher Hintergrund (Nein vs. Ja)
<b>Weitere Prädiktoren der Stimulipersonen</b>
Tragen von Make-Up (Nein vs. Ja)
Tragen eines Barts (Nein vs. Ja)
Graue Haarfarbe (Nein vs. Ja)
Glatze/Halbglatze (Nein vs. Ja)

Tabelle 1. Abhängige Variable und berücksichtigte Prädiktoren für die Analysen

Merkmale der festen Effekte, die sich auf ein zufällig variierendes Grundniveau der Fähigkeit zur Altersschätzung über die Versuchsdurchgänge hinweg beziehen, werden für die Analysen hingegen als zufällige Effekte modelliert. Die konkrete Ausprägung dieses individuellen Grundniveaus ist für die vorliegende Untersuchungsfragestellung von untergeordneter Relevanz. Dennoch muss dieses individuelle Grundniveau für die statistischen Analysen berücksichtigt werden, um den statistischen Einfluss der interessierenden festen Effekte akkurat schätzen zu können (Brown, 2021). Für die Durchführung der Analysen wurde auf die Funktionsbibliothek »lme4« (Bates et al., 2023) innerhalb der statistischen Programmiersprache R (R Core Team, 2022) zurückgegriffen.

Der multiple Regressionsansatz ermöglicht die Ermittlung der Vorhersagekraft eines Prädiktors unter bereits vorhandener Berücksichtigung weiterer Prädiktoren (sogenannte inkrementelle Vorhersagekraft). Dadurch können Vorhersagemodelle bestimmt werden, die dem Sparsamkeitsprinzip entsprechend mit möglichst wenigen Prädiktoren eine möglichst gute Vorhersage leisten. Damit geht auch die Möglichkeit der Identifikation sogenannter Scheinkorrelationen einher, d.h. die Identifikation von statistischen, aber nicht ursächlichen Zusammenhängen zwischen zwei Merkmalen/Variablen, welche durch einen (zufälligen) Zusammenhang der beiden Variablen mit einer dritten Variable zustande kommen.

Die Überprüfung der Vorhersagekraft der verschiedenen Prädiktoren erfolgte im Rahmen einer schrittweisen Entwicklung eines möglichst optimalen Vorhersagemodells zur Vorhersage der Trefferwahrscheinlichkeit in verschiedenen

Modellentwicklungsstufen. Im Rahmen einer ersten Modellentwicklungsstufe sollte zunächst überprüft werden, ob die Prädiktoren anhand der Untersuchungshypothesen 1 bis 3 eine statistisch signifikante Verbesserung der Vorhersage der Trefferwahrscheinlichkeit leisten. Die zweite und dritte Modellentwicklungsstufe diente der Überprüfung des etwaigen Einflusses von weiteren Merkmalen der Versuchspersonen beziehungsweise der Stimulipersonen auf die Vorhersage der Trefferwahrscheinlichkeit. Innerhalb der Entwicklungsstufen entsprach die Prüfreihenfolge der Variablen der in Tabelle 1 dargestellten Reihenfolge.

Im Falle einer statistisch signifikanten Vorhersageleistung wurde der jeweilige Prädiktor in das Vorhersagemodell aufgenommen. Die Überprüfung eines Prädiktors in nachfolgenden Prüfschritten erfolgte unter Berücksichtigung aller bereits in das Modell aufgenommenen Prädiktoren. Es stellte sich insofern die Frage, ob der jeweilige Prädiktor über bereits berücksichtigte Prädiktoren hinaus noch eine statistisch signifikante Verbesserung der Vorhersageleistung zu leisten vermochte. Zur Ermittlung der statistischen Signifikanz wurden Likelihood-Ratio-Tests berechnet, wobei als Signifikanzniveau (=akzeptierte Irrtumswahrscheinlichkeit) für die statistischen Tests  $p < 5\%$  festgelegt wurde (siehe Brown, 2021; Eid et al., 2017).

