

Validierung und Anwendung eines multinomialen Verarbeitungsbaummodells für Augenzeugengegenüberstellungen

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Kristina Winter
aus Münster

Düsseldorf, Juli 2023

aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Axel Buchner
2. Prof. Dr. Raoul Bell

Tag der mündlichen Prüfung: 12.09.2023

Inhaltsverzeichnis

Kurzzusammenfassung	4
Abstract	5
Einleitung	6
Ein multinomiales Verarbeitungsbaummodell für Augenzeugengegenüberstellungen	14
Besonderheiten der Methodik	19
Experiment 1	22
Experiment 2	25
Experiment 3	26
Experiment 4	28
Diskussion der Validierungsmanipulationen in den Experimenten 1 bis 4	30
Der Einfluss von simultanen und sequentiellen Gegenüberstellungsprozeduren in den Experimenten 1 bis 4	31
Der Einfluss von First-Yes-Counts-Instruktionen.....	33
Reanalyse der Daten von Horry et al. (2021).....	37
Experiment 5	39
Diskussion der Reanalyse von Horry et al. (2021) und von Experiment 5	41
Allgemeine Diskussion	42
Literatur.....	52
Einzelarbeiten.....	61
Erklärung über den Eigenanteil der in der Dissertation enthaltenen Einzelarbeiten.....	90
Erklärung an Eides Statt	91

Kurzzusammenfassung

Augenzeugengegenüberstellungen sind ein wichtiges Beweismittel, um nach einer Tat die tatbeteiligte Person zu verurteilen. Sie sind jedoch fehleranfällig. Es ist daher ausschlaggebend, sie zu optimieren. Traditionell wird die Güte einer Gegenüberstellungsprozedur z. B. durch die Verwendung von *Receiver-Operating-Characteristic-Kurven* (ROC-Kurven) ermittelt. Allerdings können ROC-Kurven Identifikationsentscheidungen nur vereinfacht modellieren. So vernachlässigen sie wichtige Informationen. Ein neues Messmodell, das die vollständige Antwortstruktur von Gegenüberstellungen berücksichtigt, wird in der vorliegenden Arbeit vorgestellt, validiert und angewendet. Dieses *Two-High Threshold Eyewitness Identification Model* (2-HT EIM) gehört zur Klasse multinomialer Verarbeitungsbaummodelle, welche speziell geeignet sind, latente kognitive Prozesse auf Basis von Antworthäufigkeiten darzustellen. Das 2-HT EIM schätzt vier kognitive Prozesse, die Augenzeugenidentifikationen zugrunde liegen können: das Erkennen der An- bzw. Abwesenheit der tatbeteiligten Person, das Erkennen, ob eine verdächtige Person aus einer Gegenüberstellung aufgrund von Unfairness hervorsticht, und Raten. In den Experimenten 1 bis 4 wurden die vier Modellparameter, die diese Prozesse abbilden, erfolgreich validiert. Dafür wurden die Prozesse selektiv durch gezielte experimentelle Manipulationen beeinflusst, nämlich durch die Manipulation der Expositionszeit, der Unfairness, der suggerierten Wahrscheinlichkeit der Anwesenheit einer tatbeteiligten Person und der Einfachheit der Zurückweisung einer Gegenüberstellung. Anschließend wurde das 2-HT EIM auf eine praktische Fragestellung angewendet: In der Forschung sind *First-Yes-Counts*-Instruktionen verbreitet, um eine Gegenüberstellung nach der ersten positiven Identifikation zu terminieren. Sie werden aber in der polizeilichen Praxis nicht verwendet. Horry et al. (2021) untersuchten erstmals den systematischen Einfluss solcher Instruktionen auf die Diskriminationsfähigkeit und Ratetendenz von Augenzeugen bzw. Augenzeuginnen. Sie fanden eine verringerte Ratetendenz durch diese Instruktionen. In einer Reanalyse ihrer Daten mit dem 2-HT EIM und in einer konzeptuellen Replikationsstudie in Experiment 5 wurden diese Schlussfolgerungen bestätigt und ihre Robustheit bekräftigt. Die Nutzung von *First-Yes-Counts*-Instruktionen in der wissenschaftlichen Forschung ist folglich problematisch, weil so die Gegebenheiten in der polizeilichen Praxis nicht ökologisch valide abgebildet werden. In Zukunft sollte die Forschung zu Augenzeugenidentifikationen adäquate Messmodelle wie das 2-HT EIM und ökologisch valide Prozeduren verwenden, damit gefundene Ergebnisse aussagekräftig und auf die polizeiliche Praxis übertragbar sind.

Abstract

Eyewitness testimonies are an important piece of evidence for convicting a culprit after a crime has been committed. However, they are prone to error. It is therefore crucial to optimize these procedures. Traditionally, the quality of a lineup procedure is determined by using, for example, *receiver operating characteristic* (ROC) curves. However, ROC curves can only model identification decisions in a simplified way. Thus, they neglect important information. A new measurement model that takes into account the full response structure of lineups is presented, validated, and applied in the present work. This *two-high threshold eyewitness identification model* (2-HT EIM) belongs to the class of multinomial processing tree models that are specifically tailored to measure latent cognitive processes based on response frequencies. The 2-HT EIM assumes four cognitive processes that can underlie eyewitness identification decisions: detection of the presence or absence of the culprit, recognizing whether a suspect stands out in a lineup due to unfairness and guessing. In Experiments 1 to 4, the four model parameters representing these processes were successfully validated. For this purpose, the processes were selectively influenced by specific experimental manipulations: the manipulation of exposure time, unfairness of a lineup, the suggested probability of the presence of a culprit and the ease of rejecting a lineup. The 2-HT EIM was then utilized for an applied research question: *First-Yes-Counts* instructions are common in research and are used to terminate a lineup after the first positive identification. However, they are not utilized in police practice. Horry et al. (2021) were the first to examine the systematic influence of such instructions on discriminability and guessing behavior in eyewitnesses. They found a reduced guessing tendency due to these instructions. In a reanalysis where the 2-HT EIM was applied to their data and in a conceptual replication study in Experiment 5, these results were confirmed and their robustness verified. Consequently, the use of *First-Yes-Counts* instructions in scientific research is problematic because it does not provide an ecologically valid representation of applied police practice. Future research on eyewitness identification decisions should use adequate measurement models such as the 2-HT EIM and ecologically valid procedures so that gained results are meaningful and transferable to police practice.

Einleitung

Für die Verurteilung einer verdächtigten Person aufgrund einer Straftat sind ausreichende und valide Beweise notwendig, die zeigen, dass die verdächtige Person tatsächlich schuldig ist. Dabei gibt es vielfältige Beweismittel, um die Schuld einer Person festzustellen. Neben Fingerabdrücken, dem Bildmaterial von Überwachungskameras oder DNA-Spuren kann auch die Aussage von Augenzeugen bzw. Augenzeuginnen, die eine Straftat und die tatbeteiligte Person beobachtet haben, genutzt werden. Typischerweise wird das Gedächtnis von Augenzeugen bzw. Augenzeuginnen dabei mit einer Gegenüberstellungsprozedur getestet.

In einer Gegenüberstellung werden dem Augenzeugen bzw. der Augenzeugin mehrere Personen präsentiert, oftmals in Form einer Wahllichtbildvorlage, die auf Fotos der Personen basiert. Eines dieser Fotos zeigt die verdächtige Person. Eine verdächtige Person kann entweder die tatsächlich tatbeteiligte Person sein, die ein Augenzeuge bzw. eine Augenzeugin während der Tat beobachtet hat, oder eine unschuldige verdächtige Person, die fälschlicherweise verdächtigt wird, weil sie beispielsweise auf die Beschreibung der tatbeteiligten Person passte und in der Nähe des Tatortes aufgegriffen wurde. Der tatsächliche Status der verdächtigten Person ist nicht bekannt, woraus sich die Notwendigkeit für eine Gegenüberstellung ergibt. Die anderen Fotos der Wahllichtbildvorlage zeigen Vergleichspersonen. Diese sind Personen, deren Schuld die Polizei sicher ausschließen kann. Ihre Fotos stammen oftmals aus einer Fotodatenbank und bzw. oder wurden digital verändert, um sie von real existierenden Personen zu entfremden (Bergold & Heaton, 2018; Colloff et al., 2016; Ministerium des Innern des Landes Nordrhein-Westfalen, 2023; U.S. Department of Justice: Office of the Deputy attorney General, 2017; Zarkadi et al., 2009). Die Vergleichspersonen sollen im Idealfall oberflächlich plausible Alternativen zur verdächtigten Person darstellen, um eine faire Gegenüberstellung zu gewährleisten und Schutz für eine unschuldige verdächtige Person zu liefern. Dafür sollen Vergleichspersonen je nach Staat hinsichtlich bestimmter Merkmalen wie beispielsweise Geschlecht, Ethnizität, Hautfarbe, Gesichtsbehaarung, Alter oder Körpergröße der Beschreibung eines Augenzeugen bzw. einer Augenzeugin ähneln und dem Bild der verdächtigten Person soweit gleichen, dass diese zwar nicht hervorsticht, aber dennoch einfach genug erkannt werden kann (Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; U.S. Department of Justice: Office of the Deputy attorney General, 2017). Die Anzahl der Vergleichspersonen variiert je nach

rechtlichen und ermittlungspraktischen Vorgaben für die Durchführung von Gegenüberstellungen und beträgt beispielsweise in den USA mindestens fünf und in Deutschland und Großbritannien mindestens sieben Vergleichspersonen (Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; U.S. Department of Justice: Office of the Deputy attorney General, 2017). Der Augenzeuge bzw. die Augenzeugin hat die Möglichkeit, eine der Personen aus der so präsentierten Gegenüberstellung als schuldig zu identifizieren oder die gesamte Gegenüberstellung zurückzuweisen. Welche Entscheidung korrekt ist, richtet sich nach der An- bzw. Abwesenheit der tatbeteiligten Person in der Gegenüberstellung. Wird die tatbeteiligte Person in einer Gegenüberstellung korrekt ausgewählt, handelt es sich um eine korrekte Entscheidung. Das gleiche gilt für eine Gegenüberstellung, in der keine tatbeteiligte Person, sondern nur eine unschuldige verdächtige Person anwesend ist, wenn diese korrekterweise zurückgewiesen wird. Es können dem Augenzeugen bzw. der Augenzeugin allerdings auch Fehler unterlaufen, indem eine der Vergleichspersonen fälschlicherweise ausgewählt wird oder eine Gegenüberstellung, in der sich eine tatbeteiligte Person befindet, zurückgewiesen wird. Im schlimmsten Fall wird eine unschuldige verdächtige Person in einer Gegenüberstellung ausgewählt, und die tatsächlich tatbeteiligte Person bleibt unentdeckt. Die möglichen Ausgänge von Identifikationen, die in einer Gegenüberstellung getroffen werden können, sind in der oberen Hälfte von Tabelle 1 zusammengefasst.

Gegenüberstellungen können simultan oder sequentiell durchgeführt werden (Andersen et al., 2014; Carlson & Carlson, 2014; Clark, 2012; Dobolyi & Dodson, 2013; Flowe et al., 2016; Gronlund et al., 2012; Horry et al., 2021; Meisters et al., 2018; Mickes et al., 2012; Neuschatz et al., 2016; Seale-Carlisle & Mickes, 2016; Seale-Carlisle et al., 2019). In simultanen Gegenüberstellungen werden dem Augenzeugen bzw. der Augenzeugin die Bilder aller Personen zeitgleich präsentiert. Der Augenzeuge bzw. die Augenzeugin muss in diesem Fall nur eine einzige Auswahlentscheidung für die gesamte Gegenüberstellung treffen. Auswahlentscheidung bedeutet hier, dass Augenzeugen sich entscheiden, ob sie eine der Personen aus der Gegenüberstellung wählen möchten, weil sie diese für die tatbeteiligte Person halten. Wird keine Auswahlentscheidung getroffen, so wird die Gegenüberstellung von dem Augenzeugen bzw. der Augenzeugin zurückgewiesen. Bei sequentiellen Gegenüberstellungen hingegen werden die Bilder der verdächtigten Person und der Vergleichspersonen einzeln nacheinander präsentiert und für jede Person muss eine separate Auswahlentscheidung getroffen werden. Eine sequentielle Gegenüberstellung gilt als

zurückgewiesen, wenn für keine der gezeigten Personen eine positive Auswahlentscheidung durch den Augenzeugen bzw. die Augenzeugin vorgenommen wurde.

Tabelle 1

Die 2 × 3-Datenstruktur einer typischen Gegenüberstellungsprozedur

	Identifikation der verdächtigen Person	Identifikation einer Vergleichsperson	Zurückweisung der Gegenüberstellung
Tatbeteiligte Person anwesend	Korrekte Identifikation der tatbeteiligten Person	Falsche Vergleichspersonenidentifikation	Falsche Zurückweisung der Gegenüberstellung
Tatbeteiligte Person abwesend	Falsche Identifikation der unschuldigen verdächtigen Person	Falsche Vergleichspersonenidentifikation	Korrekte Zurückweisung der Gegenüberstellung

Die 2 × 2-Datenstruktur einer generischen Detektionsaufgabe

	Wird die Anwesenheit eines Signals angenommen?	
Tatsächlicher Status des Signals	Ja	Nein
Signal anwesend	Treffer	Auslassung
Signal abwesend	Falscher Alarm	Korrekte Zurückweisung

Es ist in sequentiellen Gegenüberstellungen theoretisch möglich, dass der Augenzeuge bzw. die Augenzeugin mehrere Auswahlentscheidungen vornimmt, im Extremfall für jede der gezeigten Personen. Um diese Möglichkeit von vorneherein zu unterbinden werden in der wissenschaftlichen Forschung oftmals *First-Yes-Counts-Instruktionen* genutzt (Horry et al., 2021; für einen detaillierten Überblick an Studien, die First-Yes-Counts-Instruktionen nutzen, siehe auch: Brewer et al., 2012; Carlson & Carlson, 2014; Carlson et al., 2016; Carlson et al., 2008; Dobolyi & Dodson, 2013; Dunn et al., 2022; Flowe, 2011; Flowe & Ebbesen, 2007; Flowe et al., 2016; Gronlund et al., 2009; Gronlund et al., 2012; Horry, Palmer, et al., 2012; Kaesler et al., 2020; Kneller et al., 2001; Lindsay & Bellinger, 1999; McQuiston-Surrett et al., 2006; Memon & Gabbert, 2003; Mickes et al., 2012; Sauer et al., 2008; Steblay et al., 2011; Toftely, 2015), welche die Versuchspersonen einer Laborstudie instruieren, dass nur ihr erste Auswahlentscheidung zählt. Diese Instruktionen stehen allerdings im Kontrast zu den

verfügbaren Leitlinien der polizeilichen Praxis (Bill Blackwood Law Enforcement Management Institute of Texas, 2022; Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; State Bar of Michigan: Eyewitness Identification Task Force, 2015; U.S. Department of Justice: Office of the Deputy attorney General, 2017).

Für die Durchführung von Gegenüberstellungsverfahren hat sich durch moderne DNA-Analysen alter Beweismittel gezeigt, dass sie sehr fehleranfällig sind (Innocence Project, 2023). Es ist daher von größter Wichtigkeit, dass Gegenüberstellungsverfahren so durchgeführt werden, dass die Fehlerwahrscheinlichkeit minimiert wird. Forschungsergebnisse der letzten Jahrzehnte führten daher bereits zu Empfehlungsrichtlinien für die polizeiliche Praxis, für die Durchführung und die Verbesserung der Güte von Gegenüberstellungsverfahren (Wells et al., 2020; Wells et al., 1998). Für die wissenschaftliche Untersuchung von Augenzeugengegenüberstellungen im Labor, werden Versuchspersonen typischerweise fiktive Tathergänge, meist in Form eines Videos, gezeigt. Im Anschluss daran wird die Versuchsperson instruiert und soll in einer fiktiven Gegenüberstellung die tatbeteiligte Person aus dem Video wiederfinden. Im Labor ist es dabei möglich, sowohl Gegenüberstellungen mit anwesender tatbeteiligter Person als auch solche, in denen sich nur eine unschuldige verdächtige Person befindet, gezielt zu konstruieren. Im Gegensatz zu Feldstudien ist der große Vorteil, dass den Versuchsleitenden bekannt ist, ob eine tatbeteiligte Person tatsächlich in der Gegenüberstellung anwesend war. Es lässt sich also genau nachvollziehen, welche der sechs möglichen Auswahlentscheidungen von einer Versuchsperson getroffen wurde. Über viele Versuchspersonen hinweg lassen sich dann Gruppenunterschiede in diesen Auswahlentscheidungen untersuchen. So lässt sich beispielsweise die Güte verschiedener Gegenüberstellungsverfahren miteinander vergleichen. Um dabei die Güte einer Gegenüberstellungsverfahren beurteilen zu können, ist ein adäquates statistisches Maß vonnöten, welche diese Güte abbilden kann. Historisch stehen zwei Maße im Vordergrund, die im Bereich der Gegenüberstellungsforschung von Interesse waren und sind. Sie stützen sich auf die Betrachtung der Häufigkeiten der Auswahlentscheidungen einer Gegenüberstellungsverfahren. Beide dieser Maße weisen allerdings Schwächen auf. In der vorliegenden Arbeit soll daher als weiteres Maß ein neues Messmodell vorgestellt werden, das diese Schwächen nicht aufweist.

Das historisch ältere der beiden Maße im Bereich der Gegenüberstellungsforschung ist die Bestimmung des sogenannten Diagnostizitätsverhältnisses (Lindsay & Wells,

1985; Wells, 2014; Wells, Dysart, et al., 2015; Wells & Lindsay, 1980; Wells et al., 2000; Wells, Smalarz, et al., 2015; Wells et al., 1998; Wells, Smith, et al., 2015). Um das Diagnostizitätsverhältnis einer Gegenüberstellungsverfahren zu bestimmen, wird das Verhältnis der Wahrscheinlichkeit der korrekten Identifikationen der tatbeteiligten Person zu der Wahrscheinlichkeit der falschen Identifikationen der unschuldigen verdächtigten Person berechnet. Ein höheres Diagnostizitätsverhältnis wird dabei mit einer besseren Güte der entsprechenden Gegenüberstellungsverfahren assoziiert im Vergleich zu einer Gegenüberstellungsverfahren mit einem niedrigeren Diagnostizitätsverhältnis (Stebly et al., 2011). Allerdings sind im Diagnostizitätsverhältnis die Diskriminationsleistung von Augenzeugen bzw. Augenzeuginnen und ihre Antworttendenz konfundiert (Gronlund et al., 2012; Mickes et al., 2012; Wixted & Mickes, 2014). Das Diagnostizitätsverhältnis wird also durch zwei Faktoren angehoben: der Fähigkeit eines Augenzeugen bzw. einer Augenzeugin zwischen schuldigen und unschuldigen Personen zu differenzieren und konservativen Antworttendenzen (Wixted & Mickes, 2014). Konservative Antworttendenzen beschreiben die Neigung einer Person, nur eine Entscheidung zu treffen, wenn sie sich dieser sehr sicher ist (Gronlund et al., 2012; Mickes et al., 2012; Wixted & Mickes, 2014). Im Vergleich: Liberale Antworttendenzen, also die Neigung einer Person, auch bei Unsicherheit zu raten und damit eine Identifikation zu tätigen, sind mit einem niedrigeren Wert im Diagnostizitätsverhältnis assoziiert.

Aus diesem Grund wurden sogenannte *Receiver-Operating-Characteristic-Kurven* (ROC-Kurven) im Bereich der Augenzeugengegenüberstellungsforschung eingeführt (Amendola & Wixted, 2015a, 2015b; Gronlund et al., 2012; Gronlund et al., 2015; Mickes et al., 2012; Wixted & Mickes, 2014, 2015a, 2015b). Die Bildung von ROC-Kurven basiert auf der 2×2 -Datenstruktur einer generischen Detektionsaufgabe aus der Signalentdeckungstheorie (s. unterer Teil von Tabelle 1). Es muss daher zunächst die 2×3 -Datenstruktur einer Gegenüberstellungsverfahren aus dem oberen Teil von Tabelle 1 in die 2×2 -Datenstruktur aus dem unteren Teil von Tabelle 1 überführt werden, um ROC-Kurven bilden zu können. Die An- bzw. Abwesenheit eines Signals korrespondiert mit der An- bzw. Abwesenheit einer tatbeteiligten Person. Die Treffer entsprechen im Kontext von Augenzeugenidentifikationen dabei der Auswahl einer tatbeteiligten Person und die falschen Alarme der Auswahl einer unschuldigen verdächtigten Person. Auslassungen und korrekte Zurückweisungen werden gebildet, indem von den übrigen Kategorien jeweils zwei zusammengefasst werden: Dies geschieht, indem für die Auslassungen in Gegenüberstellungsverfahren mit anwesender tatbeteiligter Person die Häufigkeiten, die auf die Auswahl einer

Vergleichsperson entfallen, mit den Häufigkeiten der falschen Zurückweisungen der Gegenüberstellungsprozedur zusammengefasst werden. Für die korrekten Zurückweisungen werden in Gegenüberstellungsprozeduren mit abwesender tatbeteiligter Person die Häufigkeiten der Wahl einer Vergleichsperson mit den Häufigkeiten der korrekten Zurückweisungen der Gegenüberstellungsprozedur zusammengefasst (Wells, Smalarz, et al., 2015; Wells, Smith, et al., 2015). Die ursprüngliche Datenstruktur einer Gegenüberstellungsaufgabe wird durch diese Vereinfachung in ein Format gebracht, mit dem sich ROC-Kurven bilden lassen. Zusätzlich müssen, um ROC-Kurven bestimmen zu können, von den Versuchspersonen einer Studie zu Augenzeugenidentifikationen Antwortkonfidenzen erfragt werden. Typischerweise geschieht dies so, dass eine Versuchsperson nach einer Auswahlentscheidung oder bei Zurückweisung einer Gegenüberstellung die subjektive Sicherheit angibt, mit der sie diese Entscheidung gefällt hat. Oftmals wird die subjektive Sicherheit auf einer numerischen Beurteilungsskala angegeben. Beispielsweise handelt es sich dabei um eine Konfidenzskala, die von 1% (geringe subjektive Sicherheit, dass die ausgewählte Person die tatbeteiligte Person ist) bis 100% (absolute subjektive Sicherheit, dass die ausgewählte Person die tatbeteiligte Person ist) reicht (Mickes et al., 2012; Wixted & Mickes, 2014). Anschließend werden in einem nächsten Schritt auf Basis der Antwortkonfidenzen die Trefferrate und die Falsche-Alarm-Rate berechnet (z. B. Dobolyi & Dodson, 2013). Um den ersten Datenpunkt für die ROC-Kurve zu bestimmen werden alle Auswahlentscheidungen berücksichtigt, die mit einer Konfidenz von mindestens 90% gegeben wurden. Für diese werden Trefferrate und Falsche-Alarm-Rate bestimmt und in einem Koordinatensystem gegeneinander abgetragen. Für den zweiten Datenpunkt der ROC-Kurve werden alle Auswahlentscheidungen berücksichtigt, die mit einer Konfidenz von 70% oder mehr getroffen wurden und deren Trefferrate gegen die Falsche-Alarm-Rate im gleichen Koordinatensystem abgetragen. Die nächsten Datenpunkte der ROC-Kurve werden analog berechnet und im Koordinatensystem abgetragen, wodurch sich schließlich eine kurvenförmige Anordnung der so bestimmten Datenpunkte ergibt. (Gronlund et al., 2012; Mickes et al., 2012; Wixted & Mickes, 2014). Ein vereinfachtes Beispiel solcher ROC-Kurven findet sich in Abbildung 1.

Um letztendlich die Güte einer Gegenüberstellungsprozedur mithilfe von ROC-Kurven beurteilen zu können, wird die Fläche unter der jeweiligen Kurve einer Gegenüberstellungsprozedur, die *partial area under the curve* (pAUC), berechnet. Die Größe der pAUC verschiedener Gegenüberstellungsprozeduren lässt sich im Anschluss

miteinander vergleichen. Die Prozedur, welche die größere pAUC ergibt, wird dabei als überlegen betrachtet (Mickes et al., 2012; Mickes et al., 2014).

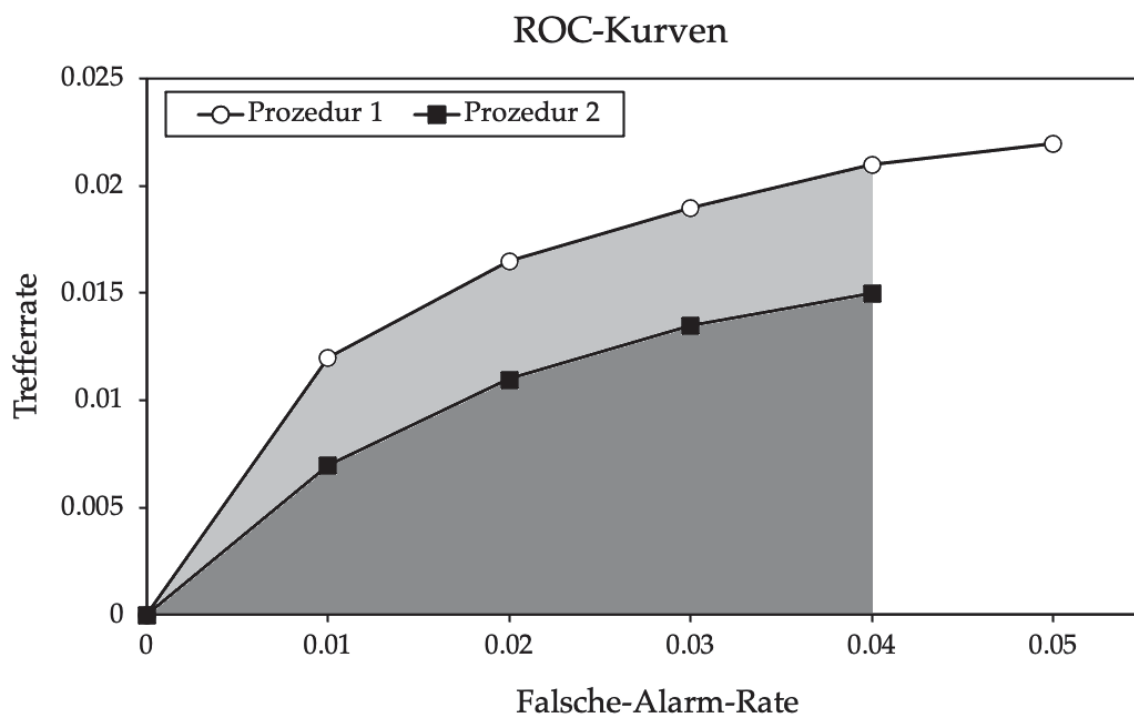


Abbildung 1. Vereinfachtes Beispiel zweier ROC-Kurven. Zur Bildung von ROC-Kurven im Kontext von Augenzeugegegenüberstellungen wird von Versuchspersonen nach einer Auswahlentscheidung die subjektive Sicherheit, mit der eine solche Entscheidung getroffen wurde, erfragt. Anschließend werden die Datenpunkte der ROC-Kurve bestimmt, indem erst die Trefferrate und die Falsche Alarm-Rate für alle Auswahlentscheidungen, die beispielsweise mit einer Konfidenz von mindestens 90% getroffen wurden, berechnet und in der Grafik abgetragen werden. Der nächste Datenpunkt wird beispielsweise für alle Auswahlentscheidungen, die mit einer Konfidenz von mindestens 70% getroffen wurden, bestimmt und abgetragen. Alle weiteren Datenpunkte werden analog bestimmt, wodurch sich die gezeigten Kurven ergeben. Um die Güte zweier Gegenüberstellungsprozeduren miteinander statistisch zu vergleichen wird jeweils die partielle Fläche unter den Kurven (pAUC) berechnet (hier dargestellt in hell- und dunkelgrau). In dem Beispiel aus dieser Abbildung würde Prozedur 1 eine höhere pAUC und somit eine höhere Güte aufweisen als Prozedur 2.

Die Nutzung von ROC-Kurven ist im Vergleich zum Diagnostizitätsverhältnis in dem Sinne vorteilhaft, dass dort nicht die Diskriminationsfähigkeit und die Antworttendenz konfundiert werden. Allerdings muss für die Bildung dieses ursprünglich aus der Signalentdeckungstheorie stammenden Maßes (Macmillan & Creelman, 2005) die Datenstruktur einer typischen Gegenüberstellungsprozedur vereinfacht werden. So gehen wichtige Informationen verloren, die in der ursprünglichen 2×3 -Datenstruktur von Augenzeugegegenüberstellungen vorhanden sind. Es wird argumentiert, dass die korrekten Identifikationen der tatbeteiligten Person und die falschen Identifikationen der unschuldigen verdächtigten Person in der Praxis die größte Relevanz aufweisen, weil sie die größten Konsequenzen verursachen und die anderen Kategorien daher weniger relevant sind (Mickes et al., 2012).

Nichtsdestotrotz verursacht die Vereinfachung der 2×3 -Datenstruktur in eine 2×2 -Datenstruktur Probleme, denn sie fasst Entscheidungen zu einer einzigen Datenkategorie zusammen, von denen angenommen werden muss, dass sie konzeptuell unterschiedlich sind: Im Kontext von Gegenüberstellungen mit anwesender tatbeteiligter Person ist dies bestenfalls nur ungünstig. Hier werden für die Bildung der 2×2 -Datenstruktur die Auswahl einer Vergleichsperson und die falsche Zurückweisung der Gegenüberstellung zusammengefasst. Jemanden fälschlich auszuwählen, weil man ihn oder sie aufgrund ähnlicher optischer Merkmale mit der tatbeteiligten Person verwechselt und eine Gegenüberstellung zurückzuweisen, weil einem beispielsweise in der Gegenüberstellung niemand vertraut genug vorkommt und dabei die tatbeteiligte Person zu übersehen, wird vermutlich jeweils von einem unterschiedlichen kognitiven Prozess begleitet. Diese Prozesse können in ROC-Kurven aber nicht differenziert abgebildet werden. Im Kontext von Gegenüberstellungen, bei denen die tatbeteiligte Person abwesend ist, verschärft sich das Problem zusätzlich. Hier werden für die Bildung der 2×2 -Datenstruktur die fälschliche Auswahl einer Vergleichsperson und die korrekte Zurückweisung der Gegenüberstellung zusammengefasst. Das bedeutet, dass hier nicht nur kognitive Prozesse zusammengefasst werden, die vermutlich unterschiedlich sind. Es wird auch eine formal korrekte Entscheidung mit einer formal falschen Entscheidung gleichgesetzt (Wells, Smalarz, et al., 2015).

Durch die Vereinfachung in eine 2×2 -Datenstruktur werden folglich wertvolle Informationen nicht berücksichtigt, die in der eigentlichen 2×3 -Datenstruktur vorhanden sind (Wells, Smalarz, et al., 2015; Wells, Smith, et al., 2015). Obgleich die korrekten Identifikationen der tatbeteiligten Person und die falschen Identifikationen der unschuldigen verdächtigten Person für die polizeiliche Praxis potentiell die größte Relevanz haben sollen (siehe Mickes et al., 2012), limitiert dies den Erkenntnisgewinn über die kognitiven Prozesse, die zu Auswahlentscheidungen während Augenzeugegegenüberstellungen führen. Es ist anzunehmen, dass durch die Vereinfachung nicht alle bei Gegenüberstellungen auftretenden kognitiven Prozesse gemessen werden können bzw. die gewonnenen Messwerte durch die Vermischung verschiedener kognitiver Prozesse kontaminiert sein könnten (Smith et al., 2018; Wells, Smalarz, et al., 2015; Wells, Smith, et al., 2015). Um Fehler bei Gegenüberstellungsverfahren zu reduzieren, sollten diese Prozesse aber verstanden werden. Die Daten der vollen 2×3 -Datenstruktur zu berücksichtigen, würde dieses Problem lösen. Außerdem werden unter Nutzung der vollen 2×3 -Datenstruktur nicht formal korrekte und formal falsche Auswahlentscheidungen miteinander gleichgesetzt. Ein alternatives Maß zu

finden, das die vollständige Datenstruktur von Gegenüberstellungsprozeduren berücksichtigt, ist daher von großer theoretischer und praktischer Relevanz.

Ein multinomiales Verarbeitungsbaummodell für Augenzeugengegenüberstellungen

In der vorliegenden Arbeit soll ein neues Messmodell für die Auswertung von Gegenüberstellungsdaten eingeführt, validiert und schließlich auf eine praktische Fragestellung angewendet werden. Das vorgestellte Modell ist in der Lage, die Informationen der vollständigen 2×3 -Datenstruktur für Gegenüberstellungen zu nutzen und auf dieser Basis vier nicht kontaminierte, latente kognitive Prozesse zu messen, die bei Entscheidungen in Gegenüberstellungsprozeduren eine Rolle spielen können.

Das in der vorliegenden Arbeit vorgestellte Messmodell gehört zur Klasse der multinomialen Verarbeitungsbaummodelle (MVB-Modelle). Anwendung finden solche multinomialen Verarbeitungsbaummodelle in diversen Feldern der kognitiven Psychologie wie beispielsweise in der Gedächtnisforschung (Bayen et al., 1996; Meiser & Bröder, 2002; Sherman et al., 2003; Smith & Bayen, 2004) oder der Entscheidungsfindung (Erdfelder & Buchner, 1998; Stahl & Degner, 2007; Unkelbach & Stahl, 2009). Übersichtsarbeiten wie beispielsweise Erdfelder (2000) und Schmidt et al. (2022) beschreiben detailliert die Verwendung von MVB-Modellen: MVB-Modelle haben eine baumartige Struktur. Dabei repräsentieren die Zweige nicht direkt beobachtbare kognitive Prozesse, die das beobachtbare, kategorial gemessene Verhalten beispielsweise als Reaktion auf Stimuli determinieren (Batchelder & Riefer, 1999; Erdfelder, 2000; Erdfelder et al., 2009; Riefer & Batchelder, 1988). Die latenten kognitiven Prozesse werden im Modell durch Parameter beschrieben, welche die Wahrscheinlichkeit für das Ablaufen der jeweiligen latenten Prozesse darstellen, wobei ihre Werte im Intervall zwischen 0 und 1 liegen. Die Nutzung von MVB-Modellen erlaubt es, die statistische Passung des Modells zu einem vorliegenden Datensatz zu bestimmen. Eine gute Modellpassung liegt vor, wenn der Unterschied zwischen beobachtbaren und vom Modell vorhergesagten Antworthäufigkeiten statistisch nicht signifikant wird. Ist Modellpassung gegeben, beschreiben die Parameterschätzer des Modells die Datenstruktur angemessen und ein so etabliertes Modell lässt sich als Basismodell festlegen. Wichtige Voraussetzung für die Etablierung eines Basismodells und die Berechnung der Modellpassung ist, dass dieses mindestens null Freiheitsgrade aufweist. Dafür müssen oftmals Restriktionen auf das Modell angewendet

werden, die theoretisch begründbar und plausibel sind. Ein genutztes Basismodell sollte im Sinne der Sparsamkeit so einfach wie möglich sein (z. B. Klauer & Kellen, 2010; Meiser & Bröder, 2002). Restriktionen finden oftmals durch die Gleichsetzung von zwei oder mehr Modellparametern statt. Da für jede Versuchsbedingung einer wissenschaftlichen Untersuchung eine eigene Instanz des multinomialen Verarbeitungsbaummodells vonnöten ist, spricht man in diesem Fall von einem verbundenen multinomialen Verarbeitungsbaummodell, bei dem die Parameter für jede Versuchsbedingung separat geschätzt werden. Zur Gewinnung von Freiheitsgraden können die Gleichsetzungen beispielsweise auch über die Versuchsbedingungen hinweg erfolgen. Auf Grundlage eines so etablierten Basismodells lassen sich statistische Tests zu einzelnen Modellparametern durchführen. Diese Tests erlauben es, inhaltliche Hypothesen über die latenten kognitiven Prozesse zu testen, welche den Parametern zugrunde liegen (Erdfelder et al., 2009). So können auch die Parameter zwischen den einzelnen Versuchsbedingungen miteinander verglichen werden, indem über die Versuchsbedingungen hinweg zusätzliche, theoretisch motivierte Restriktionen auf die im Basismodell noch freien Modellparameter angewendet werden. Anschließend wird geprüft, ob sich durch diese zusätzlichen Restriktionen die Modellpassung zum vorliegenden Datensatz im Vergleich zum vorher etablierten Basismodell signifikant verschlechtert. Wird die Modellpassung durch die zusätzlichen Restriktionen nicht signifikant schlechter, so wird kein Unterschied zwischen den Versuchsbedingungen angenommen. Verschlechtert sich die Modellpassung aber signifikant, kann ein statistisch bedeutsamer Unterschied zwischen den gleichgesetzten Modellparametern angenommen werden. Mithilfe dieses Vorgehens lassen sich Schlüsse über die zugrundeliegenden, vom Modell abgebildeten, Prozesse ziehen (Erdfelder, 2000; Schmidt et al., 2022).

Die einzelnen Modellparameter eines neues multinomialen Verarbeitungsbaummodells müssen im Sinne der Konstruktvalidität validiert werden, bevor ein MPT-Modell angewendet werden kann. Dies ist erforderlich damit das Modell genutzt werden kann, um wissenschaftliche Fragestellungen zu formalisieren und Vorhersagen psychologischer Theorien zu testen (Erdfelder, 2000; Schmidt et al., 2022). Durch die Validierung lässt sich sicherstellen, dass die einzelnen Parameter des Modells tatsächlich die vom Modell postulierten latenten kognitiven Prozesse abbilden. Für die Validierung sind experimentelle Manipulationen erforderlich, die gezielt und zuverlässig einen der postulierten Prozesse ansprechen. Im Idealfall wird dies dadurch gewährleistet, dass eine Prozedur genutzt wird, die schon nachweislich den postulierten Prozess bzw. sehr ähnliche Prozesse beeinflussen konnte (Erdfelder, 2000), sodass

die vorhergesagte Reaktion des Zielparameters offenkundig ist. Reagiert der Zielparameter dann wie vorhergesagt auf die Manipulation, ist dies ein Beleg dafür, dass der postulierte Prozess tatsächlich durch den Modellparameter abgebildet wird. Um ein neues multinomiales Verarbeitungsbaummodell zu validieren, sollte jeder Modellparameter und dessen zugrundeliegender Prozess einzeln in einem Validierungsexperiment validiert werden.

