

Relevanz von Wissen über Alltagsgegenstände und visueller Inspektionsstrategie für die Gepäckkontrolle mit Röntgengeräten

Nicole HÄTTENSCHWILER, Yanik STERCHI, Stefan MICHEL,
Adrian SCHWANINGER

*Institut Mensch in komplexen Systemen (MikS)
Hochschule für Angewandte Psychologie
Fachhochschule Nordwestschweiz (FHNW)
Riggenbachstrasse 16, Ch-4600 Olten*

Kurzfassung: Die Analyse von Gepäckstücken durch Luftsicherheitsbeauftragte (Screener) wird als visuelle Inspektionsaufgabe verstanden, bei der verbotene Gegenstände in Röntgenbildern von Gepäckstücken detektiert werden müssen. In mehreren Studien konnte gezeigt werden, dass Wissen über das Aussehen verbotener Gegenstände in Röntgenbildern wichtig für ein erfolgreiches Ausüben dieser Aufgabe ist. In der vorliegenden Studie wurde untersucht, ob auch das Wissen über das Aussehen von Alltagsgegenständen in Röntgenbildern relevant ist und ob durch die Instruktion einer anderen visuellen Inspektionsstrategie Treffer- und Fehlalarmrate beeinflusst werden können. Die Ergebnisse zeigen, dass Screener abhängig von der Instruktion das Kriterium (response bias) verändern, was sich an einer höheren Trefferrate und einer gleichzeitig erhöhten Fehlalarmrate zeigt. Gleichzeitig ergaben sich interessante Zusammenhänge mit dem Wissen über das Aussehen von Alltagsgegenständen in Röntgenbildern.

Schlüsselwörter: Luftsicherheit, Mensch-Maschinen-Interaktion, Röntgenbildanalyse, Alltagsgegenstände, Suchstrategie

1. Einleitung

Die Sicherung des Luftverkehrs ist von grosser Bedeutung für Wirtschaft und Gesellschaft. Bezugnehmend auf die terroristische Bedrohungslage wurden im letzten Jahrzehnt erhebliche Investitionen in die Weiterentwicklung von Sicherheitskontrollen an Flughäfen getätigt. Sicherheitsbeauftragte an Flughäfen (Screener) analysieren Gepäckstücke mit Röntgengeräten, um sicherzustellen, dass keine verbotenen Gegenstände (Bomben, Messer, Schusswaffen, u.a.) an Bord eines Flugzeuges gelangen.

Um bekannte und neuartige verbotene Gegenstände zuverlässig in Röntgenbildern erkennen zu können, müssen Screener gut aus- und kontinuierlich weitergebildet werden. Bisherige Studienergebnisse sprechen dafür, dass das Erlernen des Aussehens von verbotenen Gegenständen in Röntgenbildern mittels computer-basierter Training (CBT) für die Detektion eine wichtige Rolle spielt (z.B. Koller et al. 2007; Halbherr et al. 2013). Neben diesen sogenannten *wissensbasierten Faktoren* bestätigen Studien auch die Relevanz des Umgangs mit *bildbasierten Faktoren* (Rotation von Gegenständen, Verdeckung durch andere Gegenstände, Gepäckkomplexität) in Bezug auf die Röntgenbildanalyse (z.B.

Schwaninger et al. 2005; Bolfig et al. 2008). Weiter spielen Feedback durch Supervision und Testungen eine Rolle (Hardmeier et al. 2005).