Für das resultierende Gesamtmodell nach der dritten Modellentwicklungsstufe werden zur Abschätzung stichprobenbedingter Zufallsschwankungen zusätzlich 95 %-Konfidenzintervalle zu den Effektkoeffizienten berichtet. Darüber hinaus wurden die VIF-Koeffizienten zur Feststellung einer etwaigen problematischen Multikollinearität zwischen den festen Effekten berechnet (Grenzwert  $< 10$ ; Eid et al., 2017).

### III. Ergebnisse

Zu den 163 Versuchspersonen lagen in Verbindung mit den 30 Stimulibildern  $163 \cdot 30 = 4890$  Versuchsperson-Stimuliperson-Paare mit Einschätzungen zum Alter der Stimuliperson, der subjektiven Sicherheit bei der Altersschätzung sowie dem Bekanntheitsgrad bezüglich der Stimuliperson vor. In 103 Versuchsperson-Stimuliperson-Paaren lag ein näheres Bekanntschaftsverhältnis vor, welches die Altersschätzung gegebenenfalls beeinflusst hätte. Diese Paare wurden aus den weiteren Datenanalysen ausgeschlossen. Eine Prüfung des Datensatzes auf auffällige Werte identifizierte zudem ein Versuchsperson-Stimuliperson-Paar, bei welchem anstatt des Schätzalters vermutlich die Stimulipersonnummer auf dem Antwortbogen eingetragen wurde. Dieses Paar wurde ebenfalls aus den Analysen ausgeschlossen, so dass insgesamt eine Stichprobengröße mit Messwerten zu 4786 Versuchsperson-Stimuliperson-Paaren resultierte. Die Verteilungscharakteristika der abhängigen Variablen und berücksichtigten Prädiktoren ist Tabelle 2 zu entnehmen.

Im Rahmen der ersten Modellentwicklungsstufe ergab sich ein statistisch signifikanter Vorhersageeffekt des Prädiktors »Gleiche Altersgruppe von Versuchsperson und Stimuliperson« ( $LR-\chi^2 = 7,46$ ,  $df = 1$ ,  $p = 0,006$ ) für die Vorhersage der Wahrscheinlichkeit des Treffens der tatsächlichen Altersgruppe der Stimuliperson. Der Prädiktor »Gleiches Geschlecht von Versuchsperson und Stimuliperson« trug hingegen nicht zu einer statistisch signifikanten Verbesserung des Vorhersagemodells bei ( $LR-\chi^2 = 2,11$ ,  $df = 1$ ,  $p = 0,146$ ). Die »Sub-

	Anteil (Nein)	Anteil (Ja)
Treffen der Altersgruppe der Stimuliperson	44,4 %	55,6 %
Gleiche Altersgruppe von Versuchsperson und Stimuliperson	82,4 %	17,6 %
Gleiches Geschlecht von Versuchsperson und Stimuliperson	50,0 %	50,0 %
Kindliche Versuchsperson	91,5 %	8,5 %
Jugendliche Versuchsperson	97,5 %	2,5 %
Versuchsperson $\geq 60$ Jahre	80,8 %	19,2 %
Weibliches Geschlecht der Versuchsperson	46,9 %	53,1 %
Polizeilicher beruflicher Hintergrund	87,1 %	12,9 %
Tragen von Make-Up	93,5 %	6,5 %
Tragen eines Barts	76,5 %	23,5 %
Graue Haarfarbe	86,5 %	13,5 %
Glatze/Halbglatze	96,6 %	3,4 %
<b>Metrische Kennzahlen</b>		
Subjektive Sicherheit der Altersschätzung	Mittelwert = 3,3, Median = 3, Min = 1, Max = 6, SD = 1,00	

Tabelle 2. Verteilungscharakteristika der abhängigen Variablen und der berücksichtigten Prädiktorvariablen.

Hinweise: Die Kennzahlen beziehen sich auf eine Stichprobengröße von  $n = 4786$  Versuchsperson-Stimuliperson-Paaren, mit 163 unterschiedlichen Versuchspersonen und 30 unterschiedlichen Stimulibildern.

jektive Sicherheit der Altersschätzung« wies wiederum einen statistischen signifikanten Vorhersageeffekt auf ( $LR-\chi^2 = 5,46$ ,  $df = 1$ ,  $p = 0,019$ ).