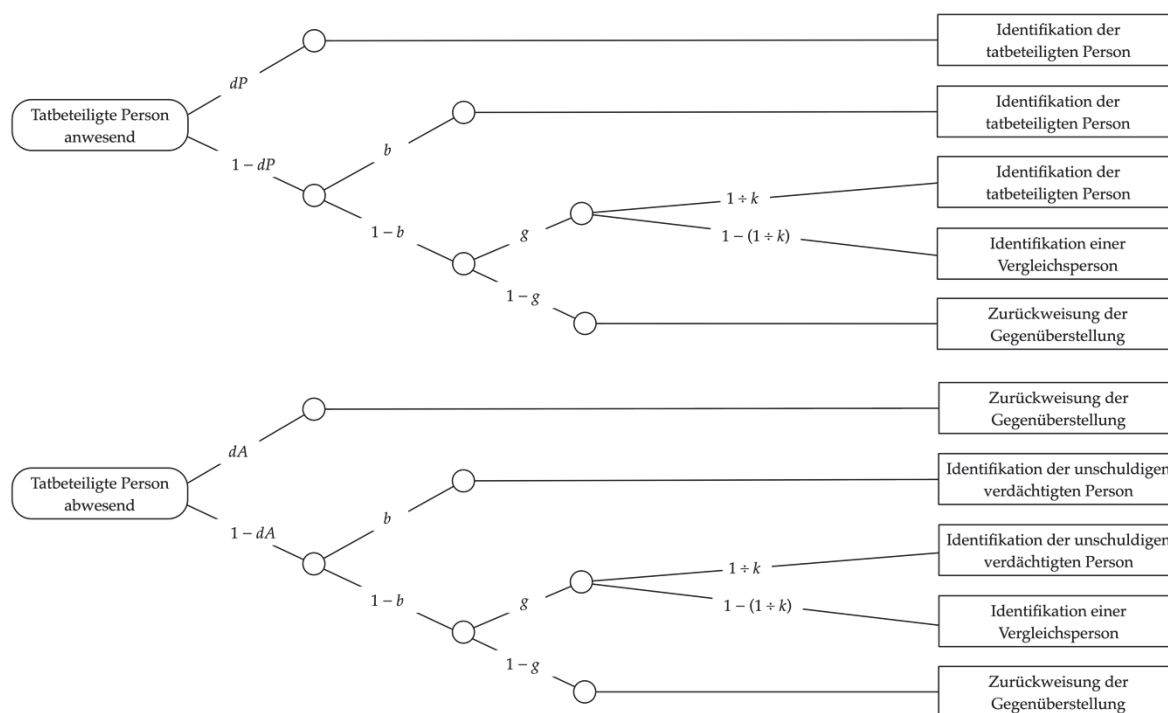


Abbildung 2. Baumstruktur des 2-HT EIMs. Die abgerundeten Rechtecke auf der linken Seite stellen die verarbeitete Reizkategorie dar, hier Gegenüberstellungen mit an- bzw. abwesender tatbeteiligter Person, auf die eine Reaktion beobachtet wird. Die Menge möglicher Reaktionen wird in Form beobachtbarer Antwortkategorien in den Rechtecken auf der rechten Seite dargestellt. Die Buchstaben an den Zweigen des Modells stellen die Wahrscheinlichkeiten für die angenommenen latenten kognitiven Prozesse dar. Parameter dP repräsentiert die Wahrscheinlichkeit, dass die Anwesenheit der tatbeteiligten Person detektiert wird. Parameter b repräsentiert die Wahrscheinlichkeit, dass eine verdächtige Person, die aufgrund von Unfairness aus der Gegenüberstellung hervorsteht, gewählt wird, unabhängig davon ob sie schuldig oder unschuldig ist. Parameter g repräsentiert die Wahrscheinlichkeit, dass eine Person aufgrund von Raten gewählt wird. Wenn eine ratebasierte Auswahl getroffen wird, so wird mit der Wahrscheinlichkeit von $1 \div k$ zufällig die verdächtige Person ausgewählt und mit der Gegenwahrscheinlichkeit von $1 - (1 \div k)$ eine Vergleichsperson. Hierbei ist k eine Konstante, die nicht vom Modell geschätzt, sondern anhand der Gegenüberstellungsgröße festgesetzt wird. Umfasst eine Gegenüberstellung z. B. sechs Personen wäre $k = 6$ und es würde mit einer Wahrscheinlichkeit von ca. 0.16667 zufällig die verdächtige Person identifiziert werden. Parameter dA repräsentiert die Wahrscheinlichkeit, dass die Abwesenheit der tatbeteiligten Person detektiert wird. Die Reihenfolge, in der die Prozesse ablaufen, ist bei der Nutzung von MVB-Modellen, wie sie hier passiert, nicht feststehend.

Für den Kontext der Augenzeugengegenüberstellungen wurde das *Two-High Threshold Eyewitness Identification Model* (2-HT EIM) vorgeschlagen (Menne et al., 2022; Winter et al., 2022). Abbildung 2 zeigt die Baumstruktur des 2-HT EIMs. Der obere Baum des 2-HT EIMs stellt die kognitiven Prozesse dar, die während

Gegenüberstellungen, in denen die tatbeteiligte Person anwesend ist, angenommen werden und gemäß der Modellstruktur zu einer Auswahlentscheidung bzw. einer Zurückweisung der Gegenüberstellung führen. Im unteren Baum werden die kognitiven Prozesse dargestellt, die während Gegenüberstellungen auftreten können, in denen die tatbeteiligte Person abwesend ist, und dort zu einer Auswahlentscheidung oder einer Zurückweisung der Gegenüberstellung führen.

In Gegenüberstellungen mit anwesender tatbeteiligter Person (oberer Baum in Abbildung 2) gibt der Anwesenheits-Detektions-Parameter dP (*culprit-presence-detection parameter*) die Wahrscheinlichkeit an, dass die Anwesenheit der tatbeteiligten Person entdeckt wird, was zu einer Identifikation der tatbeteiligten Person führt. Der gegenteilige Fall von $1 - dP$ gibt die Wahrscheinlichkeit an, dass die Anwesenheit der tatbeteiligten Person nicht erkannt wird. In diesem Fall kann die tatbeteiligte Person aufgrund zweier nicht-detektionsbasierter Prozesse immer noch gewählt werden. Eine Möglichkeit ist, dass die Gegenüberstellung unfair ist und die tatbeteiligte Person aufgrund einer optischen Unterscheidung zu den Vergleichspersonen hervorsticht. Obwohl keine Erinnerung an die tatbeteiligte Person vorliegt, kann ein Augenzeuge bzw. eine Augenzeugin dann dennoch die tatbeteiligte Person identifizieren. In diesem Fall gibt der Unfairness-Parameter b (*biased-suspect-selection parameter*) die Wahrscheinlichkeit an, dass die tatbeteiligte Person identifiziert wird. Sticht die tatbeteiligte Person aus einer Gegenüberstellung nicht hervor bzw. bemerkt der Augenzeuge bzw. die Augenzeugin das Hervorstechen nicht oder entscheidet sich dazu, es zu ignorieren, gibt $1 - b$ die Wahrscheinlichkeit an, dass keine Identifizierung auf Basis von Unfairness geschieht. In diesem Fall gibt der Rate-Parameter g (*guessing-based-selection parameter*) die Wahrscheinlichkeit an, dass ein Augenzeuge bzw. eine Augenzeugin durch Raten eine Person identifiziert. Die zufällig identifizierte Person wird mit einer Wahrscheinlichkeit von $1 \div k$, die auf der Gegenüberstellungsgröße k basiert, die tatbeteiligte Person sein und mit der Gegenwahrscheinlichkeit $1 - (1 \div k)$ eine der Vergleichspersonen. Hierbei ist k eine Konstante, die nicht aus den Daten geschätzt, sondern anhand der Anzahl der Personen in einer Gegenüberstellung im Modell festgelegt wird. In einer Gegenüberstellung mit insgesamt sechs Personen wäre $k = 6$. Die tatbeteiligte Person würde hier mit der Wahrscheinlichkeit von $1 \div 6$ zufällig identifiziert werden und eine der Vergleichspersonen mit der Gegenwahrscheinlichkeit von $1 - (1 \div 6)$, also $5 \div 6$. Entscheidet der Augenzeuge bzw. die Augenzeugin sich dagegen zu raten, so gibt $1 - g$ die Wahrscheinlichkeit an, dass die Gegenüberstellung fälschlicherweise zurückgewiesen wird.

In Gegenüberstellungen mit abwesender tatbeteiligter Person (unterer Baum in Abbildung 2) ist statt einer tatsächlich tatbeteiligten Person fälschlicherweise eine unschuldige verdächtige Person in der Gegenüberstellung anwesend. In einer solchen Situation gibt der Abwesenheits-Detektions-Parameter dA (*culprit-absence-detection parameter*) die Wahrscheinlichkeit an, dass ein Augenzeuge bzw. eine Augenzeugin erkennt, dass weder die unschuldige verdächtige Person noch eine Vergleichsperson die tatbeteiligte Person ist und folglich keine Person in der Gegenüberstellung die Tat begangen haben kann. Dies führt zu einer korrekten Zurückweisung der Gegenüberstellung. $1 - dA$ gibt die Wahrscheinlichkeit an, dass die Abwesenheit der tatbeteiligten Person nicht erkannt wird. In diesem Fall werden im unteren Baum für Gegenüberstellungen ohne tatbeteiligte Person die gleichen, nicht-detektionsbasierten Prozesse angenommen, wie im oberen Baum für die Gegenüberstellungen mit anwesender, tatbeteiligter Person. Wichtig ist anzumerken, dass die Reihenfolge, in der die vom Modell postulierten Prozesse ablaufen, bei einer Nutzung von MVB-Modellen, wie sie in der vorliegenden Arbeit geschieht, nicht feststehend ist (Schmidt et al., 2022).

In der vorliegenden Arbeit sollen in den Experimenten 1 bis 4 zunächst nacheinander die vier Modellparameter des 2-HT EIMs validiert werden. Im Anschluss an die erfolgreiche Validierung wird das Modell durch die Reanalyse bereits bestehender Daten und in Experiment 5 zusätzlich auf eine praktische Fragestellung angewendet. Wie in der wissenschaftlichen Forschung üblich wurde in den Experimenten den Versuchspersonen ein fiktives Video einer Tat gezeigt und im Anschluss Gegenüberstellungen durchgeführt. In der Hälfte aller Gegenüberstellungen wurde eine tatbeteiligte Person aus dem Video gezeigt, in der anderen Hälfte eine zuvor nicht präsentierte, unschuldige verdächtige Person. Die Auswahlentscheidungen wurden jeweils innerhalb der Versuchsbedingungen über die Versuchspersonen aggregiert und mithilfe des 2-HT EIMs analysiert.

In den Experimenten 1 bis 4 wurden experimentelle Manipulationen angewendet, die jeweils exakt einen der vier Modellparameter in erwartbare Richtung beeinflussen sollten. Um den Anwesenheits-Detektions-Parameter dP zu validieren, wurde die Fähigkeit der Augenzeugen bzw. Augenzeuginnen, die Anwesenheit der tatbeteiligten Person in einer Gegenüberstellung zu erkennen, durch die Expositionszeit zu der tatbeteiligten Person manipuliert. Um den Unfairness-Parameter b zu validieren, wurde die Fairness der Gegenüberstellungen und damit der Grad, zu dem eine verdächtige Person aus der Gegenüberstellung hervorstach durch Muttermale in

den Gesichtern der Vergleichspersonen manipuliert. Um den Rate-Parameter g zu validieren, wurde die Ratetendenz der Augenzeugen bzw. Augenzeuginnen durch Instruktionen manipuliert, die eine hohe oder niedrige Wahrscheinlichkeit der Anwesenheit einer tatbeteiligten Person in der Gegenüberstellung suggerierten. Diese drei Manipulationen orientierten sich an Beispielen aus der Literatur, die die gewünschten Effekte bereits erzielen konnten. Die Manipulation für Parameter dP orientierte sich an Memon et al. (2003), Palmer et al. (2013) und Valentine et al. (2003), die Manipulation für Parameter b an Zarkadi et al. (2009) und die Manipulation für Parameter g an Brewer und Wells (2006), Buchner et al. (1995), Clark (2005) und Keast et al. (2007). Für den Abwesenheits-Detektions-Parameter dA , der die Fähigkeit repräsentiert, die Abwesenheit einer tatbeteiligten Person in einer Gegenüberstellung zu erkennen, gab es keine solche Beispiele. Für die Validierung von Parameter dA wurde daher eine Manipulation gewählt, die es den Versuchspersonen erleichtern sollte, die Abwesenheit der tatbeteiligten Person zu erkennen, indem in Gegenüberstellungen ohne anwesende tatbeteiligte Person alle Personen Muttermale in ihren Gesichtern trugen und somit als tatbeteiligte Personen nicht in Frage kamen. In Gegenüberstellungen mit anwesender tatbeteiligter Person hingegen wurden die üblichen Gegenüberstellungen ohne Muttermale präsentiert, in denen alle Personen eine plausible Möglichkeit darstellten, die tatbeteiligte Person zu sein.

Das neue 2-HT EIM sollte sowohl für sequentielle als auch für simultane Gegenüberstellungsprozeduren geeignet sein. Um dies zu überprüfen, wurde in allen vier Validierungsexperimenten zusätzlich das Gegenüberstellungsformat manipuliert. Eine mögliche Überlegenheit von simultanen bzw. sequentiellen Gegenüberstellungsprozeduren wurde im Anschluss explorativ untersucht, da dies einen Beitrag zur viel diskutierten Frage nach der Überlegenheit einer der beiden Prozeduren leisten kann (Andersen et al., 2014; Carlson & Carlson, 2014; Clark, 2012; Dobolyi & Dodson, 2013; Flowe et al., 2016; Gronlund et al., 2012; Horry et al., 2021; Meisters et al., 2018; Mickes et al., 2012; Neuschatz et al., 2016; Seale-Carlisle & Mickes, 2016; Seale-Carlisle et al., 2019). Aus den Ergebnissen der Experimente 1 bis 4 leiten sich auch die weiteren Fragestellungen ab, die im Anschluss an die Validierungsexperimente in der praktischen Anwendung des 2-HT EIMs aufgegriffen werden.

Besonderheiten der Methodik

In den Experimenten 1 bis 5 lagen methodische Besonderheiten vor, die die Gegebenheiten der polizeilichen Praxis möglichst realitätsnah abbilden sollten, um möglichst

aussagekräftige Ergebnisse zu erzielen, in dieser Form aber selten in der Literatur zu Augenzeugenidentifikationen zu finden sind. So wurde zum einen in den Videos, die eine fiktive Tat zeigten, nicht nur eine, sondern vier tatbeteiligte Personen gezeigt. Dementsprechend wurden auch vier Gegenüberstellungen nacheinander durchgeführt. Dies hat zwei Vorteile: Es ergänzt den Erkenntnisgewinn in der wissenschaftlichen Erforschung von Augenzeugenidentifikationen, da es in der Praxis ebenfalls Taten gibt, die nicht nur von Einzelpersonen, sondern aus der Gruppe begangen werden (Bundeskriminalamt, 2021; Hobson et al., 2012; Tupper et al., 2019). Außerdem konnten so mehr Datenpunkte pro Versuchsperson erhoben werden. Bei gleichbleibender Stichprobengröße wird so eine höhere statistische Power erlangt, die für die Arbeit mit multinomialen Verarbeitungsbaummodellen benötigt wird (Heck et al., 2018).

Zum anderen wurde in den vorliegenden Experimenten erstmals eine gekreuzte Gegenüberstellungsverfahren (*crossed-lineup procedure*) verwendet (Menne et al., 2023; Winter et al., 2022, 2023). Diese gewährleistet, dass die Fotos von sowohl der tatbeteiligten Person als auch von der unschuldigen verdächtigten Person in den Gegenüberstellungen aus einer geteilten, anderen Quelle stammten, als die Fotos der Vergleichspersonen und so im gleichen Maße von den Bildern der Vergleichspersonen abwichen. Dies ist relevant, da dieses Vorgehen echten Gegenüberstellungen entspricht. Die Quelle der Fotos von Vergleichspersonen ist oftmals eine Datenbank und ihre Bilder werden teilweise auch digital verfremdet, um keine realen Personen abzubilden (Bergold & Heaton, 2018; Colloff et al., 2016; Ministerium des Innern des Landes Nordrhein-Westfalen, 2023; U.S. Department of Justice: Office of the Deputy attorney General, 2017; Zarkadi et al., 2009). Die Quelle der Fotos einer verdächtigten Person, von der unbekannt ist, ob sie unschuldig oder schuldig ist, könnte hingegen ein Social-Media-Profil sein, wenn die verdächtige Person beispielsweise auf der Flucht ist und kein Fahndungsbild angefertigt werden kann. Auch ein auf der Polizeiwache angefertigtes Fahndungsbild stammt aus einer anderen Quelle als die Fotos der Datenbank. Die so in einer Gegenüberstellung verwendeten Fotos unterscheiden sich also systematisch voneinander. In den hier aufgeführten Experimenten lag also eine ähnlich designierte unschuldige verdächtige Person vor, welche nicht aus der gleichen Datenbank stammt wie die Vergleichspersonen. Verglichen mit den Experimenten aus der Literatur ist dies ein ungewöhnliches Vorgehen, da dort oftmals die unschuldige verdächtige Person und die Vergleichspersonen aus der gleichen Datenbank stammen (z. B. Meisters et al., 2018; Wilcock & Bull, 2010). Manchmal liegt zudem gar keine designierte unschuldige verdächtige Person vor (z. B. Akan et al.,

2021; Mickes et al., 2017; Seale-Carlisle & Mickes, 2016; Zarkadi et al., 2009). Die Nutzung zweier paralleler Videos, von denen die Versuchspersonen aber nur eines präsentiert bekamen, ermöglichte es, diese gekreuzte Gegenüberstellungsprozedur umzusetzen. Es lagen dadurch zwei parallele Sets an tatbeteiligten Personen vor, die für die Versuchspersonen, je nach gesehenem Video, sowohl als tatbeteiligte Personen oder als unschuldige verdächtige Personen fungieren konnten. Für jede tatbeteiligte Person in dem ersten Video gab es im zweiten Video eine weitere tatbeteiligte Person, die dieser optisch stark ähnelte. Aus diesem Grund wurden für ein solches Paar an tatbeteiligten Personen jeweils die gleichen Vergleichsbilder genutzt. So wichen tatbeteiligte Personen und unschuldige verdächtige Personen im Mittel im gleichen Maße von den Vergleichspersonen aus der Datenbank ab. Diese Abweichung ist der polizeilichen Praxis ähnlich, da bei der Polizei nicht bekannt ist, ob die verdächtige Person tatbeteiligt oder unschuldig ist. Diese methodische Besonderheit ist für die Anwendung des 2-HT EIMs außerdem notwendig, weil der Unfairness-Parameter b für Gegenüberstellungsprozeduren mit anwesender und abwesender tatbeteiligter Person gleich ist und somit eine gleichförmige Abweichung dieser Personen von den Vergleichsbildern annimmt (Menne et al., 2022).

Eine dritte Besonderheit der in der vorliegenden Arbeit angewendeten Methodik ist der Verzicht auf First-Yes-Counts-Instruktionen in sequentiellen Gegenüberstellungen, wie sie in unterschiedlichen Formen in vielen veröffentlichten Studien Anwendung finden (z. B. Horry et al., 2021; Kaesler et al., 2020; Memon & Gabbert, 2003; Sauer et al., 2008). Eine Ausnahme bildet hier lediglich Experiment 5, in dem First-Yes-Counts-Instruktionen explizit manipuliert wurden. Da es in der polizeilichen Praxis Augenzeugen bzw. Augenzeuginnen erlaubt ist, mehrere Identifikationen vorzunehmen und diese lediglich für den Gerichtsprozess dokumentiert bzw. von dem Augenzeugen bzw. der Augenzeugin in einer zweiten Runde selbst aufgelöst werden müssen (Bill Blackwood Law Enforcement Management Institute of Texas, 2022; Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; Horry, Memon, et al., 2012; State Bar of Michigan: Eyewitness Identification Task Force, 2015; Steblay et al., 2011; U.S. Department of Justice: Office of the Deputy attorney General, 2017; Wells et al., 2011; Wells, Steblay, et al., 2015), ist es im Rahmen der ökologischen Validität sinnvoll und wichtig, eine solche Prozedur auch in der wissenschaftlichen Praxis umzusetzen. In der vorliegenden Arbeit wurde daher in den Experimenten 1 bis 4 immer die letzte Identifikation in einer sequentiellen Gegenüberstellung ausgewertet. Das gleiche gilt für die „Keine-First-Yes-Counts-Instruktionen“-Bedingung von Experiment 5. Die letzte Identifikation entspricht dabei

der Möglichkeit eines Augenzeugen bzw. einer Augenzeugin, eine bereits getroffene Entscheidung in einer sequentiellen Gegenüberstellung später zu revidieren, wenn eine neue Person gezeigt wird, welche der Erinnerung des Augenzeugen bzw. der Augenzeugin besser entspricht. Dies ist parallel zu simultanen Gegenüberstellungen, bei denen alle Personen einer Gegenüberstellung gleichzeitig gesehen werden und potentielle Auswahlentscheidungen so vor einer finalen Entscheidung noch revidiert werden können.

In allen Experimenten der vorliegenden Arbeit waren außerdem die Positionen der sechs Personen in den jeweiligen Gegenüberstellungen sowie die Reihenfolge, in der diese vier Gegenüberstellungen präsentiert wurden, randomisiert. Die Anzahl der Gegenüberstellungen und die Länge der sequentiellen Gegenüberstellungen war den Versuchspersonen jeweils unbekannt. Weitere Details zur Methodik der einzelnen Experimente können in den beiden Einzelarbeiten nachgelesen werden, denen diese Arbeit zugrunde liegt (s. Winter et al., 2022, 2023).

Experiment 1

In Experiment 1 sollte die Validität des Anwesenheits-Detektions-Parameters dP des neuen 2-HT EIMs geprüft werden. Parameter dP soll die Fähigkeit von Augenzeugen bzw. Augenzeuginnen abbilden, in einer Gegenüberstellung mit anwesender tatbeteiligter Person diese Anwesenheit zu erkennen. Bestehende Studien zeigen, dass die tatbeteiligte Person öfters korrekt identifiziert wird, wenn eine lange Expositionszeit der Versuchsperson zur tatbeteiligten Person vorlag als wenn eine kurze Expositionszeit vorlag (Memon et al., 2003; Palmer et al., 2013; Valentine et al., 2003). Analog dazu wurde in Experiment 1 die Expositionszeit der Versuchspersonen zur tatbeteiligten Person durch die Länge des präsentierten Videos der fiktiven Tat manipuliert. Ein langes Video – im Fall von Experiment 1 eine lange Expositionszeit von 130 Sekunden – sollte zu einer besseren Erinnerungsleistung der Versuchspersonen an die tatbeteiligten Personen führen als ein kurzes Video - eine kurze Expositionszeit von 13 Sekunden. Unter Annahme der Validität des Anwesenheits-Detektions-Parameters dP sollte dieser in der Versuchsbedingung mit langer Expositionszeit signifikant größer ausfallen als in der Versuchsbedingung mit kurzer Expositionszeit. Zusätzlich zur Expositionszeit (lang vs. kurz) wurde auch die Art der Gegenüberstellungsverfahren (simultan vs. sequentiell) manipuliert, wodurch sich ein 2×2 -Design ergab. In Experiment 1, sowie in allen Experimenten der vorliegenden Arbeit, wurden die

Versuchspersonen zufällig den Versuchsbedingungen zugeteilt. Das Signifikanzniveau wurde für alle Analysen auf $\alpha = .05$ festgesetzt.

Die Auswertung der in Experiment 1 gewonnenen Daten und aller Daten der folgenden Experimente sowie der Reanalyse erfolgte mit der Software multiTree (Moshagen, 2010). Für jede der Versuchsbedingungen in Experiment 1 wurde jeweils ein Exemplar des 2-HT EIMs aus Abbildung 3 konstruiert, sodass das Basismodell aus insgesamt acht Bäumen bestand. Damit das genutzte Basismodell so einfach wie möglich war, wurden alle Modellrestriktionen, die vor Durchführung des Experiments auf Basis theoretischer Überlegungen sinnvoll waren, gesetzt. Da in allen vier Versuchsbedingungen von Experiment 1 das gleiche Bildmaterial für die Gegenüberstellungen genutzt wurde, sollte es keine Unterschiede in der Unfairness der Gegenüberstellung sowie in der Fähigkeit der Versuchspersonen, die Abwesenheit einer tatbeteiligten Person zu erkennen, zwischen den Versuchsbedingungen geben. Sowohl der Unfairness-Parameter b als auch der Abwesenheits-Detektions-Parameter dA wurden deswegen jeweils über alle vier Versuchsbedingungen hinweg gleichgesetzt. Das so restringierte Basismodell passte zu den gewonnenen Daten [$G^2(6) = 3.91$, $p = .689$]. Die Parameterschätzer des Anwesenheits-Detektions-Parameters dP für alle vier Versuchsbedingungen sind in Abbildung 2 dargestellt. Die Schätzer der verbleibenden drei Parameter können in Winter et al. (2022) nachgelesen werden. Zur Testung, ob die Modellparameter sich hinsichtlich der Validierungsmanipulation unterschieden, wurden sie jeweils in einem zweiten Schritt in den Experimenten 1 bis 4 sowohl für die simultanen als auch für die sequentiellen Gegenüberstellungen über die Versuchsbedingungen der Validierung hinweg gleichgesetzt.

In Experiment 1 wurde die Validierungshypothese, dass der Anwesenheits-Detektions-Parameter dP in der Versuchsbedingung mit langer Expositionszeit größer ist als in der Versuchsbedingung mit kurzer Expositionszeit getestet. Dafür wurde Parameter dP sowohl für die simultanen als auch für die sequentiellen Gegenüberstellungen über die Versuchsbedingungen mit langer und kurzer Expositionszeit hinweg gleichgesetzt. Diese neue Modellpassung wurde mit der des Basismodells verglichen. Parameter dP war in den Versuchsbedingungen mit langer Expositionszeit signifikant größer als in den Versuchsbedingungen mit kurzer Expositionszeit [$\Delta G^2(2) = 72.21$, $p < .001$]. Dieser Befund zeigt, dass die Validierung des Anwesenheits-Detektions-Parameters dP erfolgreich war.

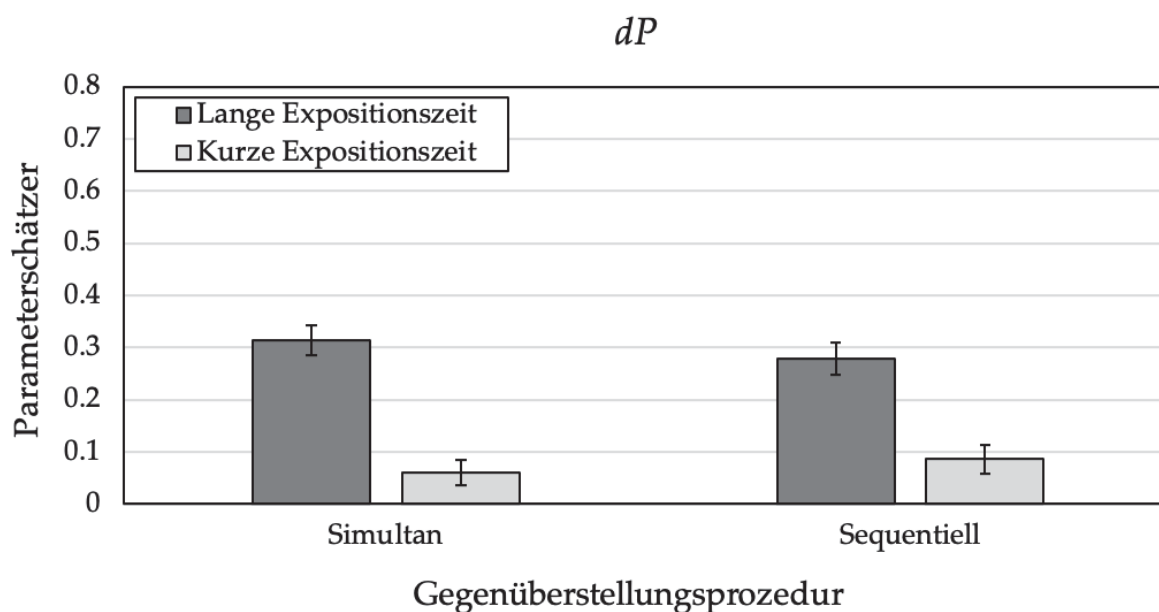


Abbildung 3. Parameterschätzer des Anwesenheits-Detektions-Parameters dP des 2-HT EIMs in Experiment 1. Die Balken repräsentieren die Wahrscheinlichkeit der Auswahl einer tatbeteiligten Person aufgrund der Detektion ihrer Anwesenheit in der Gegenüberstellung als Funktion der Manipulation der Expositionszeit (lange Expositionszeit vs. kurze Expositionszeit) und der Gegenüberstellungsprozedur (simultan vs. sequentiell). Die Fehlerbalken zeigen den Standardfehler.

Zusätzlich war der Rate-Parameter g in simultanen Gegenüberstellungen in der Versuchsbedingung mit langer Expositionszeit signifikant kleiner ($g = .45$, Standardfehler = $.02$) als in der Versuchsbedingung mit kurzer Expositionszeit ($g = .52$, Standardfehler = $.02$) [$\Delta G^2(1) = 6.47$, $p = .011$]. Dies zeigte sich nicht für die sequentiellen Gegenüberstellungen [$\Delta G^2(1) = 0.03$, $p = .871$]. Dieser Befund lässt sich beispielsweise mit kompensatorischem Raten erklären (Hirshman, 1995; Küppers & Bayen, 2014; Meiser et al., 2007; Riefer et al., 1994). Eine Versuchsperson, die nur eine schwache Erinnerung an die tatbeteiligte Person hat, wird in Gegenüberstellungen öfter dazu tendieren, zu raten, als eine Versuchsperson, die eine bessere Erinnerung an die tatbeteiligte Person aufweist und diese wahrscheinlicher auf Basis der eigenen Erinnerungen identifizieren kann. Sequentielle Gegenüberstellungen scheinen einen gewissen Schutz gegen kompensatorisches Raten aufzuweisen. Eine mögliche Erklärung wäre, dass es für einen Augenzeugen bzw. eine Augenzeugin in sequentiellen Gegenüberstellungen nicht direkt ersichtlich ist, dass keine der gezeigten Personen identifiziert werden kann. Die Tatsache, dass die Personen nacheinander präsentiert werden und die genaue Länge einer Gegenüberstellung unbekannt ist, eröffnet die Möglichkeit, dass erst die nächste gezeigte Person den Erinnerungen an die tatbeteiligte Person entspricht und ein Augenzeuge bzw. eine Augenzeugin im Falle von nicht ausreichenden Erinnerungen an die tatbeteiligte Person weniger zu einer auf

Raten basierten Auswahl verleitet wird. In einer simultanen Gegenüberstellung ist hingegen auf den ersten Blick ersichtlich, dass eine tatbeteiligte Person nicht detektiert werden kann, was die Tendenz zu kompensatorischem Raten begünstigen kann.

Experiment 2

Experiment 2 diente dazu, die Validität des Unfairness-Parameters b zu prüfen. Parameter b soll die Wahrscheinlichkeit abbilden, dass ein Augenzeuge oder eine Augenzeugin eine verdächtige Person in einer Gegenüberstellung auswählt, weil diese aufgrund von Unfairness hervorsteht. Unfairness in Gegenüberstellungen wurde bereits zuverlässig durch große Muttermale im Gesicht, durch die eine verdächtige Person zwischen den Vergleichspersonen hervorsteht, erzeugt (Zarkadi et al., 2009). In Experiment 2 wurden daher in einer unfairen Versuchsbedingung den Bildern der Vergleichspersonen Muttermale hinzugefügt. Die verdächtige Person (egal ob schuldig oder unschuldig) war somit die einzige Person in der Gegenüberstellung, die keine Muttermale aufwies. Deswegen sollte sie stärker aus dieser Gegenüberstellung hervorstechen als eine verdächtige Person in einer fairen Gegenüberstellung, in der niemand Muttermale im Gesicht aufwies. Unter Annahme der Validität des Unfairness-Parameters b sollte dieser in der unfairen Versuchsbedingung mit den Muttermalen signifikant größer ausfallen als in der fairen Versuchsbedingung ohne Muttermale. Die Methodik in Experiment 2 entsprach mit Ausnahme der veränderten Bilder der Vergleichspersonen in den unfairen Versuchsbedingung der Methodik aus Experiment 1. In Experiment 2 wurden ausschließlich lange Videos präsentiert.

Die Restriktionen des Basismodells und die Datenauswertung für Experiment 2 wurden analog zu Experiment 1 vorgenommen. Da der Unfairness-Parameter b in Experiment 2 manipuliert werden sollte, wurde für die Findung des Basismodells nur der Abwesenheits-Detektions-Parameter dA über alle Versuchsbedingungen hinweg gleichgesetzt. Das so restringierte Basismodell passte zu den Daten [$G^2(3) = 2.32$, $p = .509$]. Die Parameterschätzer des Unfairness-Parameters b sind in Abbildung 4 abgebildet. Alle weiteren deskriptiven Statistiken finden sich in Winter et al. (2022). Parameter b war in den unfairen Versuchsbedingungen signifikant größer als in den fairen Versuchsbedingungen [$\Delta G^2(2) = 28.59$, $p < .001$]. Dieser Befund zeigt, dass die Validierung des Unfairness-Parameters b erfolgreich war.

Der Anwesenheits-Detektions-Parameter dP unterschied sich nicht zwischen fairen und unfairen Gegenüberstellungen [$\Delta G^2(2) = 0.12$, $p = .940$]. Der Rate-Parameter g

war in simultanen Gegenüberstellungen für die unfaire Versuchsbedingung signifikant kleiner ($g = .40$, Standardfehler = .02) als in der fairen Bedingung ($g = .49$, Standardfehler = .02) [$\Delta G^2(1) = 7.93$, $p = .005$]. Dies zeigte sich nicht in sequentiellen Gegenüberstellungen [$\Delta G^2(1) = 0.09$, $p = .760$]. Eine mögliche Erklärung für diesen Befund könnte sein, dass Raten durch die Tatsache, dass eine Person aus der Gegenüberstellung hervorstach und diese dadurch leichter zu identifizieren war, in der unfairen Bedingung geringer ausfiel als in der fairen.

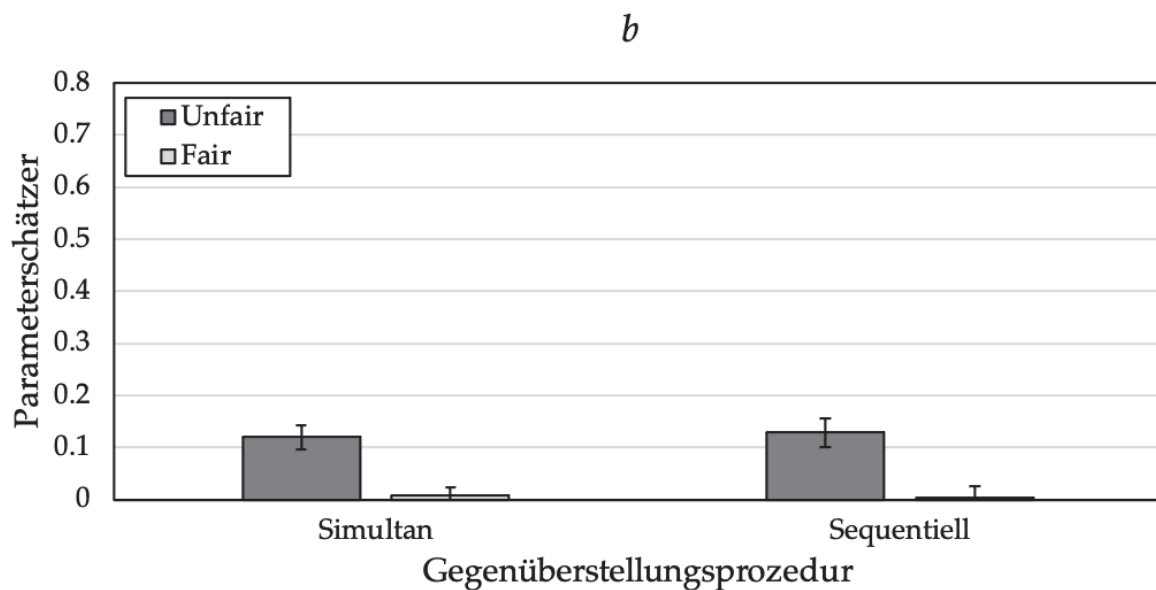


Abbildung 4. Parameterschätzer des Unfairness-Parameters b des 2-HT EIMs in Experiment 2. Die Balken repräsentieren die Wahrscheinlichkeit der Auswahl einer verdächtigten Person, weil diese aufgrund von Unfairness aus der Gegenüberstellung hervorsticht, als Funktion der Manipulation der Unfairness (fair vs. unfair) und der Gegenüberstellungsverfahren (simultan vs. sequentiell). Die Fehlerbalken zeigen den Standardfehler.

Experiment 3

Experiment 3 diente dazu, die Validität des Rate-Parameters g zu prüfen. Parameter g soll dabei die Tendenz eines Augenzeugen bzw. einer Augenzeugin abbilden, eine Auswahl aufgrund von Raten zu tätigen. Die Ratetendenz einer Person lässt sich beispielsweise mit der Manipulation von Instruktionen und dadurch entstehenden Erwartungen manipulieren: In klassischen Alt-Neu-Rekognitionsaufgaben wird dabei die Stärke der Erwartung manipuliert, dass ein nicht erinnertes Stimulus wahrscheinlich alt bzw. neu ist. Dies geschieht, ohne dass die tatsächliche Wahrscheinlichkeit des Auftretens der alten und neuen Items verändert wird. Eine Häufung von alten bzw. neuen Items wird also nur durch die gegebenen Instruktionen suggeriert

(Buchner et al., 1995). Ähnliche Befunde gibt es in der Literatur für Augenzeu-
genidentifikationen: Einseitige Instruktionen, die die Anwesenheit einer tatbeteiligten
Person in der Gegenüberstellung suggerieren, erhöhen die Tendenz zu raten, dass
die tatbeteiligte Person in der Gegenüberstellung anwesend ist gegenüber zweiseiti-
gen Instruktionen, die auch die Möglichkeit unterstreichen, dass die tatbeteiligte Per-
son nicht in der Gegenüberstellung zu finden sein könnte (Brewer & Wells, 2006;
Clark, 2005; Keast et al., 2007). Um den Rate-Parameter g zu validieren, wurden in
Experiment 3 die Instruktionen manipuliert, die den Versuchspersonen nach dem Vi-
deo des Tathergangs, aber vor den Gegenüberstellungen präsentiert wurden. In der
Versuchsbedingung mit hoher suggerierter Anwesenheitswahrscheinlichkeit wurde
den Versuchspersonen suggeriert, dass die tatbeteiligte Person mit hoher Wahr-
scheinlichkeit in der Gegenüberstellung zu finden sein würde. In dieser Versuchsbe-
dingung sollte die Wahrscheinlichkeit von Identifikationen, die auf Raten basieren,
erhöht sein. In der Versuchsbedingung mit niedriger suggerierter Anwesenheits-
wahrscheinlichkeit wurde den Versuchspersonen hingegen suggeriert, dass die tat-
beteiligte Person mit niedriger Wahrscheinlichkeit in der Gegenüberstellung zu fin-
den sein würde. In dieser Versuchsbedingung sollte die Wahrscheinlichkeit von
Identifikationen, die auf Raten basieren, erniedrigt sein. Die tatsächlichen Anwesen-
heitswahrscheinlichkeiten waren in allen Versuchsbedingungen gleich. Unter An-
nahme der Validität des Rate-Parameters g sollte dieser in der Versuchsbedingung
mit hoher suggerierter Anwesenheitswahrscheinlichkeit größer sein als in der Ver-
suchsbedingung mit niedriger suggerierter Anwesenheitswahrscheinlichkeit. Die
Methodik von Experiment 3 glich der Methodik aus Experiment 1, mit Ausnahme
der veränderten Instruktionen. Es wurden immer die langen Videos verwendet.