Die Analyse von Gepäckstücken kann als visuelle Inspektionsaufgabe verstanden werden, welche aus visueller Suche und Entscheidung besteht (Koller et al. 2007; Wales et al. 2009). Die Entscheidung, ob ein verbotener Gegenstand vorhanden ist, kann anhand der Signaldetektionstheorie beschrieben werden (Green & Swets 1966; Macmillan & Creelman 2005). Die Signaldetektionstheorie besagt vereinfacht, dass sich Trefferrate und Fehlalarmrate einerseits durch die *Sensitivität* ergeben: die Fähigkeit, *Rauschen* (Gepäckstücke mit Alltagsgegenständen) von *Signal und Rauschen* (hier Gepäckstücke mit Alltagsgegenständen und verbotenem Gegenstand) unterscheiden zu können. Andererseits spielt auch das sogenannte *Kriterium* eine Rolle, also die Tendenz, ein Gepäckstück als *OK* oder *nicht OK* zu bewerten. Es wird angenommen, dass unter konstanter Sensitivität das Kriterium geändert werden kann, womit sich Treffer- und Fehlalarmrate gleichgerichtet verändern. Die dabei möglichen Wertepaare werden durch die Receiver-Operating-Characteristic-Kurve (ROC-Kurve) beschrieben.

Zur Schätzung von Sensitivität und Kriterium werden häufig Masse verwendet, die sich aus einem Wert für Treffer- und Fehlalarmrate berechnen (z.B. d' und A'). Diese gehen von einer bestimmten Form der ROC-Kurve aus (für mehr Informationen siehe Macmillan & Creelman 2005). Einige Studien weisen nun darauf hin, dass diese Annahmen in der Röntgenbildanalyse manchmal verletzt werden (z.B. Wolfe et al. 2007; Godwin et al. 2010). Alternativ erlauben Konfidenzratings, empirische ROC-Kurven zu bestimmen und die Fläche unter der Kurve (Englisch: *area under the curve*: AUC) als Sensitivitätsmass verwenden zu können (Macmillan & Creelman 2005).

Die Sensitivität ist einerseits hoch, wenn viel Wissen über das Aussehen verbotener Gegenstände in Röntgenbildern vorhanden ist. Gleichzeitig müsste aber auch Wissen über das Aussehen von Alltagsgegenständen in Röntgenbildern die Unterscheidung zwischen harmlos und verboten vereinfachen, was gemäss der Konzeptionierung mittels Signaldetektionstheorie als Reduktion von Rauschen verstanden werden kann. Hättenschwiler et al. (2015) konnten zeigen, dass mehr Wissen über Alltagsgegenstände mit weniger Fehlalarmen korreliert. Dies lässt sich intuitiv folgendermassen erklären: Wird ein Alltagsgegenstand als harmlos identifiziert, entfällt das Risiko, dass er fälschlicherweise als verboten eingestuft wird (Fehlalarm). Aus Sicht der Effizienz ist es wünschenswert, die Fehlalarme möglichst zu minimieren. Dennoch stellt sich die Frage, ob Wissen über Alltagsgegenstände auch dazu genutzt werden könnte, die Trefferrate zu steigern. Gemäss Signalentdeckungstheorie müsste dies möglich sein, wenn es gelingt, das Kriterium liberaler zu setzen (also die Tendenz zu steigern, ein Gepäckstück als *nicht OK* zu bewerten): Treffer- und Fehlalarmrate würden beide zunehmen. Die potentielle Rolle von Wissen über Alltagsgegenstände könnte insbesondere bei der Erkennung von neuen, noch nie zuvor gesehenen verbotene Gegenständen relevant sein. Wegen dem fehlenden Wissen über deren Aussehen (wissensbasierte Faktoren), sind neue verbotene Gegenstände umso schwieriger zu erkennen, je unähnlicher sie zu bekannten verbotenen Gegenständen sind. Es kann aber angenommen werden, dass Screener mit gutem Wissen über Alltagsgegenstände die Möglichkeit haben, neue verbotene Gegenstände nach dem Ausschlussprinzip zu erkennen. Eine entsprechende Suchstrategie (fortan Verifikationssuche genannt) müsste also beinhalten, dass nur Gepäckstücke als OK bewertet werden, bei denen sämtliche beinhalteten Gegenstände als Alltagsgegenstände identifiziert werden können. Wenn

Screeener instruiert werden können, Verifikationssuche anzuwenden, und dadurch ihr Kriterium verschieben, hätte dies interessante Anwendungsmöglichkeiten in der Praxis. So könnten z.B. Gepäckstücke von bestimmten, als potentiell gefährlich eingestuft Passagieren mit dieser Suchstrategie untersucht und somit die Wahrscheinlichkeit, verbotene Gegenstände zu entdecken, erhöht werden.