Unter Berücksichtigung der beiden Prädiktoren mit einer statistisch signifikanten Vorhersageleistung aus der ersten Modellentwicklungsstufe führte in Bezug auf die untersuchten Prädiktoren der Versuchspersonen (zweite Modellentwicklungsstufe) die Hinzunahme des Prädiktors »Kindliche Versuchsperson« zu einer weiteren statistisch signifikanten Verbesserung der Vorhersageleistung des Modells ( $LR-\chi^2 = 29,71$ ,  $df = 1$ ,  $p < 0,001$ ). Die galt jedoch nicht für die Prädiktoren »Jugendliche Versuchsperson« ( $LR-\chi^2 = 0,64$ ,  $df = 1$ ,  $p = 0,424$ ), »Versuchsperson  $\geq 60$  Jahre« ( $LR-\chi^2 = 3,67$ ,  $df = 1$ ,  $p = 0,055$ ), »Weibliches Geschlecht der Versuchsperson« ( $LR-\chi^2 = 0,01$ ,  $df = 1$ ,  $p = 0,909$ ) und »Polizeilicher beruflicher Hintergrund« ( $LR-\chi^2 = 3,74$ ,  $df = 1$ ,  $p = 0,053$ ).

Im Rahmen der dritten Modellentwicklungsstufe in Bezug auf bestimmte Merkmale der Stimulipersonen erbrachte der Prädiktor »Tragen von Make-Up« eine statistisch signifikante Verbesserung der Vorhersageleistung des Modells ( $LR-\chi^2 = 84,05$ ,  $df = 1$ ,  $p < 0,001$ ), ebenso wie nachfolgend der Prädiktor »Tragen eines Barts« ( $LR-\chi^2 = 14,90$ ,  $df = 1$ ,  $p < 0,001$ ). Keine statistisch signifikante Verbesserung der Vorhersageleistung resultierte durch die Prädiktoren »Graue Haarfarbe« ( $LR-\chi^2 = 0,72$ ,  $df = 1$ ,  $p = 0,396$ ) und »Glatze/Halbglatze« ( $LR-\chi^2 < 0,00$ ,  $df = 1$ ,  $p = 0,988$ ).

Das aus den Analysen resultierende Gesamtmodell zur Vorhersage der Wahrscheinlichkeit des Treffens der tatsächlichen Altersgruppe der Stimuliperson wird in Tabelle 3 abgebildet,

Prädiktor	Beta	95 %-KI		p-Wert (LR-Test)
		Untere Grenze	Obere Grenze	
(Achsenabschnitt)	0,006	-0,208	0,219	
Gleiche Altersgr. von Versuchsp. und Stimulip.	0,221	0,064	0,379	0,006 **
Subjektive Sicherheit der Altersschätzung	0,073	0,013	0,134	0,018 *
Kindliche Versuchsperson	-0,659	-0,892	-0,429	< 0,001 ***
Tragen von Make-Up	-1,052	-1,308	-0,803	< 0,001 ***
Tragen eines Barts	0,274	0,135	0,414	< 0,001 ***

Tabelle 3. Kennwerte des resultierenden Vorhersagemodells zur Vorhersage der Wahrscheinlichkeit des Treffens der tatsächlichen Altersgruppe der Stimuliperson.  
Hinweise: Der p-Wert bezieht sich auf einen LR-Test unter Vergleich eines Modells mit und ohne den jeweiligen Prädiktor (insofern sind die p-Werte bis auf den letzten Prädiktor nicht identisch mit den berichteten p-Werten im Kontext des schrittweisen Aufbaus des Vorhersagemodells). Positive Beta-Koeffizienten stehen für eine Erhöhung der Trefferwahrscheinlichkeit durch das Vorliegen bzw. eine höhere Ausprägung des Prädiktors, negative Beta-Koeffizienten für eine Verringerung der Trefferwahrscheinlichkeit durch das Vorliegen des Prädiktors. Eine Analyse der VIF-Koeffizienten der Prädiktoren verwies auf keine problematische Multikollinearität (alle VIF-Werte deutlich unter dem Grenzwert von 10). \*\*\*p < 0,001, \*\*p < 0,01, \*p < 0,05.