Die Restriktionen des Basismodells und die Datenauswertung für Experiment 3 fand
analog zu der in Experiment 1 statt. Das so restringierte Basismodell passte zu den
Daten [$G^2(6) = 3.85, p = .698$]. Die Parameterschätzer des Rate-Parameters g sind in
Abbildung 5 dargestellt. Alle weiteren deskriptiven Statistiken finden sich in Winter
et al. (2022). Parameter g war in den Versuchsbedingungen mit hoher suggerierter
Anwesenheitswahrscheinlichkeit signifikant größer als in den Versuchsbedingungen
mit niedriger suggerierter Anwesenheitswahrscheinlichkeit [$\Delta G^2(2) = 137.36,$
 $p < .001$]. Dieser Befund zeigt, dass die Validierung des Rate-Parameters g erfolgreich
war. Der Anwesenheits-Detektions-Parameter dP unterschied sich nicht hinsichtlich
der suggerierten Anwesenheitswahrscheinlichkeit der tatbeteiligten Person
[$\Delta G^2(2) = 4.13, p = .127$].

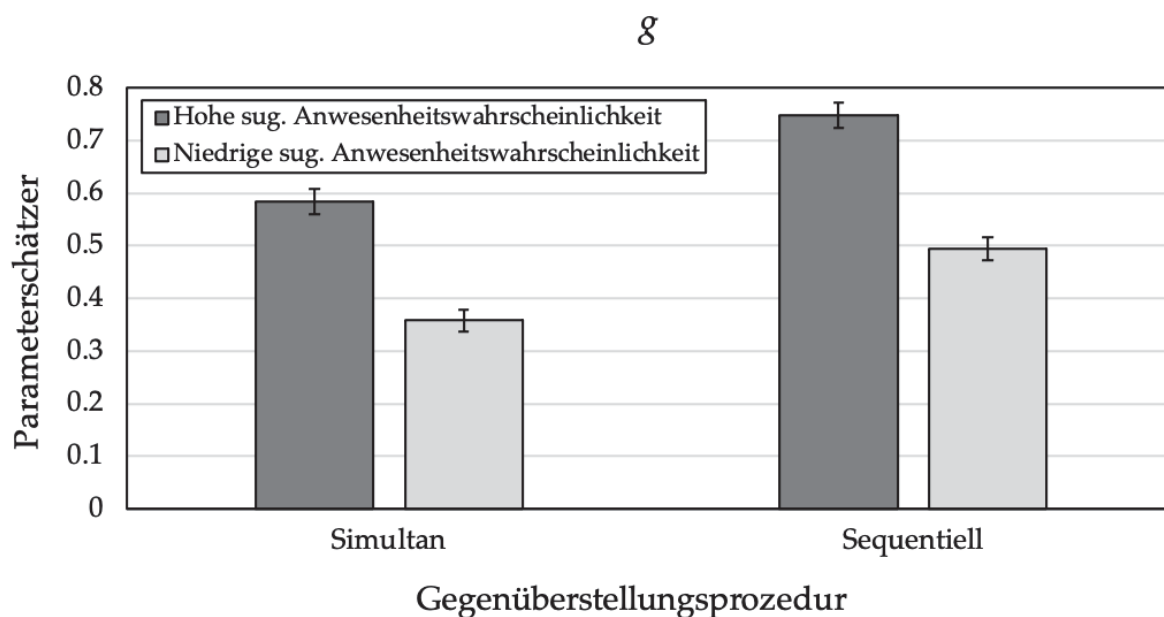


Abbildung 5. Parameterschätzer des Rate-Parameters g des 2-HT EIMs in Experiment 3. Die Balken repräsentieren die Wahrscheinlichkeit der Auswahl einer verdächtigten Person aufgrund von Raten, als Funktion der Manipulation der suggerierten (sug.) Wahrscheinlichkeit der Anwesenheit einer tatbeteiligten Person (hohe suggerierte Anwesenheitswahrscheinlichkeit vs. niedrige suggerierte Anwesenheitswahrscheinlichkeit) und der Gegenüberstellungsverfahren (simultan vs. sequentiell). Die Fehlerbalken zeigen den Standardfehler.

Experiment 4

Experiment 4 diente dazu, die Validität des Abwesenheits-Detektions-Parameters dA zu prüfen. Parameter dA sollte die Fähigkeit von Augenzeugen bzw. Augenzeuginnen abbilden, die Abwesenheit einer tatbeteiligten Person in einer Gegenüberstellung zu erkennen. In der Literatur gibt es keine direkten Beispiele für eine mögliche Validierungsmanipulation für den Abwesenheits-Detektions-Parameter dA . Es gibt aber durchaus Hinweise darauf, dass gewisse Verfahren die Fähigkeit von Augenzeugen bzw. Augenzeuginnen verbessern können, die Abwesenheit einer tatbeteiligten Person zu erkennen (s. Menne et al., 2022). Das Ziel der Manipulation in Experiment 4 war es, dass die Personen in einer Gegenüberstellung ohne tatbeteiligte Person dieser möglichst unähnlich sehen sollten. Wie in Experiment 2 wurden dazu große Muttermale in den Gesichtern der Vergleichspersonen platziert; allerdings nur in den Gegenüberstellungen, in denen die tatbeteiligte Person abwesend war. Zusätzlich zu diesen wiesen auch die unschuldigen verdächtigten Personen bei Gegenüberstellungen mit abwesender tatbeteiligter Person Muttermale im Gesicht auf. Im Vergleich zu Gegenüberstellungen ohne tatbeteiligte Person und ohne diese Muttermale sollte es den Versuchspersonen in den Gegenüberstellungen mit Muttermalen

aufgrund dieser neuartigen Merkmale im Gesicht leichter fallen, die Abwesenheit der tatbeteiligten Person zu erkennen und eine Gegenüberstellung zurückzuweisen. Die Versuchspersonen sollten es folglich in der schwer-zurückzuweisen-Versuchsbedingung ohne Muttermale schwerer haben die Abwesenheit einer tatbeteiligten Person zu detektieren als die Versuchspersonen in der einfach-zurückzuweisen-Versuchsbedingung mit Muttermalen. Unter Annahme der Validität des Abwesenheits-Detektions-Parameters dA wurde erwartet, dass dieser in der Schwer-zurückzuweisen-Versuchsbedingung kleiner sein würde als in der Einfach-zurückzuweisen-Versuchsbedingung. Die Methodik in Experiment 4 glich der Methodik aus Experiment 1 mit Ausnahme der mit Muttermalen versehenen Bilder in den Gegenüberstellungen ohne tatbeteiligte Person der Einfach-zurückzuweisen-Versuchsbedingungen. Es wurden immer die langen Videos verwendet.

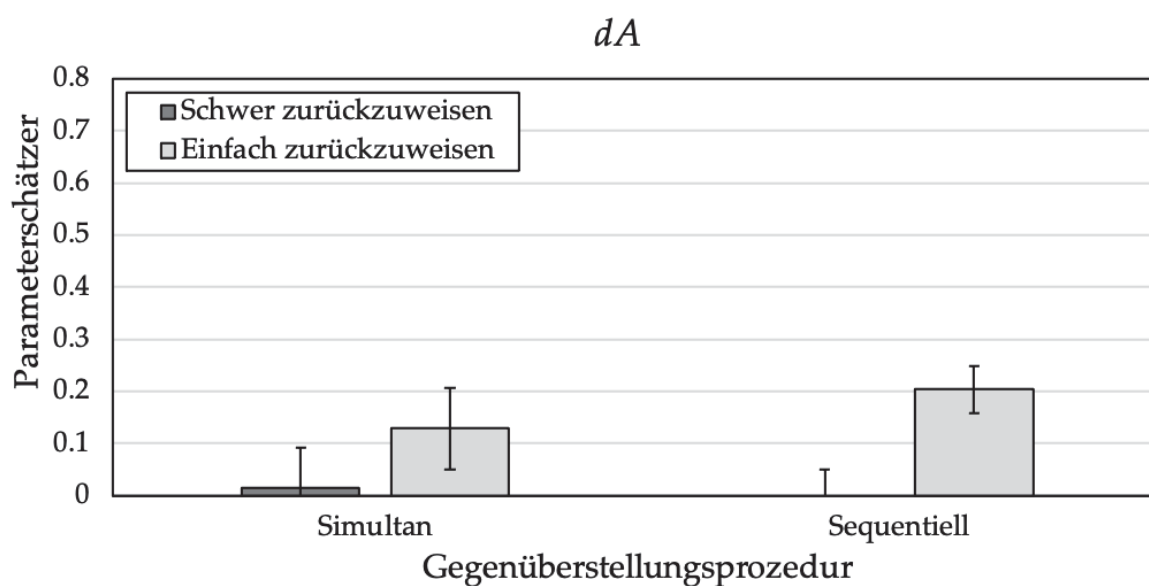


Abbildung 6. Parameterschätzer des Abwesenheits-Detektions-Parameters dA des 2-HT EIMs in Experiment 4. Die Balken repräsentieren die Wahrscheinlichkeit der Zurückweisung einer Gegenüberstellung ohne tatbeteiligte Person aufgrund der Detektion ihrer Abwesenheit in der Gegenüberstellung als Funktion der Manipulation der Einfachheit, mit der eine Gegenüberstellung zurückzuweisen ist (schwer zurückzuweisen vs. einfach zurückzuweisen) und der Gegenüberstellungsprozedur (simultan vs. sequentiell). Die Fehlerbalken zeigen den Standardfehler.

Die Restriktionen des Basismodells und die Datenauswertung für Experiment 4 fand analog zu der in Experiment 1 statt. Da der Abwesenheits-Detektions-Parameter dA in Experiment 4 manipuliert werden sollte, wurde für das Basismodells nur der Unfairness-Parameter b über alle Versuchsbedingungen hinweg gleichgesetzt. Das so restringierte Basismodell passte zu den Daten [$G^2(3) = 5.00, p = .172$]. Die Parameterschätzer des Abwesenheits-Detektions-Parameters dA sind in Abbildung 6

dargestellt. Alle weiteren deskriptiven Statistiken finden sich in Winter et al. (2022). Parameter dA war in der Schwer-zurückzuweisen-Versuchsbedingung kleiner als in der Einfach-zurückzuweisen-Versuchsbedingung [$\Delta G^2(2) = 11.95, p = .003$]. Dieser Befund zeigt, dass die Validierung des Abwesenheits-Detektions-Parameters dA erfolgreich war. Die Manipulation der Einfachheit der Zurückweisung der Gegenüberstellungen hatte keinen Effekt auf den Anwesenheits-Detektions-Parameter dP [$\Delta G^2(2) = 2.93, p = .231$] oder den Rate-Parameter g [$\Delta G^2(2) = 0.38, p = .827$].

Diskussion der Validierungsmanipulationen in den Experimenten 1 bis 4

Die Validierungsmanipulationen der Experimente 1 bis 4, die jeweils einen bestimmten Zielparameter ansprechen sollten, beeinflussten die Modellparameter jeweils konsistent in vorhergesagter Richtung. Die Ergebnisse der vier Validierungsexperimente zeigen daher insgesamt, dass die Parameter des 2-HT EIMs in der Lage sind die postulierten Prozesse abzubilden. Eine Validierung des 2-HT EIMs auf Basis der Ergebnisse der Experimente 1 bis 4 kann daher als erfolgreich angesehen werden. In Experiment 1 konnte der Anwesenheits-Detektions-Parameter dP durch die Manipulation der Länge des Tatvideos wie vorhergesagt beeinflusst werden: Eine lange Expositionszeit erhöhte im Vergleich zu einer kurzen Expositionszeit die Fähigkeit der Versuchspersonen, die tatbeteiligte Person später in einer Gegenüberstellung zu detektieren. In Experiment 2 konnte der Unfairness-Parameter b durch das Hinzufügen von Muttermalen in den Gesichtern der Vergleichspersonen, welche die Unfairness in einer Gegenüberstellung erhöhten, beeinflusst werden: Die Präsentation einer unfairen Gegenüberstellung erhöhte im Vergleich zur Präsentation einer fairen Gegenüberstellung die Wahrscheinlichkeit, dass eine Versuchsperson eine verdächtige Person auswählte, weil diese aus der Gegenüberstellung hervorstach. In Experiment 3 konnte der Rate-Parameter g durch die Präsentation einseitiger Instruktionen, welche entweder eine hohe oder eine niedrige Wahrscheinlichkeit suggerierten, dass eine tatbeteiligte Person in einer Gegenüberstellung vorhanden war, beeinflusst werden: Versuchspersonen entschieden sich für mehr Identifikationen basierend auf Raten, wenn sie glaubten, dass es sehr wahrscheinlich war, dass eine tatbeteiligte Person in einer Gegenüberstellung anwesend war, als wenn sie glaubten, dass dies unwahrscheinlich war. In Experiment 4 konnte der Abwesenheits-Detektions-Parameter dA durch die Platzierung von Muttermalen in den Gesichtern der Personen in Gegenüberstellungen ohne tatbeteiligte Person beeinflusst werden: Es fiel

Versuchspersonen leichter, die Abwesenheit einer tatbeteiligten Person in einer Gegenüberstellung zu erkennen, wenn eine Gegenüberstellung aufgrund auffälliger Merkmale in den Gesichtern aller Personen leichter zurückzuweisen war.

In den Experimenten 1 und 2 reagierte zusätzlich jeweils der Rate-Parameter g in den simultanen Gegenüberstellungsprozeduren auf die Validierungsmanipulationen. In der allgemeinen Diskussion werden mögliche Limitationen für das 2-HT EIM durch dieses Befundmusters erörtert.

Der Einfluss von simultanen und sequentiellen Gegenüberstellungsprozeduren in den Experimenten 1 bis 4

Neben den vier Validierungsmanipulationen wurde in der vorliegenden Arbeit auch der Einfluss simultaner und sequentieller Gegenüberstellungsprozeduren auf die Modellparameter des 2-HT EIMs untersucht. Primär diente die Manipulation der Gegenüberstellungsprozedur in den Experimenten 1 bis 4 dem Ziel, das Modell für beide Prozeduren zu validieren. Gleichzeitig erlaubt diese Manipulation es aber auch, dass der Einfluss der Gegenüberstellungsprozedur auf die vier Modellparameter explorativ untersucht werden kann. Weder der Unfairness-Parameter b noch der Abwesenheits-Detektions-Parameter dA wurden von der Manipulation der Gegenüberstellungsprozedur beeinflusst.

In den einzelnen Experimenten 1 bis 4 gab es keinen signifikanten Unterschied zwischen simultanen und sequentiellen Gegenüberstellungen in Parameter dP . Da die simultane Gegenüberstellungsprozedur aber immer zuverlässig deskriptiv höhere dP -Werte erzielte, wurde eine kombinierte Analyse durchgeführt. Dafür wurde ein Basismodell definiert, das alle Daten der Experimente 1 bis 4 zusammenfasste und die gleichen Restriktionen wie zuvor aufwies. Die Details der kombinierten Analyse finden sich in Winter et al. (2022). In einem zweiten Schritt wurde Parameter dP sowohl für die simultanen als auch für die sequentiellen Gegenüberstellungen über alle Bedingungen der vier Validierungsexperimente mit langen Videos hinweg gleichgesetzt. Daraus ergaben sich insgesamt zwei Parameterschätzer für den Anwesenheits-Detektions-Parameter dP ; einen für die simultanen Gegenüberstellungen ($dP_{\text{simultan}} = .31$, Standardfehler = .01) und einen für die sequentiellen Gegenüberstellungen ($dP_{\text{sequentiell}} = .26$, Standardfehler = .01). Ein statistischer Vergleich über alle vier Validierungsexperimente hinweg zeigte einen signifikanten Unterschied zwischen den Gegenüberstellungsprozeduren im Anwesenheits-Detektions-Parameter dP

zugunsten der simultanen Gegenüberstellungen [$\Delta G^2(1) = 10.13, p = .001$]. Parameter dP ist dem Konzept der Diskriminationsfähigkeit, das beispielsweise mit pAUCs gemessen wird, relativ ähnlich (Winter et al., 2022). In der Literatur gibt es unter Verwendung von pAUCs Evidenz dafür, dass mit simultanen Gegenüberstellungen eine höhere Diskriminationsfähigkeit einher geht, als mit sequentiellen Gegenüberstellungen (Carlson & Carlson, 2014; Dobolyi & Dodson, 2013; Horry et al., 2021; Mickes et al., 2012; Neuschatz et al., 2016; Seale-Carlisle & Mickes, 2016; Seale-Carlisle et al., 2019). In anderen Quellen findet man hingegen keinen Unterschied zwischen beiden Gegenüberstellungsprozeduren hinsichtlich der Diskriminationsfähigkeit (Andersen et al., 2014; Clark, 2012; Flowe et al., 2016; Gronlund et al., 2012; Meisters et al., 2018; Sučić et al., 2015). Das gefundene Ergebnismuster, fügt sich folglich gut in die gemischten Befunde ein, die auf Basis von ROC-Kurven gefunden wurden. Es scheint eine immense statistische Power vonnöten zu sein, um den Überlegenheitseffekt der Diskriminationsfähigkeit von simultanen Gegenüberstellungsprozeduren im Vergleich zu sequentiellen Gegenüberstellungsprozeduren zu finden.

Entgegen der Befunde in der Literatur (z. B. Clark, 2012; Gronlund et al., 2009; Palmer & Brewer, 2012; Steblay et al., 2011) zeigte sich in der vorliegenden Arbeit in sequentiellen Gegenüberstellungen eine höhere Tendenz zu Raten als in simultanen Gegenüberstellungen. Der Rate-Parameter g fiel in allen vier Validierungsexperimenten für die sequentiellen Gegenüberstellungen signifikant größer aus als für die simultanen Gegenüberstellungen. In der Literatur werden aber sequentielle Gegenüberstellungen regelmäßig mit einer geringeren Ratetendenz assoziiert (Clark, 2012; Gronlund et al., 2009; Palmer & Brewer, 2012; Steblay et al., 2011). Ein Artefakt des 2-HT EIMs, das diesen überraschenden Befund produzierte, lässt sich ausschließen, da in den Experimenten 1 bis 4 auch die Identifikationsraten der tatbeteiligten Person und der unschuldigen verdächtigten Person für sequentielle Gegenüberstellungen jeweils höher waren als für simultane Gegenüberstellungen. Der gefundene Effekt findet sich daher schon auf Basis der rohen Antwortraten der einzelnen Experimente. Dies weist darauf hin, dass die Versuchspersonen in den sequentiellen Gegenüberstellungen generell mehr Identifikationen tätigten als in den simultanen. Eine Erklärung dieses Phänomens ist daher nicht im angewendeten Messmodell zu finden. Plausibler ist, dass eine Besonderheit der gewählten Methodik dieses überraschende Befundmuster verursacht haben könnte. In den vier aktuellen Validierungsexperimenten wurde die Methodik eng an der polizeilichen Praxis orientiert, um eine hohe ökologische Validität zu gewährleisten. Es wurde deswegen von der Nutzung von

First-Yes-Counts-Instruktionen in sequentiellen Gegenüberstellungen abgesehen. Versuchspersonen konnten in den sequentiellen Gegenüberstellungen aller vier Validierungsexperimente mehrfache Identifikationen vornehmen, um eine frühere Entscheidung zu revidieren, falls später eine Person in der Gegenüberstellung gezeigt wurde, die ihrer Erinnerung an die tatbeteiligte Person besser entsprach. Die letzte Identifikation einer Versuchsperson wurde schließlich für die Datenauswertung verwendet. Dies steht im Gegensatz zu vielen Laborstudien in der Literatur, die First-Yes-Counts-Instruktionen verwenden (z. B. Horry et al., 2021; Kaesler et al., 2020; Memon & Gabbert, 2003; Sauer et al., 2008).

Horry et al. (2021) untersuchten als erste systematisch den Einfluss von First-Yes-Counts-Instruktionen auf Augenzeugenidentifikationen. In den Versuchsbedingungen mit First-Yes-Counts-Instruktionen teilten sie ihren Versuchspersonen mit, dass nur ihre erste Identifikation zählen würde und verglichen diese mit Versuchsbedingungen, in denen die gleichen Instruktionen ohne den Abschnitt mit den First-Yes-Counts-Instruktionen gegeben wurden. Ihre Ergebnisse legen nahe, dass die Verwendung von First-Yes-Counts-Instruktionen keinen Einfluss auf die Diskriminationsfähigkeit eines Augenzeugen bzw. einer Augenzeugin hat, wohl aber die Tendenz zu Raten reduziert. Um den Einfluss von First-Yes-Counts-Instruktionen auf Augenzeugen bzw. Augenzeuginnen besser zu verstehen, wird im Folgenden eine erneute Auswertung der Daten von Horry et al. (2021) mithilfe des 2-HT EIMs sowie eine konzeptuelle Replikation ihrer Befunde in Experiment 5 vorgestellt.

Der Einfluss von First-Yes-Counts-Instruktionen

Die genauere Betrachtung der Prozesse, die Einfluss auf sequentielle Gegenüberstellungen haben, ist interessant und wichtig, da diese im Zuge von Empfehlungen aus der Wissenschaft (Wells et al., 1998) weltweit in vielen Polizeidistrikten als Standardprozedur eingeführt wurden (Bill Blackwood Law Enforcement Management Institute of Texas, 2022; Bundesministerium des Innern und für Heimat [BMI], 2021; Horry, Memon, et al., 2012; Seale-Carlisle & Mickes, 2016). Wie bereits genannt ist es bei sequentiellen im Gegensatz zu simultanen Gegenüberstellungsverfahren prinzipiell möglich, mehr als eine Auswahlentscheidung vorzunehmen. Dies erschwert die Auswertung der Daten einer Gegenüberstellungsstudie, da nicht eindeutig ist, ob Versuchspersonen beispielsweise ihre vorherige Auswahl geändert haben (Stebly et al., 2011) und wirft die Frage auf, welche der Entscheidungen als „die richtige“ zu

werten ist. Um dieses Problem von vorneherein zu unterbinden, werden in vielen wissenschaftlichen Untersuchungen daher First-Yes-Counts-Instruktionen genutzt (z. B. Horry et al., 2021; Kaesler et al., 2020; Memon & Gabbert, 2003; Sauer et al., 2008).

First-Yes-Counts-Instruktionen finden auf unterschiedliche Arten und Weisen Anwendung: In einigen Studien wird nur die erste Identifikation einer Versuchsperson gezählt, aber die sequentielle Gegenüberstellung läuft dennoch weiter (Carlson et al., 2016; Carlson et al., 2008; Flowe, 2011; Gronlund et al., 2009; Gronlund et al., 2012; Horry, Palmer, et al., 2012; Lindsay & Bellinger, 1999; Mickes et al., 2012; Toftely, 2015). In anderen Studien bricht die sequentielle Gegenüberstellung hingegen tatsächlich nach der ersten Identifikation ab (Brewer et al., 2012; Carlson & Carlson, 2014; Dobolyi & Dodson, 2013; Dunn et al., 2022; Flowe & Ebbesen, 2007; Flowe et al., 2016; Kaesler et al., 2020; Kneller et al., 2001; Memon & Gabbert, 2003; Sauer et al., 2008). Diese Verfahren stehen in einem starken Kontrast zu echten Gegenüberstellungen, die in der polizeilichen Praxis durchgeführt werden und in denen normalerweise keine solche Abbruchregel genutzt wird (Bill Blackwood Law Enforcement Management Institute of Texas, 2022; Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; Horry, Memon, et al., 2012; State Bar of Michigan: Eyewitness Identification Task Force, 2015; Steblay et al., 2011; U.S. Department of Justice: Office of the Deputy attorney General, 2017; Wells et al., 2011; Wells, Steblay, et al., 2015). Diese Diskrepanz zwischen der wissenschaftlichen Forschung und der Anwendung von Gegenüberstellungsprozeduren in der Praxis macht es erforderlich, die Prozesse zu untersuchen und zu verstehen, die First-Yes-Counts-Instruktionen auslösen.

Neben weiteren Manipulationen präsentierten Horry et al. (2021) in ihrer Studie ihren Versuchspersonen sequentielle Gegenüberstellungen mit bzw. ohne First-Yes-Counts-Instruktionen. Sie fanden eine deskriptiv geringere Identifikationsrate der verdächtigten Person in der Bedingung mit First-Yes-Counts-Instruktionen im Vergleich zu der Bedingung ohne solche Instruktionen (.26 vs. .29) sowie eine signifikant höhere Zurückweisungsrate von Gegenüberstellungen mit First-Yes-Counts-Instruktionen im Vergleich zu der Versuchsbedingung ohne First-Yes-Counts-Instruktionen (.53 vs. .43). Diese Ergebnisse wurden von Horry et al. (2021) als eine geringere Tendenz zu raten unter Anwendung von First-Yes-Counts-Instruktionen als ohne diese Instruktionen interpretiert. Diese Schlussfolgerung erscheint plausibel, da Versuchspersonen, die nur eine einzige Möglichkeit hatten, eine Identifikation vorzunehmen,

vermutlich vorsichtiger waren, diese Identifikation zu tätigen, da die nächste, noch nicht gezeigte Person potentiell eine bessere Passung zur Erinnerung der Versuchsperson aufweisen konnte als die aktuelle. Eine weitere Identifikation war aber in diesem Fall nicht mehr möglich. Eine Versuchsperson, die beliebig viele Identifikationen vornehmen konnte, musste diese Vorsicht hingegen nicht zeigen. Horry et al. (2021) fanden außerdem keinen signifikanten Unterschied zwischen den Versuchsbedingungen mit und ohne First-Yes-Counts-Instruktionen in pAUCs. Dies interpretierten sie als keinen Einfluss von First-Yes-Counts-Instruktionen auf die Diskriminationsfähigkeit ihrer Versuchspersonen.

Bezogen auf die wissenschaftliche Forschung bedeuten diese Befunde, vor allem die reduzierte Ratetendenz, dass viele Studien zu Augenzeugenidentifikationen eine geringe ökologische Validität aufweisen könnten. Die grundsätzliche Annahme, sequentielle Gegenüberstellungsverfahren würden ein konservatives Antwortverhalten und damit eine geringere Ratetendenz induzieren als simultane (Clark, 2012; Gronlund et al., 2009; Palmer & Brewer, 2012; Steblay et al., 2011), könnte ein Artefakt der angewendeten First-Yes-Counts-Instruktionen sein. Aussagen, die daher auf der Basis solcher Befunde getroffen werden, reflektieren ggf. also gar nicht die Prozesse, die während Gegenüberstellungsverfahren in der Praxis ablaufen.

Allerdings weist auch die Studie von Horry et al. (2021) Probleme auf, die die Validität ihrer Ergebnisse bedrohen könnten. Das genutzte Messmodell war, erstens, vorab nicht validiert. Zur Betrachtung der Ratetendenz der Versuchspersonen verglichen Horry et al. (2021) zwei verschiedene Identifikationsraten zwischen den Versuchsbedingungen: die Rate der Identifikationen einer verdächtigten Person und die Rate der Zurückweisungen der Gegenüberstellung. Es wurde angenommen, dass eine Erhöhung der Rate von Identifikationen einer verdächtigten Person sowie eine Verringerung der Rate der Zurückweisungen einer Gegenüberstellung durch eine erhöhte Ratetendenz zustande kommen könnte. Diese Annahme weist eine hohe Augenscheinvalidität auf. Es ist aber nicht abschließend spezifiziert, wie sich die gemessenen Größen absolut und relativ zueinander verändern sollen, wenn die Wahrscheinlichkeit von Identifikationsentscheidungen auf der Basis von Rateprozessen sich verändert. Die beiden Raten wurden also nur implizit kombiniert, um Schlussfolgerungen über die Ratetendenz der Versuchspersonen zu ermöglichen. Im Idealfall ist ein genutztes Maß aber experimentell validiert, damit nachzuweisen ist, dass seine Ergebnisse verlässlich sind und es wirklich die Prozesse abbildet, die es abbilden soll.

Eng damit verbunden ist der zweite Kritikpunkt zur Validität der Ergebnisse von Horry et al. (2021): Nur der Unterschied in den Raten der Zurückweisung der Gegenüberstellungen zwischen den Versuchsbedingungen mit und ohne First-Yes-Counts-Instruktionen wurde signifikant und war für die Versuchsbedingung mit First-Yes-Counts-Instruktionen niedriger als für die Versuchsbedingung ohne solche Instruktionen. Die Rate der Identifikationen der verdächtigten Person unterschied sich hingegen nicht zwischen den Versuchsbedingungen. Die Interpretation, dass in der Bedingung mit First-Yes-Counts-Instruktionen weniger geraten wurde als in der Versuchsbedingung ohne solche Instruktionen, wurde ausschließlich auf dem Unterschied in der Rate der Zurückweisungen basiert, der signifikant wurde. Würde man hingegen aber nur die Identifikationsrate der verdächtigten Person alleine betrachten, würde sich ein anderes Bild zeigen, nämlich dass kein signifikanter Unterschied in der Raten tendenz der Versuchspersonen vorlag. Die Ergebnisse sind hier folglich uneindeutig, da nur einer der zwei Unterschiede signifikant wurde und der andere nicht. Es ist durch die fehlende Validierung unklar, wie sich diese Ergebnisse insgesamt integrieren lassen. Um dieses Problem zu vermeiden, wäre eine einzelne Messvariable, die die Raten tendenz abbilden kann, und ein einzelner, auf dieser Messvariablen basierender statistischer Test, hilfreich.

Drittens wird von Horry et al. (2021) für die Berechnung der pAUCs, mit welchen die Diskriminationsfähigkeit der Versuchspersonen untersucht wurde, ein gänzlich anderes Maß verwendet als für die Untersuchung der Raten tendenz, bei welcher Identifikationsraten verglichen wurden. Dies ist ungünstig, weil beispielsweise beiden Maßen unterschiedliche Annahmen zugrunde liegen. Besser wäre ein Messmodell, das alle Prozesse von Augenzeugenidentifikationen und somit verschiedene Maße integriert. Diesen Maßen würden dann die gleichen Annahmen zugrunde liegen. Im Idealfall sollte sich zudem auch noch prüfen lassen, ob ein verwendetes Messmodell zu den Daten passt, auf die es angewendet wird.

Das in dieser Arbeit in Experiment 1 bis 4 validierte 2-HT EIM ist in der Lage, auf diese drei Bedrohungen der Validität der Schlussfolgerungen von Horry et al. (2021) zu reagieren: die Modellparameter sind validiert, es gibt ein einzelnes Maß für die Raten tendenz von Versuchspersonen und die Fähigkeit der Versuchspersonen, die An- und Abwesenheit der tatbeteiligten Person zu detektieren, wird im gleichen Messmodell gemessen. Zusätzlich lässt sich die Passung des Modells zu den Daten bestimmen. Eine Reanalyse der Daten von Horry et al. (2021) mit dem 2-HT EIM ist daher ein erster Schritt, um die Validität der gefundenen Ergebnisse zu überprüfen.

Wichtig ist außerdem, dass die Annahmen, die auf Basis der Befunde von Horry et al. (2021) getroffen werden können, sich zusätzlich auch auf reliable Befunde stützen (Open Science Collaboration, 2015). Eine unabhängige Replikation ist dabei der Goldstandard, um das Vertrauen in neue Befunde zu erhöhen (Simons, 2014). Es ist daher ein naheliegender weiterer Schritt, eine konzeptionelle Replikation der sequentiellen Versuchsbedingungen von Horry et al. (2021) durchzuführen, um zu prüfen, ob sich ihre Ergebnisse bestätigen lassen. Im Folgenden werden zuerst die Ergebnisse der Reanalyse und dann in Experiment 5 die Ergebnisse der konzeptionellen Replikation vorgestellt. Der Fokus lag hierbei für beide Analysen auf den Ergebnissen des Rate-Parameters g und des Anwesenheits-Detektions-Parameters dP , da diese konzeptuell den Prozessen der Ratetendenz und der Diskriminationsfähigkeit am nächsten kommen (Winter et al., 2022). Basierend auf den Befunden von Horry et al. (2021) wurde dabei in beiden Untersuchungen kein Unterschied im Anwesenheits-Detektions-Parameter dP zwischen den Versuchsbedingungen erwartet, aber ein verringerter Rate-Parameter g in der Versuchsbedingung mit Fist-Yes-Counts-Instruktionen im Vergleich zu der Bedingung ohne solche Instruktionen.

Reanalyse der Daten von Horry et al. (2021)

Zur Überprüfung der Validität der Ergebnisse von Horry et al. (2021) wurden die Daten ihrer sequentiellen Versuchsbedingungen mit und ohne First-Yes-Counts-Instruktionen, die in der Originalveröffentlichung im Format der 2×3 -Datenstruktur vorlagen, mithilfe des 2-HT EIMs ausgewertet. Da Horry et al. (2021) außerdem den Faktor der Position der verdächtigten Person in der Gegenüberstellung manipulierten, was für die hier verfolgte Fragestellung irrelevant ist, wurden die Antworthäufigkeiten der sequentiellen Versuchsbedingungen über diesen zusätzlichen Faktor zusammengefasst.

Um den Einfluss von First-Yes-Counts-Instruktionen auf Augenzeugenidentifikationen zu untersuchen, präsentierten Horry et al. (2021) ihren Versuchspersonen jeweils eines von vier Tatvideos und im Anschluss daran eine von vier Gegenüberstellungen. Die Gegenüberstellung konnte dabei entweder eine tatbeteiligte Person aufweisen, die zuvor in dem Tatvideo gesehen wurde, oder eine unschuldige verdächtige Person. Ähnlich zu der genutzten Methodik der weiter oben beschriebenen Experimente 1 bis 4 handelte es sich hier um eine tatbeteiligte Person aus einem der anderen, nicht gesehenen Videos. Indem Horry et al. (2021) jeweils zwei der vier tatbeteiligten Personen zueinander zuordneten und diese basierend auf dem zuvor

gesehenen Video für eine Versuchsperson entweder tatbeteiligt oder unschuldig verdächtig waren, wies auch ihre Methodik eine designierte unschuldige verdächtige Person auf. Jede der vier Gegenüberstellungen besaß eine eigene Menge an Vergleichspersonen, was die Methodik von Horry et al. (2021) von der gekreuzten Gegenüberstellungsprozedur unterscheidet, bei der sich ein Paar tatbeteiligter Personen die gleichen Vergleichspersonen teilt (Menne et al., 2022, 2023; Winter et al., 2022).

Den Versuchspersonen in Horry et al. (2021) wurden für Laborstudien zu Gegenüberstellungen übliche, zweiseitige Instruktionen präsentiert. Die Versuchspersonen der First-Yes-Counts-Versuchsbedingung wurden zusätzlich instruiert, dass nur ihre erste Identifikation zählen würde. Das restliche Vorgehen war für beide sequentiellen Versuchsbedingungen gleich: Unabhängig von der Versuchsbedingung war es allen Versuchspersonen möglich, multiple Identifikationen zu tätigen. Wenn multiple Identifikationen getätigt wurden, so mussten diese in einer zweiten Runde von den Versuchspersonen selbst aufgelöst werden. In der Versuchsbedingung mit First-Yes-Counts-Instruktionen wurde bei Vorlage multipler Identifikationen allerdings nur die erste Identifikation der Versuchspersonen für die Datenauswertung gezählt, in der Versuchsbedingung ohne First-Yes-Counts-Instruktionen hingegen die Identifikation, die in der zweiten Runde final getätigt wurde. Weitere Details zur Methodik finden sich in Horry et al. (2021).

Die Datenauswertung fand analog zu Experiment 1 statt. Da auch Horry et al. (2021) in allen Versuchsbedingungen das gleiche Material für die Gegenüberstellungen verwendeten, wurden in der Reanalyse der Unfairness-Parameter b und der Abwesenheits-Detektions-Parameter dA jeweils über beide betrachtete Versuchsbedingungen hinweg gleichgesetzt, um ein Basismodell zu erhalten, das so einfach wie möglich war. Das so restringierte Basismodell passte zu den Daten [$G^2(2) = 2.42, p = .298$]. Die deskriptiven Statistiken zum Anwesenheits-Detektions-Parameter dP und zum Rate-Parameter g sind in Abbildung 7 dargestellt. Alle weiteren deskriptiven Statistiken finden sich in Winter et al. (2023).

Der Anwesenheits-Detektions-Parameter dP unterschied sich nicht signifikant zwischen den sequentiellen Gegenüberstellungen mit und ohne First-Yes-Counts-Instruktionen [$\Delta G^2(1) = 3.05, p = .081$]. Der Rate-Parameter g war in der sequentiellen Versuchsbedingung mit First-Yes-Counts-Instruktionen signifikant kleiner als in der Versuchsbedingung ohne solche Instruktionen [$\Delta G^2(1) = 5.31, p = .021$]. Diese Ergebnisse passen zu den Ergebnissen aus der Originalarbeit von Horry et al. (2021). Sie unterstützen die Validität der Befunde, dass sich unter Verwendung von First-Yes-

Counts-Instruktionen kein Einfluss auf die Diskriminationsleistung, wohl aber auf die Ratetendenz der Versuchspersonen zeigt. Versuchspersonen rieten unter Verwendung von First-Yes-Counts-Instruktionen weniger als Versuchspersonen, denen keine First-Yes-Counts-Instruktionen präsentiert wurden.