Bis heute gibt es gemäss unserem Kenntnisstand noch keine Studie, welche untersucht, welche Effekte sich durch die Instruktion einer solchen Suchstrategie (Verifikationssuche) auf die Detektionsleistung ergeben. In dieser ersten explorativen Studie wurde dieser Fragestellung nachgegangen.

2. Methoden

2.1 Teilnehmende

31 erfahrene Screeener eines internationalen Flughafens nahmen am Experiment teil. Das durchschnittliche Alter der Screeener betrug 45.42 Jahre (SD = 8.92) und die durchschnittliche Arbeitserfahrung 8.40 Jahre (SD = 5.56). 64.5% waren weiblich.

2.2 Design

Das Experiment wurde als Within-Subjects-Design umgesetzt. Als unabhängige Variablen wurden verbotene Gegenstände (bekannte vs. neue) und die Suchstrategie (normale Suche vs. Verifikationssuche) manipuliert. Die abhängigen Variablen waren Treffer- und Fehlalarmrate sowie AUC, welche mit einer simulierten Gepäckscreeningaufgabe (SGSA) erhoben wurden. Wissen über Alltagsgegenstände wurde mit einem Alltagsgegenstandstest (AGT) als potentielles Korrelat zur Detektionsleistung erhoben.

2.3 Material

Der AGT umfasste 32 Röntgenbilder von Handgepäckstücken, wobei auf jedem Bild drei Gegenstände durch einen roten Rahmen markiert waren. Der Test enthielt insgesamt 17 verbotene Gegenstände und 79 Alltagsgegenstände. Die Gegenstände entstammten den Kategorien *Alltagsgegenstand*, *unkonventionelle Spreng- und Brandvorrichtung (USBV)* oder *anderer verbotener Gegenstand*.

Die SGSA enthielt 128 Röntgenbilder, wovon die Hälfte einen verbotenen Gegenstand aus den Kategorien *Messer*, *Schusswaffen*, *USBV* oder *andere verbotene Gegenstände* enthielt. Von den 64 verbotenen Gegenständen war die Hälfte auch im computerbasierten Training der Screeener (CBT) enthalten, die andere Hälfte wurde neu aufgenommen.

2.4 Vorgehen

Die Erhebung fand an zwei verschiedenen Zeitpunkten statt. Bei der ersten Testung führten die Screeener den Alltagsgegenstandstest (AGT) durch, bei der zweiten die SGSA.

Beim AGT wurden die Screeener instruiert, alle markierten Gegenstände zuerst zu kategorisieren (siehe Material) und anschliessend möglichst treffend zu bezeichnen. Die Bearbeitung des AGT dauerte ca. 45-60 Minuten.

Bei der SGSA mussten die Screener bei jedem von 128 Röntgenbildern von Gepäckstücken entscheiden, ob dieses einen verbotenen Gegenstand enthielt (*nicht OK*) oder harmlos (*OK*) war. Nach dieser Entscheidung gaben sie zusätzlich an, wie sicher sie sich waren. Die SGSA war in vier Blöcke aufgeteilt, wobei die Screener bei zwei dieser Blöcke instruiert wurden, eine bestimmte Suchstrategie, die sogenannte Verifikationssuche (VS), anzuwenden. Die Screener erhielten bei VS die Instruktion, für jedes Objekt im Röntgenbild zu entscheiden, ob sie dieses kennen oder nicht, und das Gepäckstück als "Nicht OK" einzustufen, falls sie mindestens einen Gegenstand nicht als harmlosen Alltagsgegenstand identifizieren konnten. Bei den restlichen zwei Blöcken sollten die Screener wie im Berufsalltag arbeiten (normale Suche, NS). Die Reihenfolge der Blöcke wurde über die Versuchspersonen ausbalanciert. Die Bearbeitung der SGSA dauerte ca. 30 Minuten.