wobei ausschließlich die festen Effekte des Modells berichtet werden. Die ermittelten Beta-Effektkoeffizienten lassen sich wie folgt interpretieren. Für die metrische Variable »Subjektive Sicherheit« der Altersschätzung signalisiert der positive Effektkoeffizient ein Ansteigen der Trefferwahrscheinlichkeit mit zunehmender Sicherheit bei der Altersschätzung. Die Effektkoeffizienten für die anderen, dichotomen Merkmale mit Ja- vs. Nein-Ausprägung stehen bei positivem Vorzeichen für ein Ansteigen, bei negativem Vorzeichen für ein Abfallen der Trefferwahrscheinlichkeit bei Vorliegen der Ja-Ausprägung. Die absolute Höhe des Effektkoeffizienten weist dabei eine Beziehung zur Höhe des Ansteigens/Abfallens der Trefferwahrscheinlichkeit durch das jeweilige Merkmal auf, wobei diese Beziehung aufgrund des gewählten logistischen Regressionsansatzes allerdings nicht linear ausfällt (siehe Eid et al., 2017).

Vernachlässigt man aus Gründen der Einfachheit die Höhe der Effektkoeffizienten und fokussiert auf das Vorzeichen der Effektkoeffizienten, so wird Folgendes deutlich: Die Trefferwahrscheinlichkeit wird durch die Zugehörigkeit der Versuchsperson zur Altersgruppe der Stimuliperson (Own-Age-Effekt), eine subjektiv höhere Sicherheit bei der Altersschätzung und das Tragen eines Barts seitens der Stimuliperson erhöht. Demgegenüber reduziert sich die Trefferwahrscheinlichkeit bei kindlichen Versuchspersonen und Stimulipersonen, die Make-Up tragen.

#### IV. Diskussion und Schlussfolgerungen

##### 1. Zentrale Ergebnisse der Studie

Zur Beantwortung der forschungsinitiierenden Fragestellungen »Wie gut können Menschen das Alter von fremden Personen einschätzen?« und »Von welchen Faktoren ist die Güte der Altersschätzung abhängig?« wurde ein Experiment durchgeführt, um Einflussfaktoren auf die Güte der Altersschätzung näher zu bestimmen. Hierfür schätzten Versuchspersonen das Alter einer Serie verschiedener Personen, deren Gesicht in Frontalansicht präsentiert wurde. Das Experiment knüpfte an bereits bestehende Erkenntnisse im Bereich der Personenidentifizierung und Personenbeschreibung an, insbesondere den sogenannten *Own-Age-Effekt* und *Own-Sex-Effekt*. Bisherige Forschungs- und Studienergebnisse verwiesen diesbezüglich auf eine heterogene, teils widersprüchliche Befundlage beziehungsweise wurden entsprechende Effekte vorrangig im Bereich der Personenidentifizierung untersucht. Gerade im Bereich der Personenbeschreibung resultierte insofern eine Forschungslücke, welcher die vorliegende Unter-

suchung begegnet. Anhand der Ergebnisse ergeben sich folgende zentrale Feststellungen.

Im Einklang mit der im Vorfeld des Experiments aufgestellten Hypothese trafen die Versuchspersonen die Altersgruppe der Stimulipersonen (Altersabstand max. 5 Jahre) mit einer höheren Wahrscheinlichkeit, wenn Versuchs- und Stimulipersonen derselben Altersgruppe entstammen. Der Own-Age-Effekt erfuhr insofern auch für die innerhalb des Experiments abgeforderte Personenbeschreibung eine Bestätigung, wobei bereits die erwähnte Studie des LKA NRW diesen Effekt mit Bezug zu Personenbeschreibungen feststellen konnte. In dieser Studie wurden Alters- und Größeneinschätzungen zu Bankräubern mit den im Nachgang erkennungsdienstlich erhobenen tatsächlichen Kennwerten der Täter verglichen (Landeskriminalamt Nordrhein-Westfalen, 2011).