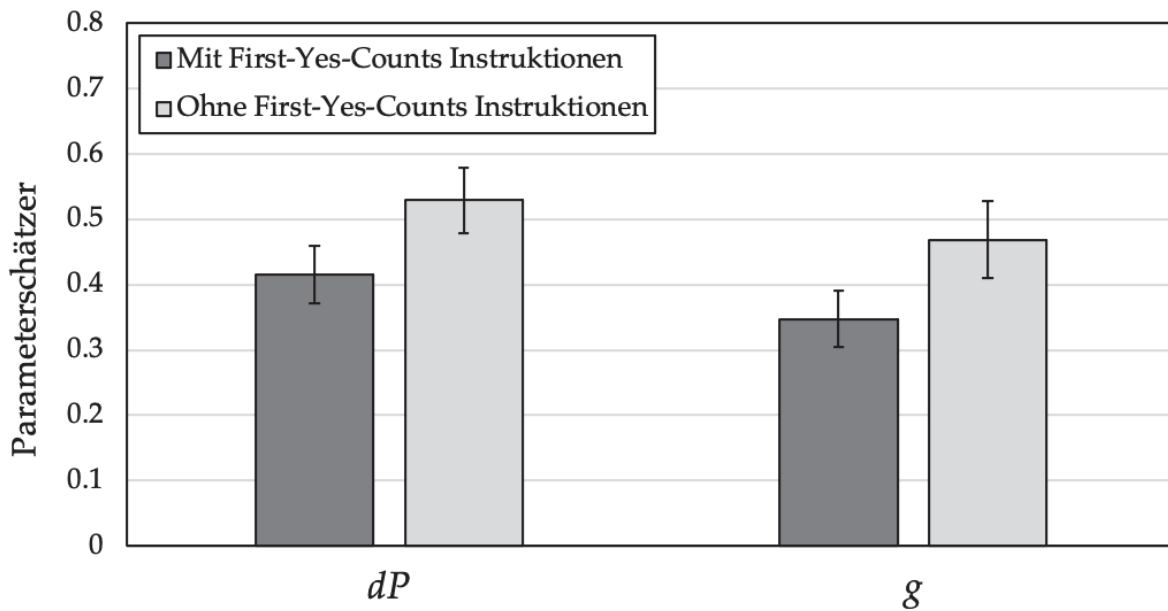


Abbildung 7. Parameterschätzer des Anwesenheits-Detektions-Parameters dP und des Rate-Parameters g des 2-HT EIMs in der Reanalyse von Horry et al. (2021). Die Balken repräsentieren die Wahrscheinlichkeit der Selektion einer tatbeteiligten Person aufgrund der Detektion ihrer Anwesenheit in der Gegenüberstellung und die Wahrscheinlichkeit einer rate-basierten Auswahl einer Person in der Gegenüberstellung als Funktion der Manipulation der präsentierten Instruktionen (mit First-Yes-Counts-Instruktionen vs. ohne First-Yes-Counts-Instruktionen). Die Fehlerbalken zeigen den Standardfehler.

Experiment 5

In Experiment 5 wurde die Reliabilität der Ergebnisse durch eine konzeptuelle Replikation der Studie von Horry et al. (2021) geprüft. Um dies zu erreichen, wurden in Experiment 5 sequentielle Gegenüberstellungen mit bzw. ohne First-Yes-Counts-Instruktionen präsentiert. Die Methodik war eng an der von Experiment 1 orientiert, mit folgenden Ausnahmen: In der First-Yes-Counts-Versuchsbedingung wurden vor den Gegenüberstellungen Instruktionen präsentiert, die die Versuchspersonen darauf hinwiesen, dass nur ihre erste Antwort zählte. Diese Instruktionen wurden vor jeder der vier Gegenüberstellungen wiederholt. Zudem war es den Versuchspersonen in der First-Yes-Counts-Versuchsbedingung nicht möglich, mehr als eine Identifikation zu treffen, da nach der ersten Identifikation zwar die Gegenüberstellung

nicht abbrach, aber der Button für die „Ja, diese Person war dabei“-Auswahl deaktiviert wurde. Es wurden zudem nur die langen Videos präsentiert.

Anders als bei Horry et al. (2021) wurden alle Personen einer Gegenüberstellung nur einmal präsentiert, selbst bei mehrfachen Identifikationen gab es keine zweite Gegenüberstellungsrunde. Berücksichtigt wurde in der Versuchsbedingung mit First-Yes-Counts-Instruktionen die erste (und einzige) Identifikation, die getroffen wurde und in der Versuchsbedingung ohne First-Yes-Counts-Instruktionen die letzte Identifikation. Es wurden wieder vier anstelle von nur einer tatbeteiligten Person im Video präsentiert und vier Gegenüberstellungen durchgeführt. Die Positionen der Personen in einer Gegenüberstellung und die Reihenfolge der Gegenüberstellungen selbst waren erneut randomisiert. Weitere Details zur Methode finden sich in Winter et al. (2023). Unter Annahme der Reliabilität der Ergebnisse von Horry et al. (2021) sollten sich in Experiment 5 die gleichen Ergebnisse wie in der Reanalyse zeigen.

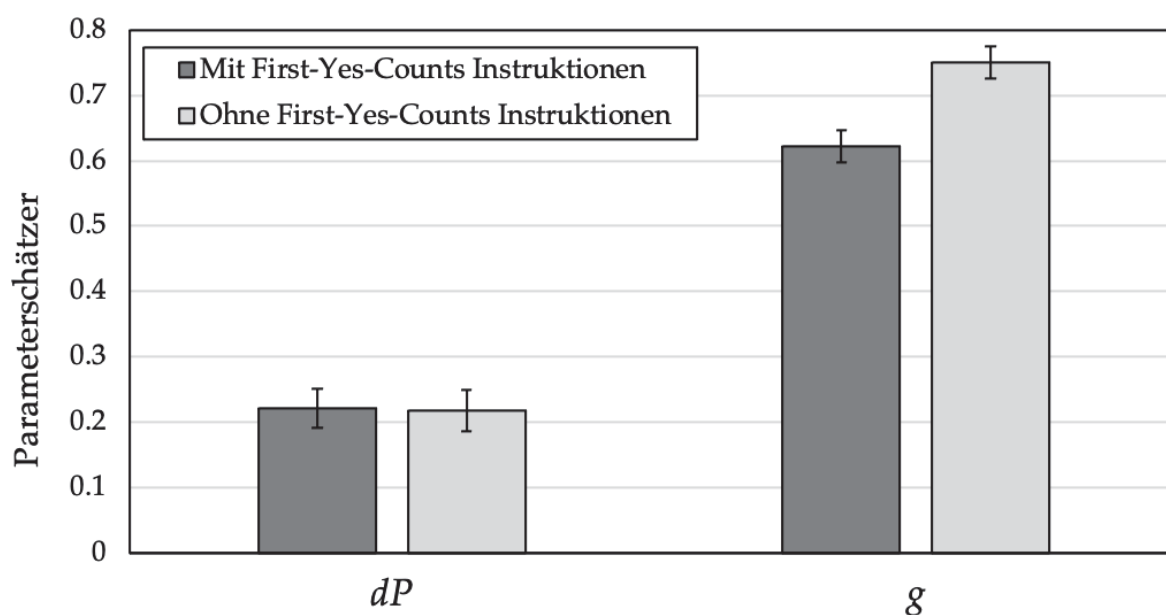


Abbildung 8. Parameterschätzer des Anwesenheits-Detektions-Parameters dP und des Rate-Parameters g des 2-HT EIMs in Experiment 5. Die Balken repräsentieren die Wahrscheinlichkeit der Selektion einer tatbeteiligten Person aufgrund der Detektion ihrer Anwesenheit in der Gegenüberstellung und die Wahrscheinlichkeit einer rate-basierten Auswahl einer Person in der Gegenüberstellung als Funktion der Manipulation der präsentierten Instruktionen (mit First-Yes-Counts-Instruktionen vs. ohne First-Yes-Counts-Instruktionen). Die Fehlerbalken zeigen den Standardfehler.

Die Restriktionen des Basismodells und die Datenauswertung für Experiment 5 fanden analog zu Experiment 1 statt. Das so restringierte Basismodell passte zu den Daten [$G^2(2) = 3.09, p = .214$]. Die Parameterschätzer des Anwesenheits-Detektions-Parameters dP und des Rate-Parameters g sind in Abbildung 8 dargestellt. Alle weiteren deskriptiven Statistiken finden sich in Winter et al. (2023). Parameter dP unterschied

sich nicht zwischen den Versuchsbedingungen mit und ohne First-Yes-Counts-Instruktionen [$\Delta G^2(1) = 0.01, p = .936$]. Parameter g war in der Versuchsbedingung mit First-Yes-Counts-Instruktionen signifikant kleiner als in der Versuchsbedingung ohne diese Instruktionen [$\Delta G^2(1) = 21.16, p < .001$]. Diese Befunde stützen die Reliabilität der Ergebnisse der Reanalyse und von den Originaldaten von Horry et al. (2021).

Diskussion der Reanalyse von Horry et al. (2021) und von Experiment 5

In der Reanalyse von Horry et al. (2021) und in Experiment 5 wurden sowohl die Validität als auch die Reliabilität der Ergebnisse von Horry et al. (2021) geprüft. Dabei konnte in beiden Untersuchungen das gleiche Ergebnismuster beobachtet werden. Sowohl in der Reanalyse als auch in Experiment 5 wurde kein Unterschied im Anwesenheits-Detektions-Parameter dP zwischen den sequentiellen Versuchsbedingungen mit und ohne First-Yes-Counts-Instruktionen gefunden. Der Rate-Parameter g war in beiden Analysen für die Versuchsbedingungen mit First-Yes-Counts-Instruktionen signifikant kleiner als für die Versuchsbedingungen ohne solche Instruktionen. Dies entspricht auch der Interpretation der originalen Befunde von Horry et al. (2021). Zudem unterstützen die Ergebnisse der Reanalyse und von Experiment 5, die mithilfe des 2-HT EIMs erzielt wurden, die Validität und Reliabilität der Ergebnisse von Horry et al. (2021).

Wie bereits erwähnt ist eine Analyse der Daten mit dem 2-HT EIM relevant, weil ihre Validität nicht durch die drei auf Horry et al. (2021) zutreffenden Kritikpunkte bedroht ist. Das 2-HT EIM ist ein validiertes Modell, das in der Lage ist, den Rateprozess in einer Augenzeugengegenüberstellung innerhalb eines Messwertes zu integrieren, der nur einen einzigen statistischen Signifikanztests erfordert. Außerdem integriert das 2-HT EIM auch die Messung der Fähigkeit, die An- und Abwesenheit der tatbeteiligten Person zu detektieren und die Messung der Ratetendenz in ein einziges Messmodell. Es lässt sich feststellen, dass sich konsistent die gleichen Ergebnisse finden lassen wie in Horry et al. (2021), obwohl unterschiedliche Maße und eine nicht identische Methode verwendet wurden. Diese Tatsache spricht für die Robustheit des Effektes, den First-Yes-Counts-Instruktionen auf Augenzeugenidentifikationen haben und für die Fähigkeit des 2-HT EIMs, kognitive Prozesse im Auswahlverhalten von Augenzeugen bzw. Augenzeuginnen zuverlässig zu reflektieren.

Allgemeine Diskussion

In der vorliegenden Arbeit wurde das 2-HT EIM als ein neues multinomiales Verarbeitungsbaummodell für die Datenauswertung von Augenzeugenidentifikationen vorgestellt, validiert und zur Untersuchung des Einflusses von First-Yes-Counts-Instruktionen auf sequentielle Gegenüberstellungen angewendet. Augenzeugenidentifikationen weisen generell eine hohe Fehlerwahrscheinlichkeit auf (Innocence Project, 2023) und die bisherigen Messmodelle zur Auswertung der gewonnenen Daten haben eindeutige Schwächen. Im Maß des Diagnostizitätsverhältnisses ist die Diskriminationsfähigkeit mit der Ratetendenz konfundiert (Gronlund et al., 2012; Mickes et al., 2012; Wixted & Mickes, 2014). ROC-Kurven umgehen dieses Problem zwar, aber es muss für die Bildung der Kurven auf einen vereinfachten Teil der Datenstruktur von Augenzeugenidentifikationen zurückgegriffen werden. Folglich gehen Informationen verloren und kognitive Prozesse können nicht oder nur kontaminiert geschätzt werden (Smith et al., 2018; Wells, Smalarz, et al., 2015; Wells, Smith, et al., 2015). Die Einführung eines neuen statistischen Messmodells war daher erforderlich, um zuverlässige Aussagen über die kognitiven Prozesse treffen zu können, die Augenzeugenidentifikation zu Grunde liegen. Nur wenn diese Prozesse angemessen abgebildet werden, lassen sich zuverlässige Aussagen über die Güte einer Gegenüberstellungsprozedur treffen. Dies ist eine wichtige Voraussetzung, um die generell geringe Genauigkeit von Augenzeugenidentifikationen optimieren zu können.

In der vorliegenden Arbeit wurde als Lösungsvorschlag das neue 2-HT EIM vorgestellt und durch die Experimente 1 bis 4 validiert. Durch die Validierung konnte sichergestellt werden, dass die Modellparameter die latenten kognitiven Prozesse abbilden, für die sie intendiert sind. Mithilfe des 2-HT EIM lässt sich die vollständige 2×3 -Datenstruktur von Augenzeugenidentifikationsaufgaben nutzen. So können die korrekten Identifikationen einer tatbeteiligten Person, die falschen Identifikationen einer unschuldigen verdächtigten Person, die Identifikationen von Vergleichspersonen in und die Zurückweisungen von Gegenüberstellungen mit an- bzw. abwesender tatbeteiligter Person berücksichtigt werden. Das 2-HT EIMs ist daher in der Lage, vier latente kognitive Prozesse abzubilden, die während Augenzeugenidentifikationen auftreten können.

Jedes der vier Validierungsexperimente diente dazu, einen der vier Modellparameter zu validieren. In Experiment 1 wurde der Anwesenheits-Detektions-Parameter dP validiert, der die Fähigkeit eines Augenzeugen bzw. einer Augenzeugin abbildet,

eine anwesende tatbeteiligte Person korrekt zu identifizieren. In Experiment 2 wurde der Unfairness-Parameter b validiert, der die Wahrscheinlichkeit repräsentiert, dass ein Augenzeuge bzw. eine Augenzeugin eine verdächtige Person in einer Gegenüberstellung aufgrund hervorstechender Merkmale auswählt. In Experiment 3 wurde der Rate-Parameter g validiert, der die Tendenz eines Augenzeugen bzw. einer Augenzeugin abbildet, eine Person aus der Gegenüberstellung aufgrund von Raten zu wählen. In Experiment 4 wurde der Abwesenheits-Detektions-Parameter dA validiert, der in Gegenüberstellungen ohne anwesende tatbeteiligte Person die Fähigkeit des Augenzeugen bzw. der Augenzeugin abbildet, zu erkennen, dass keine der Personen in der Gegenüberstellung die tatbeteiligte Person gewesen sein kann und die Gegenüberstellung daher zurückzuweisen ist. Die Ergebnisse der Experimente 1 bis 4 zeigen, dass alle vier Modellparameter valide auf die experimentellen Manipulationen reagierten, wodurch das 2-HT EIM zuverlässig auf neue Fragestellungen angewendet werden kann. Die Validität des 2-HT EIMs wird zusätzlich dadurch gestützt, dass das Modell auch auf bereits bestehende Daten aus der Literatur angewendet werden konnte, wobei die Ergebnisse der hier vorliegenden Validierungsexperimente repliziert wurden (Menne et al., 2022). Interessant zu betrachten war außerdem, dass der vom Modell geschätzte Wert des Unfairness-Parameters b , der einen direkten Schätzer für die Unfairness einer Gegenüberstellung ausgibt, in den Experimenten 1 bis 5 der vorliegenden Arbeit, sowie in der Reanalyse der Daten von Horry et al. (2021) immer sehr klein war. Dies zeigt, dass alle präsentierten Gegenüberstellungen ein hohes Maß an Fairness aufwiesen. Eine Ausnahme bilden hier die unfairen Bedingungen von Experiment 2, wo die Unfairness der Gegenüberstellungen explizit manipuliert wurde.

In den Experimenten 1 bis 4 zeigte sich zusätzlich, dass der Rate-Parameter g für sequentielle Gegenüberstellungen größer war als für simultane. Versuchspersonen rieten folglich in sequentiellen Gegenüberstellungen mehr als Versuchspersonen in simultanen Gegenüberstellungen. Dies ist überraschend, weil sequentielle Gegenüberstellungen eher mit einer geringeren Ratetendenz assoziiert werden als simultane Gegenüberstellungen (Clark, 2012; Gronlund et al., 2009; Palmer & Brewer, 2012; Steblay et al., 2011). Eine mögliche Ursache dieses Befundmusters wurde in der Tatsache vermutet, dass Versuchspersonen in den Experimenten 1 bis 4 in der Lage dazu waren mehr als eine Person in sequentiellen Gegenüberstellungen auszuwählen und keine First-Yes-Counts-Instruktionen verwendet wurden. Horry et al. (2021) untersuchten als erste systematisch den Einfluss von First-Yes-Counts-Instruktionen auf die Diskriminationsfähigkeit und das Rateverhalten von Augenzeugen bzw.

Augenzeuginnen. Sie fanden in ihrer Originalstudie keinen Einfluss dieser Instruktionen auf die Diskriminationsfähigkeit ihrer Versuchspersonen, aber eine verringerte Ratetendenz unter First-Yes-Counts Instruktionen. Um die Validität und Reliabilität dieser Befunde zu untersuchen, wurden eine Reanalyse der Daten von Horry et al. (2021) mithilfe des 2-HT EIMs, sowie in Experiment 5 eine konzeptuelle Replikation ihrer Studie durchgeführt. Diese zeigten, dass sich die von Horry et al. (2021) gefundenen Ergebnisse bestätigen ließen. Der Abwesenheits-Detektions-Parameter wurde nicht von First-Yes-Counts-Instruktionen beeinflusst, aber der Rate-Parameter war für die Versuchsbedingungen mit First-Yes-Counts-Instruktionen kleiner als für die Versuchsbedingungen ohne solche Instruktionen.

Die Robustheit des Effekts, dass First-Yes-Counts-Instruktionen das Rateverhalten von Augenzeugen bzw. Augenzeuginnen verringern ist relevant, da diese Instruktionen nur in der wissenschaftlichen Forschung verwendet werden (z. B. Horry et al., 2021; Kaesler et al., 2020; Memon & Gabbert, 2003; Sauer et al., 2008), in der Praxis aber keine Anwendung finden (Bill Blackwood Law Enforcement Management Institute of Texas, 2022; Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; Horry, Memon, et al., 2012; State Bar of Michigan: Eyewitness Identification Task Force, 2015; Steblay et al., 2011; U.S. Department of Justice: Office of the Deputy attorney General, 2017; Wells et al., 2011; Wells, Steblay, et al., 2015). In der polizeilichen Praxis ist es üblich, entweder alle Identifikationen zuzulassen und diese schließlich vor Gericht zu präsentieren, oder eine zweite Gegenüberstellungsrunde durchzuführen, in der multiple Identifikationen vom Augenzeugen bzw. der Augenzeugin selbst aufgelöst werden müssen (z. B. Horry, Memon, et al., 2012; Wells et al., 2011). Dies wird aus mehreren Gründen praktiziert. Der erste Grund ist, dass eine frühe Identifikation die funktionale Gegenüberstellungsgröße verringern würde, wenn die Gegenüberstellung nach dieser Identifikation abbricht (Wells et al., 2011; Wells, Steblay, et al., 2015). Typischerweise gibt es Regeln, wie viele Personen eine Gegenüberstellung umfassen muss, damit sie gültig ist. Die zulässige Gegenüberstellungsgröße kann dabei je nach Staat unterschiedlich sein (Bundesministerium des Innern und für Heimat [BMI], 2021; Home Office, 2017; U.S. Department of Justice: Office of the Deputy attorney General, 2017). Ein verfrühter Abbruch der Gegenüberstellung durch eine Identifikation könnte außerdem juristische Probleme aufwerfen, da ein Verfahren zur Beweissicherung nicht korrekt angewendet wurde: Es ist schließlich möglich, dass eine verdächtige Person erst spät in einer Gegenüberstellung präsentiert wird. Wird dann verfrüht eine Vergleichsperson identifiziert, würde die Gegenüberstellung abbrechen, ohne dass die verdächtige

Person dem Augenzeugen bzw. die Augenzeugin je präsentiert wurde. Die Reaktion eines Augenzeugen bzw. einer Augenzeugin auf die verdächtige Person zu prüfen, ist aber zentral für eine Gegenüberstellung, selbst wenn zuvor eine falsche Identifikation getätigt wurde (Wells et al., 2011; Wells, Steblay, et al., 2015). Mit Hinblick auf die verwendeten Instruktionen könnte die ökologische Validität vieler Studien in der Augenzeugenidentifikationsforschung folglich verringert sein, da deren Methodik nicht mit der polizeilichen Praxis übereinstimmt. Die Ergebnisse von Horry et al. (2021), der Reanalyse ihrer Daten und von Experiment 5 weisen darauf hin, dass die Häufigkeit von Raten in sequentiellen Gegenüberstellungen der polizeilichen Praxis durch die Erkenntnisse der wissenschaftlichen Forschung generell unterschätzt werden könnte. Gerade bei Fragestellungen, die strengen juristischen Regeln unterliegen, wie sie beispielsweise für die Verwendung von First-Yes-Counts-Instruktionen vorliegen, ist es jedoch sehr wichtig zu verstehen, welchen Einfluss diese Limitationen auf das Auswahlverhalten von Augenzeugen bzw. Augenzeuginnen besitzen, damit die realen Prozesse, die nicht oder nur schwer veränderbar sind, besser verstanden und berücksichtigt werden. Nur eine mit der Realität vergleichbare Methodik von Laborstudien erlaubt es, dass Aussagen über die Güte einer echten Gegenüberstellungsprozedur getroffen werden können. Das gilt über den Einsatz von First-Yes-Counts-Instruktionen, der in der vorliegenden Arbeit explizit untersucht wurde, hinaus für alle weiteren methodischen Eigenheiten, die von der polizeilichen Praxis abweichen. Im Folgenden werden daher Implikationen diskutiert, wie sich die Erforschung von Augenzeugenidentifikationen hinsichtlich der Verbesserung der Vergleichbarkeit von wissenschaftlicher Forschung und polizeilicher Praxis optimieren lassen könnte.

Auf Basis der Befunde der vorliegenden Arbeit lassen sich einige Implikationen ableiten, wie die Forschung zu Augenzeugenidentifikation und Gegenüberstellungsprozeduren gestaltet werden sollte um die Vergleichbarkeit mit der polizeilichen Praxis zu erhöhen. Für die Forschung ist es, erstens, essentiell, dass ein angewendetes Messmodell möglichst alle Informationen berücksichtigt, die bei einer Studie zum Thema Augenzeugenidentifikationen anfallen. Ein Rückgriff auf eine 2×2 -Datenstruktur ist nicht vertretbar, wenn es eine Möglichkeit gibt, auch die volle 2×3 -Datenstruktur zu nutzen. Außerdem lassen sich mithilfe des 2-HT EIMs basierend auf dieser 2×3 -Datenstruktur mehr der an Augenzeugenidentifikationen beteiligten kognitiven Prozesse in einem in sich geschlossenen kohärenten Messmodell messen. Zweitens sollte es zur Ergänzung der bereits vorliegenden Forschungserkenntnisse, die oftmals auf Basis von einer einzelnen tatbeteiligten Person gewonnen wurden,

mehr Untersuchungen mit mehreren tatbeteiligten Personen geben. Es kommt in der polizeilichen Praxis regelmäßig vor, dass Taten aus einer Gruppe heraus begangen werden (Bundeskriminalamt, 2021; Hobson et al., 2012; Tupper et al., 2019). Es ist in der polizeilichen Praxis also nicht unüblich, mehrere Gegenüberstellungen zu einer Tat durchzuführen. Welchen Einfluss dies im Gegensatz zu nur einer durchgeführten Gegenüberstellung auf das Auswahlverhalten von Augenzeugen bzw. Augenzeuginnen hat und welche Möglichkeiten es gibt, multiple tatbeteiligte Personen identifizieren zu lassen, sollte in zukünftiger Forschung stärker berücksichtigt werden. Drittens sollten in Zukunft beispielsweise mehr Laborstudien eine designierte unschuldige verdächtige Person verwenden, die sich im gleichen Maße von den Vergleichspersonen unterscheidet, wie eine tatbeteiligte Person. Oftmals wird in Laborstudien ein weiteres Vergleichsbild aus einer Datenbank als unschuldige verdächtige Person genutzt (z. B. Akan et al., 2021; Meisters et al., 2018; Mickes et al., 2017; Seale-Carlisle & Mickes, 2016; Wilcock & Bull, 2010; Zarkadi et al., 2009). Dies trägt aber der Tatsache, dass sich in der polizeilichen Praxis die Bilder von unschuldiger verdächtigter Person und tatbeteiligter Person systematisch von den Vergleichsbildern unterscheiden, nicht Rechnung. Der Polizei ist nicht bekannt, ob eine verdächtige Person unschuldig oder schuldig ist. Das Bild einer unschuldig verdächtigten Person wird also immer, so wie die Bilder von tatsächlich tatbeteiligten Personen, aus einer anderen Quelle, beispielsweise von Social Media, stammen als die Bilder der Vergleichspersonen, welche typischerweise aus einer Datenbank stammen und bzw. oder digital, verfremdet wurden (Bergold & Heaton, 2018; Colloff et al., 2016; Ministerium des Innern des Landes Nordrhein-Westfalen, 2023; U.S. Department of Justice: Office of the Deputy attorney General, 2017; Zarkadi et al., 2009). Es existieren folglich immer systematische Abweichungen zwischen dem Bild der verdächtigten Person und den Bildern der Vergleichspersonen. Diese systematische Abweichung sollte auch in der Forschung entsprechend berücksichtigt werden, beispielsweise durch die Nutzung der gekreuzten Gegenüberstellungsverfahren (Menne et al., 2023; Winter et al., 2022, 2023).

Besonderes Augenmerk legt die vorliegende Arbeit aber auf, viertens, die Abschaffung von First-Yes-Counts-Instruktionen in der wissenschaftlichen Forschung. Die Ergebnisse von Horry et al. (2021), die Reanalyse ihrer Daten und die konzeptuelle Replikation in Experiment 5 haben deutlich veranschaulicht, dass die Verwendung dieser Instruktionen das Raten in Gegenüberstellungen nachweislich verringert. Da in der polizeilichen Praxis keine solchen Instruktionen genutzt werden, spiegeln viele wissenschaftliche Untersuchungen daher nicht die Prozesse der Realität wider.

Es kann angenommen werden, dass das Ausmaß von Raten während einer sequentiellen Gegenüberstellung bisher systematisch unterschätzt wurde. Multiple Identifikationen könnten darauf hindeuten, dass ein Augenzeuge bzw. eine Augenzeugin sich während einer Gegenüberstellung unentschieden hat und deswegen eine weitere Person auswählt, die ihrer Erinnerung an die tatbeteiligte Person besser entspricht. Dies ist ein Prozess, der auch in einer simultanen Gegenüberstellung möglich ist. Eine bessere Alternative für die wissenschaftliche Forschung wäre daher die Verwendung der letzten Identifikation einer Versuchsperson, wie es auch in den Experimenten 1 bis 4 umgesetzt wurde. Alternativ kann auch eine zweite Gegenüberstellungsrunde genutzt werden, bei der die Versuchsperson etwaige multiple Identifikationen eigenständig auflöst (z. B. Home Office, 2017; Horry, Memon, et al., 2012). Von einer Nutzung von First-Yes-Counts-Instruktionen sollte in der zukünftigen wissenschaftlichen Forschung jedoch abgesehen werden, weil ihre Nutzung die ökologische Validität von so getroffenen Aussagen über Augenzeugegegensicherungen einschränken würde. Die Eigenschaft von First-Yes-Counts-Instruktionen, Raten in sequentiellen Gegenüberstellungen zu verringern ohne die Diskriminationsfähigkeit zu beeinflussen, ist allerdings interessant. Es ist wünschenswert, dass Augenzeeugen bzw. Augenzeeuginnen möglichst wenig raten, weil dann die Wahrscheinlichkeit der fälschlichen Identifikation einer unschuldigen verdächtigten Person reduziert wird. Es ist daher ein weiterer, interessanter Schritt, Mechanismen zu untersuchen, die sich in der polizeilichen Praxis umsetzen lassen und ebenfalls selektiv das Raten verringern. Eine solche Möglichkeit liefert beispielsweise die Verwendung von Instruktionen, ähnlich zu denen aus Experiment 3, die einen Augenzeeugen bzw. eine Augenzeeugin darauf hinweisen, dass vermutlich keine tatbeteiligte Person in der Gegenüberstellung anwesend ist.

Limitationen des 2-HT EIMs werden in den Originalarbeiten ausführlicher diskutiert. Sie umfassen beispielsweise die Tatsache, dass die Parameter des Modells nur auf Gruppenebene schätzbar sind, da mit nur vier Datenpunkten pro Versuchsperson, wie in der vorliegenden Arbeit, keine Schätzungen auf einer individuellen Ebene möglich sind (Heck et al., 2018). In anderen Paradigmen der wissenschaftlichen Forschung mit multinomialen Verarbeitungsbaummodellen ist es üblich, mehr Datenpunkte pro Versuchsperson zu erheben (z. B. Bayen et al., 1996; Bröder et al., 2013; Buchner et al., 1995; Schmidt et al., 2022) was eine individuelle Schätzung ermöglicht. Dies würde es erlauben, die individuellen Unterschiede zwischen einzelnen Versuchspersonen zu untersuchen. Ohne die Schätzungen auf individueller Ebene wird statistisch hingegen schlicht angenommen, dass sich alle Versuchspersonen in

allen Prozessen mit dem gleichen Wert beschreiben lassen. Es ist prinzipiell sinnvoll, eine solche Annahme vorab beispielsweise mithilfe hierarchischer Modellierungen zu testen. Im Kontext von Augenzeugengegenüberstellungen würde eine Messung von mehr als vier Datenpunkten pro Versuchsperson aber in einem direkten Konflikt mit der ökologischen Validität der Augenzeugenidentifikationsforschung stehen, da in der Realität eher eine geringe Anzahl von Personen gleichzeitig an einer Tat beteiligt sind (Tupper et al., 2019). Aus diesem Grund wurden für die aktuelle Arbeit nur vier Datenpunkte pro Versuchsperson erhoben. Die Limitation, dass die Daten nicht auf individueller Ebene der Versuchsperson betrachtet werden können, gilt im Übrigen auch beispielsweise für die Analyse mit ROC-Kurven, für welche ebenfalls große Datenmengen benötigt werden.

Das 2-HT EIM erfasst außerdem keine Antwortkonfidenzen von Augenzeugen bzw. Augenzeuginnen, welche in der wissenschaftlichen Forschung zu Augenzeugenidentifikationen und auch in der polizeilichen Praxis typischerweise erhoben werden (Mickes et al., 2012; Wixted & Mickes, 2014). Eine Berücksichtigung von Antwortkonfidenzen war kein Kernziel der vorliegenden Arbeit. Von Interesse war es zunächst, ein neues, adäquates statistisches Messmodell zu etablieren, das auf die volle 2×3 -Datenstruktur von Augenzeugenidentifikationen zugreifen kann und in der Lage ist, die vier vorgestellten latenten kognitiven Prozesse, die an solchen Identifikationen beteiligt sind, zu messen. Eine Berücksichtigung von Antwortkonfidenzen ist grundsätzlich in der Modellierung mit multinomialen Verarbeitungsbaummodellen möglich, würde aber einer Erweiterung des 2-HT EIMs bedürfen (Erdfelder et al., 2009). Diese Erweiterung müsste anschließend ein erneutes Validierungsverfahren durchlaufen.

Eine weitere Limitation des 2-HT EIMs umfasst die Tatsache, dass in den Experimenten 1 und 2 nicht nur der Zielparameter auf die Validierungsmanipulation reagierte. In den simultanen Versuchsbedingungen beider Experimente wurde der Rate-Parameter g zusätzlich von der Validierungsmanipulation beeinflusst. Bei der Validierung von MVB-Modellen kann sowohl ein starkes als auch ein schwaches Validierungskriterium angewendet werden (Erdfelder, 2000). Das schwache Kriterium besagt, dass es mindestens eine Manipulation gibt, die einen Parameter in der vorhergesagten Weise beeinflussen kann (Erdfelder, 2000). Diese Aussage kann für alle vier Parameter getroffen werden. Nach dem strengen Validierungskriterium darf nur der explizit manipulierte Parameter auf die Manipulation reagieren, alle anderen Parameter sollten unbeeinflusst bleiben (Erdfelder, 2000). Erst dann, wenn eine starke

Modellvalidierung für alle Modellparameter gefunden ist, lässt sich die Annahme treffen, dass alle Prozesse in einem MVB-Modell tatsächlich unabhängig und nicht kontaminiert voneinander gemessen werden können. Das strenge Validierungskriterium lässt sich nur auf den Rate-Parameter g und den Abwesenheits-Detektions-Parameter dA anwenden. Allerdings kann es generell schwierig sein, eine experimentelle Manipulation zu finden, auf die nur genau ein Zielparameter reagiert (Schmidt et al., 2022). Relevant ist daher, dass es plausible Erklärungen für die Reaktion vom Rate-Parameter g auf die Validierungsmanipulationen der Expositionszeit und Unfairness geben muss. Im Fall von Experiment 1 lässt sich die Reaktion des Rate-Parameters g durch kompensatorisches Raten erklären (Hirshman, 1995; Küppers & Bayen, 2014; Meiser et al., 2007; Riefer et al., 1994). In Experiment 2 lässt sich die Reaktion des Rate-Parameters g durch die Tatsache erklären, dass eine Person aus der Gegenüberstellung hervorstach und diese dadurch leichter zu identifizieren war. Die grundsätzliche Validität des 2-HT EIMs ist durch den Umstand einer schwachen Validierung nicht gefährdet. Das Modell lässt sich dennoch auf neue Fragestellungen anwenden. Generell gilt jedoch, dass bei einer schwachen Validierung *mindestens* eine experimentelle Manipulation gefunden werden muss, um eine Parametervalidierung zu stützen (Erdfelder, 2000). Deshalb sollte in Folgestudien beispielsweise eine Abschwächung der verwendeten Validierungsmanipulationen untersucht werden, um zu überprüfen, ob die Zielparameter weiterhin angesprochen werden und ob möglicherweise die Nebeneffekte auf den Rate-Parameter g verschwinden. Dies könnte beispielsweise bedeuten, dass zwei Expositionszeiten genutzt werden, die weniger stark divergieren, oder Gesichter mit unauffälligeren Muttermalen. So ließe sich möglicherweise auch eine Validierung des Anwesenheits-Detektions-Parameters dP und des Unfairness-Parameters b nach dem starken Validierungskriterium erzielen.

Auf Basis der hier vorgestellten Ergebnisse lässt sich dennoch schließen, dass die Nutzung des 2-HT EIMs viele Vorteile aufweist: Durch das 2-HT EIM ist es möglich, Prozesse, die bei Augenzeugenidentifikationen passieren, und die vorher nicht messbar waren, zu erfassen: Das Modell ist beispielsweise in der Lage, zwischen den beiden detektionsbasierten Prozessen der Anwesenheits- bzw. Abwesenheitsdetektion zu unterscheiden, was mit traditionellen Maßen zur Diskriminationsfähigkeit nicht möglich war. Zuvor, beispielsweise bei der Nutzung von ROC-Kurven, konnte nur in einem gemeinsamen Maß untersucht werden, wie gut Augenzeugen bzw. Augenzeuginnen zwischen tatbeteiligter und unschuldiger verdächtigter Person unterscheiden können. Ein separates Maß für die Abwesenheits-Detektion existierte bisher

nicht. Die Nutzung der vollen 2×3 -Datenstruktur erlaubt es, diesen Prozess aus den Daten zu messen und zu untersuchen. Dass in Experiment 1 nur der Anwesenheits-Detektions-Parameter und in Experiment 4 nur der Abwesenheits-Detektions-Parameter auf die experimentelle Manipulation reagiert haben, aber nicht umgekehrt, ist ein Beleg dafür, dass beide Prozesse eigenständig und getrennt voneinander stattfinden. Dies wird durch Befunde aus der Literatur gestützt, denn dort finden sich experimentelle Manipulationen, die gezielt nur den neuartigen Abwesenheits-Detektions-Parameter dA ansprechen können. Die Reanalysen der Daten von Karageorge und Zajac (2011) und Wilcock und Bull (2010) in Menne et al. (2022) zeigen, dass es Prozeduren gibt, die es beispielsweise Kindern und älteren Erwachsenen vereinfachen, eine Gegenüberstellung zurückzuweisen, wenn keine tatbeteiligte Person anwesend war, ohne die Fähigkeit, die Anwesenheit einer tatbeteiligten Person zu detektieren, zu beeinflussen. Die Auftrennung der Prozesse zur Detektion der An- bzw. Abwesenheit einer tatbeteiligten Person hat daher auch praktische Relevanz und eröffnet den Rahmen für ihre gezielte und individuelle Erforschung.

Des Weiteren ist das 2-HT EIM dazu in der Lage, zwischen den zwei nicht-detektionsbasierten Prozessen der Unfairness und des Ratens zu unterscheiden und sie gemeinsam in einem kohärenten Modell zu messen. Unfairness ist definiert als das Hervorstechen einer verdächtigten Person aus der Gegenüberstellung aufgrund bestimmter Merkmale. Beim Raten wird zufällig eine Person aus einer Gegenüberstellung ausgewählt, weil der Augenzeuge bzw. die Augenzeugin jemanden identifizieren möchte, aber keine Erinnerung an die verdächtige Person hat. Diese Trennung ist hierbei nicht neu. Es existieren Studien, die sowohl Unfairness (z. B. Colloff et al., 2016; Wetmore et al., 2015), als auch Raten (z. B. Lampinen et al., 2020; Malpass & Devine, 1981) in Gegenüberstellungen betrachten. Allerdings wurden Unfairness und Raten bisher nicht in einem gemeinsamen Messmodell erfasst und zeitgleich zu den detektionsbasierten Prozessen gemessen. Für die Messung von Unfairness wurden bisher immer externe Maße und teilweise auch Prozeduren genutzt, die von klassischen Gegenüberstellungsprozeduren abweichen (s. Menne et al., 2023 für weiterführende Informationen). Für die Messung der Ratetendenz wurde historisch bisher beispielsweise auf den Vergleich von den Identifikationsraten der tatbeteiligten Person und den Zurückweisungsraten der Gegenüberstellungen zwischen den Versuchsbedingungen zurückgegriffen, welche, wie weiter oben bereits erläutert wurde, ein separates Maß neben der Messung der Diskriminationsfähigkeit darstellen. So sind die Annahmen, die zur Messung des Rateprozesses getroffen werden müssen, u. U. nicht die gleichen, unter denen die Diskriminationsfähigkeit mit ROC-

Kurven gemessen wird. Die Nutzung eines einzigen Messmodells, das alle gemessenen Prozesse in sich vereint und ihnen die gleichen Annahmen zugrunde legt, umgeht dieses Problem. Das 2-HT EIM vereinfacht also die Messung dieser zwei nicht-detektionsbasierten Prozesse, indem es sie in ein kohärentes Messmodell integriert. Zusammenfassend wurde in der vorliegenden Arbeit ein neues Messmodell für Augenzeugenidentifikationen vorgestellt, validiert und angewendet, welches die Schwächen der traditionell genutzten Messmethoden ausgleichen kann. Außerdem wurde eine Forschungsmethodik gewählt, die eine hohe Ähnlichkeit zur polizeilichen Praxis aufweist. Die Ergebnisse der Experimente 1 bis 4 zeigen, dass das Modell valide ist. Die Ergebnisse der Reanalyse der Daten von Horry et al. (2021) und von Experiment 5 zeigen zusätzlich, dass das Modell auf empirische Daten aus fingierten Gegenüberstellungen passt und auf praktisch relevante Fragestellungen angewandt werden kann. Außerdem bietet es Ansatzpunkte für die Erforschung neuer Fragestellungen, beispielsweise des Abwesenheits-Detektions-Prozesses, der sich durch das 2-HT EIM erstmalig messen lässt, und die Verbesserung der Vergleichbarkeit zwischen wissenschaftlicher Forschung und polizeilicher Praxis. Es lässt sich daher schlussfolgern, dass die Verwendung des 2-HT EIMs das Potential hat, neue Impulse für die wissenschaftliche Erforschung von Augenzeugenidentifikationen zu schaffen und so mit neuen Erkenntnissen das Forschungsfeld als solches voranzubringen.