2.5 Analysen

Anhand der Konfidenzratings wurde mit dem R-Paket pROC (R Core Team 2016; Robin et al. 2011) für jeden Screener die AUC bestimmt. Vergleiche zwischen den beiden visuellen Inspektionsstrategien wurden wegen fehlender Normalverteilung der abhängigen Variablen mit dem Wilcoxon-Rangsummentest auf Signifikanz geprüft.

3. Resultate und Diskussion

Abbildung 1 zeigt die Mittelwerte der Trefferrate abhängig von der Instruktion (NS, VS) für neue und bekannte verbotene Gegenstände sowie die Fehlalarmrate.

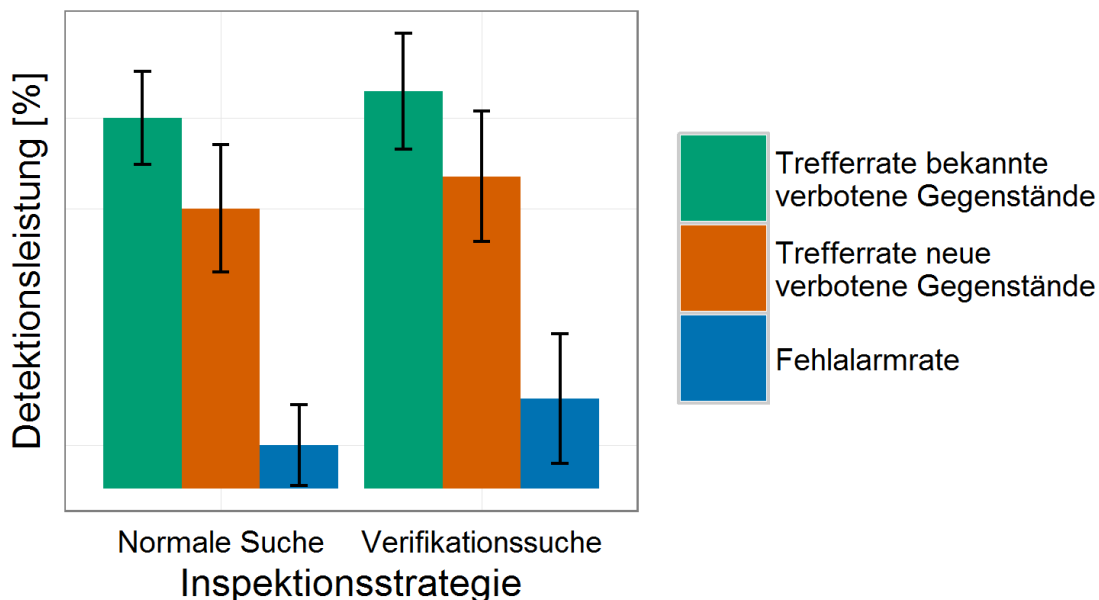


Abbildung 1: Trefferrate und Fehlalarmrate aufgegliedert nach Suchstrategie (normale Suche vs. Verifikationssuche) und der Art von verbotenen Gegenständen (bekannte vs. neue). Die Fehlerbalken entsprechen Standardabweichungen. Die Trefferraten sind vertraulich, weshalb die Ordinatenwerte nicht angegeben sind.

Wie erwartet, war die Trefferrate für bekannte verbotene Gegenstände höher als für neue verbotene Gegenstände (d.h. verbotene Gegenstände, welche nicht im CBT

der Screener vorhanden waren), $W = 1934$, $p < .001$. Die Screener erzielten mit der Verifikationssuche im Vergleich zur normalen Suche eine höhere Trefferrate, sowohl für bekannte verbotene Gegenstände ($W = 85$, $p = .02$) als auch für neue verbotene Gegenstände ($W = 95.5$, $p = .02$). Gleichzeitig war auch die Fehlalarmrate signifikant höher ($W = 60.5$, $p < .001$).