Ebenso in Übereinstimmung mit der zitierten Studie des LKA NRW, allerdings entgegen der eigenen Forschungshypothese, zeigte sich kein Einfluss eines Geschlechtsunterschiedes zwischen Versuchsperson und Stimuliperson (Own-Sex-Effekt). Somit deuten die bisherigen Forschungsergebnisse insgesamt an, dass der Own-Sex-Effekt vorrangig für Personenidentifizierungen und weniger für Personenbeschreibungen eine Rolle zu spielen scheint. Weiterhin konnte ein statistisch signifikanter Einfluss des Grades der subjektiven Sicherheit auf die Güte der Altersschätzung ermittelt werden, wobei unsere Forschungshypothese einen solchen Einfluss auf Basis vorangegangener Untersuchungen nicht annahm. Personen, die sich bei ihren Einschätzungen subjektiv sicherer waren, wiesen im Durchschnitt tatsächlich eine höhere Trefferwahrscheinlichkeit hinsichtlich der Altersgruppe der Stimuliperson auf. Die Heterogenität bisheriger Befunde zur Relevanz der subjektiven Sicherheit bei Augenzeugen könnte gegebenenfalls durch bis dato noch nicht ausreichend explorierte Drittvariablen moderiert werden (vgl. Kaminski & Sporer, 2016; Martschuk & Sporer, 2021), was in zukünftigen Untersuchungen näher betrachtet werden sollte.

Über die genannten Effekte hinaus verwiesen die Untersuchungsergebnisse auf bedeutsame Einflüsse von weiteren Merkmalen der Stimulipersonen (Bart, Graue Haare und das Tragen von Makeup) auf die Güte der Altersschätzung. Bereits Wells (1985) stellte in seiner Studie eine starke Abhängigkeit der Qualität der Personenbeschreibung sowie Personenidentifizierung von den Merkmalen des Stimuligesichtes fest, insbesondere zeigte sich ein Vorteil von markanten Gesichtern. Im Hinblick auf die Erweiterung des bisherigen

Forschungsstands im Bereich der Personenbeschreibung durch vorliegende Untersuchung erscheint vor allem bedeutsam, dass die identifizierten Einflussfaktoren bei den Stimulipersonen veränderbar sind und somit auch einer gezielten Manipulation durch Täter unterliegen können.

Im Unterschied zu bisherigen Untersuchungen (vgl. den Abschnitt Einleitung), die den Fokus allerdings auf Identifizierungsleistungen beziehungsweise andere Wahrnehmungs-/Erinnerungsleistungen außerhalb von Altersschätzungen gelegt haben, ergab sich keine allgemeine Überlegenheit des weiblichen Geschlechts bei der Akkuratheit von Altersschätzungen, ebenso wenig wie pauschal schlechtere Altersschätzungen durch Senioren. Im Weiteren fielen die Einschätzungen von Zeugen mit einem polizeilichen beruflichen Hintergrund nicht statistisch signifikant besser oder schlechter aus als bei Versuchspersonen ohne diesen Hintergrund.

## 2. Schlussfolgerungen für die polizeiliche Praxis

Die polizeiliche Forschung auf dem Gebiet der Altersschätzung fällt limitiert aus, gerade im deutschsprachigen Raum. Diesbezüglich belastbare empirische Erkenntnisse rekurrieren bis dato vor allem auf die Untersuchung des LKA NRW zu Alters- und Größenschätzungen im speziellen Kontext der Beschreibung von Bankräubern. Insgesamt konzentrierte sich die bisherige Forschung vor allem auf das Gebiet der Personenidentifizierung. Diese hat im Strafverfahren sicherlich erhebliche Relevanz, weil sie unmittelbar zur Identifizierung des Täters sowie dessen Verurteilung führen kann und sich gleichzeitig eine hohe Brisanz bei fehlerhaften Personenidentifizierungen ergibt.

Personenbeschreibungsvorgänge haben allerdings ebenfalls eine hohe Bedeutung bei der Tataufklärung. Im Unterschied zu Identifizierungsvorgängen wird die Konsequenz von fehlerhaften Beschreibungen gegebenenfalls jedoch schwerer sichtbar. Wird auf Basis der Personenbeschreibung kein Tatverdächtiger ermittelt, weiß man im Anschluss nicht, ob dies durch fehlerhafte Beschreibungsmerkmale oder anderweitige Ursachen bedingt war. Wird hingegen ein Tatverdächtiger ermittelt und der Tat überführt, lässt der Gesamterfolg etwaige fehlerhafte Beschreibungen des Täters in den Hintergrund rücken.