Literatur

- Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T. & Benjamin, A. S. (2021). The effect of lineup size on eyewitness identification. *Journal of Experimental Psychology: Applied*, 27(2), 369-392. <https://dx.doi.org/10.1037/xap0000340>
- Amendola, K. L. & Wixted, J. T. (2015a). Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in a randomized field trial. *Journal of Experimental Criminology*, 11(2), 263-284. <https://doi.org/10.1007/s11292-014-9219-2>
- Amendola, K. L. & Wixted, J. T. (2015b). No possibility of a selection bias, but direct evidence of a simultaneous superiority effect: A reply to Wells et al. *Journal of Experimental Criminology*, 11(2), 291-294. <https://doi.org/10.1007/s11292-015-9227-x>
- Andersen, S. M., Carlson, C. A., Carlson, M. A. & Gronlund, S. D. (2014). Individual differences predict eyewitness identification performance. *Personality and Individual Differences*, 60, 36-40. <https://doi.org/10.1016/j.paid.2013.12.011>
- Batchelder, W. H. & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review*, 6(1), 57-86. <https://doi.org/10.3758/BF03210812>
- Bayen, U. J., Murnane, K. & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(1), 197-215. <https://doi.org/10.1037/0278-7393.22.1.197>
- Bergold, A. N. & Heaton, P. (2018). Does filler database size influence identification accuracy? *Law and Human Behavior*, 42(3), 227-243. <https://doi.org/10.1037/lhb0000289>
- Bill Blackwood Law Enforcement Management Institute of Texas (2022). *Model Policy on Eyewitness Identification*. Sam Houston State University. Abgerufen am 29.09.2022 unter http://www.lemionline.org/resources/documents/ewid_final.pdf
- Brewer, N., Weber, N., Wooton, D. & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgements under deadline pressure. *Psychological Science*, 23(10), 1208-1214. <https://doi.org/10.1177/0956797612441217>
- Brewer, N. & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12(1), 11-30. <https://doi.org/10.1037/1076-898X.12.1.11>
- Bröder, A., Kellen, D., Schütz, J. & Rohrmeier, C. (2013). Validating a two-high-threshold measurement model for confidence rating data in recognition. *Memory*, 21(8), 916-944. <https://doi.org/10.1080/09658211.2013.767348>

- Buchner, A., Erdfelder, E. & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, 124(2), 137-160. <https://doi.org/10.1037/0096-3445.124.2.137>
- Bundeskriminalamt (2021). *PKS 2021 Bund - Tatverdächtige insgesamt - T22 Sonstige Angaben zum Tatverdächtigen (V1.0)*. Abgerufen am 06.07.2023 unter <https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/PKS2021/PKSTabellen/BundTV/bundTV.html?nn=194208>
- Bundesministerium des Innern und für Heimat [BMI] (2021). *Richtlinien für das Strafverfahren und das Bußgeldverfahren (RiStBV)*. Bundesamt für Justiz. Abgerufen am 17.07.2023 unter https://www.verwaltungsvorschriften-im-internet.de/bsvwvbund_28032023_BMJRB3313104000060001.htm
- Carlson, C. A. & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory and Cognition*, 3(2), 45-53. <https://doi.org/10.1016/j.jarmac.2014.03.004>
- Carlson, C. A., Carlson, M. A., Weatherford, D. R., Tucker, A. & Bednarz, J. (2016). The effect of backloading instructions on eyewitness identification from simultaneous and sequential lineups. *Applied Cognitive Psychology*, 30(6), 1005-1013. <https://dx.doi.org/10.1002/acp.3292>
- Carlson, C. A., Gronlund, S. D. & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 14(2), 118-128. <https://doi.org/10.1037/1076-898X.14.2.118>
- Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, 29(5), 575-604. <https://doi.org/10.1007/s10979-005-7121-1>
- Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, 7(3), 238-259. <https://doi.org/10.1177/1745691612439584>
- Colloff, M. F., Wade, K. A. & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, 27(9), 1227-1239. <https://doi.org/10.1177/0956797616655789>
- Dobolyi, D. G. & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied*, 19(4), 345-357. <https://doi.org/10.1037/a0034596>
- Dunn, J. C., Kaesler, M. & Semmler, C. (2022). A model of position effects in the sequential lineup. *Journal of Memory and Language*, 122, 104297. <https://doi.org/10.1016/j.jml.2021.104297>
- Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie* [unveröffentlichte Habilitationsschrift]. Rheinische Friedrich-Wilhelms

- Universität Bonn. Abgerufen am 17.07.2023 unter <https://madoc.bib.uni-mannheim.de/63897/1/MMKP.pdf>
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M. & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, 217(3), 108-124. <https://doi.org/10.1027/0044-3409.217.3.108>
- Erdfelder, E. & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(2), 387-414. <https://doi.org/10.1037/0278-7393.24.2.387>
- Flowe, H. D. (2011). An exploration of visual behaviour in eyewitness identification tests. *Applied Cognitive Psychology*, 25(2), 244-254. <https://doi.org/10.1002/acp.1670>
- Flowe, H. D. & Ebbesen, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law and Human Behavior*, 31(1), 33-52. <https://dx.doi.org/10.1007/s10979-006-9045-9>
- Flowe, H. D., Smith, H. M. J., Karoğlu, N., Onwuegbusi, T. O. & Rai, L. (2016). Configural and component processing in simultaneous and sequential lineup procedures. *Memory*, 24(3), 306-314. <https://doi.org/10.1080/09658211.2015.1004350>
- Gronlund, S. D., Carlson, C. A., Dailey, S. B. & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 15(2), 140-152. <https://doi.org/10.1037/a0015082>
- Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A. & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, 1(4), 221-228. <https://doi.org/10.1016/j.jarmac.2012.09.003>
- Gronlund, S. D., Mickes, L., Wixted, J. T. & Clark, S. E. (2015). Chapter one - conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Hrsg.), *Psychology of Learning and Motivation* (Band 63, S. 1-43). Elsevier Inc. <https://doi.org/10.1016/bs.plm.2015.03.003>
- Heck, D. W., Hoffmann, A. & Moshagen, M. (2018). Detecting nonadherence without loss in efficiency: A simple extension of the crosswise model. *Behavior Research Methods*, 50(5), 1895-1905. <https://doi.org/10.3758/s13428-017-0957-8>
- Hirshman, E. (1995). Decision process in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(2), 302-313. <https://doi.org/10.1037/0278-7393.21.2.302>
- Hobson, Z., Wilcock, R. & Valentine, T. (2012). Multiple suspect showing: A survey of police identification officers. *Policing*, 7(1), 79-87. <https://doi.org/10.1093/police/pas021>
- Home Office (2017). *Police and Criminal Evidence Act 1984 (PACE). Codes of Practice, Code D*. The Stationery Office. Abgerufen am 21.07.2022 unter

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/903812/pace-code-d-2017.pdf

- Horry, R., Fitzgerald, R. J. & Mansour, J. K. (2021). "Only your first yes will count": The impact of pre-lineup instructions on sequential lineup decisions. *Journal of Experimental Psychology: Applied*, 27(1), 170-186. <https://doi.org/10.1037/xap0000337>
- Horry, R., Memon, A., Wright, D. B. & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior*, 36(4), 257-265. <https://doi.org/10.1037/h0093959>
- Horry, R., Palmer, M. A. & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied*, 18(4), 346-360. <https://doi.org/10.1037/a0029779>
- Innocence Project (2023). *Exonerate the Innocent*. Abgerufen am 19.01.2023 unter <https://innocenceproject.org/exonerate/>
- Kaesler, M., Dunn, J. C., Ransom, K. & Semmler, C. (2020). Do sequential lineups impair underlying discriminability? *Cognitive Research: Principles and Implications*, 5(35), 1-21. <https://doi.org/10.1186/s41235-020-00234-5>
- Karageorge, A. & Zajac, R. (2011). Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of wildcards. *British Journal of Psychology*, 102(2), 161-183. <https://doi.org/10.1348/000712610X507902>
- Keast, A., Brewer, N. & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology*, 97(4), 286-314. <https://doi.org/10.1016/j.jecp.2007.01.007>
- Klauer, K. C. & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review*, 17(4), 465-478. <https://doi.org/10.3758/PBR.17.4.465>
- Kneller, W., Memon, A. & Stevenage, S. (2001). Simultaneous and sequential lineups: decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, 15(6), 659-671. <https://doi.org/10.1002/acp.739>
- Küppers, V. & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *The Quarterly Journal of Experimental Psychology*, 67(10), 2042-2059. <https://doi.org/10.1080/17470218.2014.904914>
- Lampinen, J. M., Race, B., Wolf, A. P., Phillips, P., Moriarty, N. & Smith, A. M. (2020). Comparing detailed and less detailed pre-lineup instructions. *Applied Cognitive Psychology*, 34(2), 409-424. <https://doi.org/10.1002/acp.3627>
- Lindsay, R. C. L. & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology*, 84(3), 315-321. <https://dx.doi.org/10.1037/0021-9010.84.3.315>

- Lindsay, R. C. L. & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, 70(3), 556-564. <https://doi.org/10.1037/0021-9010.70.3.556>
- Macmillan, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates. Abgerufen am 17.07.2023 unter <https://psycnet.apa.org/record/2004-19022-000>
- Malpass, R. S. & Devine, P. G. (1981). Eyewitness identification: Lineup constructions and the absence of the offender. *Journal of Applied Psychology*, 66(4), 482-489. <https://doi.org/10.1037/0021-9010.66.4.482>
- McQuiston-Surrett, D., Malpass, R. S. & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, and Law*, 12(2), 137-169. <https://doi.org/10.1037/1076-8971.12.2.137>
- Meiser, T. & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 116-137. <https://doi.org/10.1037//0278-7393.28.1.116>
- Meiser, T., Sattler, C. & Von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *The Quarterly Journal of Experimental Psychology*, 60(7), 1015-1040. <https://doi.org/10.1080/17470210600875215>
- Meisters, J., Diederhofen, B. & Musch, J. (2018). Eyewitness identification in simultaneous and sequential lineups: An investigation of position effects using receiver operating characteristics. *Memory*, 26(9), 1297-1309. <https://doi.org/10.1080/09658211.2018.1464581>
- Memon, A. & Gabbert, F. (2003). Unravelling the effects of sequential presentation in culprit-present lineups. *Applied Cognitive Psychology*, 17(6), 703-714. <https://doi.org/10.1002/acp.909>
- Memon, A., Hope, L. & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, 94, 339-354. <https://doi.org/10.1348/000712603767876262>
- Menne, N. M., Winter, K., Bell, R. & Buchner, A. (2022). A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Scientific Reports*, 12, 13379. <https://doi.org/10.1038/s41598-022-17400-y>
- Menne, N. M., Winter, K., Bell, R. & Buchner, A. (2023). Measuring lineup fairness from eyewitness identification data using a multinomial processing tree model. *Scientific Reports*, 13, 6290. <https://doi.org/10.1038/s41598-023-33424-4>
- Mickes, L., Flowe, H. D. & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, 18(4), 361-376. <https://doi.org/10.1037/a0030609>
- Mickes, L., Moreland, M. B., Clark, S. E. & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute d' , not the

- diagnosticity ratio. *Journal of Applied Research in Memory and Cognition*, 3(2), 58-62. <https://doi.org/10.1016/j.jarmac.2014.04.007>
- Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., Goodsell, C. A., Weatherford, D. & Wixted, J. T. (2017). ROCs in eyewitness identification: Instructions versus confidence ratings. *Applied Cognitive Psychology*, 31(5), 467-477. <https://doi.org/10.1002/acp.3344>
- Ministerium des Innern des Landes Nordrhein-Westfalen (2023). *Wahllichtbildvorlage im Strafverfahren*. Abgerufen am 06.07.2023 unter https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=1&bes_id=9147&aufgehoben=N
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42(1), 42-54. <https://doi.org/10.3758/BRM.42.1.42>
- Neuschatz, J. S., Wetmore, S. A., Key, K. N., Cash, D. K., Gronlund, S. D. & Goodsell, C. A. (2016). A comprehensive evaluation of showups. In B. H. Bornstein & M. K. Miller (Hrsg.), *Advances in Psychology and Law* (Band 5, S. 43-69). NY: Springer. https://doi.org/10.1007/978-3-319-29406-3_2
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Palmer, M. A. & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, 36(3), 247-255. <https://doi.org/10.1037/h0093923>
- Palmer, M. A., Brewer, N., Weber, N. & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, 19(1), 55-71. <https://dx.doi.org/10.1037/a0031602>
- Riefer, D. M. & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95(3), 318-339. <https://doi.org/10.1037/0033-295X.95.3.318>
- Riefer, D. M., Hu, X. & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(3), 680-693. <https://doi.org/10.1037/0278-7393.20.3.680>
- Sauer, J. D., Brewer, N. & Wells, G. L. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal and Criminological Psychology*, 13(1), 123-135. <https://doi.org/10.1348/135532506X159203>
- Schmidt, O., Erdfelder, E. & Heck, D. W. (2022). Tutorial on multinomial processing tree modeling: How to develop, test, and extend MPT models. *PsyArXiv*. <https://doi.org/10.31234/osf.io/gh8md>
- Seale-Carlisle, T. M. & Mickes, L. (2016). US line-ups outperform UK line-ups. *Royal Society Open Science*, 3(9), 2-12. <https://doi.org/10.1098/rsos.160300>

- Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D. & Mickes, L. (2019). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied*, 25(3), 410-430. <https://doi.org/10.1037/xap0000222>
- Sherman, J. W., Groom, C. J., Ehrenberg, K. & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition*, 21(3), 213-246. <https://doi.org/10.1521/soco.21.3.213.25340>
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76-80. <https://doi.org/10.1177/1745691613514755>
- Smith, A. M., Wells, G. L., Smalarz, L. & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves the applied value of lineups without improving memory performance: Commentary on Colloff, Wade, and Strange (2016). *Psychological Science*, 29(9), 1548-1551. <https://doi.org/10.1177/0956797617698528>
- Smith, R. E. & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(4), 756-777. <https://doi.org/10.1037/0278-7393.30.4.756>
- Stahl, C. & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology*, 54(2), 99-112. <https://doi.org/10.1027/1618-3169.54.2.99>
- State Bar of Michigan: Eyewitness Identification Task Force (2015). *Law Enforcement and eyewitness identifications: A policy writing guide*. Abgerufen am 29.09.2022 unter https://www.michigan.gov/-/media/Project/Websites/mcoles/TD-Webpage/state_bar_eyewitness_law_enforcement_policy_writing_guide.pdf?rev=6e19807bd78447d2b54e6f9cc7ab186b
- Stebly, N. K., Dysart, J. E. & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, 17(1), 99-139. <https://doi.org/10.1037/a0021650>
- Sučić, I., Tokić, D. & Ivešić, M. (2015). Field study of response accuracy and decision confidence with regard to lineup composition and lineup presentation. *Psychology, Crime & Law*, 21(8), 798-819. <https://dx.doi.org/10.1080/1068316X.2015.1054383>
- Toftely, S. M. (2015). *Eyewitness identification: An investigation of the feature-detection hypothesis* [unveröffentlichte Doktorarbeit]. University of Oklahoma, Graduate college. Abgerufen am 17.07.2023 unter <http://hdl.handle.net/11244/14621>
- Tupper, N., Sauerland, M., Sauer, J. D. & Hope, L. (2019). Eyewitness identification procedures for multiple perpetrator crimes: a survey of police in Sweden, Belgium, and the Netherlands. *Psychology, Crime & Law*, 25(10), 992-1007. <https://doi.org/10.1080/1068316x.2019.1611828>
- U.S. Department of Justice: Office of the Deputy attorney General (2017). *Eyewitness identification: Procedures for conducting photo arrays*. Sally Q. Yates. Abgerufen am 29.09.2022 unter <https://www.justice.gov/file/923201/download>

- Unkelbach, C. & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, 18(1), 22-38. <https://doi.org/10.1016/j.concog.2008.09.006>
- Valentine, T., Pickering, A. & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, 17(8), 969-993. <https://doi.org/10.1002/acp.939>
- Wells, G. L. (2014). Eyewitness Identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science*, 23(1), 11-16. <https://doi.org/10.1177/0963721413504781>
- Wells, G. L., Dysart, J. E. & Steblay, N. K. (2015). The flaw in Amendola and Wixted's conclusion on simultaneous versus sequential lineups: Comment. *Journal of Experimental Criminology*, 11(2), 285-289. <https://doi.org/10.1007/s11292-014-9225-4>
- Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A. & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, 44(1), 3-36. <https://doi.org/10.1037/lhb0000359>
- Wells, G. L. & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 88(3), 776-784. <https://doi.org/10.1037/0033-2909.88.3.776>
- Wells, G. L., Malpass, R. S., Lindsay, R. C., Fisher, R. P., Turtle, J. W. & Fulero, S. M. (2000). From the lab to the police station. A successful application of eyewitness research. *American Psychologist*, 55(6), 581-598. <https://doi.org/10.1037//0003-066x.55.6.581>
- Wells, G. L., Smalarz, L. & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory and Cognition*, 4(4), 313-317. <https://doi.org/10.1016/j.jarmac.2015.08.008>
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M. & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, 22(6), 603-647. <https://doi.org/10.1023/A:1025750605807>
- Wells, G. L., Smith, A. M. & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory and Cognition*, 4(4), 324-328. <https://doi.org/10.1016/j.jarmac.2015.08.010>
- Wells, G. L., Steblay, N. K. & Dysart, J. E. (2011). A test of the simultaneous vs. sequential lineup methods: An initial report of the ASJ national eyewitness identification field studies. In *American Judicature Society*. https://popcenter.asu.edu/sites/default/files/library/reading/PDFs/lineup_methods.pdf
- Wells, G. L., Steblay, N. K. & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous

- lineup procedure. *Law and Human Behavior*, 39(1), 1-14.
<https://doi.org/10.1037/lhb0000096>
- Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A. & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, 4(1), 8-14.
<https://doi.org/10.1016/j.jarmac.2014.07.003>
- Wilcock, R. & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology*, 24(5), 718-736.
<https://doi.org/10.1002/acp.1582>
- Winter, K., Menne, N. M., Bell, R. & Buchner, A. (2022). Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Scientific Reports*, 12, 15571. <https://doi.org/10.1038/s41598-022-19513-w>
- Winter, K., Menne, N. M., Bell, R. & Buchner, A. (2023). Evaluating the impact of First-Yes-Counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Scientific Reports*, 13, 6572.
<https://doi.org/10.1038/s41598-023-33424-4>
- Wixted, J. T. & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, 121(2), 262-276. <https://doi.org/10.1037/a0035940>
- Wixted, J. T. & Mickes, L. (2015a). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition*, 4(4), 318-323. <https://doi.org/10.1016/j.jarmac.2015.08.009>
- Wixted, J. T. & Mickes, L. (2015b). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, 4(4), 329-334.
<https://doi.org/10.1016/j.jarmac.2015.08.007>
- Zarkadi, T., Wade, K. A. & Stewart, N. (2009). Creating fair lineups for suspects with distinctive features. *Psychological Science*, 20(12), 1448-1453.
<https://doi.org/10.1111/j.1467-9280.2009.02463.x>

Einzelarbeiten

Experimente 1 – 4:

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2022). Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Scientific Reports*, 12, 15571. <https://doi.org/10.1038/s41598-022-19513-w>

Reanalyse & Experiment 5:

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2023). Evaluating the impact of First-Yes-Counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Scientific Reports*. 13, 6572. <https://doi.org/10.1038/s41598-023-33424-4>



OPEN

Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions

Kristina Winter [✉], Nicola M. Menne , Raoul Bell  & Axel Buchner 

To improve police protocols for lineup procedures, it is helpful to understand the processes underlying eyewitness identification performance. The *two-high threshold (2-HT) eyewitness identification model* is a multinomial processing tree model that measures four latent cognitive processes on which eyewitness identification decisions are based: two detection-based processes (the detection of culprit presence and absence) and two non-detection-based processes (biased and guessing-based selection). The model takes into account the full 2×3 data structure of lineup procedures, that is, suspect identifications, filler identifications and rejections in both culprit-present and culprit-absent lineups. Here the model is introduced and the results of four large validation experiments are reported, one for each of the processes specified by the model. The validation experiments served to test whether the model's parameters sensitively reflect manipulations of the processes they were designed to measure. The results show that manipulations of exposure duration of the culprit's face at encoding, lineup fairness, pre-lineup instructions and ease of rejection of culprit-absent lineups were sensitively reflected in the parameters representing culprit-presence detection, biased suspect selection, guessing-based selection and culprit-absence detection, respectively. The results of the experiments thus validate the interpretations of the parameters of the 2-HT eyewitness identification model.

Eyewitness identifications often are essential for convicting culprits. Given that eyewitness identification decisions are made after the crime has occurred, it is necessary to assess the witness's memory to test the hypothesis that a suspect is the culprit. The memory test is usually done in the form of a lineup. Lineup procedures have to follow strict protocols¹. The suspect's face is to be presented among the faces of known innocent fillers. The witness's task is to identify the culprit or reject the lineup. However, instead of correctly identifying the culprit or rejecting the lineup with an innocent suspect, the witness may falsely reject a lineup even though the culprit is present or select a known innocent filler. In the worst case, the witness identifies an innocent suspect².

To improve police protocols for lineup procedures, it is helpful to understand the cognitive processes upon which these eyewitness identification decisions are based. Here we introduce the *two-high threshold (2-HT) eyewitness identification model* that measures four different latent processes underlying eyewitness identification performance: detection of the presence or absence of the culprit, biased suspect selection and guessing-based selection. To validate this model, we tested in four large experiments whether the four model parameters respond sensitively to manipulations that can be expected to influence the postulated latent processes represented by the parameters.

Many studies have examined how police lineup procedures affect the quality of eyewitness identification decisions^{1,3}. At first glance, it may seem desirable that lineup procedures lead to a high rate of correct culprit identifications. However, this is so only if all culprit identifications are based on the detection of the culprit. Unfortunately, however, correct culprit identifications may also be caused by the biased selection of the suspect due to an unfair lineup procedure in which the suspect stands out from the fillers or by a tendency to select one of the lineup members based on guessing. Both of these processes do not only lead to correct culprit identifications but also to false innocent-suspect identifications. To understand eyewitness identification performance, it is thus important to distinguish between different latent processes that may contribute to eyewitness identification decisions.

Department of Experimental Psychology, Heinrich Heine University Düsseldorf, 40204 Düsseldorf, Germany. ✉email: kristina.winter@hhu.de

The 2 × 2 data structure of the standard signal-detection task			
	Yes	No	
Signal present	Hit	Miss	
Signal absent	False alarm	Correct rejection	
The 2 × 3 data structure of the typical lineup identification task			
	Suspect identification	Filler identification	Lineup rejection
Culprit present	Correct culprit identification	False filler identification	False lineup rejection
Culprit absent	False innocent-suspect identification	False filler identification	Correct lineup rejection

Table 1. The data structures of the standard signal-detection task and the typical lineup identification task.

To measure eyewitness identification performance, researchers have often relied on the *diagnosticity ratio*^{3–9} which is defined as the ratio between the proportion of correct culprit identifications and the proportion of false innocent-suspect identifications¹⁰. A higher diagnosticity ratio was taken to indicate superior eyewitness identification performance. However, the diagnosticity ratio attains higher values not only when a procedure is associated with better discrimination between culprits and innocent suspects but also when witnesses exhibit a conservative response bias^{11–13}. Therefore, it has been suggested to apply *Receiver Operating Characteristic* (ROC) curves to the analysis of eyewitness identification decisions^{11,13–18}. Derived from *signal detection theory* (SDT)¹⁹, ROC curves are constructed by plotting the hit rate (the proportion of correct detections of a signal) against the false alarm rate (the proportion of false positive responses to noise) at different levels of response bias. In lineup research, the hit rate corresponds to the rate of correct culprit identifications. The false alarm rate corresponds to the rate of false innocent-suspect identifications. The levels of response bias are usually derived from confidence judgements that witnesses are asked to assign to their identification decisions. To measure the degree to which the procedure allows witnesses to discriminate between culprits and innocent suspects, the *partial area under the curve* (*pAUC*) is calculated. The lineup procedure with the greater *pAUC* is to be preferred^{11,20}.

The ROC analysis was derived from SDT¹⁹ that was originally proposed for detection problems with a 2 × 2 data structure (Table 1a): participants respond “yes” or “no” to a stimulus in which the signal is present or absent, respectively. There are four data categories in such tasks: hits, false alarms, misses and correct rejections. The analysis of performance is based on hits and false alarms because the remaining data categories are redundant (miss rate = 1 – hit rate; correct rejection rate = 1 – false alarm rate). By contrast, the detection problem in lineups has a 2 × 3 data structure (Table 1b) with other, non-redundant data categories. In culprit-present lineups, one may correctly identify the culprit (correct culprit identification), falsely reject the lineup (false lineup rejection) and falsely identify a filler (false filler identification). In culprit-absent lineups, one may falsely identify the innocent suspect (false innocent-suspect identification), correctly reject the lineup (correct lineup rejection) and falsely identify a filler (false filler identification).

It has been argued that, for the purpose of deciding which of two lineup procedures is superior, it is sufficient to focus only on correct culprit identifications and false innocent-suspect identifications because these data categories have the most far-reaching consequences in practice¹¹. However, if the goal is to distinguish between the latent cognitive processes underlying eyewitness identification decisions, filler identifications can provide useful information. For instance, rejecting a culprit-absent lineup represents a correct decision whereas falsely identifying a filler represents an error. Hence, the underlying cognitive processes differ^{8,9}. The 2-HT eyewitness identification model introduced here is thus based on the full range of data categories available from lineup procedures: correct culprit identifications, false innocent-suspect identifications, false filler identifications in culprit-present and culprit-absent lineups as well as correct and false lineup rejections. The model belongs to the class of multinomial processing tree (MPT) models that has proven useful to analyze, and to test hypotheses about, the latent processes underlying observable behavior in various fields of psychology, including memory (e.g.^{21–24}) and decision making (e.g.^{25–27}). There are several excellent introductions to MPT modeling^{28,29}. Parameter estimation and statistical tests on the model parameters can be performed with freely available computer programs^{30–32}.

A graphical illustration of the 2-HT eyewitness identification model is displayed in Fig. 1. The upper tree represents the processes leading to identifications and rejections in culprit-present lineups. The witness may detect the presence of the culprit with probability dP (for *detection of the presence of the culprit*). If culprit detection fails (with probability $1 - dP$), the witness may identify the culprit based on two types of non-detection-based processes. If the lineup is unfair, the culprit stands out from the fillers so that it can be inferred who the suspect is without relying on memory. In such cases, biased selection of the suspect occurs with probability b (for *biased suspect selection*). With probability $1 - b$, no biased suspect selection occurs (e.g., if the lineup is fair or the witness does not attend to the features creating the unfairness). In this case, it is still possible to select one of the lineup members as the culprit based on guessing. Guessing-based selection differs from biased suspect selection in that, when the witness selects one of the lineup members based on guessing, the culprit is selected among the fillers with a probability that is equal to $1 \div$ lineup size. With probability $1 - (1 \div$ lineup size), one of the fillers is selected. To illustrate, in a six-person lineup, the probability that selecting one of the lineup members based on guessing results in the identification of the suspect among the fillers is $1/6$ while the probability of selecting one of the fillers is $5/6$. The probability with which guessing leads to a suspect identification is a function of the size of the lineup and does not depend on the witnesses’ internal cognitive representation which is why this element

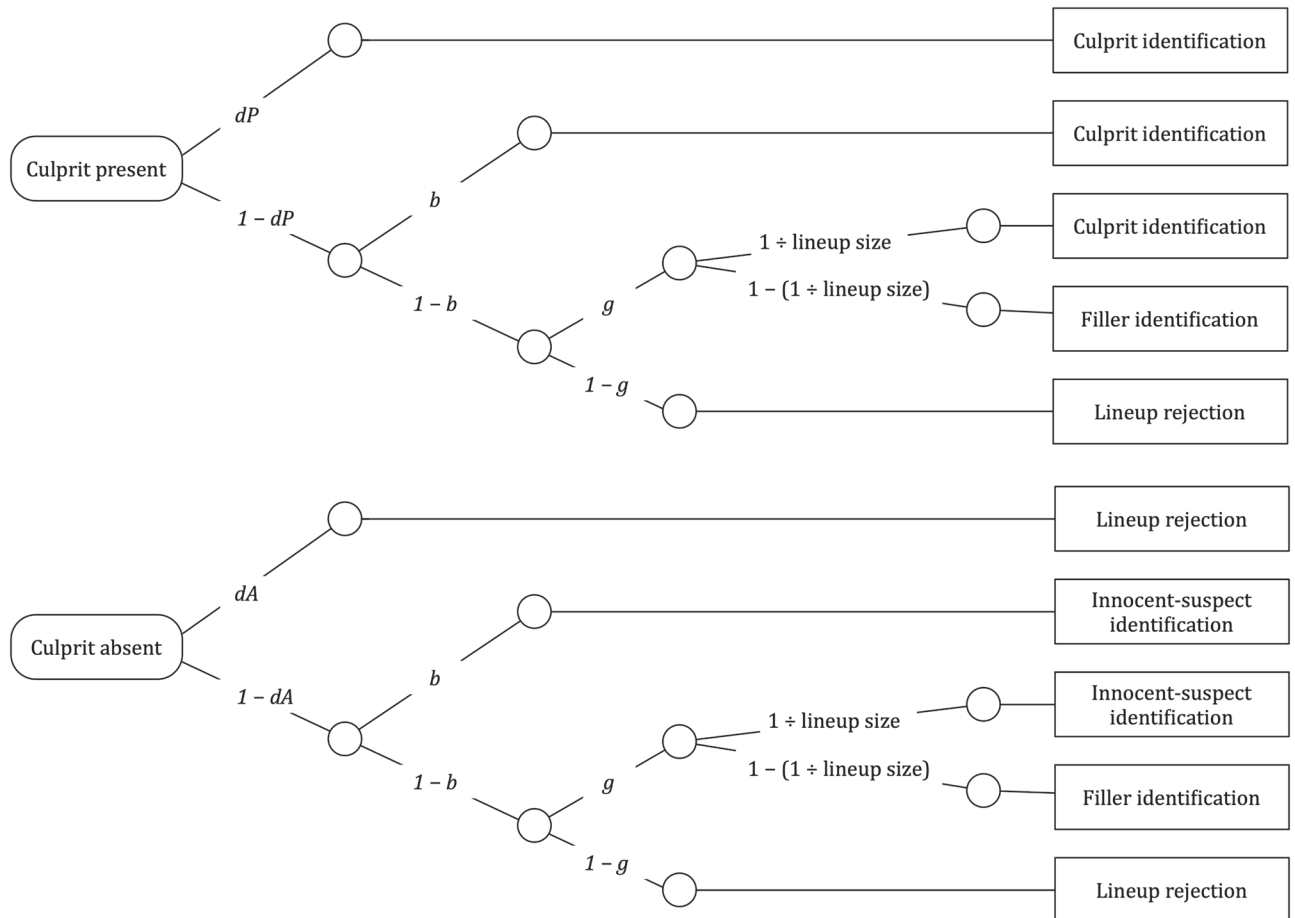


Figure 1. Graphical illustration of the two-high threshold (2-HT) eyewitness identification model. The rounded rectangles on the left represent culprit-present and culprit-absent lineups that have to be processed in order to arrive at an observable response. The rectangles on the right represent the observable response categories. The letters attached to the branches represent the probabilities of the latent cognitive processes postulated by the model. Parameter dP represents the probability of detecting the presence of the culprit. Parameter b represents the probability of biased selection of a suspect who stands out from the other lineup members. Parameter g represents the probability of guessing-based selection among the lineup members, as a consequence of which the suspect might be chosen based on chance with a probability of $1 \div \text{lineup size}$ (approximately 0.16667 in the present case of six lineup members). Parameter dA represents the probability of detecting the absence of the culprit. Identifiability was ensured using multiTree's³⁰ repeated analysis module in which the expectation-maximization algorithm is repeatedly applied to identical category frequencies (1000 times in the present case for every dataset presented here). Identical parameter estimates were obtained in all cases, confirming that the model is identifiable. There are as many free parameters as there are independent data categories to fit so that the unrestricted model has zero degrees of freedom.

of the model is a constant and not a parameter that has to be estimated from data. If none of the lineup members is selected based on guessing (with probability $1 - g$), the lineup is falsely rejected.

For culprit-absent lineups in which the suspect is innocent (lower tree of Fig. 1), parallel cognitive processes are postulated to occur: the absence of the culprit and the fact that no one else in the lineup can possibly be the culprit is detected with probability dA (for *detection of the absence of the culprit*), leading to the correct rejection of the lineup. With probability $1 - dA$, the absence of the culprit is not detected, in which case the same non-detection-based processes are postulated as in culprit-present lineups. An obvious characteristic of guessing-based selection (parameter g) is that it must not differ between culprit-absent and culprit-present lineups. The same applies to biased suspect selection (parameter b): the police do not know whether their suspect is the culprit or innocent. Therefore, if a lineup is unfair, it is equivalently unfair for culprits and innocent suspects. For reasons of ecological validity, it is thus optimal to have a designated innocent suspect who deviates from the fillers in a way that is similar to how the culprit deviates from the fillers when studying the effects of lineup unfairness.

The 2-HT eyewitness identification model is a two-high threshold model such as, for example, the widely used two-high threshold model of source memory^{21,33}. The model includes parameter dP for the detection of the presence of the culprit and parameter dA for the detection of the absence of the culprit. While in other two-high threshold models the assumption that both detection parameters are equal is a requirement for identifiability (e.g.^{21,34}), they can be estimated separately in the 2-HT eyewitness identification model (cf.³⁵). Given that the

model does not have to be restricted by setting the dA parameter equal to another parameter, it is possible to determine at an empirical level whether eyewitnesses are able to detect the absence of the culprit. To anticipate, the results of the present Experiment 4 and reanalyses of published studies³⁶ suggest that people detect the absence of the culprit in some situations but fail to detect the absence of the culprit in others.

In a new MPT model it is necessary to validate the interpretation of the model parameters²⁸. A successful validation requires that the model fits the empirical data and that the model parameters respond sensitively to manipulations of the processes they were designed to measure. For this purpose, validation experiments are needed in which well-established or outright trivial manipulations of the latent cognitive processes are used to test whether the changes in the latent cognitive processes are reflected in the model's parameters. Here, we report four validation experiments, one for each of the four parameters of the model. To influence the detection of the presence of the culprit, reflected in parameter dP , we manipulated the exposure duration of the culprits' faces at encoding. To influence biased suspect selection, reflected in parameter b , we manipulated the unfairness of the lineup. To influence guessing-based selection, reflected in parameter g , we manipulated whether the pre-lineup instructions suggested either a high or a low probability of the culprit being in the lineup. To influence the detection of the absence of the culprit, reflected in parameter dA , we compared the standard lineup procedure with a condition in which all lineup members in culprit-absent lineups could easily be ruled out as the culprit. Given the lively discussion about whether lineups should be simultaneous or sequential^{11,13,37–39}, one group of participants saw simultaneous lineups while another group of participants saw sequential lineups in each of the four experiments.

Experiment 1: Detection of the presence of the culprit (parameter dP)

Experiment 1 served to test the validity of parameter dP which is assumed to represent the probability of detecting the presence of the culprit. A culprit's presence in a lineup is easier to detect when the culprit's face was exposed for a long than for a short duration during a crime^{40–42}. We thus presented long or short crime videos. If the 2-HT eyewitness identification model is valid, parameter dP should be higher in the long-exposure condition than in the short-exposure condition.

Method. *Sample.* Participants were recruited using the *Gapfish* research panel (<https://gapfish.com>). Participants had to be of legal age (≥ 18 years) and had to have good eyesight and German language skills. A power analysis with *multiTree*³⁰ showed that, given $\alpha = \beta = 0.05$, at least $N = 628$ participants were needed to detect a difference in the dP parameter between the long- and short-exposure condition of $\Delta dP = 0.12$ (corresponding to an effect size of $w = 0.07$), estimated from a pilot laboratory study. A somewhat larger sample was recruited to compensate for the loss of data that can be expected in online studies. In total, 932 participants gave informed consent, but 135 of them did not complete the study. A total of 33 data sets had to be excluded because of multiple participations (i.e., participants saw the crime video more than once). Another 22 data sets had to be excluded because participants had failed the attention check (see below) or because they had reported technical problems. The final sample included 742 participants (327 women and 415 men) with a mean age of 44 years ($SD = 15$).

Ethical approval. The following applies to all experiments reported here. Prior to participation, informed consent was obtained from all participants. Participants were warned that they would see a video containing verbal and physical abuse. They were asked not to continue if they felt uncomfortable when anticipating to view such a video. After the experiment, participants were informed about the purpose of the study and assured that the crime was staged. Ethical approval had been received from the ethics committee of the Faculty of Mathematics and Natural Sciences of Heinrich Heine University Düsseldorf. The study was run in accordance with the declaration of Helsinki. Informed consent for the publication was obtained from the actors of the staged-crime videos to show the stimulus material.

Materials and procedure. The experiment was conducted online using *SoSci Survey*⁴³. Participants were only allowed to participate with computers or laptops, not with smartphones or tablets. Participants were asked to complete the study alone in a quiet environment. Prior to the experiment proper, participants were asked to bring their browsers into full-screen mode.