Die Analyse der Sensitivität (AUC-Werte) zeigte keine Unterschiede zwischen den beiden Suchstrategien, weder für bekannte, $W = 240$, $p = .88$ (NS: $M = .889$, $SD = .066$; VS: $M = .890$, $SD = .068$) noch für neue verbotene Gegenstände, $W = 262.5$, $p = .78$ (NS: $M = .794$, $SD = .079$; VS: $M = .789$, $SD = .067$). Die veränderte Suchstrategie (NS vs. VS) hat also nicht die Sensitivität beeinflusst und die Unterschiede in Treffer- und Fehlalarmrate müssen durch eine Veränderung des Kriteriums entstanden sein. Demnach sind Screener also grundsätzlich in der Lage, aufgrund einer Instruktion ihr Entscheidungskriterium zu verschieben.

In einem nächsten Schritt wurde untersucht, ob mehr Wissen über das Aussehen von Alltagsgegenständen in Röntgenbildern mit einer höheren Trefferrate bzw. einer tieferen Fehlalarmrate bei der SGSA einhergeht (wie einleitend beschrieben, würde man einen dieser Zusammenhänge oder eine Kombination der beiden erwarten). Wie Tabelle 1 anhand von Rangkorrelationen (wegen fehlender Linearität der Beziehung wurde auf Pearson-Korrelationen verzichtet) zeigt, waren Screener mit einer hohen Leistung im AGT bei NS besser im Erkennen von bekannten verbotenen Gegenständen und weisen eine marginal signifikant tiefere Fehlalarmrate auf. Sie waren jedoch nicht besser im Erkennen neuer verbotener Gegenstände. Dass Screener mit mehr Wissen über Alltagsgegenstände weniger Fehlalarme erzielen (wenn auch hier bloss marginal signifikant), stimmt mit den Ergebnissen aus Hättenschwiler et al. (2015) überein. Das bessere Abschneiden bei bekannten, aber nicht bei neuen verbotenen Gegenständen, könnte daran liegen, dass Screener mit mehr Wissen über Alltagsgegenstände auch besseres Wissen über die aus dem Training bekannten verbotenen Gegenstände besitzen.

Tabelle 1: Rangkorrelationen zwischen der Anzahl richtig benannter Gegenstände im Alltagsgegenstandstest (AGT) und Ergebnissen aus der SGSA (Trefferrate, TR, sowie Fehlalarmrate). Signifikanzwerte wurden einseitig berechnet.

Korrelationen AGT und SGSA Werte (TR und FAR)	TR bekannte verbotene Gegenstände	TR neue verbotene Gegenstände	Fehlalarmrate
Normale Suche	$r_s = .430$ $p = .008$	$r_s = -.117$ $p = .735$	$r_s = -.298$ $p = .052$
Verifikationssuche	$r_s = .391$ $p = .015$	$r_s = .322$ $p = .038$	$r_s = -.018$ $p = .462$

Bemerkenswert ist, dass unter der Bedingung der VS das Wissen über Alltagsgegenstände nicht mehr mit der Fehlalarmrate sondern nun positiv mit der Trefferrate bei neuen verbotenen Gegenständen korrelierte. Dieses Ergebnis liefert einen ersten Hinweis, dass Wissen über Alltagsgegenstände auch zur Steigerung der Trefferrate für neue verbotene Gegenstände eingesetzt werden kann. Da aber das Wissen über Alltagsgegenstände nicht manipuliert wurde, kann nicht ausgeschlossen werden, dass dem gefundenen Zusammenhang eine weitere Variable zu Grunde lag (z.B. allgemein bessere Kompetenz in der Erkennung verbotener Gegenstände).