Es ist davon auszugehen, dass viele polizeiliche und justizielle Entscheidungen hinsichtlich der Validität von Altersschätzungen aufgrund eines auf Erfahrungswissen beruhenden intuitiven »Bauchgefühls« getätigt werden. Wissenschaftliche Untersuchungen verweisen allerdings eindeutig auf die Fehleranfälligkeit und mangelnde Validität solcher Einschätzungen (Egisdóttir et al., 2006; Eid et al., 2017). Für eine überzeugende, objektiv nachvollziehbare polizeiliche Arbeit erscheint daher essentiell, auf wissenschaftlichen Methoden beruhende Indikatoren für die Beurteilung der Zuverlässigkeit von Altersschätzungen durch Zeugen zu nutzen. Derartige Indikatoren konnten in vorliegendem empirischem Experiment identifiziert werden, indem der Einfluss von Stimuli- und Personenmerkmalen auf die Wahrscheinlichkeit korrekter Altersschätzungen bestimmt wurde. Die gewonnenen Erkenntnisse können polizeilichen Ermittlerinnen und Ermittlern in praktischen Anwendungsfällen wertvolle Anknüpfungspunkte liefern, um wissenschaftlich begründet Rückschlüsse auf die Zuverlässigkeit einer Altersschätzung zu

ziehen. Gleichwohl sind in diesem Kontext auch die Limitationen dieser Studie und die Notwendigkeit weiterer Untersuchungen zu berücksichtigen, auf die im Folgeabschnitt eingegangen werden soll.

## 3. Limitationen der Untersuchung und Notwendigkeit weiterer Untersuchungen

Bei der verwendeten Stichprobe für die Untersuchung handelt es sich nicht um eine bevölkerungsrepräsentative Stichprobe, sondern um eine sogenannte Gelegenheitsstichprobe (vgl. unter dem Abschnitt Methode, siehe Döring & Bortz, 2016). Abweichungen der Charakteristika dieser Stichprobe vom Bevölkerungsdurchschnitt sind allerdings nicht pauschal als problematisch anzusehen. Stattdessen sind entsprechende Abweichungen vor allem dann mit Problemen behaftet, falls diese zusätzlich in einem systematischen Zusammenhang mit der abhängigen Variable der Altersschätzung stehen und keiner Kontrolle innerhalb der statistischen Modellierungen unterliegen (Eid et al., 2017). Auf den ersten Blick könnte diesbezüglich vor allem das eher überdurchschnittliche Bildungsniveau der erwachsenen Personen in dieser Stichprobe zu Verzerrungen der erzielten Stichprobenergebnisse führen. Im Rahmen einer gesonderten Überprüfung wiesen die (beschränkt) vorhandenen Variationen des Bildungsgrades in der Stichprobe allerdings keinen statistisch signifikanten Einfluss auf die Güte der Altersschätzung auf. Aufgrund der Stichprobenzusammensetzung bleibt dennoch ungewiss, inwieweit die erzielten Ergebnisse auf niedrigere Bildungsgrade/Personen mit Lerneinschränkungen übertragbar sind.

Im Rahmen der experimentellen Untersuchung wurden die Stimulipersonen in der Form standardisierter fotografischer Porträtaufnahmen präsentiert. Durch die hohe Standardisierung der Stimulibilder lag der Fokus des Untersuchungsdesigns darauf, Variationen der Altersgüteschätzung möglichst eindeutig auf die gezielten Variationen des Stimulimerkmale zurückführen zu können (und damit auf einer *hohen internen Validität*, siehe Eid et al., 2017). Gleichzeitig resultieren hieraus naturgemäß Einschränkungen der Übertragbarkeit der erzielten Erkenntnisse auf praktische Konstellationen der polizeilichen Ermittlungstätigkeit (im Sinne der sogenannten *externen Validität*, ebd.). So beobachten Zeugen andere (tatverdächtige) Personen in der Regel aus einer räumlichen Perspektive und gegebenenfalls zusätzlich in Bewegung. Darüber hinaus achten die Zeugen in der Wahrnehmungssituation mitunter nicht speziell auf das Alter der anderen Personen und bei zeitlich verzögerten Sachverhaltsschilderungen sind neben wahrnehmungsbezogenen Faktoren auch gedächtnispsychologische Faktoren zu berücksichtigen (vgl. Greuel, 2016). Ein anderer Untersuchungsansatz könnte daher im Rahmen tatsächlicher Ermittlungsverfahren gewonnene Altersschätzungen mit später objektiv ermittelten Altersangaben vergleichen und im Hinblick auf mögliche Einflussfaktoren hinsichtlich der Güte der Altersschätzung untersuchen.