Staged-crime videos. In lineup research, the standard procedure involves exposing each participant to a single culprit at encoding and to a single (culprit-present or culprit-absent) lineup at test (see e.g.^{13,44–48}). This procedure generates only a single data point per participant. In order to increase the efficiency of data collection, researchers have introduced between two and at least 14 to-be-identified persons in their lineup studies^{41,49–58}. We followed this lead and presented our participants with a crime video containing four culprits. Specifically, the crime video showed four alleged hooligans of the German soccer club FC Bayern München attacking an alleged fan of a rivaling soccer club, Borussia Dortmund, at a bus station (Fig. 2). The culprits and the victim wore fan clothing (shirts, scarfs and caps) of their respective soccer clubs. The culprits insulted the victim, poked fun at him and tossed his belongings around. At the end of the video, the victim got knocked to the ground. The culprits continued to physically abuse the victim before they apparently noticed another person approaching (not visible in the video) and ran away shouting loudly.

Two parallel videos were created, henceforth called Video 1 and Video 2. The videos had the same content (i.e., the same verbal abuse and the same acts of violence in the same sequence and with the same timing), the important exception being that the victims and the culprits were different persons. However, care was taken that the victim in Video 1 was as similar as possible to the victim in Video 2 and that each of the four culprits in Video 1 was as similar as possible to one of the four culprits in Video 2 (i.e., Hooligan A in Video 1 matched Hooligan A



Figure 2. Parallel scenes from the two versions of the video. The videos depicted a staged crime. The culprits and the victim were volunteer actors.

in Video 2, Hooligan B in Video 1 matched Hooligan B in Video 2 and so on). It was randomly determined which of the two parallel versions of the video was shown. The videos were shown in a resolution of 885×500 pixels.

For each of the two parallel videos, we created a long version that lasted about 130 s and a short version that lasted only 13 s (the rest of the video was cut off). In both versions of the videos, the culprits' faces were clearly visible from a frontal view. Participants were randomly assigned to one of the two exposure-duration conditions. About half of the participants saw the long video while the other half of the participants saw the short video.

Participants could start the video by clicking on a "Start" button. The control elements of the video player were disabled so that participants could not fast-forward, replay or stop the video. After the video had been presented, participants received an attention-check question about the persons they saw in the video. They passed the attention check when they were able to remember that "soccer fans" had been shown in the video by choosing this option among ten alternatives. The option "athletes" (intended to be one of the nine distractors) was frequently chosen instead of the nominally correct option which suggests that this option was unintentionally ambiguous. Therefore, selecting this option was counted as correct in Experiment 1. This option was replaced by the less ambiguous option "waiters" in all subsequent experiments.

Lineup procedures. Participants were informed that they had to identify the FC Bayern München hooligans from the video they had just seen. Participants received two-sided lineup instructions that emphasized both the need to identify the culprit if the culprit was present and the need to reject the lineup if the culprit was absent. Participants were not informed about how many lineups were about to follow.

Participants saw four separate lineups, two culprit-present lineups and two culprit-absent lineups, in random order. Each of the four lineups consisted of the facial photographs of six persons, one suspect and five fillers. In each of the two culprit-present lineups, a randomly selected face of one of the hooligans of the video the participants had seen was presented among the fillers. In each of the two culprit-absent lineups, the face of an innocent suspect was presented that the participants had not witnessed committing a crime. The innocent suspect was one of the hooligans from the video that the participants had not seen. The innocent suspect had the same hair color, hair style and stature as one of the hooligans from the video the participants had seen. For instance, if participants had seen Video 1, two randomly selected hooligans (e.g., Hooligan B and Hooligan C) from Video 1 served as the culprits in the two culprit-present lineups, while two of the hooligans from Video 2 (Hooligan A and Hooligan D in this example) served as the innocent suspects in the culprit-absent lineups. The pictures of the fillers were taken from the CVL database of Minear and Park⁵⁹. Given that all suspects were young male adults, 20 pictures of male adults aged between 18 and 29 years were chosen as fillers. The fillers roughly resembled the culprit or innocent suspect in hair color, hairstyle and stature. Together with the fact that it was randomly determined whether participants saw Video 1 or Video 2, this procedure served to ensure that the culprits and innocent suspects differed to the same degree, on average, from the fillers in the lineup. In this way the procedure is highly ecologically valid because the situation corresponds closely to that of a real-world lineup in which the photograph of the suspect (whose status as culprit or innocent suspect is unknown to the police), stems from a different source (e.g., social media) than the photographs of the fillers (e.g., a database) and may thus differ to some degree from that of the fillers. All photographs showed the faces from a frontal view with a neutral facial expression against a black background with no clothes visible. The photographs were edited to harmonize face sizes and lighting conditions and were presented in a resolution of 142×214 pixels. The order of the lineups was randomized, as was the position of the suspect and the fillers in each lineup.

It was randomly determined whether participants saw simultaneous or sequential lineups (Fig. 3). In a simultaneous lineup, the photographs of the suspect and the filler faces were presented in a row. Participants clicked on the "Yes, was present" button below a photograph if they thought that it showed one of the culprits. They clicked on the "No, none of these persons was present" button to reject the lineup. To make the procedure similar to that of a typical police lineup, participants also indicated how confident they were that their decision was correct. The participants initiated the presentation of the next lineup by clicking on a "Continue" button. In a sequential lineup, the photographs were presented successively. Participants decided for each photograph whether it showed one of the culprits or not by clicking on the "Yes, was present" button below the photograph or the "No, this person was not present" button presented at the right side of the screen, respectively. Participants

Example of a simultaneous lineup:



Example of a sequential lineup:

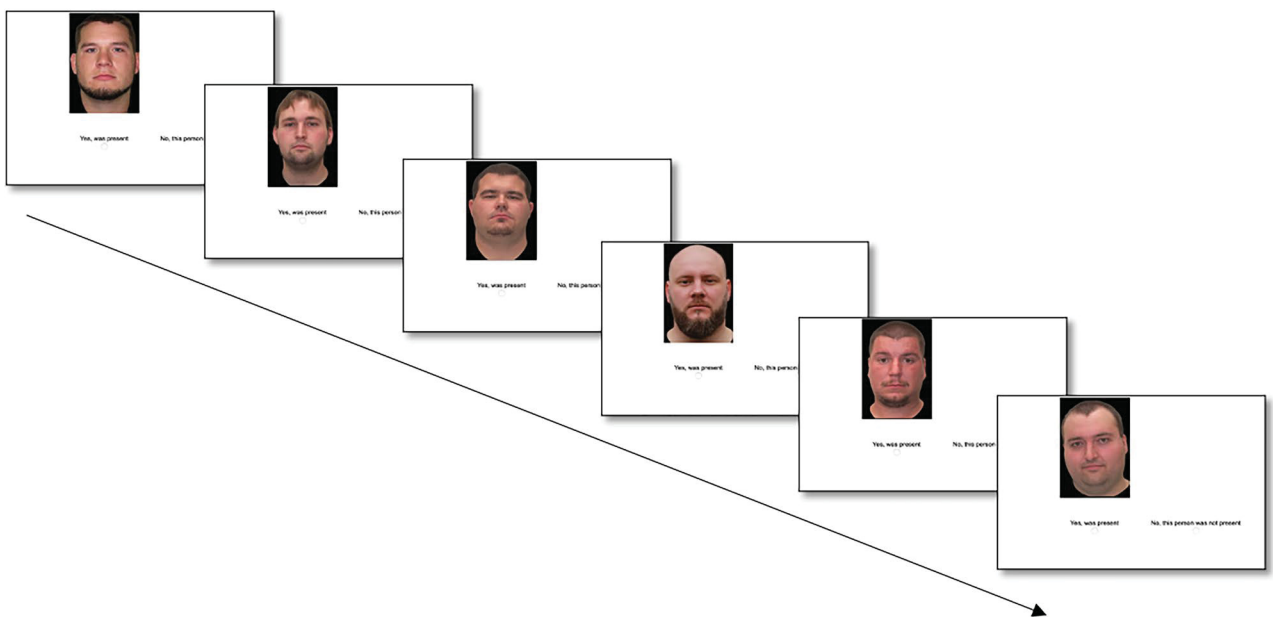


Figure 3. Examples of the simultaneous and sequential lineup procedures. The figure shows English translations of the German labels shown to the German participants.

had to make a decision for each of the six photographs. If a participant identified more than one person in a lineup, the last decision was used in the analysis, in accordance with standard police procedures in Germany⁶⁰ and other jurisdictions (cf.^{61–63}) and parallel to the simultaneous lineups in which participants were able to revise their decision before they clicked on the “Continue” button. To make the procedure similar to that of a typical police lineup, participants also indicated how confident they were that their decision was correct. The lineup was rejected if none of the members of the lineup was identified as one of the culprits.

Results. For all analyses reported in this article, model fits and parameter estimates were obtained using *multiTree*³⁰. One instance of the model illustrated in Fig. 1 was needed for each cell of the 2 (exposure duration: long exposure vs. short exposure) × 2 (lineup procedure: simultaneous vs. sequential) between-subjects design. The raw response frequencies are reported in Table 2.

Our goal was to begin with a base model that was as simple as possible. Therefore, we used whatever we could derive from the design of the studies to impose restrictions onto the 2-HT eyewitness identification model. First, it is obvious that lineup fairness must necessarily be the same across conditions because the same lineups were used in all conditions. Therefore, we set the biased-suspect-selection parameter *b* to be equal across all conditions. Second, for the same reason parameter *dA* was also set to be equal across all conditions. The base model incorporating these trivial restrictions fit the data, $G^2(6) = 3.91, p = 0.689$. Multinomial processing-tree models allow to test hypotheses directly at the level of the postulated processes. For instance, the hypothesis that the culprit-presence detection parameter *dP* is higher in the long-exposure condition than in the short-exposure condition can be implemented by setting parameter *dP* to be equal between these conditions. If the model including this equality restriction provides a significantly worse fit to the data than the base model, then it is necessary to conclude that parameter *dP* differs between conditions. The estimates of the culprit-presence

	Culprit-present lineups			Culprit-absent lineups		
	Culprit identifications	Filler identifications	Lineup rejections	Innocent-suspect identifications	Filler identifications	Lineup rejections
Experiment 1						
Simultaneous						
Long exposure	147	94	141	38	138	206
Short exposure	64	152	166	44	157	181
Sequential						
Long exposure	140	164	58	46	212	104
Short exposure	80	193	85	57	213	88
Experiment 2						
Simultaneous						
Fair	156	99	129	30	140	214
Unfair	172	71	135	60	100	218
Sequential						
Fair	137	164	65	43	213	110
Unfair	177	155	52	81	182	121
Experiment 3						
Simultaneous						
Low culprit probability	128	78	190	28	119	249
High culprit probability	141	128	107	44	167	165
Sequential						
Low culprit probability	104	119	151	38	144	192
High culprit probability	136	164	62	56	199	107
Experiment 4						
Simultaneous						
Difficult to reject	142	121	135	40	163	195
Easy to reject	136	103	119	19	140	199
Sequential						
Difficult to reject	148	165	85	52	233	113
Easy to reject	125	183	82	35	191	164

Table 2. Raw response frequencies observed in Experiments 1 to 4.

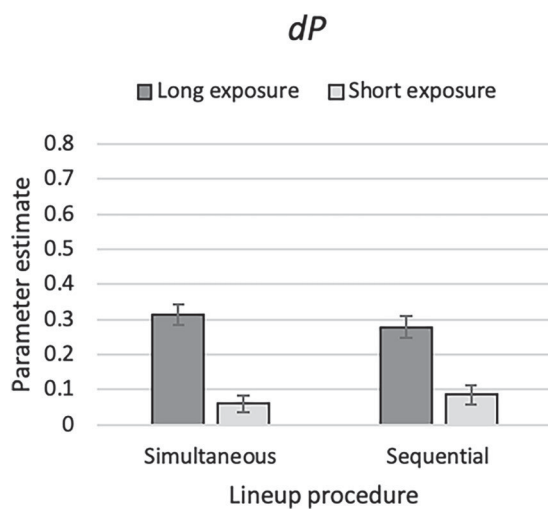


Figure 4. Estimates of parameter dP of the 2-HT eyewitness identification model representing the probability of detecting the presence of the culprit in Experiment 1 as a function of exposure duration (long exposure or short exposure) and lineup procedure (simultaneous or sequential). The error bars represent the standard errors.

Lineup procedure	Exposure duration	Parameter		
		<i>b</i>	<i>g</i>	<i>dA</i>
Simultaneous	Long exposure	0.03 (0.01)	0.45 (0.02)	0.03 (0.03)
	Short exposure		0.52 (0.02)	
Sequential	Long exposure		0.75 (0.02)	
	Short exposure		0.75 (0.02)	

Table 3. Parameter estimates and standard errors (in parentheses) of parameters *b*, *g* and *dA* in Experiment 1.

detection parameter *dP* are shown in Fig. 4. Parameter *dP* was significantly higher in the long-exposure condition than in the short-exposure condition, $\Delta G^2(2) = 72.21$, $p < 0.001$.

The estimates of parameters *b*, *g* and *dA* are reported in Table 3. In the simultaneous lineups, the probability of guessing-based selection (captured by parameter *g*) was significantly higher when exposure duration was short than when it was long, $\Delta G^2(1) = 6.47$, $p = 0.011$, which is to be expected given the well-known phenomenon of compensatory guessing^{64–67}. Interestingly, exposure duration had no effect on guessing-based selection in the sequential lineups, $\Delta G^2(1) = 0.03$, $p = 0.871$, suggesting that the sequential presentation of the faces in the lineup prevented compensatory guessing.

Discussion. Long exposure of the culprit's face provides ample opportunity for encoding which should increase the probability of detecting the culprit's face in a lineup. Experiment 1 confirmed this prediction. Parameter *dP* was significantly higher in the long-exposure condition than in the short-exposure condition. The validation of parameter *dP* was thus successful; parameter *dP* sensitively reflected the manipulation of culprit-presence detection in the predicted direction.

In simultaneous lineups, parameter *g* was also affected by exposure duration. In the short-exposure condition in which culprit-presence detection was poor, guessing-based selection was more prevalent than in the long-exposure condition in which culprit-presence detection was better. This effect on parameter *g* is expected given the well-known phenomenon of compensatory guessing: people often rely on guessing to compensate for poor memory^{64–67}. Note that the simplest situation in model validation is one in which only the target parameter (*dP* in the present experiment) is affected by the manipulation. This is not always possible because manipulations may have side effects. The ideal situation in such cases is one in which there is an obvious and plausible explanation of these side effects—such as compensatory guessing in the present instance. Interestingly, evidence of compensatory guessing was absent in the sequential-lineup condition. The sequential presentation of the faces may thus protect against compensatory guessing: in sequential lineups, participants may not increase their reliance on guessing when culprit-presence detection is poor because they cannot know whether the face of the culprit is yet to be presented. In simultaneous lineups, it is more obvious that the culprit cannot be detected, which may provide more favorable conditions for compensatory guessing. This hypothesis can be further explored in future research.

Experiment 2: Biased selection of the suspect (parameter *b*)

Experiment 2 served to test the validity of parameter *b* which is assumed to represent the probability of biased selection of a culprit or an innocent suspect who stands out from the fillers. Biased suspect selection is thus assumed to occur when lineups are unfair. In Experiment 1, care was taken to create fair lineups in which the suspect did not stand out from the fillers. The fact that the estimate of parameter *b* was extremely low suggests that biased suspect selection occurred with a very low probability. In Experiment 2, this fair-lineup condition was contrasted with an unfair-lineup condition. A straightforward method of creating unfair lineups is to add a conspicuous feature such as a birthmark to the suspect's face that distinguishes this face from the filler faces⁶⁸. In line with this approach, unfair lineups were created by adding large birthmarks to all filler faces but not to the faces of the culprits and innocent suspects. The suspect's face thus stood out from these fillers in that it was the only face without a birthmark. If the 2-HT eyewitness identification model is valid, then the biased-suspect-selection parameter *b* should be higher in the unfair-lineup condition than in the fair-lineup condition.

Method. *Sample.* Participants who had not participated in Experiment 1 were recruited using the *Gapfish* research panel. We aimed at achieving about the same sample size as in Experiment 1. In total, 930 participants gave informed consent but 139 of them did not complete the study. A total of 19 datasets had to be excluded because of multiple participations. Another 16 datasets had to be excluded because participants had failed the attention check or because they had reported technical problems. The final sample included 756 participants (355 women and 401 men) with a mean age of 44 years ($SD = 15$). A sensitivity analysis with *G*Power*⁶⁹ showed that with a sample size of $N = 756$, four eyewitness identification decisions, $\alpha = 0.05$ and a statistical power of $1 - \beta = 0.95$, it was possible to detect even small effects of the unfairness on the biased-selection parameter *b* of size $w = 0.07$.

Materials and procedure. Experiment 2 was identical to Experiment 1, with the following exceptions. Participants saw the full-length staged-crime video used in the long-exposure condition of Experiment 1. Unfair lineups were created by digitally manipulating the filler photographs to create lineups in which all lineup members

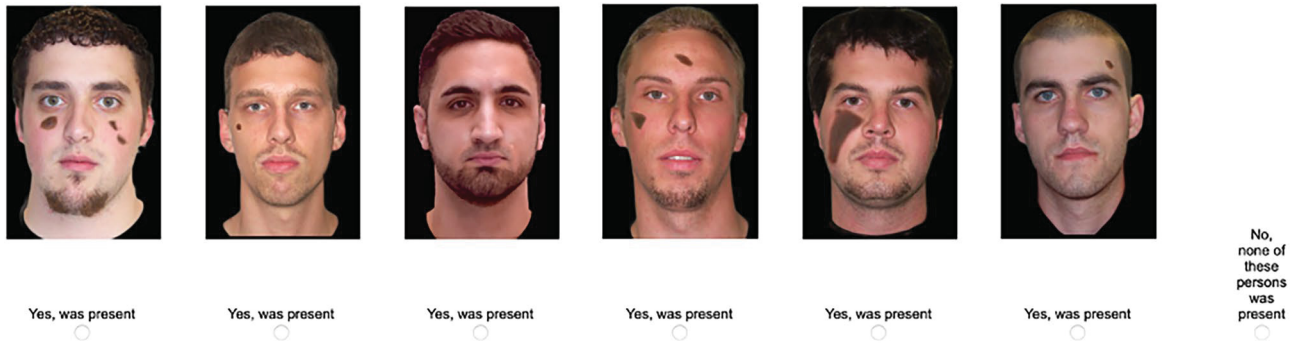


Figure 5. Example of an unfair simultaneous lineup used in Experiment 2. The figure shows English translations of the German labels shown to the German participants.

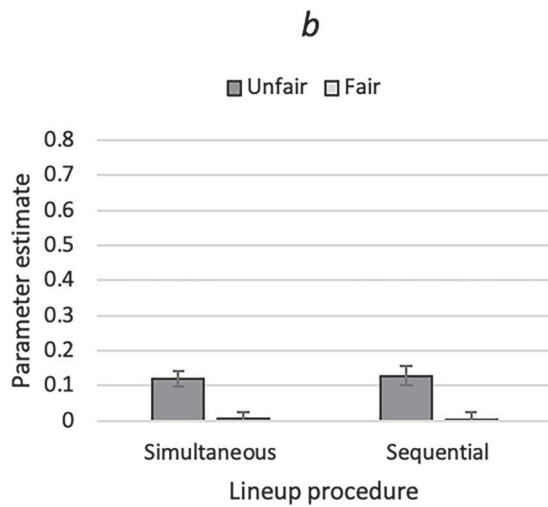


Figure 6. Estimates of parameter *b* of the 2-HT eyewitness identification model representing the probability of biased suspect selection in Experiment 2 as a function of the unfairness of the lineup (unfair or fair) and lineup procedure (simultaneous or sequential). The error bars represent the standard errors.

Lineup procedure	Lineup fairness	Parameter		
		<i>dP</i>	<i>g</i>	<i>dA</i>
Simultaneous	Unfair	0.34 (0.04)	0.40 (0.02)	0.11 (0.03)
	Fair	0.35 (0.03)	0.49 (0.02)	
Sequential	Unfair	0.29 (0.04)	0.76 (0.02)	
	Fair	0.28 (0.03)	0.77 (0.02)	

Table 4. Parameter estimates and standard errors (in parentheses) of parameters *dP*, *g* and *dA* in Experiment 2.

except the suspect had large birthmarks on their faces (Fig. 5). In the fair-lineup condition, participants saw the same lineups as in Experiment 1.

Results. One instance of the model illustrated in Fig. 1 was needed for each cell of the 2 (lineup fairness: unfair vs. fair) × 2 (lineup procedure: simultaneous vs. sequential) design. The raw response frequencies are reported in Table 2. As in Experiment 1 and for the reason given there, parameter *dA* was set to be equal across all conditions. The base model incorporating these restrictions fit the data, $G^2(3) = 2.32, p = 0.509$. The estimates of the biased-suspect-selection parameter *b* are shown in Fig. 6. Parameter *b* was significantly higher when the lineup was unfair than when it was fair, $\Delta G^2(2) = 28.59, p < 0.001$.

The estimates of parameters *dP*, *g* and *dA* are reported in Table 4. Culprit-presence detection (captured by parameter *dP*) did not differ as a function of lineup fairness, $\Delta G^2(2) = 0.12, p = 0.940$. Guessing-based selection

(captured by parameter g) was less likely when the lineup was unfair than when it was fair, but this difference was significant only in the simultaneous lineup condition, $\Delta G^2(1) = 7.93$, $p = 0.005$, and not in the sequential lineup condition, $\Delta G^2(1) = 0.09$, $p = 0.760$.

Discussion. The results support the validity of the interpretation that parameter b represents the process of biased suspect selection: parameter b was higher in the unfair conditions than in the fair conditions. In simultaneous lineups, lineup fairness also affected parameter g . A possible explanation is that guessing-based selection may have been reduced by the relative ease with which the fillers could be ruled out due to their large birthmarks in the unfair-lineup condition.

Experiment 3: Guessing-based selection among the lineup members (parameter g)

Experiment 3 served to test the validity of parameter g which is assumed to reflect the probability of guessing-based selection among the lineup members. A well-established way to manipulate guessing in old-new recognition paradigms is to create the expectation (e.g., via instructions) that a stimulus is more likely to be old than new or vice versa⁷⁰. Similarly, one-sided pre-lineup instructions insinuating that the culprit is in the lineup increase the witnesses' willingness to guess that one of the lineup members is the culprit compared to two-sided pre-lineup instructions that emphasize the possibility that the culprit may not be in the lineup (e.g.^{50,71,72}). For instance, Lindsay et al.⁷³ manipulated guessing by providing either one-sided pre-lineup instructions stating that "The guilty party is in the lineup, all you have to do is pick him out" or two-sided pre-lineup instructions stating that "the guilty party may or may not be in the lineup" (p. 799). Similarly, half of the participants in Experiment 3 received the instruction that the probability of the culprit being in the lineup was high which should encourage participants to select one of the lineup members as the culprit based on guessing. The other half of the participants received the instruction that the probability of the culprit being in the lineup was low which should discourage participants from selecting one of the lineup members as the culprit based on guessing. If parameter g of the 2-HT eyewitness identification model is valid, then parameter g should be higher in the high-culprit-probability condition than in the low-culprit-probability condition.

Method. *Sample.* Participants who had not participated in Experiments 1 and 2 were recruited using the *Gapfish* research panel. We aimed at achieving about the same sample size as in Experiments 1 and 2. In total, 896 participants gave informed consent but 108 of them did not complete the study. A total of 23 datasets had to be excluded because of multiple participations. Another 11 datasets had to be excluded because participants had failed the attention check or because they had reported technical problems. The final sample included 754 participants (336 women, 417 men, 1 diverse) with a mean age of 45 years ($SD = 15$). A sensitivity analysis with *G*Power*⁶⁹ showed that with a sample size of $N = 754$, four eyewitness identification decisions, $\alpha = 0.05$ and a statistical power of $1 - \beta = 0.95$, it was possible to detect even small effects of the pre-lineup instructions on the guessing-based-selection parameter g of size $w = 0.07$.

Materials and procedure. Experiment 3 was identical to Experiment 1 with the following exceptions. Participants saw the full-length staged-crime video used in the long-exposure condition of Experiment 1. In the high-culprit-probability condition, the instructions read: "It is **likely** that there is one of the culprits in each of the lineups. Therefore, you should select the 'Yes, was present' button that belongs to the recognized face if one of the persons feels familiar". Before each lineup, participants in the high-culprit-probability condition received the following reminder: "It is likely that one of the culprits is in the lineup. You simply have to choose him". In the low-culprit-probability condition, the instructions read: "It is **unlikely** that one of the culprits is in the lineup. Therefore, you should select the 'Yes, was present' button that belongs to the recognized face only if you are very certain that you have recognized the right person". Before each lineup, participants in the low-culprit-probability condition received the following reminder: "It is unlikely that one of the culprits is in the lineup. Please choose someone only if you are very certain".

Results. One instance of the model illustrated in Fig. 1 was needed for each cell of the 2 (pre-lineup instructions: high culprit probability vs. low culprit probability) \times 2 (lineup procedure: simultaneous vs. sequential) design. The raw response frequencies are reported in Table 2. As in Experiment 1 and for the reason given there, parameters b and dA were set to be equal across all conditions. The base model incorporating these restrictions fit the data, $G^2(6) = 3.85$, $p = 0.698$. The estimates of the guessing-based selection parameter g are shown in Fig. 7. Parameter g was significantly higher in the high-culprit-probability condition than in the low-culprit-probability condition, $\Delta G^2(2) = 137.36$, $p < 0.001$.

The estimates of parameters dP , b and dA are reported in Table 5. Culprit-presence detection (captured by parameter dP) did not differ as a function of the pre-lineup instructions, $\Delta G^2(2) = 4.13$, $p = 0.127$.

Discussion. The results support the validity of the interpretation that parameter g represents the probability of guessing-based selection among the lineup members. Parameter g was higher when participants expected culprits in the lineups with a high rather than low probability.

Experiment 4: Detection of the absence of the culprit (parameter dA)

Experiment 4 served to test the validity of parameter dA which is assumed to represent the probability of detecting the absence of the culprit. To manipulate parameter dA , the standard lineups were compared against lineups in which it could easily be detected that none of the lineup members was the culprit. In these easy-to-reject

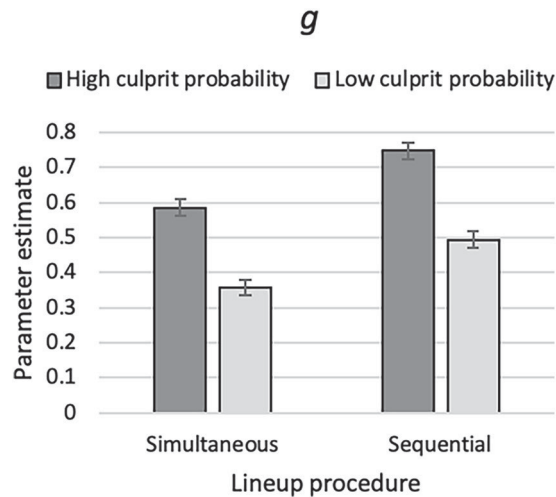


Figure 7. Estimates of parameter g of the 2-HT eyewitness identification model representing the probability of guessing-based selection among the lineup members in Experiment 3 as a function of the pre-lineup instructions (high culprit probability or low culprit probability) and lineup procedure (simultaneous or sequential). The error bars represent the standard errors.

Lineup procedure	Pre-lineup instructions	Parameter		
		dP	b	dA
Simultaneous	High culprit probability	0.29 (0.03)	0.03 (0.01)	0.05 (0.03)
	Low culprit probability	0.26 (0.03)		
Sequential	High culprit probability	0.27 (0.03)		
	Low culprit probability	0.19 (0.03)		

Table 5. Parameter estimates and standard errors (in parentheses) of parameters dP , b and dA in Experiment 3.

culprit-absent lineups, all lineup members, including the innocent suspect, shared a conspicuous facial feature that could be used to rule out the lineup members as culprits. These easy-to-reject lineups were compared to standard culprit-absent lineups that were difficult to reject because all lineup members, including the innocent suspect, matched the culprit in appearance. If the 2-HT eyewitness identification model is valid, the culprit-absence detection parameter dA should be higher when the lineup is easy to reject than when it is difficult to reject.

Method. *Sample.* Participants who had not participated in Experiments 1 to 3 were recruited using the *Gapfish* research panel. We aimed at achieving about the same sample size as in Experiments 1 to 3. In total, 929 participants gave informed consent but 117 of them did not complete the study. A total of 22 datasets had to be excluded because of multiple participations. Another 18 datasets had to be excluded because participants had failed the attention check or because they had reported technical problems. The final sample included 772 participants (331 women, 437 men, 4 diverse) with a mean age of 46 years ($SD=15$). A sensitivity analysis with *G*Power*⁶⁹ showed that with a sample size of $N=772$, four eyewitness identification decisions, $\alpha=0.05$ and a statistical power of $1-\beta=0.95$, it was possible to detect even small effects of the ease of lineup rejection on the detection of the absence of the culprit parameter dA of size $w=0.07$.

Materials and procedure. Experiment 4 was identical to Experiment 1 with the following exceptions. Participants saw the full-length staged-crime video used in the long-exposure condition of Experiment 1. For the easy-to-reject condition, culprit-absent lineups were created by digitally manipulating the facial photographs such that all lineup members, including the innocent suspect, had large birthmarks on their faces, just like the filler faces in the unfair-lineup condition of Experiment 2 (Fig. 5). This manipulation was designed to facilitate the detection of culprit absence. In the difficult-to-reject condition, participants saw the same culprit-absent lineups as in Experiment 1.

Results. One instance of the model illustrated in Fig. 1 was needed for each cell of the 2 (ease of rejection: difficult to reject vs. easy to reject) \times 2 (lineup procedure: simultaneous vs. sequential) between-subjects design.

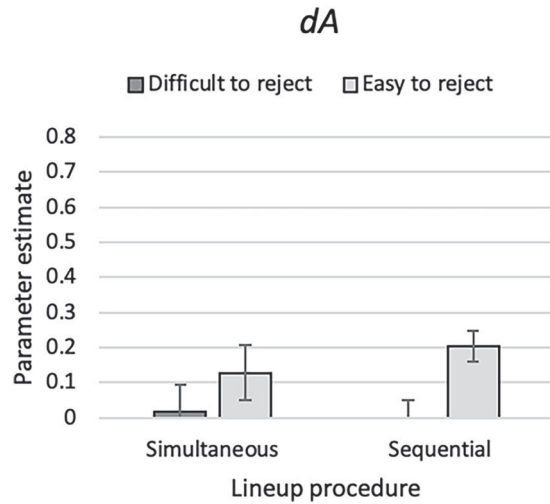


Figure 8. Estimates of parameter dA of the 2-HT eyewitness identification model representing the probability of culprit-absence detection in Experiment 4 as a function of ease of rejection of the culprit-absent lineups (difficult to reject or easy to reject) and lineup procedure (simultaneous or sequential). The error bars represent the standard errors.

Lineup procedure	Ease of rejection	Parameter		
		dP	b	g
Simultaneous	Difficult to reject	0.30 (0.03)	0.00 (0.01)	0.52 (0.03)
	Easy to reject	0.32 (0.03)		0.51 (0.03)
Sequential	Difficult to reject	0.29 (0.03)		0.71 (0.03)
	Easy to reject	0.23 (0.03)		0.73 (0.03)

Table 6. Parameter estimates and standard errors (in parentheses) of parameters dP , b and g in Experiment 4.

The raw response frequencies are reported in Table 2. As in Experiment 1 and for the reason given there, parameter b was set to be equal across all conditions. The base model incorporating these restrictions fit the data, $G^2(3) = 5.00, p = 0.172$. The estimates of the culprit-absence detection parameter dA are shown in Fig. 8. Parameter dA was significantly higher when the culprit-absent lineup was easy to reject than when it was difficult to reject, $\Delta G^2(2) = 11.95, p = 0.003$.

The estimates of parameters dP , b and g are reported in Table 6. The ease-of-rejection manipulation affected neither culprit-presence detection (captured by parameter dP), $\Delta G^2(2) = 2.93, p = 0.231$, nor guessing-based selection (captured by parameter g), $\Delta G^2(2) = 0.38, p = 0.827$.

Discussion. The results support the validity of the interpretation that parameter dA represents the probability of detecting the absence of the culprit. Parameter dA was enhanced when the culprit-absent lineups were easy to reject because all lineup members, including the innocent suspect, had the same conspicuous facial feature which could be used to rule them out as culprits.

General discussion

Here we introduce the 2-HT eyewitness identification model that serves to measure latent processes underlying eyewitness identification performance. Before such a measurement model is applied to novel and unresolved research questions, it is important to empirically assess whether the model's parameters sensitively reflect the processes they were designed to measure. This was done here by testing whether manipulations that can reasonably be expected to target culprit-presence detection, biased suspect selection, guessing-based selection and culprit-absence detection are sensitively reflected in the model's parameters. In Experiment 1, parameter dP sensitively reflected the manipulation of exposure duration that can be assumed to affect the detection of the presence of the culprit; parameter dP was larger when the culprits' faces were visible for a long time than when they were visible for only a short time. In Experiment 2, the manipulation of lineup fairness was sensitively reflected in parameter b that was designed to capture the biased selection of the suspect which occurs in unfair lineups; parameter b was larger when a conspicuous facial feature distinguished the filler faces from the suspects' faces than when this conspicuous facial feature was absent in all faces of the lineups. In Experiment 3, the manipulation of the pre-lineup instructions was sensitively reflected in parameter g that was designed to measure guessing-based selection; parameter g was larger when the participants were led to expect culprits in the lineups with a high rather

than low probability. The effects demonstrated in Experiments 1 to 3 are closely aligned with well-established findings in the literature^{40,44,46,48,50,53,68,70,72–81}. This reflects the deliberate strategy to use, for the purpose of validation experiments, manipulations that are well understood so that one can be certain as to which processes and, hence, which model parameters they should affect. This strategy was not readily available for the process of the detection of culprit absence, reflected in parameter dA , possibly because other approaches for analyzing lineup data do not specifically postulate a separate process for culprit-absence detection so that this process is less well studied. Therefore, an intentionally trivial manipulation was used in Experiment 4: culprit-absence detection should be easier when members of culprit-absent lineups can be easily rejected based on a conspicuous facial feature that distinguishes them from the culprit. This ease-of-rejection manipulation was reflected in parameter dA that was designed to capture the detection of the absence of the culprit; parameter dA was higher when the members of the culprit-absent lineups were easy to reject than when they were difficult to reject.

It is a distinguishing feature of the 2-HT eyewitness identification model that it differentiates between two types of detection processes: the detection of culprit presence and the detection of culprit absence. Typical analyses of lineup data provide a single accuracy measure that is intended to reflect the witnesses' ability to distinguish between culprit presence and absence. The 2-HT eyewitness identification model goes one step further by allowing to measure these two processes separately. The results of the present series of validation experiments suggest that these two types of detection processes can be manipulated independently of each other. The manipulation of exposure duration affected the culprit-presence detection parameter dP but not the culprit-absence detection parameter dA , whereas the manipulation of the ease with which culprit-absent lineups can be rejected influenced dA but not dP . Including a separate parameter for culprit-absence detection in a measurement model may help to stimulate thinking about how lineup procedures may help to improve the process of culprit-absence detection.

Another distinguishing feature of the present model is that it differentiates between two forms of non-detection-based processes. Parameter g reflects the process of selecting a lineup member based on guessing, while parameter b reflects the biased selection of a suspect who stands out from the fillers in an unfair lineup. This distinction is, of course, not entirely new, but the qualitative distinction between the two types of non-detection-based selection processes may be emphasized if they are distinguished in a measurement model. In the lineups used in Experiments 1, 3 and 4 and in the fair-lineup condition of Experiment 2, care was taken to create fair lineups by selecting fillers that resembled the culprits and the innocent suspects in hair color, hairstyle and stature and by making sure that the photographs of the fillers could not be distinguished from those of the culprit and the innocent suspect based on head size or lighting conditions. When lineups are fair, the biased-suspect-selection parameter b can be expected to be close to zero, which was confirmed by the present experiments. Experiment 2 shows that the 2-HT eyewitness identification model is also able to detect biased suspect selection in unfair lineups. We consider it a strength of this model that it allows measuring biased suspect selection in unfair lineups directly based on the eyewitness identification data of interest as opposed to having to rely on indirect data from a separate mock-witness task for determining lineup fairness⁸².

The main purpose of including both simultaneous and sequential lineups in the present experiments was to test whether the 2-HT eyewitness identification model validly reflects eyewitness identification processes in both types of lineups. Indeed, manipulations of exposure duration, lineup fairness, pre-lineup instructions and ease of rejection were sensitively reflected in culprit-presence detection, biased suspect selection, guessing-based selection and culprit-absence detection for simultaneous and sequential lineups, which supports the model's validity for measuring the latent processes underlying eyewitness identification decisions in simultaneous and sequential lineups. Beyond this main purpose it seems interesting to explore which type of lineup procedure is superior in terms of culprit-presence detection. A superiority of sequential lineups has been postulated based on the diagnosticity ratio⁴⁴ which, however, has been criticized for confounding the ability to distinguish between a culprit and an innocent suspect and response bias²⁰. SDT-based analyses have sometimes shown equivalent performance of simultaneous and sequential lineups^{13,37,39,47,83,84} and sometimes a superiority of simultaneous over sequential lineups^{11,38,62,85–88}. In all experiments reported here, the estimates of the culprit-presence detection parameter dP were slightly but consistently higher in the simultaneous-lineup conditions than in the sequential-lineup conditions. This difference in parameter dP was not statistically significant in any of the individual experiments, but when we combined the data from all experiments (Fig. 9), dP was significantly larger in simultaneous than in sequential lineups ($\Delta G^2[1] = 10.13, p = 0.001$; details of this analysis are reported in the Supplementary Materials that can be found in the OSF repository). This result nicely converges with those of SDT-based analyses: there is an advantage of simultaneous compared to sequential lineups in terms of culprit-presence detection, but the advantage is small and can be reliably observed only when the sample sizes are large enough to guarantee sufficient statistical power to detect such small effects.