4. Schlussfolgerung

Die Ergebnisse zeigen, dass Screener aufgrund der Instruktion einer neuen Suchstrategie ihre Trefferrate steigern können, was allerdings auch mehr Fehlalarme mit sich bringt. Da die AUC durch die Instruktion nicht beeinflusst wird, handelt es sich dabei um eine Veränderung des Kriteriums - konsistent mit der Betrachtung der Röntgenbildinspektion als visuelle Suche und Entscheidungsaufgabe gemäss der Signalentdeckungstheorie. Die Verifikationssuche eignet sich wegen den operativen Kosten durch die erhöhte Fehlalarmrate zwar nicht für die Kontrolle sämtlicher Röntgenbilder. Aber die Verifikationssuche könnte in Zukunft genutzt werden, um basierend auf einer Risikoabschätzung für bestimmte Flüge, Passagiere oder eine kurzzeitig erhöhte Bedrohungslage die Trefferrate in der Gepäckkontrolle zu steigern.

Falls zukünftige Forschung bestätigen kann, dass Wissen über Alltagsgegenstände unter Anwendung der normalen Suche die Fehlalarmrate senkt und unter Anwendung der Verifikationssuche die Erkennung neuer verbotener Gegenstände erhöht, könnte zukünftig spezifisches Training von Alltagsgegenständen sowohl die Luftsicherheit erhöhen als auch die Kosten der Sicherheitskontrolle senken.

5. Literatur

- Bolfing, A., Halbherr, T., & Schwaninger, A. (2008). How image based factors and human factors contribute to threat detection performance in x-ray aviation security screening. *HCI and Usability for Education and Work, Lecture Notes in Computer Science*, 5298, 419-438.
- Godwin, H.J., Menneer, T., Cave, K.R. & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, 18 (10), 1439–1463.
- Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Hättenschwiler, N., Michel, S., Ritzmann, S., Schwaninger, A. (2015). A First Exploratory Study on the Relevance of Everyday Object Knowledge and Training for Increasing Efficiency in Airport Security X-ray Screening. *Proceedings of the 49th IEEE International Carnahan Conference on Security Technology*, Taipei Taiwan, September 21-24, 2015, 25-30.
- Halbherr, T., Schwaninger, A., Budgell, G., & Wales, A. (2013). Airport security screener competency: a cross-sectional and longitudinal analysis. *International Journal of Aviation Psychology*, 23(2), 113-129.
- Hardmeier D, Hofer F, Schwaninger A (2005) The X-ray object recognition test (X-ray ORT) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening. *IEEE ICCST Proceedings*, 39: 189-192.
- Koller, S., Hardmeier, D., Michel, S., & Schwaninger, A. (2007). Investigating training, transfer and viewpoint effects resulting from recurrent CBT of x-ray image interpretation. *Journal of Transportation Security*, 1(2), 81-106.
- Macmillan, N.A., Creelman, C.D. (2005). *Detection theory: A users guide* (2nd ed.). New York: Cambridge University Press.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12 (1), 77. BioMed Central. doi:10.1186/1471-2105-12-77
- Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners visual abilities & visual knowledge measurement. *IEEE Aerospace and Electronic Systems*, 20(6), 29-35.
- Wales, A.W.J., Anderson, C., Jones, K.L., Schwaninger, A., Horne, J.A., 2009. Evaluating the two-component inspection model in a simplified luggage search task. *Behavior Research Methods*, 41(3), 937-943. doi:10.3758/BRM.41.3.937
- Wolfe, J.M., Horowitz, T.S., Van Wert, M.J., Kenner, N.M., Place, S.S., Kibbi, N., 2007. Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, 136(4), 623–638.



Gesellschaft für
Arbeitswissenschaft e.V.

Soziotechnische Gestaltung des digitalen Wandels – kreativ, innovativ, sinnhaft

63. Kongress der
Gesellschaft für Arbeitswissenschaft

FHNW Brugg-Windisch, Schweiz

15. – 17. Februar 2017

GfA Press

Bericht zum 63. Arbeitswissenschaftlichen Kongress vom 15. – 17. Februar 2017

FHNW Brugg-Windisch, Schweiz

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.

Dortmund: GfA-Press, 2017

ISBN 978-3-936804-22-5

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Dortmund**

Schriftleitung: Matthias Jäger

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet, den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen.

USB-Print: Dr. Philipp Baumann, Olten

Screen design und Umsetzung

© 2017 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de