Hinsichtlich des Zusammenhangs zwischen der subjektiven Sicherheit und der Güte von Altersschätzungen verweisen andere Studien mit Bezug zur Personenidentifizierung auf die Bedeutung von Drittvariablen, die diesen Zusammenhang moderieren (bspw. das Alter der Versuchspersonen oder die Zugehörigkeit zur Polizei, vgl. Kaminski & Sporer, 2016; Martschuk & Sporer, 2021). Derartige Interaktionseffekte wurden in vorliegender Untersuchung nicht untersucht und

sollten in zukünftigen Untersuchungen adressiert werden. Darüber hinaus verwiesen vorliegende Studienergebnisse darauf, dass kindliche Zeugen als Versuchspersonen im Durchschnitt niedrigere Trefferwahrscheinlichkeiten bezüglich der tatsächlichen Altersgruppe der Stimulipersonen aufwiesen. Das Fehlen von Bildern mit kindlichen Stimulipersonen verhinderte allerdings eine direkte Überprüfung, ob Kinder im Sinne des Own-Age-Effektes Stimulipersonen der eigenen Altersgruppe besser als erwachsene Stimulipersonen wahrnehmen können.

Das entwickelte statistische Vorhersagemodell zeigt die Relevanz von Merkmalen der Stimulipersonen auf die Güte der Altersschätzung auf (Tragen von Make-Up, Tragen eines Barts). Bei der Interpretation dieser Einflüsse gilt allerdings zu berücksichtigen, dass Variationen dieser Merkmale an verschiedene Personen gekoppelt waren und keiner systematischen Kontrolle innerhalb des Experiments unterlagen (d.h. die Versuchspersonen wurden diesbezüglich nicht beeinflusst). Insofern waren diese Variationen gegebenenfalls an weitere, innerhalb der statistischen Modelle nicht berücksichtigte personenspezifische Merkmale gekoppelt, die sich als unkontrollierte Störvariablen auf die Güte der Altersschätzung auswirkten. Ein alternativer Untersuchungsansatz könnte insofern darin bestehen, dass dieselben Stimulipersonen mehrmals mit unterschiedlichen Merkmalsausprägungen präsentiert werden (ggf. auch mittels digitaler Nachbearbeitung der Fotos), in diesem Fall aber sinnvollerweise nicht mehrfach den gleichen, sondern unterschiedlichen Versuchspersonen.

Die vorliegende Untersuchung stellt also lediglich einen weiteren Mosaikstein dar, um wissenschaftlich validierte Einflussfaktoren auf die Güte von Altersschätzungen durch Zeugen zu ermitteln. Weitere Studien sind notwendig, damit sich die moderne polizeiliche Ermittlungsarbeit zukünftig verstärkt auf eine empirisch belastbare Faktenlage und nicht allein subjektive Theorien der Ermittlerinnen und Ermittler stützen kann.

## V. Literatur

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist, 34*(3), 341–382. <https://doi.org/10.1177/0011000005285875>

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12*(6), 1043–1047. <https://doi.org/10.3758/BF03206441>

Bates, D., Maechler, M., Bolker, B., cre, Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2023). *lme4: Linear Mixed-Effects Models using »Eigen« and S4* (1.1–32). <https://CRAN.R-project.org/package=lme4>

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science, 4*(1), 1–19. <https://doi.org/10.1177/2515245920960351>

Buchner, A., & Brandt, M. (2017). Gedächtniskonzeptionen und Wissensrepräsentationen. In J. Müsseler & M. Rieger (Hrsg.), *Allgemeine Psychologie (3. Aufl.)* (S. 401–434). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-53898-8\\_12](https://doi.org/10.1007/978-3-642-53898-8_12)

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41089-5>

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Beltz.

Ellis, H., Shepherd, J., & Bruce, A. (1973). The effects of age and sex upon adolescents' recognition of faces. *The Journal of genetic psychology, 123*(1st Half), 173–174. <https://doi.org/10.1080/00221325.1973.10533202>

Greuel, L. (2016). Rechts- und Aussagepsychologie. In T. Porsch & B. Werdes (Hrsg.), *Polizei-psychologie: Ein Lehrbuch für das Bachelorstudium Polizei* (1. Auflage, S. 291–318). Hogrefe.

Guski, R. (1996). *Wahrnehmen—Ein Lehrbuch* (DM 54.-). Kohlhammer.

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition, 21*(9–10), 1306–1336. <https://doi.org/10.1080/13506285.2013.823140>

Hills, P. J., & Lewis, M. B. (2011). The own-age face recognition bias in children and adults. *Quarterly Journal of Experimental Psychology (2006), 64*(1), 17–23. <https://doi.org/10.1080/17470218.2010.537926>

Kaminski, K. S., & Sporer, S. L. (2016). Sind Polizisten/-innen bessere Augenzeugen/-innen? Ein Vergleich zwischen Polizisten/-innen und Zivilpersonen hinsichtlich Beschreibungen und Identifizierungsaussagen. *Recht & Psychiatrie, 34*(1), 18–26.

Landeskriminalamt Nordrhein-Westfalen. (2011). *Alters- und Größenschätzungen durch Tatenzeugen (Stand: August 2011)*. [https://polizei.nrw/sites/default/files/2016-11/Alters\\_Groesseneinschaetzung.pdf](https://polizei.nrw/sites/default/files/2016-11/Alters_Groesseneinschaetzung.pdf)

Lindholm, T. (2005). Own-age biases in verbal person memory. *Memory (Hove, England), 13*(1), 21–30. <https://doi.org/10.1080/09658210344000549>

Martschuk, N., & Sporer, S. L. (2021). Tatsächliche und wahrgenommene Richtigkeit von Personenidentifizierungen älterer Augenzeugen. *Forensische Psychiatrie, Psychologie, Kriminologie, 15*(2), 188–197. <https://doi.org/10.1007/s11757-020-00636-7>

Megreya, A. M., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica, 137*(1), 83–89. <https://doi.org/10.1016/j.actpsy.2011.03.003>

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Hrsg.), *The handbook of eyewitness psychology, Vol II: Memory for people*. (2007-04817-001; S. 3–34). Lawrence Erlbaum Associates Publishers.

Milne, R., & Bull, R. (2003). *Psychologie der Vernehmung: Die Befragung von Tatverdächtigen, Zeugen und Opfern (I. Erckenbrecht, Übers.)* (29,95 €). Huber.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Schwöbel, A.-K. (2014). *Merkmalsbasierte Gesichtswahrnehmung bei jüngeren und älteren Erwachsenen (Bachelorarbeit)*. Abteilung Psychologie in den Bildungswissenschaften der Johannes Gutenberg-Universität Mainz.

Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces—A comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20*(2), 125–133. <https://doi.org/10.1111/j.2044-8309.1981.tb00485.x>

Sporer, S. L. (1992). *Das Wiedererkennen von Gesichtern* (DM 58.-). Psychologie Verlags Union.

Sporer, S. L., & Sauerland, M. (2008). Personenidentifizierung. *Forensische Psychiatrie, Psychologie, Kriminologie, 2*(1), 28–36. <https://doi.org/10.1007/s11757-008-0057-9>

Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnen (1989). *European Psychologist, 19*(3), 207–220.

Wells, G. L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology, 70*(4), 619–626. <https://doi.org/10.1037/0021-9010.70.4.619>

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness Evidence: Improving its Probative Value. *Psychological Science in the Public Interest, 7*(2), 45–75. <https://doi.org/10.1111/j.1529-1006.2006.00027.x>

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8*(5), 453–464. <https://doi.org/10.1002/acp.2350080504>