This convergence is not too surprising because the present 2-HT eyewitness identification model was designed to separate, among other things, culprit-presence detection (parameter dP) from guessing-based selection (parameter g), which is conceptually similar to the distinction between sensitivity (d') and response bias (c) in SDT⁸⁹. A more surprising finding is that, in the experimental paradigm used here, the estimate of the guessing-based selection parameter g was consistently higher in the sequential than in the simultaneous lineups. At first glance, this may seem unexpected because it has often been found that sequential lineups lead to more conservative responding than simultaneous lineups (e.g.⁸³). In the present experiments, the higher tendency towards guessing that the culprit is in the lineup is already apparent at the level of the raw response probabilities: the mean rate of lineup identifications was consistently higher in the sequential lineups than in the simultaneous lineups (0.77 to 0.55 in Experiment 1, 0.65 to 0.54 in Experiment 2, 0.77 to 0.54 in Experiment 3 and 0.72 to 0.57 in Experiment 4). The most plausible explanation of the discrepancy of the data reported here to the data obtained in previous studies is that we explicitly followed standard police protocols by continuing the presentation of the sequential lineup even after an early identification^{60–63}; only the last identification was used. By contrast, in many previous studies the *only-the-first-yes-counts* rule was used. Horry et al.⁶² have demonstrated

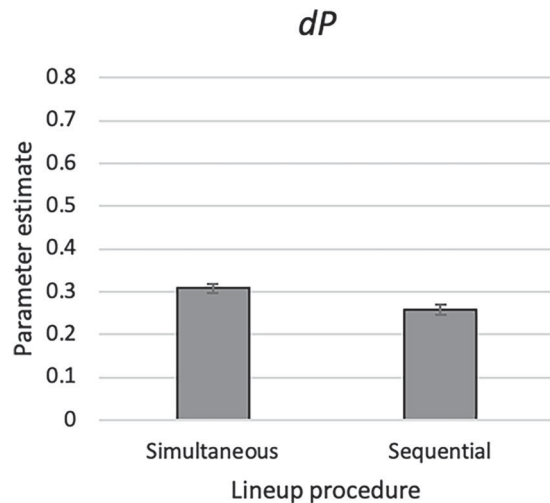


Figure 9. Estimates of parameter dP of the 2-HT eyewitness identification model representing the probability of detecting the presence of the culprit in simultaneous and sequential lineups, respectively. The estimates are based on the data of all four experiments, excluding the short-duration condition of Experiment 1. The difference in dP between the simultaneous conditions and the sequential conditions was small ($\Delta dP = 0.05$). The error bars represent the standard errors.

that only-the-first-yes-counts instructions systematically reduce the rate of positive identifications in sequential lineups by discouraging participants from guessing. This is plausible at a psychological level: participants may well shy away from using their only identification option when they cannot know, at the time of their decision, whether better options will follow. These findings suggest that the standard police protocol applied here may induce a particularly pronounced tendency to select a lineup member based on guessing in sequential lineups: even if a lineup member was selected at some point, a later lineup member could still be selected if the witness came to the conclusion that the later lineup member provided a better match to their memory for the culprits.

A limitation which the present approach has in common with other approaches is that it allowed us to estimate the different processes underlying eyewitness identification decisions only at the group level. Due to the small number of responses, it would be difficult to estimate parameters at the individual level⁹⁰. As the participants saw four lineups each, only four data points per participant were collected. Compared to other paradigms designed to collect data for MPT models (e.g.,^{21,22,25,27,91}), this is a comparatively small number of data points per participant. However, using a larger number of lineups would go against ecological validity because crimes with multiple culprits typically involve only a small number of culprits⁹².

Another limitation is that the 2-HT eyewitness identification model in the present version accounts only for lineup identifications and rejections but not for confidence judgements. Accounting for confidence judgements would thus require an extension of the model (cf.^{35,93–98}). A more general point is that MPT models are based on a threshold concept, assuming that observed behaviors are the result of discrete mental states rather than continuously distributed variables such as an assumed memory strength variable which is posited to underlie recognition decisions in SDT-based approaches²⁹. With respect to such recognition decisions, some researchers have argued that the threshold assumption of MPT models is inconsistent with the available empirical evidence^{99,100}, but others have shown that both SDT-based and threshold models can account for recognition memory performance^{34,35,89,93–98,101}. On the whole, it is important to note that all models are necessarily simplifications of the complexities of reality. The assumption of different mental states and of thresholds that have to be crossed to get from one state to another (inherent in all threshold models and, thus, in the 2-HT eyewitness identification model) is a simplification, but so is, for instance, the assumption of equal-variance, normal distribution of an assumed signal strength variable in models based on signal detection theory⁸⁹. The question is not which set of simplifications is correct because all simplifications of the complexities of reality must necessarily be false. Instead, the question is which set of simplifications is sufficiently useful. Here we introduced a novel MPT model to the field of eyewitness identification research which we hope may turn out to be useful in that it makes it possible to measure the four latent cognitive processes of culprit-presence and culprit-absence detection, biased selection and guessing-based selection using the *whole* 2×3 data matrix of the typical lineup identification task (Table 1). When introducing a novel MPT model, it is an important first step to experimentally validate the interpretation of the model parameters by providing an empirical test of whether the model parameters capture the processes they were designed to measure, which is the purpose of the present series of experiments. The theoretical and practical usefulness of threshold models in the field of eyewitness identification decisions can only be determined in the long run and thus has to be evaluated based on future research.

Conclusions

The purpose of the present study was to introduce, and to provide an experimental validation of, the novel 2-HT eyewitness identification model that takes into account the full 2×3 structure of lineup data. The present results support the validity of the model for analyzing lineup data. Now that the model parameters have been demonstrated to sensitively reflect the effects of well-established and obvious manipulations of the latent processes underlying eyewitness identification decisions in simultaneous and sequential lineups, the model can be used in future studies to measure the latent processes underlying the witnesses' decisions in lineups, for example, to test novel hypotheses.

Data availability

The datasets generated and analyzed during the current study, as well as supplementary materials are available in the OSF repository, <https://osf.io/qbzs2/>.

Received: 18 February 2022; Accepted: 30 August 2022

Published online: 16 September 2022

References

- Wells, G. L. *et al.* Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law Hum. Behav.* **44**, 3–36. <https://doi.org/10.1037/lhb0000359> (2020).
- Innocence Project. *Exonerate the Innocent*. <https://innocenceproject.org/exonerate/> (2022).
- Wells, G. L. *et al.* Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law Hum. Behav.* **22**, 603–647. <https://doi.org/10.1023/A:1025750605807> (1998).
- Lindsay, R. C. L. & Wells, G. L. Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *J. Appl. Psychol.* **70**, 556–564. <https://doi.org/10.1037/0021-9010.70.3.556> (1985).
- Wells, G. L. Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Curr. Dir. Psychol. Sci.* **23**, 11–16. <https://doi.org/10.1177/0963721413504781> (2014).
- Wells, G. L., Dysart, J. E. & Steblay, N. K. The flaw in Amendola and Wixted's conclusion on simultaneous versus sequential lineups: Comment. *J. Exp. Criminol.* **11**, 285–289. <https://doi.org/10.1007/s11292-014-9225-4> (2015).
- Wells, G. L. *et al.* From the lab to the police station. A successful application of eyewitness research. *Am. Psychol.* **55**, 581–598. <https://doi.org/10.1037//0003-066x.55.6.581> (2000).
- Wells, G. L., Smalarz, L. & Smith, A. M. ROC analysis of lineups does not measure underlying discriminability and has limited value. *J. Appl. Res. Mem. Cogn.* **4**, 313–317. <https://doi.org/10.1016/j.jarmac.2015.08.008> (2015).
- Wells, G. L., Smith, A. M. & Smalarz, L. ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *J. Appl. Res. Mem. Cogn.* **4**, 324–328. <https://doi.org/10.1016/j.jarmac.2015.08.010> (2015).
- Wells, G. L. & Lindsay, R. C. On estimating the diagnosticity of eyewitness nonidentifications. *Psychol. Bull.* **88**, 776–784 (1980).
- Mickes, L., Flowe, H. D. & Wixted, J. T. Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *J. Exp. Psychol. Appl.* **18**, 361–376. <https://doi.org/10.1037/a0030609> (2012).
- Wixted, J. T. & Mickes, L. A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychol. Rev.* **121**, 262–276. <https://doi.org/10.1037/a0035940> (2014).
- Gronlund, S. D. *et al.* Showups versus lineups: An evaluation using ROC analysis. *J. Appl. Res. Mem. Cogn.* **1**, 221–228. <https://doi.org/10.1016/j.jarmac.2012.09.003> (2012).
- Amendola, K. L. & Wixted, J. T. Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in a randomized field trial. *J. Exp. Criminol.* **11**, 263–284. <https://doi.org/10.1007/s11292-014-9219-2> (2015).
- Amendola, K. L. & Wixted, J. T. No possibility of a selection bias, but direct evidence of a simultaneous superiority effect: A reply to Wells *et al.* *J. Exp. Criminol.* **11**, 291–294. <https://doi.org/10.1007/s11292-015-9227-x> (2015).
- Gronlund, S. D., Mickes, L., Wixted, J. T. & Clark, S. E. In *Psychology of Learning and Motivation* Vol 63 (ed. Brian, H. S. R.) 1–43 (Elsevier, 2015).
- Wixted, J. T. & Mickes, L. Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *J. Appl. Res. Mem. Cogn.* **4**, 318–323. <https://doi.org/10.1016/j.jarmac.2015.08.009> (2015).
- Wixted, J. T. & Mickes, L. ROC analysis measures objective discriminability for any eyewitness identification procedure. *J. Appl. Res. Mem. Cogn.* **4**, 329–334. <https://doi.org/10.1016/j.jarmac.2015.08.007> (2015).
- Macmillan, N. A. & Creelman, C. D. *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, 2005).
- Mickes, L., Moreland, M. B., Clark, S. E. & Wixted, J. T. Missing the information needed to perform ROC analysis? Then compute d' , not the diagnosticity ratio. *J. Appl. Res. Mem. Cogn.* **3**, 58–62. <https://doi.org/10.1016/j.jarmac.2014.04.007> (2014).
- Bayen, U. J., Murnane, K. & Erdfelder, E. Source discrimination, item detection, and multinomial models of source monitoring. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 197–215. <https://doi.org/10.1037/0278-7393.22.1.197> (1996).
- Meiser, T. & Bröder, A. Memory for multidimensional source information. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 116–137. <https://doi.org/10.1037/0278-7393.28.1.116> (2002).
- Sherman, J. W., Groom, C. J., Ehrenberg, K. & Klauer, K. C. Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Soc. Cogn.* **21**, 213–246. <https://doi.org/10.1521/soco.21.3.213.25340> (2003).
- Smith, R. E. & Bayen, U. J. A multinomial model of event-based prospective memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **30**, 756–777. <https://doi.org/10.1037/0278-7393.30.4.756> (2004).
- Erdfelder, E. & Buchner, A. Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *J. Exp. Psychol. Learn. Mem. Cogn.* **24**, 387–414. <https://doi.org/10.1037/0278-7393.24.2.387> (1998).
- Stahl, C. & Degner, J. Assessing automatic activation of valence: A multinomial model of EAST performance. *Exp. Psychol.* **54**, 99–112. <https://doi.org/10.1027/1618-3169.54.2.99> (2007).
- Unkelbach, C. & Stahl, C. A multinomial modeling approach to dissociate different components of the truth effect. *Conscious. Cogn.* **18**, 22–38. <https://doi.org/10.1016/j.concog.2008.09.006> (2009).
- Erdfelder, E. *et al.* Multinomial processing tree models: A review of the literature. *Z. Psychol. / J. Psychol.* **217**, 108–124. <https://doi.org/10.1027/0044-3409.217.3.xxx> (2009).
- Schmidt, O., Erdfelder, E. & Heck, D. W. *Tutorial on multinomial processing tree modeling: How to develop, test, and extend MPT models*. PsyArXiv. <https://doi.org/10.31234/osf.io/gh8md> (2022).
- Moshagen, M. multiTree: A computer program for the analysis of multinomial processing tree models. *Behav. Res. Methods* **42**, 42–54. <https://doi.org/10.3758/BRM.42.1.42> (2010).

31. Singmann, H. & Kellen, D. MPTinR: Analysis of multinomial processing tree models in R. *Behav. Res. Methods* **45**, 560–575. <https://doi.org/10.3758/s13428-012-0259-0> (2013).
32. Stahl, C. & Klauer, K. C. HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behav. Res. Methods* **39**, 267–273. <https://doi.org/10.3758/BF03193157> (2007).
33. Bell, R., Mieth, L. & Buchner, A. Emotional memory: No source memory without old-new recognition. *Emotion* **17**, 120–130. <https://doi.org/10.1037/emo0000211> (2017).
34. Snodgrass, J. G. & Corwin, J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *J. Exp. Psychol. Gen.* **117**, 34–50. <https://doi.org/10.1037/0096-3445.117.1.34> (1988).
35. Bröder, A., Kellen, D., Schütz, J. & Rohrmeier, C. Validating a two-high-threshold measurement model for confidence rating data in recognition. *Memory* **21**, 916–944. <https://doi.org/10.1080/09658211.2013.767348> (2013).
36. Menne, N. M., Winter, K., Bell, R. & Buchner, A. A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Sci. Rep.* **12**, 13379. <https://doi.org/10.1038/s41598-022-17400-y> (2022).
37. Andersen, S. M., Carlson, C. A., Carlson, M. A. & Gronlund, S. D. Individual differences predict eyewitness identification performance. *Person. Individ. Differ.* **60**, 36–40. <https://doi.org/10.1016/j.paid.2013.12.011> (2014).
38. Carlson, C. A. & Carlson, M. A. An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *J. Appl. Res. Mem. Cogn.* **3**, 45–53. <https://doi.org/10.1016/j.jarmac.2014.03.004> (2014).
39. Flowe, H. D., Smith, H. M. J., Karoğlu, N., Onwuegbusi, T. O. & Rai, L. Configural and component processing in simultaneous and sequential lineup procedures. *Memory* **24**, 306–314. <https://doi.org/10.1080/09658211.2015.1004350> (2016).
40. Memon, A., Hope, L. & Bull, R. Exposure duration: Effects on eyewitness accuracy and confidence. *Br. J. Psychol.* **94**, 339–354. <https://doi.org/10.1348/000712603767876262> (2003).
41. Palmer, M. A., Brewer, N., Weber, N. & Nagesh, A. The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *J. Exp. Psychol. Appl.* **19**, 55–71. <https://doi.org/10.1037/a0031602> (2013).
42. Valentine, T., Pickering, A. & Darling, S. Characteristics of eyewitness identification that predict the outcome of real lineups. *Appl. Cogn. Psychol.* **17**, 969–993. <https://doi.org/10.1002/acp.939> (2003).
43. Leiner, D. J. SoSci Survey. <https://www.soscisurvey.de> (2021).
44. Colloff, M. F., Wade, K. A. & Strange, D. Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychol. Sci.* **27**, 1227–1239. <https://doi.org/10.1177/0956797616655789> (2016).
45. Colloff, M. F., Wade, K. A., Wixted, J. T. & Maylor, E. A. A signal-detection analysis of eyewitness identification across the adult lifespan. *Psychol. Aging* **32**, 243–258. <https://doi.org/10.1037/pag0000168> (2017).
46. Karageorge, A. & Zajac, R. Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of wildcards. *Br. J. Psychol.* **102**, 161–183. <https://doi.org/10.1348/000712610X507902> (2011).
47. Meisters, J., Diedenhofen, B. & Musch, J. Eyewitness identification in simultaneous and sequential lineups: An investigation of position effects using receiver operating characteristics. *Memory* **26**, 1297–1309. <https://doi.org/10.1080/09658211.2018.1464581> (2018).
48. Smith, A. M. Why do mistaken identification rates increase when either witnessing or testing conditions get worse?. *J. Appl. Res. Mem. Cogn.* **9**, 495–507. <https://doi.org/10.1016/j.jarmac.2020.08.002> (2020).
49. Brewer, N., Caon, A., Todd, C. & Weber, N. Eyewitness identification accuracy and response latency. *Law Hum. Behav.* **30**, 31–50. <https://doi.org/10.1007/s10979-006-9002-7> (2006).
50. Brewer, N. & Wells, G. L. The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *J. Exp. Psychol. Appl.* **12**, 11–30. <https://doi.org/10.1037/1076-898X.12.1.11> (2006).
51. Lindsay, D. S., Read, J. D. & Sharma, K. Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychol. Sci.* **9**, 215–218. <https://doi.org/10.1111/1467-9280.00041> (1998).
52. Saraiva, R. B. *et al.* Eyewitness metamemory predicts identification performance in biased and unbiased line-ups. *Leg. Criminol. Psychol.* **25**, 111–132. <https://doi.org/10.1111/lcrp.12166> (2020).
53. Wilcock, R. & Bull, R. Novel lineup methods for improving the performance of older eyewitnesses. *Appl. Cogn. Psychol.* **24**, 718–736. <https://doi.org/10.1002/acp.1582> (2010).
54. Brewer, N., Keast, A. & Sauer, J. D. Children's eyewitness identification performance: Effects of a Not Sure response option and accuracy motivation. *Leg. Criminol. Psychol.* **15**, 261–277. <https://doi.org/10.1348/135532509X474822> (2010).
55. Brigham, J. C. Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. *Basic Appl. Soc. Psychol.* **11**, 105–115. https://doi.org/10.1207/s15324834basp1101_7 (1990).
56. Flowe, H. D. An exploration of visual behaviour in eyewitness identification tests. *Appl. Cogn. Psychol.* **25**, 244–254. <https://doi.org/10.1002/acp.1670> (2011).
57. Flowe, H. D. & Ebbesen, E. B. The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law Hum. Behav.* **31**, 33–52. <https://doi.org/10.1007/s10979-006-9045-9> (2007).
58. Sauerland, M. & Sporer, S. L. Fast and confident: Postdicting eyewitness identification accuracy in a field study. *J. Exp. Psychol. Appl.* **15**, 46–62. <https://doi.org/10.1037/a0014560> (2009).
59. Minear, M. & Park, D. C. A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* **36**, 630–633. <https://doi.org/10.3758/BF03206543> (2004).
60. German Federal Ministry of the Interior and Community. *Richtlinien für das Strafverfahren und das Bußgeldverfahren (RiStBV) [Guidelines for criminal proceedings and summary proceedings]*. https://www.verwaltungsvorschriften-im-internet.de/bsvwv_bund_01011977_420821R5902002.htm (2021).
61. Steblay, N. K., Dysart, J. E. & Wells, G. L. Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychol. Public Policy Law* **17**, 99–139. <https://doi.org/10.1037/a0021650> (2011).
62. Horry, R., Fitzgerald, R. J. & Mansour, J. K. “Only your first yes will count”: The impact of pre-lineup instructions on sequential lineup decisions. *J. Exp. Psychol. Appl.* **27**, 170–186. <https://doi.org/10.1037/xap0000337> (2021).
63. Carlson, C. A., Carlson, M. A., Weatherford, D. R., Tucker, A. & Bednarz, J. The effect of backloading instructions on eyewitness identification from simultaneous and sequential lineups. *Appl. Cogn. Psychol.* **30**, 1005–1013. <https://doi.org/10.1002/acp.3292> (2016).
64. Küppers, V. & Bayen, U. J. Inconsistency effects in source memory and compensatory schema-consistent guessing. *Q. J. Exp. Psychol.* **67**, 2042–2059. <https://doi.org/10.1080/17470218.2014.904914> (2014).
65. Meiser, T., Sattler, C. & Von Hecker, U. Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *Q. J. Exp. Psychol.* **60**, 1015–1040. <https://doi.org/10.1080/17470210600875215> (2007).
66. Hirshman, E. Decision process in recognition memory: Criterion shifts and the list-strength paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 302–313. <https://doi.org/10.1037/0278-7393.21.2.302> (1995).
67. Riefer, D. M., Hu, X. & Batchelder, W. H. Response strategies in source monitoring. *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 680–693. <https://doi.org/10.1037/0278-7393.20.3.680> (1994).
68. Zarkadi, T., Wade, K. A. & Stewart, N. Creating fair lineups for suspects with distinctive features. *Psychol. Sci.* **20**, 1448–1453. <https://doi.org/10.1111/j.1467-9280.2009.02463.x> (2009).
69. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191. <https://doi.org/10.3758/BF03193146> (2007).

70. Buchner, A., Erdfelder, E. & Vaterrodt-Plünnecke, B. Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *J. Exp. Psychol. Gen.* **124**, 137–160. <https://doi.org/10.1037/0096-3445.124.2.137> (1995).
71. Clark, S. E. A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law Hum. Behav.* **29**, 575–604. <https://doi.org/10.1007/s10979-005-7121-1> (2005).
72. Keast, A., Brewer, N. & Wells, G. L. Children's metacognitive judgments in an eyewitness identification task. *J. Exp. Child Psychol.* **97**, 286–314. <https://doi.org/10.1016/j.jecp.2007.01.007> (2007).
73. Lindsay, R. C. L. *et al.* Biased lineups: Sequential presentation reduces the problem. *J. Appl. Psychol.* **76**, 796–802. <https://doi.org/10.1037/0021-9010.76.6.796> (1991).
74. Bornstein, B. H., Deffenbacher, K. A., Penrod, S. D. & McGorty, E. K. Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychol. Crime Law* **18**, 473–490. <https://doi.org/10.1080/1068316x.2010.508458> (2012).
75. Flowe, H. D. & Humphries, J. E. An examination of criminal face bias in a random sample of police lineups. *Appl. Cogn. Psychol.* **25**, 265–273. <https://doi.org/10.1002/acp.1673> (2011).
76. Lampinen, J. M. *et al.* Comparing detailed and less detailed pre-lineup instructions. *Appl. Cogn. Psychol.* **34**, 409–424. <https://doi.org/10.1002/acp.3627> (2020).
77. Malpass, R. S. & Devine, P. G. Eyewitness identification: Lineup constructions and the absence of the offender. *J. Appl. Psychol.* **66**, 482–489. <https://doi.org/10.1037/0021-9010.66.4.482> (1981).
78. Murnane, K. & Shiffrin, R. M. Interference and the representation of events in memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **17**, 855–874. <https://doi.org/10.1037/0278-7393.17.5.855> (1991).
79. Ratcliff, R., Sheu, C.-F. & Gronlund, S. D. Testing global memory models using ROC curves. *Psychol. Rev.* **99**, 518–535. <https://doi.org/10.1037/0033-295X.99.3.518> (1992).
80. Shapiro, P. N. & Penrod, S. D. Meta-analysis of facial identification studies. *Psychol. Bull.* **100**, 139–156. <https://doi.org/10.1037/0033-2909.100.2.139> (1986).
81. Wetmore, S. A. *et al.* Effect of retention interval on showup and lineup performance. *J. Appl. Res. Mem. Cogn.* **4**, 8–14. <https://doi.org/10.1016/j.jarmac.2014.07.003> (2015).
82. Doob, A. N. & Kirshenbaum, H. M. Bias in police lineups—partial remembering. *J. Police Sci. Admin.* **1**, 287–293 (1973).
83. Clark, S. E. Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspect. Psychol. Sci.* **7**, 238–259. <https://doi.org/10.1177/1745691612439584> (2012).
84. Sučić, I., Tokić, D. & Ivešić, M. Field study of response accuracy and decision confidence with regard to lineup composition and lineup presentation. *Psychol. Crime Law* **21**, 798–819. <https://doi.org/10.1080/1068316X.2015.1054383> (2015).
85. Dobolyi, D. G. & Dodson, C. S. Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *J. Exp. Psychol. Appl.* **19**, 345–357. <https://doi.org/10.1037/a0034596> (2013).
86. Neuschatz, J. S. *et al.* In *Advances in Psychology and Law* Vol 5 Chapter 2 (eds Brian, H. B. & Monica, K. M.) 43–69 (Springer, 2016).
87. Seale-Carlisle, T. M. & Mickes, L. US line-ups outperform UK line-ups. *R. Soc. Open Sci.* **3**, 2–12. <https://doi.org/10.1098/rsos.160300> (2016).
88. Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D. & Mickes, L. Designing police lineups to maximize memory performance. *J. Exp. Psychol. Appl.* **25**, 410–430. <https://doi.org/10.1037/xap0000222> (2019).
89. Bröder, A. & Schütz, J. Recognition ROCs are curvilinear-or are they? On premature arguments against the two-high-threshold model of recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 587–606. <https://doi.org/10.1037/a0015279> (2009).
90. Heck, D. W., Arnold, N. R. & Arnold, D. TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behav. Res. Methods* **50**, 264–284. <https://doi.org/10.3758/s13428-017-0869-7> (2018).
91. Schaper, M. L., Mieth, L. & Bell, R. Adaptive memory: Source memory is positively associated with adaptive social decision making. *Cognition* **186**, 7–14. <https://doi.org/10.1016/j.cognition.2019.01.014> (2019).
92. Tupper, N., Sauerland, M., Sauer, J. D. & Hope, L. Eyewitness identification procedures for multiple perpetrator crimes: A survey of police in Sweden, Belgium, and the Netherlands. *Psychol. Crime Law* **25**, 992–1007. <https://doi.org/10.1080/1068316x.2019.1611828> (2019).
93. Erdfelder, E. & Buchner, A. Process-dissociation measurement models: Threshold theory or signal-detection theory?. *J. Exp. Psychol. Gen.* **127**, 83–96. <https://doi.org/10.1037/0096-3445.127.1.83> (1998).
94. Klauer, K. C. & Kellen, D. Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychon. Bull. Rev.* **17**, 465–478. <https://doi.org/10.3758/PBR.17.4.465> (2010).
95. Schütz, J. & Bröder, A. Signal detection and threshold models of source memory. *Exp. Psychol.* **58**, 293–311. <https://doi.org/10.1027/1618-3169/a000097> (2011).
96. Kellen, D. & Klauer, K. C. Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychol. Rev.* **122**, 542–557. <https://doi.org/10.1037/a0039251> (2015).
97. Kellen, D., Singmann, H., Vogt, J. & Klauer, K. C. Further evidence for discrete-state mediation in recognition memory. *Exp. Psychol.* **62**, 40–53. <https://doi.org/10.1027/1618-3169/a000272> (2015).
98. Province, J. M. & Roudier, J. N. Evidence for discrete-state processing in recognition memory. *Proc. Natl. Acad. Sci.* **109**, 14357–14362. <https://doi.org/10.1073/pnas.1103880109> (2012).
99. Wixted, J. T. Dual-process theory and signal-detection theory of recognition memory. *Psychol. Rev.* **114**, 152–176. <https://doi.org/10.1037/0033-295X.114.1.152> (2007).
100. Yonelinas, A. P. & Parks, C. M. Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychol. Bull.* **133**, 800–832. <https://doi.org/10.1037/0033-2909.133.5.800> (2007).
101. Malmberg, K. J. On the form of ROCs constructed from confidence ratings. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 380–387. <https://doi.org/10.1037/0278-7393.28.2.380> (2002).

Acknowledgements

We want to thank Tobias Koch for his technical assistance in presenting the videos as part of the online data collection.

Author contributions

K.W., N.M.M., R.B. and A.B. contributed to the study conception, design, material preparation and data analysis. K.W. collected the data and wrote the first draft of the manuscript with subsequent input from all co-authors who gave final approval for publication.

Funding

Open Access funding enabled and organized by Projekt DEAL. The work reported herein was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-BU 945/10-1, Project Number 456214986.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19513-w>.

Correspondence and requests for materials should be addressed to K.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022



OPEN

Evaluating the impact of first-yes-counts instructions on eyewitness performance using the two-high threshold eyewitness identification model

Kristina Winter , Nicola Marie Menne , Raoul Bell  & Axel Buchner 

In eyewitness research, multiple identification decisions in sequential lineups are typically prevented by telling participants that only their first identification decision counts. These first-yes-counts instructions are incompatible with standard police protocols prescribing that witnesses shall see the entire lineup. Horry et al. were the first to experimentally test how this discrepancy between eyewitness research and standard police protocols affects eyewitness identification decisions. Here, the two-high threshold eyewitness identification model was used to disentangle the effect of the first-yes-counts instructions on the detection and guessing processes underlying eyewitness identification decisions. We report both a reanalysis of Horry et al.'s data and a conceptual replication. Both the reanalysis and the results of the conceptual replication confirm that first-yes-counts instructions do not affect the detection of the culprit but decrease the probability of guessing-based selections. To improve the ecological validity, research on sequential lineups should avoid first-yes-counts instructions.

Following science-based recommendations (e.g.,¹), many jurisdictions have replaced simultaneous lineups in which eyewitnesses see photographs of a single suspect and a finite number of fillers at once with sequential lineups (e.g.,^{2–6}). In a typical sequential lineup, photographs of one suspect and a finite number of fillers are shown in a random sequence and an eyewitness makes an identification decision or a non-identification decision for each photograph separately. If an eyewitness makes only non-identification decisions, the lineup is rejected. At the other extreme, the sequential presentation implies, at least in principle, that an eyewitness can make as many identification decisions as there are photographs in the lineup. In eyewitness research, multiple identification decisions in sequential lineups are typically prevented by first-yes-counts instructions^{7,8}. In these instructions, participants are informed that only their first identification decision counts. Sometimes, the lineup presentation continues after a first identification decision has been made and participants are shown the complete lineup (e.g.,^{9–17}). In other studies, lineup presentation is terminated immediately after the first identification decision (e.g.,^{18–27}). In sharp contrast, standard police protocols do not use first-yes-counts instructions^{2–5,8,28–31}. Instead, lineup presentation continues after identification decisions and eyewitnesses can select further photographs if these provide a better match to the memory of the culprit. Given this discrepancy between research and police practice, it is important to understand the effects of first-yes-counts instructions on the processes involved in eyewitness identification decisions.

Horry et al.³² were the first to evaluate this discrepancy experimentally. Among other things, they compared a sequential-lineup condition in which participants were told that only their first yes-response would count as an identification decision to a sequential-lineup condition without first-yes-counts instructions. The suspect identification rate was numerically but not significantly lower (0.26 compared to 0.29) and the lineup rejection rate was significantly higher (0.53 compared to 0.43) in the condition with first-yes-counts instructions compared to the condition without such instructions. This result was interpreted to indicate that first-yes-counts instructions increase participants' response criterion. Such an interpretation seems plausible given that participants in

Department of Experimental Psychology, Heinrich Heine University Düsseldorf, 40204 Düsseldorf, Germany. ✉email: kristina.winter@hhu.de

the condition with first-yes-counts instructions knew that they had only one chance to make an identification decision. These participants may have been particularly hesitant to select a lineup member, not knowing whether another lineup member would follow which might be an even better match to their memory of the culprit and to which they would then no longer be able to make an identification decision. Whereas two response rates were used by Horry et al.³² as a basis for drawing conclusions about participants' response criterion, receiver operating characteristics were used as a measurement model to assess whether the instructions affected participants' ability to discriminate the culprit from the innocent suspect. In terms of the partial area under the curve, there was a numerical, but not statistically significant, difference between the sequential-lineup condition with first-yes-counts instructions (partial area under the curve = 0.014) and the sequential-lineup condition without such instructions (partial area under the curve = 0.017). Based on these two types of analyses, it is possible to draw the conclusions that first-yes-counts instructions induce a conservative response bias while leaving the overall ability to discriminate the culprit from the innocent suspect unaffected.

If these conclusions are valid, then they have important implications in that they suggest that a sizeable number of previous eyewitness identification studies may lack ecological validity. Specifically, if first-yes-counts instructions affect the processes underlying eyewitness identification decisions, inferences based on data from laboratory studies using first-yes-counts instructions may not apply to lineups that follow standard police protocols. For instance, sequential lineups have been associated with particularly conservative responding⁷ but this finding may only be an artifact of first-yes-counts instructions. However, a necessary condition for the validity of these conclusions is that the findings on which they are based are reliable³³. Therefore, an independent replication of Horry et al.'s³² sequential-lineup conditions is needed before drawing such far-reaching conclusions.

Independent of the question of reliability, there are at least three threats to the validity of Horry et al.'s³² conclusions. First, Horry et al.³² used two separate types of response rates—the suspect identification rate and the lineup rejection rate—which they implicitly combined in their conclusion about the effect of first-yes-counts instructions into a two-component response-bias measure, albeit a measure that is not readily quantifiable. At this point it is important to realize that there is no such thing as a measure without a measurement model. This means that by implicitly combining the suspect identification rate and the lineup rejection rate into a (not readily quantifiable) two-component response-bias measure, an ad-hoc measurement model was applied. In a verbal approximation, this measurement model might look something like this: A decreasing suspect identification rate and an increasing lineup rejection rate may be assumed to indicate a decreasing willingness to make an identification decision which may be seen as an index of an increasingly conservative response bias. This measurement model may seem intuitively plausible and thus may have some face validity. Nevertheless, it is preferable to rely on a measurement model that is not only intuitively plausible but one that has also been validated experimentally, that is, for which it has been demonstrated experimentally that it really measures the processes it is intended to measure. To anticipate, here we will use such a validated measurement model. Second, the effect of first-yes-counts instructions on the implicit, not readily quantifiable response-criterion measure was assessed by performing two separate statistical tests of which one showed no significant difference between the instruction conditions whereas the other did. Specifically, the suspect identification rate was only numerically but not significantly lower in the condition with first-yes-counts instructions compared to the condition without such instructions while the lineup rejection rate was significantly higher in the condition with first-yes-counts instructions compared to the condition without such instructions. The fact that only one of the two tests was statistically significant creates some ambiguity as to the difference in response bias between the instruction conditions: According to the first test with a nonsignificant result one has to conclude that there is no difference in response bias between the instruction conditions whereas according to the second test with a statistically significant result one has to conclude that such a difference in response bias between test conditions exists. To arrive at unambiguous conclusions, it is better to have a single measure that can be compared between conditions using a single statistical test. To anticipate, the measurement model we use here provides such a measure. Third, a completely different measurement model based on receiver operating characteristics was used for assessing the effect of the first-yes-counts instructions on the ability to discriminate the culprit from the innocent suspect. Instead of relying on two entirely different measurement models with different sets of underlying assumptions, it seems desirable to use one single measurement model based on one coherent set of assumptions for measuring various types of processes underlying eyewitness identification decisions. Ideally it should also be possible to demonstrate that this measurement model fits the data before measures are interpreted.

The two-high threshold (2-HT) eyewitness identification model is such a measurement model. The model has been experimentally validated: Multiple experiments have shown that the model's measures indeed reflect the processes they were designed to measure^{34,35}. The 2-HT eyewitness identification model belongs to the class of multinomial processing tree models^{36–38}. These models are used to provide estimates of measures representing the probabilities of latent processes from categorical data. Statistical hypotheses are tested directly at the level of these parameters. Based on the entire 2×3 data structure of eyewitness identification decisions (i.e., correct culprit identifications, false innocent-suspect identifications, false filler identifications in culprit-present and culprit-absent lineups, correct and false lineup rejections), the 2-HT eyewitness identification model provides measures of the detection of the presence of the culprit (dP) and of the detection of the absence of the culprit in the lineup (dA), a measure of the biased selection of the suspect in unfair lineups (b) and a measure of selections based on guessing processes (g).

Figure 1 shows a graphical illustration of the model. The upper tree represents the processes leading to culprit and filler identifications as well as lineup rejections in culprit-present lineups. The eyewitness detects the presence of the culprit with probability dP . If culprit detection fails, which occurs with probability $1 - dP$, then the eyewitness may identify the culprit based on two types of non-detection-based processes. If the lineup is unfair in that the culprit stands out from the fillers so that it can be inferred who the culprit is without relying on memory, then biased selection of the culprit occurs with probability b . With probability $1 - b$, no biased selection of the

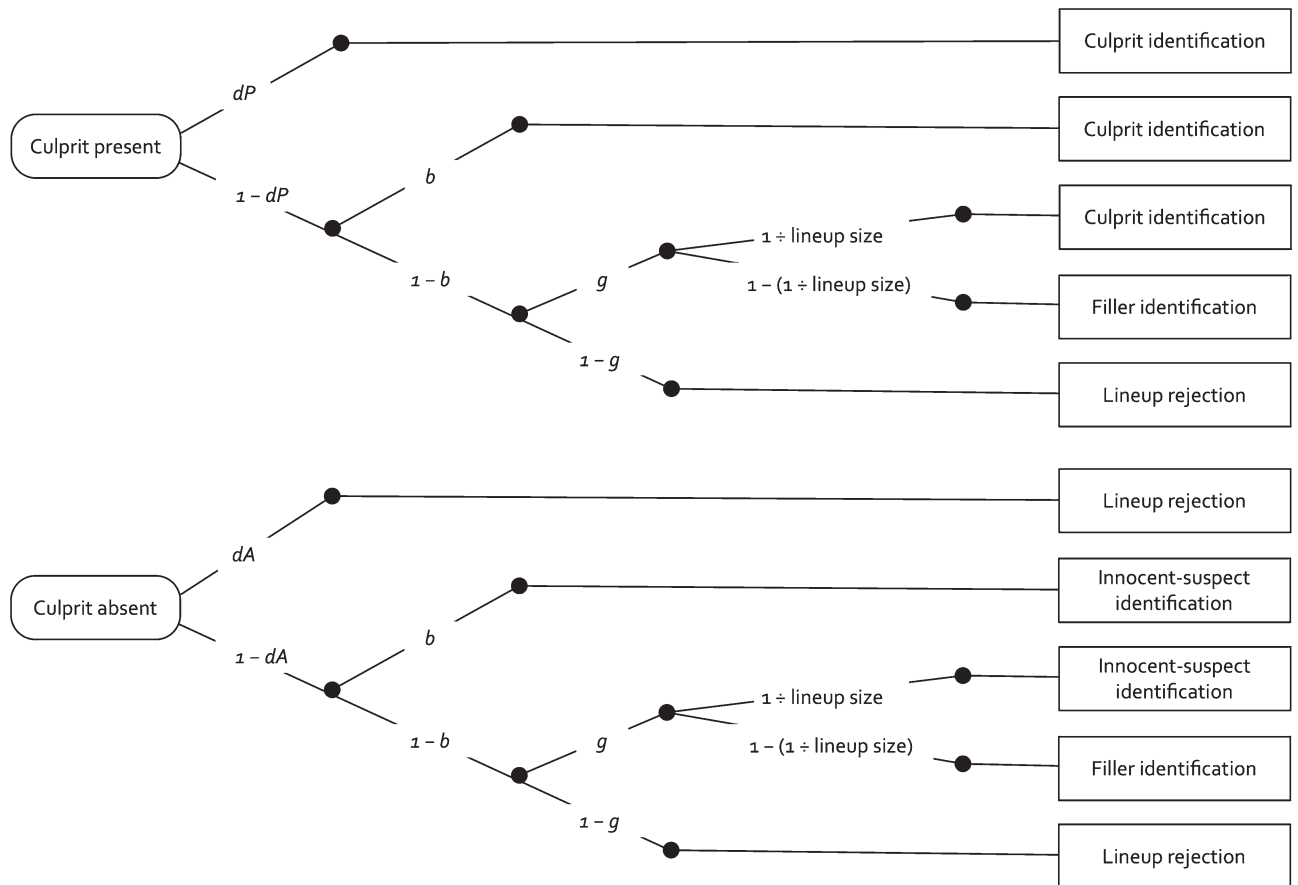


Figure 1. Graphical illustration of the 2-HT eyewitness identification model. The rounded rectangles on the left represent lineups in which the culprit is either present or absent. The rectangles on the right represent the possible responses of an eyewitness in a culprit-present or a culprit-absent lineup. The letters attached to the branches represent the probabilities of the latent cognitive processes postulated by the model that lead to the responses of the eyewitness (dP : detection of the presence of the culprit; b : biased selection of a suspect who stands out from the fillers; g : guessing-based selection among the lineup members; dA : detection of the absence of the culprit). The *lineup size* is a known constant: the number of persons in the lineup.

culprit occurs (e.g., if the lineup is fair or the eyewitness does not attend to the features creating the unfairness or chooses to ignore these features). In this case, it is still possible to select one of the lineup members as the culprit based on guessing with probability g . The culprit is then selected among the fillers with probability $1 \div \text{lineup size}$ (e.g., $1 \div 6$ in a six-person lineup). With probability $1 - (1 \div \text{lineup size})$, one of the fillers is selected (e.g., $5 \div 6$ in a six-person lineup). The probability with which guessing leads to a culprit identification thus depends only on the lineup size and not on any cognitive process which is why this element of the model is a known constant and not a parameter that has to be estimated from data. In the present context it is important to realize that this constant is the same irrespective of whether first-yes-counts instructions are used, provided a proper randomization procedure is in place. Specifically, if an eyewitness selects one person out of N persons in the lineup based on guessing, then the sampling probability that the selected person, out of the N persons in the lineup, is the culprit is always $1 \div N$, irrespective of the presence or absence of first-yes-counts instructions. If none of the lineup members is selected based on guessing, which occurs with probability $1 - g$, then the lineup is falsely rejected.

The lower tree of the model represents the processes that occur in culprit-absent lineups. The absence of the culprit and the fact that no one else in the lineup can possibly be the culprit is detected with probability dA , leading to the correct rejection of the lineup. With probability $1 - dA$, the absence of the culprit is not detected, in which case the same non-detection-based processes occur with the same probabilities as in culprit-present lineups, the difference being that guessing and biased selection in culprit-absent lineups may lead to the selection of the innocent suspect rather than to the selection of the culprit (see³⁴ and³⁵ for more details).

Here we focus on the culprit-presence-detection parameter dP which is conceptually similar, but not identical, to the discrimination between culprits and innocent suspects that is hypothesized to be reflected in the partial area under the curve in analyses based on receiver operating characteristics and on the guessing-based-selection parameter g which provides a single, validated and unambiguous measure of the selection of lineup members based on guessing.

In a first step, the 2-HT eyewitness identification model was used to reanalyze the sequential-lineup data reported by Horry et al.³² to assess the effects of the first-yes-counts instructions on culprit-presence detection

and guessing-based selection within a single coherent and experimentally validated measurement model. In addition to this reanalysis, an independent conceptual replication of the sequential-lineup conditions of Horry et al.³² was conducted, the data of which were also analyzed using the 2-HT eyewitness identification model. The central conclusion drawn from the analyses of Horry et al.³² was that the first-yes-counts instructions left detection-based processes unaffected but decreased the participants' willingness to make identification decisions. If this conclusion is valid, then the first-yes-counts instructions should not affect the probability with which the presence of the culprit is detected whereas the probability of guessing-based selection should be lower in the condition with first-yes-counts instructions than in the condition without such instructions. In other words, the culprit-presence-detection parameter dP of the 2-HT eyewitness identification model should not differ between the condition with first-yes-counts instructions and the condition without such instructions whereas the guessing-based-selection parameter g of the 2-HT eyewitness identification model should be significantly lower in the condition with first-yes-counts instructions than in the condition without such instructions.

Reanalysis of the sequential-lineup data reported by Horry et al.³²

Horry et al.³² randomly assigned 896 participants to the conditions of a 3 (lineup type: sequential lineup with first-yes-counts instructions vs. sequential lineup without such instructions vs. simultaneous lineup) \times 2 (culprit presence: culprit present vs. culprit absent) \times 2 (suspect position in the lineup: 2 vs. 5) between-subjects design. Here we focus on the effect of the first-yes-counts instructions. Therefore, we reanalyzed the identification decisions of the 559 participants in the two sequential-lineup conditions (with first-yes-counts instructions and without such instructions), aggregated across suspect position.

Method. Participants watched one out of four possible mock-crime videos with the same content but a different culprit. Lineups were composed of one suspect and five fillers who matched the description of the suspect. In culprit-present lineups, the suspect was the culprit from the one video that had been seen. In culprit-absent lineups, the suspect was a culprit from one of the other videos that had not been seen. Participants were asked to identify a culprit in one lineup but they knew that the culprit did not have to be in the lineup. The instructions were identical for all participants except that the participants in the condition with first-yes-counts instructions received the following additional information: "If you respond 'yes' to a photo, you will not be able to change that decision, and you will not be able to respond 'yes' to any later photos"³². Independent of the instructions, participants were able to make multiple identification decisions in both sequential-lineup conditions and if they did, they were presented with a second lap of the lineup to clarify their decision. In the condition with first-yes-counts instructions, each participant's first identification decision in the initial lineup was used for the analyses. In the condition without such instructions, this was different. If participants made multiple identification decisions in the initial lineup, the unique identification decision given in the second lap was used for the analyses. For more details, see Horry et al.³².

Results. For all analyses reported in this article, multiTree³⁹ was used to obtain parameter estimates and to perform likelihood-ratio goodness-of-fit tests for which the α level was set to 0.05. The response frequencies underlying the analyses reported here are shown in Table 1. These response frequencies were taken from the penultimate and the last lines of Horry et al.'s³² Table 3 in which the data are already aggregated across suspect position.

We needed two instances of the model depicted in Fig. 1, one for the condition with first-yes-counts instructions and one for the condition without such instructions. Given that the lineups comprised six persons, the sampling probability with which guessing-based selections lead to identifications of the suspect, given by $1 \div$ lineup size, was set to 0.16667 as an approximation to $1 \div 6$. We wanted to begin with a base model that was as simple as possible. Therefore, we used whatever we could derive from the design of the study to impose restrictions onto the 2-HT eyewitness identification model. First, given that the same, well-constructed lineups were used in both conditions (which were essentially fair, see below), we set the biased-suspect-selection parameter b to be equal across the two instruction conditions. Second, for the same reason the culprit-absence-detection parameter dA was also set to be equal across the two conditions. The base model incorporating these two restrictions fit the data, $G^2(2) = 2.42$, $p = 0.298$. The estimates of parameters b and dA were 0.00 ($SE = 0.02$) and 0.13 ($SE = 0.12$), respectively. The estimates of the culprit-presence-detection parameter dP and of the guessing-based-selection parameter g are displayed in Fig. 2.

In multinomial processing-tree models, hypotheses can be tested directly at the level of the postulated processes. For instance, the hypothesis that the culprit-presence-detection parameter dP does not differ between the condition with first-yes-counts instructions and the condition without such instructions can be implemented by setting parameter dP to be equal between these conditions. If the model including this equality restriction

	Culprit-present lineups			Culprit-absent lineups		
	Culprit identifications	Filler identifications	Lineup rejections	Innocent-suspect identifications	Filler identifications	Lineup rejections
First-yes-counts instructions	71	26	61	7	38	102
No first-yes-counts instructions	69	23	30	5	48	79

Table 1. Response frequencies from Table 3 of Horry et al.³².

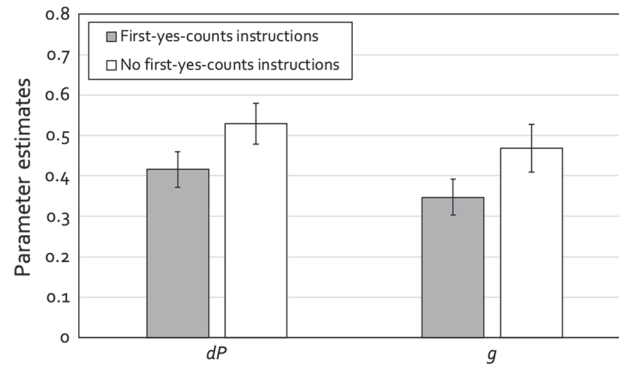


Figure 2. Estimates of parameters dP (detection of the presence of the culprit) and g (guessing-based selection) of the 2-HT eyewitness identification model as a function of lineup instructions (first-yes-counts instructions vs. no first-yes-counts instructions) for the sequential-lineup data of Horry et al.³². The error bars represent the standard errors.

provides a significantly worse fit to the data than the base model, then it is necessary to conclude that parameter dP differs between conditions. Parameter dP did not differ significantly between the condition with first-yes-counts instructions and the condition without such instructions, $\Delta G^2(1) = 3.05$, $p = 0.081$. In contrast, parameter g was significantly lower in the condition with first-yes-counts instructions than in the condition without first-yes-counts instructions, $\Delta G^2(1) = 5.31$, $p = 0.021$.

Discussion. The results of this reanalysis are compatible with the conclusions reached by Horry et al.³². The presence or absence of first-yes-counts instructions did not affect the probability with which the culprit was detected whereas it affected the probability of guessing-based selection in the expected direction. Specifically, the culprit-presence-detection parameter dP of the 2-HT eyewitness identification model did not differ significantly between the condition with first-yes-counts instructions and the condition without such instructions whereas the guessing-based-selection parameter g of the 2-HT eyewitness identification model was significantly lower in the condition with first-yes-counts instructions than in the condition without such instructions.

However, at a descriptive level the culprit-presence-detection parameter dP was lower in the condition with first-yes-counts instructions than in the condition without such instructions. This observation may fuel the suspicion that there might in fact be a difference in dP which could turn out to be statistically significant if the relevant statistical test were even more sensitive. This was one reason as to why we aimed at replicating the sequential-lineup conditions of Horry et al.³² and to test, based on even more datapoints and thus with even more sensitivity, whether first-yes-counts instructions have indeed no effect on culprit-presence detection while decreasing the probability of guessing-based selections.

Conceptual replication of the sequential-lineup conditions of Horry et al.³²

The reliability of findings is important³³ because conclusions can only be valid if they are based on reliable data. What is more, an independent replication of critical findings is regarded as the gold standard in assessing the reliability of findings and as a necessary step in science to increase the trust in new findings⁴⁰. We therefore tested the reliability of the findings of Horry et al.³² in an independent conceptual replication of their sequential-lineup conditions. We chose to perform a conceptual replication because, if successful, a test of the replicability of findings is even more powerful if one can show that the core finding does not depend on details of the experimental implementation. Our implementation differed in the following aspects from that used in the study of Horry et al.³². First, in the condition with first-yes-counts instructions, only one identification was technically possible. Second, participants saw each lineup only once. Third, if multiple identifications occurred in the condition without first-yes-counts instructions, we considered the last decision as a revision of any previous decisions; therefore, the last identification was used in the analysis. Fourth, we used a multiple-culprit mock crime; every participant was presented with four different lineups, two of which were culprit-present lineups and two were culprit-absent lineups. Finally, the positions of all lineup members, including the culprit and the innocent suspect, were randomly selected. Given these differences between the two studies, a replication of the results of Horry et al.³² would indicate that their results are reliable and robust in the sense that they do not depend on the details of the experimental implementation. In this case, the culprit-presence-detection parameter dP of the 2-HT eyewitness identification model should not differ between the condition with first-yes-counts instructions and the condition without such instructions whereas the guessing-based-selection parameter g should be significantly lower in the condition with first-yes-counts instructions than in the condition without such instructions.

Method. *Sample.* Participants were recruited using the Cint research panel (<https://www.cint.com>). Criteria for participation were legal age (≥ 18 years), good eyesight and German language skills. Given that the design of the present experiment comprised two groups, we aimed at recruiting about half as many participants as in the four-group experiments of Winter et al.³⁵ in which about 750 participants were recruited. We thus aimed at collecting about 375 data sets and terminated the data collection at the end of the day at which this criterion

was surpassed. Of the 421 participants who gave informed consent, 52 did not complete the study. Another 14 participants were excluded because they had failed to respond correctly to the attention check question, because they had indicated technical problems or because they had participated more than once which implies that they had watched the mock-crime video more than once. The final sample comprised 355 participants (142 women, 210 men, 3 diverse) with a mean age of 46 years ($SD=15$). A sensitivity analysis using G*Power⁴¹ showed that given $\alpha = \beta = 0.05$, a sample size of $N=355$ and four identification decisions per participant, it was possible to detect an effect of the first-yes-counts instructions of size $w=0.10$ on the model parameters.

Ethics statement. All participants gave informed consent prior to beginning the experiment. They were informed that during the experiment they would see a video in which a group of persons would attack another person verbally and physically. Participants were asked not to participate if they felt uncomfortable when anticipating to watch such a video. Ethical approval had been obtained from the ethics committee of the Faculty of Mathematics and Natural Sciences of Heinrich Heine University Düsseldorf for a series of experiments to which the current experiment belongs. The experiment was run in accordance with the declaration of Helsinki.

Materials and procedure. Materials and procedure were essentially the same as those of Winter et al.³⁵ with the exception of the manipulation of the presence or absence of the first-yes-counts instructions that are described in detail below. The experiment was run online using SoSci Survey⁴². It was possible to participate with a computer and a laptop but not with a mobile device such as a smartphone. Participants were asked to complete the study alone and in a quiet environment and to put their browser into full-screen mode prior to starting the experiment.

Mock-crime videos. Participants watched one of two mock-crime videos. In each of the two videos, four alleged hooligans of the German soccer club FC Bayern München—henceforth referred to as the culprits—attacked an alleged fan of a rivaling soccer club Borussia Dortmund—henceforth referred to as the victim—at a bus station. The culprits and the victim wore typical fan clothing (shirts, scarfs and caps) of their respective soccer clubs. The culprits insulted the victim, poked fun at him and tossed his belongings around. At the end of the video, the victim got knocked to the ground. The culprits continued to physically abuse the victim before they apparently noticed another person approaching (not visible in the video) and ran away shouting loudly. The two videos (henceforth Video 1 and Video 2) showed the same events (i.e., the same verbal abuse and the same acts of violence in the same sequence and with the same timing), but the victims and the culprits were played by different actors. However, the victim in Video 1 matched the victim in Video 2 in terms of age, hair color, hair style and stature, as determined by the present authors. The same was true for the four culprits, that is, Hooligan A in Video 1 matched Hooligan A in Video 2, Hooligan B in Video 1 matched Hooligan B in Video 2 and so on. It was randomly determined which of the two parallel versions of the video was shown to a particular participant. The videos were about 130 s long and shown in a resolution of 885×500 pixels.

Participants started the video by clicking on a “Start” button. Participants could not stop, fast forward or replay the video. Once the video had ended, participants were asked to answer an attention check question probing for the type of persons seen in the video. The correct response was to select “soccer fans” from a list of ten alternatives.

Lineup procedures. Participants were informed that they had to identify the FC Bayern München hooligans from the video they had just seen. Participants received two-sided lineup instructions that emphasized both the need to identify the culprit if the culprit was present and the need to reject the lineup if the culprit was absent. Participants were not informed about how many lineups were about to follow. Each lineup consisted of the facial photographs of six persons, one culprit or innocent suspect and five fillers. Participants saw two culprit-present lineups and two culprit-absent lineups.

To manipulate the suspect’s guilt, the crossed-lineup procedure³⁵ was used. In each of the two culprit-present lineups, a randomly selected face of one of the hooligans of the video the participants had seen was presented among the fillers. In each of the two culprit-absent lineups, the face of an innocent suspect was presented that the participants had not witnessed committing a crime. The innocent suspect was one of the hooligans from the video that the participants had not seen. For instance, if participants had seen Video 1, two randomly selected hooligans (e.g., Hooligan B and Hooligan C) from Video 1 served as the culprits in the two culprit-present lineups, while two of the hooligans from Video 2 (Hooligan A and Hooligan D in this example) served as the innocent suspects in the culprit-absent lineups. The same set of fillers was used in culprit-present and culprit-absent lineups. For each of the four lineups, five photographs of male adults aged between 18 and 29 years were selected from the Center for Vital Longevity Face Database of Minear and Park⁴³ to serve as fillers. The fillers resembled the culprit and the matching innocent suspect in age, hair color, hair style and stature. Together with the fact that it was randomly determined whether participants saw Video 1 or Video 2, the crossed-lineup procedure ensures that the culprits and innocent suspects differ to the same degree, on average, from the fillers in the lineup. The crossed-lineup procedure is ecologically valid because the situation it creates corresponds closely to that of a real-world lineup in which the photograph of the suspect—whose status as culprit or innocent suspect is unknown to the police—stems from a different source (e.g., social media) than the photographs of the fillers (e.g., a database). Therefore, the photographs of the fillers may differ to some degree from the photograph of the suspect in terms of variables such as tint, resolution or softness, particularly in cases in which the photographs of the fillers are digitally manipulated^{31,44–46}. All photographs showed the faces from a frontal view with a neutral facial expression against a black background with no clothes visible. The photographs were edited to harmonize

face sizes and lighting conditions and were presented in a resolution of 142×214 pixels. The order of the lineups was randomized, as was the position of the culprit or innocent suspect and the fillers in each lineup.

Participants were randomly assigned either to the condition with first-yes-counts instructions or to the condition without such instructions. In the condition with first-yes-counts instructions, each lineup was preceded by a screen announcing the next lineup as the “1st lineup”, “2nd lineup” and so on and the following statement (the following quotations are translations of text originally presented in German): “If you choose ‘Yes, was present’ for one face, this decision is irrevocable. You will not be able to select ‘Yes was present’ for any other face in the lineup”. A click on the “I am ready” button started the successive presentation of the photographs. Participants decided for each photograph whether it showed one of the culprits or not by clicking on the “Yes, was present” button or the “No, this person was not present” button, respectively. Either response initiated the presentation of the next photograph. Immediately after participants’ first and only identification decision in a particular lineup, the ‘Yes, was present’ button was disabled and its label changed to ‘A yes response is no longer possible’ for all remaining photographs of the lineup for which participants were only able to select the ‘No, this person was not present’ button.

In the condition with no first-yes-counts instructions, each lineup was preceded by a screen announcing the next lineup as the “1st lineup”, “2nd lineup” and so on, without the first-yes-counts statement. A click on the “I am ready” button started the successive presentation of the photographs. Within each lineup, participants decided for each photograph whether it showed one of the culprits or not by clicking on the “Yes, was present” button or the “No, this person was not present” button, respectively. Either response initiated the presentation of the next photograph. In case of more than one identification decision in a lineup, the last decision was used in the analysis.

A lineup was counted as rejected if none of the members of the lineup was identified as one of the culprits. To make the procedure similar to the procedure used by Horry et al.³² and to that of a typical police lineup, participants also indicated how confident they were that their decision was correct. After every lineup, participants had to click on a “Continue” button to start the next lineup or, after the final lineup, to reach a screen on which questions were asked about possible technical difficulties during the online experiment. At the end of the experiment, all participants were informed that the crime shown in the video had been staged and they were informed about the purpose of the experiment.

Results. The response frequencies underlying the analysis reported here are shown in Table 2. We needed two instances of the model depicted in Fig. 1, one for the condition with first-yes-counts instructions and one for the condition without such instructions. Given that the lineups comprised six persons, $1 \div$ lineup size was set to 0.16667 as an approximation to $1 \div 6$. As before, we wanted to begin with a base model that was as simple as possible. The design of the present experiment makes it possible to derive the same restrictions for the 2-HT eyewitness identification model that we derived from the design of the experiment of Horry et al.³² in the reanalysis presented above. The biased-suspect-selection parameter b and the culprit-absence-detection parameter dA were each set to be equal across the two conditions. The base model incorporating these restrictions fit the data, $G^2(2) = 3.09$, $p = 0.214$. The estimates of parameters b and dA were 0.01 ($SE = 0.01$) and 0.06 ($SE = 0.04$), respectively. The estimates of the culprit-presence-detection parameter dP and the guessing-based-selection parameter g are displayed in Fig. 3.

Parameter dP did not differ significantly between the condition with first-yes-counts instructions and the condition without such instructions, $\Delta G^2(1) = 0.01$, $p = 0.936$. In contrast, parameter g was significantly lower in the condition with first-yes-counts instructions than in the condition without such instructions, $\Delta G^2(1) = 21.16$, $p < 0.001$.

Discussion. The results of this conceptual replication of the sequential-lineup conditions of Horry et al.³² nicely fit with the results of the original study. The presence or absence of first-yes-counts instructions did not affect the probability with which the culprit was detected whereas guessing-based selection was significantly less likely to occur in the condition with first-yes-counts instructions than in the condition without such instructions. This confirms the conclusions drawn from the reanalysis of the data reported by Horry et al.³². The fact that there was again no difference between conditions in parameter dP may be seen as particularly remarkable given that the relevant statistical test here was based on more data points (cf. Tables 1 and 2) and thus more sensitive than the corresponding test in the reanalysis of the data reported by Horry et al.³².

	Culprit-present lineups			Culprit-absent lineups		
	Culprit identifications	Filler identifications	Lineup rejections	Innocent-suspect identifications	Filler identifications	Lineup rejections
First-yes-counts instructions	112	147	107	31	184	151
No first-yes-counts instructions	110	167	67	50	194	100

Table 2. Response frequencies obtained in the conceptual replication of the sequential-lineup conditions of Horry et al.³².

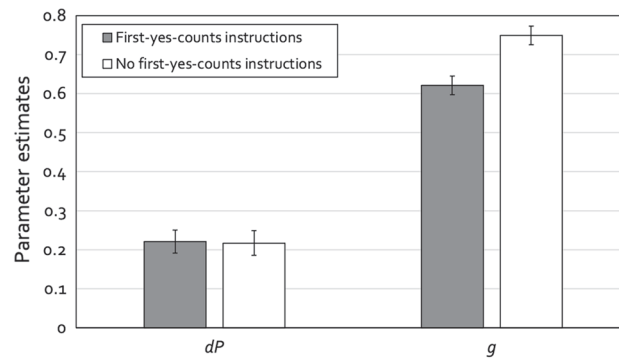


Figure 3. Estimates of parameters dP (detection of the presence of the culprit) and g (guessing-based selection) of the 2-HT eyewitness identification model as a function of lineup instructions (first-yes-counts instructions vs. no first-yes-counts instructions) for the conceptual replication of the sequential-lineup conditions of Horry et al.³². The error bars represent the standard errors.

General discussion

Horry et al.³² were the first to demonstrate experimentally that the presence or absence of first-yes-counts instructions in sequential lineups affects eyewitness identification decisions. In order to assess the reliability of the sequential-lineup findings of Horry et al.³² and to deal with the three threats to the validity of their conclusions explicated in the introduction above, we reanalyzed their data using the 2-HT eyewitness identification model and conducted a conceptual replication of the sequential-lineup conditions of their experiment. The results of both the reanalysis and the conceptual replication are well in line with the conclusions that were drawn from the analyses of Horry et al.³²: First-yes-counts instructions in sequential lineups do not affect the ability of eyewitnesses to detect the culprit but first-yes-counts instructions decrease the probability of guessing-based selection compared to instructions in which it is not mentioned that only the first identification decision counts. This conclusion was reached here using a single coherent and experimentally validated measurement model in which the complete 2×3 data structure of eyewitness identification decisions is used to separately measure the detection and guessing processes involved in eyewitness identification decisions. On the one side, we regard using a single coherent measurement model as advantageous compared to the use of two separate and conceptually different measurement models by Horry et al.³²—receiver operating characteristics for eyewitnesses' ability to discriminate between culprits and innocent suspects and an implicit ad-hoc measurement model for eyewitnesses' response criterion based on both the suspect identification rate and the lineup rejection rate. On the other side, it seems quite impressive that two such different approaches to measuring the processes underlying eyewitness identification decisions lead to the same basic conclusions. Not only did the two measurement approaches differ, but the identification conditions in Horry et al.³² and the replication study also differed in many respects such as in the number of culprits, the length of the mock-crime video and the setting in which participants were tested (online vs. laboratory). As an aside, these procedural differences may be the reason why the probability of culprit-presence detection was lower and the probability of guessing-based selection was higher in the present conceptual replication than in the reanalysis of the Horry et al.³² data. However, the important point is that the relevant comparisons between the conditions with and without first-yes-counts instructions led to the same conclusions irrespective of these procedural changes. The conclusion about the effect of first-yes-counts instructions on the processes underlying eyewitness identification decisions thus do not depend on the details of the experimental implementation.

With respect to guessing-based selection, the conclusion drawn from both datasets analyzed here is that first-yes-counts instructions serve to reduce the probability with which this process occurs. Should one thus use first-yes-counts instructions if the goal is to reduce guessing-based selection? There is in fact a far simpler and more straightforward method for reducing the probability of guessing-based selection without affecting the probability of culprit-presence detection. As Winter et al.³⁵ have shown in their Experiment 3, using pre-lineup instructions that discourage all but the very certain identifications suffices to reach this goal.

Given that the lineups used in both the experiment of Horry et al.³² and the present conceptual replication were essentially fair (parameter b was essentially zero in both analyses), there was no basis for biased-suspect selection to differ between the condition with first-yes-counts instructions and the condition without such instructions. Therefore, the existing data do not allow us to draw firm conclusions about whether, and if so how, first-yes-counts instructions might affect biased-suspect selection in unfair lineups. The effect of first-yes-counts instructions on biased-suspect selection in unfair lineups is thus an issue for future research to explore.

Knowing that and how first-yes-counts instructions affect the processes underlying eyewitness identification decisions is important. This is so because one may now conclude that a sizeable number of previous eyewitness identification studies indeed lack ecological validity in that they underestimate the tendency of eyewitnesses to select the culprit based on guessing relative to real police lineups (e.g.,^{10–27,47}). Parallel to what Steblay et al.⁸ have stated, we know of no jurisdiction that has implemented a rule to take into account only the witnesses' first identification decision in sequential lineups^{3–5,29–31}. The reason as to why such a rule is avoided seems obvious.

Legal systems typically come with a regulation as to how many fillers have to be in a lineup together with the suspect. For instance, in the United States a lineup usually consists of five fillers and the suspect⁴⁸, in the United Kingdom a lineup usually consists of eight fillers and the suspect²⁹ and in Germany a lineup usually consists of seven fillers and the suspect². A first-yes-counts rule could be seen to undermine these regulations. Specifically, whenever an eyewitness would make an identification decision prior to seeing the last photo in the lineup and thus terminate the lineup presentation, it could be argued by defense lawyers that the actual lineup size was lower than the lineup size stipulated by the legal system. The most straightforward solution to the problem is to assume that if multiple identification decisions occur, later identification decisions imply a revision of earlier identification decisions which implies that the last identification decision in a lineup is the one to be used. Alternatively, in some legal systems eyewitnesses are obliged to perform multiple laps through a lineup before they are allowed to make one final identification (e.g.,^{3,29}). Either way, research on sequential lineups needs to implement one of the rules instead of the first-yes-counts rule if the aim is not to decrease the ecological validity of this research.

Data availability

The datasets generated and analyzed during the current study are available in the OSF repository, <https://osf.io/pxhk8/>.

Received: 9 January 2023; Accepted: 12 April 2023

Published online: 21 April 2023

References

- Wells, G. L. *et al.* Eyewitness identification procedures: recommendations for lineups and photospreads. *Law Hum Behav.* **22**, 603–647. <https://doi.org/10.1023/A:1025750605807> (1998).
- German Federal Ministry of the Interior and Community. *Richtlinien für das Strafverfahren und das Bußgeldverfahren (RiStBV) [Guidelines for criminal proceedings and summary proceedings]*, https://www.verwaltungsvorschriften-im-internet.de/bsvwvbund_01011977_420821R5902002.htm (2021).
- Horry, R., Memon, A., Wright, D. B. & Milne, R. Predictors of eyewitness identification decisions from video lineups in England: a field study. *Law Hum Behav.* **36**, 257–265. <https://doi.org/10.1037/h0093959> (2012).
- Bill Blackwood Law Enforcement Management Institute of Texas. *Model Policy on Eyewitness Identification*, http://www.lemiitonline.org/resources/documents/ewid_final.pdf (2022).
- Wells, G. L., Steblay, N. K. & Dysart, J. E. Double-blind photo lineups using actual eyewitnesses: an experimental test of a sequential versus simultaneous lineup procedure. *Law Human Behav.* **39**, 1–14. <https://doi.org/10.1037/lhb0000096> (2015).
- Seale-Carlisle, T. M. & Mickes, L. US line-ups outperform UK line-ups. *R. Soc. Open Sci.* **3**, 2–12. <https://doi.org/10.1098/rsos.160300> (2016).
- McQuiston-Surrett, D., Malpass, R. S. & Tredoux, C. G. Sequential versus simultaneous lineups: a review of methods, data, and theory. *Psychol. Public Policy Law* **12**, 137–169. <https://doi.org/10.1037/1076-8971.12.2.137> (2006).
- Stebly, N. K., Dysart, J. E. & Wells, G. L. Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. *Psychol. Public Policy Law* **17**, 99–139. <https://doi.org/10.1037/a0021650> (2011).
- Carlson, C. A., Carlson, M. A., Weatherford, D. R., Tucker, A. & Bednarz, J. The effect of backloading instructions on eyewitness identification from simultaneous and sequential lineups. *Appl. Cogn. Psychol.* **30**, 1005–1013. <https://doi.org/10.1002/acp.3292> (2016).
- Carlson, C. A., Gronlund, S. D. & Clark, S. E. Lineup composition, suspect position, and the sequential lineup advantage. *J. Exp. Psychol. Appl.* **14**, 118–128. <https://doi.org/10.1037/1076-898X.14.2.118> (2008).
- Gronlund, S. D., Carlson, C. A., Dailey, S. B. & Goodsell, C. A. Robustness of the sequential lineup advantage. *J. Exp. Psychol. Appl.* **15**, 140–152. <https://doi.org/10.1037/a0015082> (2009).
- Horry, R., Palmer, M. A. & Brewer, N. Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *J. Exp. Psychol. Appl.* **18**, 346–360. <https://doi.org/10.1037/a0029779> (2012).
- Gronlund, S. D. *et al.* Showups versus lineups: an evaluation using ROC analysis. *J. Appl. Res. Mem. Cogn.* **1**, 221–228. <https://doi.org/10.1016/j.jarmac.2012.09.003> (2012).
- Mickes, L., Flowe, H. D. & Wixted, J. T. Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous versus sequential lineups. *J. Exp. Psychol. Appl.* **18**, 361–376. <https://doi.org/10.1037/a0030609> (2012).
- Toftely, S. M. *Eyewitness Identification: An Investigation of the Feature-Detection Hypothesis* (University of Oklahoma, 2015).
- Flowe, H. D. An exploration of visual behaviour in eyewitness identification tests. *Appl. Cogn. Psychol.* **25**, 244–254. <https://doi.org/10.1002/acp.1670> (2011).
- Lindsay, R. C. L. & Bellinger, K. Alternatives to the sequential lineup: The importance of controlling the pictures. *J. Appl. Psychol.* **84**, 315–321. <https://doi.org/10.1037/0021-9010.84.3.315> (1999).
- Dunn, J. C., Kaesler, M. & Semmler, C. A model of position effects in the sequential lineup. *J. Memory Lang.* **122**, 104297. <https://doi.org/10.1016/j.jml.2021.104297> (2022).
- Flowe, H. D. & Ebbesen, E. B. The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law Hum Behav.* **31**, 33–52. <https://doi.org/10.1007/s10979-006-9045-9> (2007).
- Kaesler, M., Dunn, J. C., Ransom, K. & Semmler, C. Do sequential lineups impair underlying discriminability? *Cogn. Res. Princ. Impl.* **5**, 1–21. <https://doi.org/10.1186/s41235-020-00234-5> (2020).
- Kneller, W., Memon, A. & Stevenage, S. Simultaneous and sequential lineups: decision processes of accurate and inaccurate eyewitnesses. *Appl. Cogn. Psychol.* **15**, 659–671. <https://doi.org/10.1002/acp.739> (2001).
- Memon, A. & Gabbert, F. Unravelling the effects of sequential presentation in culprit-present lineups. *Appl. Cogn. Psychol.* **17**, 703–714. <https://doi.org/10.1002/acp.909> (2003).
- Sauer, J. D., Brewer, N. & Wells, G. L. Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Leg. Criminol. Psychol.* **13**, 123–135. <https://doi.org/10.1348/135532506X159203> (2008).
- Carlson, C. A. & Carlson, M. A. An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *J. Appl. Res. Mem. Cogn.* **3**, 45–53. <https://doi.org/10.1016/j.jarmac.2014.03.004> (2014).
- Dobolyi, D. G. & Dodson, C. S. Eyewitness confidence in simultaneous and sequential lineups: a criterion shift account for sequential mistaken identification overconfidence. *J. Exp. Psychol. Appl.* **19**, 345–357. <https://doi.org/10.1037/a0034596> (2013).
- Flowe, H. D., Smith, H. M. J., Karoğlu, N., Onwuegbusi, T. O. & Rai, L. Configural and component processing in simultaneous and sequential lineup procedures. *Memory* **24**, 306–314. <https://doi.org/10.1080/09658211.2015.1004350> (2016).
- Brewer, N., Weber, N., Wooton, D. & Lindsay, D. S. Identifying the bad guy in a lineup using confidence judgements under deadline pressure. *Psychol. Sci.* **23**, 1208–1214. <https://doi.org/10.1177/09567976124412> (2012).

28. State Bar of Michigan: Eyewitness Identification Task Force. *Law Enforcement and Eyewitness Identifications: A Policy Writing Guide*, https://www.michigan.gov/-/media/Project/Websites/mcoles/TD-Webpage/state_bar_eyewitness_law_enforcement_policy_writing_guide.pdf?rev=6e19807bd78447d2b54e6f9cc7ab186b (2015).
29. Home Office. *Police and Criminal Evidence Act 1984 (PACE). Codes of Practice, Code D*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/903812/pace-code-d-2017.pdf (2017).
30. Wells, G. L., Steblay, N. K. & Dysart, J. E. in *American Judicature Society* (2011).
31. U.S. Department of Justice: Office of the Deputy attorney General. *Eyewitness Identification: Procedures for Conducting Photo Arrays*, <https://www.justice.gov/file/923201/download> (2017).
32. Horry, R., Fitzgerald, R. J. & Mansour, J. K. “Only your first yes will count”: the impact of pre-lineup instructions on sequential lineup decisions. *J. Exp. Psychol. Appl.* **27**, 170–186. <https://doi.org/10.1037/xap0000337> (2021).
33. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349**, aac4716. <https://doi.org/10.1126/science.aac4716> (2015).
34. Menne, N. M., Winter, K., Bell, R. & Buchner, A. A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Sci. Rep.* **12**, 13379. <https://doi.org/10.1038/s41598-022-17400-y> (2022).
35. Winter, K., Menne, N. M., Bell, R. & Buchner, A. Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Sci. Rep.* **12**, 15571. <https://doi.org/10.1038/s41598-022-19513-w> (2022).
36. Erdfelder, E. *et al.* Multinomial processing tree models: a review of the literature. *Z. Psychol./J. Psychol.* **217**, 108–124. <https://doi.org/10.1027/0044-3409.217.3.xxx> (2009).
37. Schmidt, O., Erdfelder, E. & Heck, D. W. Tutorial on multinomial processing tree modeling: how to develop, test, and extend MPT models. *PsyArXiv*, <https://doi.org/10.31234/osf.io/gh8md> (2022).
38. Riefer, D. M. & Batchelder, W. H. Multinomial modeling and the measurement of cognitive processes. *Psychol. Rev.* **95**, 318–339. <https://doi.org/10.1037/0033-295X.95.3.318> (1988).
39. Moshagen, M. multiTree: a computer program for the analysis of multinomial processing tree models. *Behav. Res. Methods* **42**, 42–54. <https://doi.org/10.3758/BRM.42.1.42> (2010).
40. Simons, D. J. The value of direct replication. *Perspect. Psychol. Sci.* **9**, 76–80. <https://doi.org/10.1177/1745691613514755> (2014).
41. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191. <https://doi.org/10.3758/BF03193146> (2007).
42. Leiner, D. J. SoSci Survey <https://www.sosicisurvey.de> (2022).
43. Minear, M. & Park, D. C. A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* **36**, 630–633. <https://doi.org/10.3758/BF03206543> (2004).
44. Federal Ministry of the Interior of the State of North-Rhine-Westphalia. *Wahllichtbildvorlage im Strafverfahren [Photo lineups in criminal proceedings]*, https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=1&bes_id=91478&aufgehoben=N (2022).
45. Zarkadi, T., Wade, K. A. & Stewart, N. Creating fair lineups for suspects with distinctive features. *Psychol. Sci.* **20**, 1448–1453. <https://doi.org/10.1111/j.1467-9280.2009.02463.x> (2009).
46. Colloff, M. F., Wade, K. A. & Strange, D. Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychol. Sci.* **27**, 1227–1239. <https://doi.org/10.1177/09567976166655789> (2016).
47. Lindsay, R. C. L. & Wells, G. L. Improving eyewitness identifications from lineups: simultaneous versus sequential lineup presentation. *J. Appl. Psychol.* **70**, 556–564. <https://doi.org/10.1037/0021-9010.70.3.556> (1985).
48. Police Executive Research Forum. *A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies*, https://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf (2013).

Author contributions

K.W., N.M.M., R.B. and A.B. contributed to the study conception, design, material preparation and data analysis. K.W. collected the data and wrote the first draft of the manuscript with subsequent input from all co-authors who gave final approval for publication.

Funding

Open Access funding enabled and organized by Projekt DEAL. The work reported herein was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BU 945/10–1, Project Number 456214986.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Erklärung über den Eigenanteil der in der Dissertation enthaltenen Einzelarbeiten

Die vorliegende Dissertation umfasst zwei Fachartikel mit insgesamt fünf Experimenten und einer Reanalyse bestehender Daten. Für jeden Fachartikel wird im Folgenden aufgeführt, welche Personen bei der Planung und Umsetzung der Experimente, bei der Datenauswertung und beim Verfassen der Manuskripte mitgearbeitet haben. Der überwiegende Teil der Arbeit lag für beide Fachartikel jeweils bei der Erstautorin.

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2022). Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Scientific Reports*, 12, 15571. <https://doi.org/10.1038/s41598-022-19513-w>

Planung: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Umsetzung: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Auswertung: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Manuskript: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2023). Evaluating the impact of First-Yes-Counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Scientific Reports*. 13, 6572. <https://doi.org/10.1038/s41598-023-33424-4>

Planung: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Umsetzung: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Auswertung: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Manuskript: Winter, K., Menne, N. M., Bell, R., & Buchner, A.

Erklärung an Eides Statt

Hiermit versichere ich an Eides Statt, dass ich die Dissertation mit dem Titel »Validierung und Anwendung eines multinomialen Verarbeitungsbaummodells für Augenzeugengegenüberstellungen« selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der »Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf« erstellt habe.

Ich versichere insbesondere:

- (1) Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.
- (2) Alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen habe ich als solche kenntlich gemacht; dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen.
- (3) Die Arbeit habe ich in der vorliegenden oder einer modifizierten Form noch nicht als Dissertation vorgelegt – sei es an der Heinrich-Heine-Universität Düsseldorf oder an einer anderen Universität.

Datum: 18. Juli 2023

Name: Kristina Winter

Unterschrift: