

Zusammenhang von Vertrauen und mentaler Beanspruchung in automatisierten Systemen

Gloria PÖHLER, Barbara DEML, Simon DILLMANN

*Institut für Arbeitswissenschaft und Betriebsorganisation
Karlsruher Institut für Technologie
Kaiserstr. 12, D-76131 Karlsruhe*

Kurzfassung: Automatisierte Systeme nehmen einen zunehmend präsenteren Anteil in unserem täglichen Leben ein. Das Konstrukt Vertrauen spielt im Umgang mit automatisierten Systemen eine wichtige Rolle, da es eine maßgebliche Determinante für deren Nutzung bildet. Die vorliegende Untersuchung beschäftigt sich mit der Wechselwirkung von Nutzervertrauen und mentaler Beanspruchung bei zwei verschiedenen Reliabilitäten des automatisierten Systems. Zu diesem Zweck wurde eine Studie mit $N = 30$ Personen durchgeführt, wobei 15 Personen jeder Versuchsgruppe zugeordnet wurden. Die Studie ergab signifikante Unterschiede der subjektiven Einschätzung von Vertrauen zwischen beiden Versuchsgruppen. Ein negativer Zusammenhang zwischen mentaler Beanspruchung und Vertrauen konnte nicht bestätigt werden.

Schlüsselwörter: Automation, Vertrauen, mentale Beanspruchung, Systemreliabilität, Flugsimulation, Fragebogenmaße

1. Einleitung

In den vergangenen Jahrzehnten gewannen automatisierte Systeme in den verschiedensten Lebensbereichen zunehmend an Relevanz. Hierbei meint Automatisierung bzw. Automation Tätigkeiten, die ein Mensch ausführen könnte, die zu seiner Entlastung jedoch auf eine Maschine übertragen werden (Moray et al. 2000). Die Automatisierung einzelner Tätigkeiten oder auch ganzer Prozesse bringt zahlreiche Vorteile mit sich, beispielsweise die Erhöhung von Präzision und Ökonomie bei gleichzeitiger Reduktion von Beanspruchung und Fehlern des Operateurs (Sarter et al. 1997). Einhergehend mit diesen Effekten verändert sich auch die Rolle des Operateurs innerhalb des Mensch-Maschine-Regelkreises: vom aktiv handelnden Akteur in gering automatisierten Umgebungen hin zu einem passiven Überwacher, der nur dann aktiv in das System eingreift, wenn dieses einen abnormen Zustand aufweist (Sheridan & Parasuraman 2006).

Ein wichtiges Konstrukt, das die Qualität der Überwachung automatisierter Systeme maßgeblich bestimmt, stellt Vertrauen dar (Manzey 2012). Das Konstrukt Vertrauen wird von Lee & See (2004) dabei definiert als die „Überzeugung, dass ein System dem Individuum dabei hilft, bestimmte Ziele zu erreichen in einer Situation, die von Ungewissheit und Verletzlichkeit geprägt ist“. Bei einem hohen Vertrauen in die Automation erfolgt hierbei eine Abnahme der Überwachungsleistung, das jedoch in extremen Ausmaßen zu einer missbräuchlichen Nutzung (Misuse) führen kann (Parasuraman & Riley 1997). Demgegenüber kann ein unterdurchschnittlich ausgeprägtes Vertrauen zu einer Nichtnutzung des Systems (Disuse) führen, was wiederum der beanspruchungsreduzierenden Intention der Automation zuwiderläuft.

Beide Fälle markieren eine Fehlnutzung der Automation, ausgelöst durch übersteigertes bzw. mangelndes Vertrauen.

Eine Reihe von Forschungsarbeiten widmeten sich bereits dem Themenbereich Vertrauen in der Automationsforschung, wobei jeweils unterschiedliche Fragestellungen in das Zentrum der Betrachtungen gerückt wurden. Ein prominentes Thema bildete beispielsweise die Untersuchung des Zusammenhangs zwischen der Reliabilität eines Systems und des Ausmaßes an Vertrauen in eben dieses. Studien hierzu offenbarten eine Wechselwirkung zwischen Vertrauen und dem Auftreten von Fehlern der Automation, wobei die Nutzung von automatisierten Systemen bzw. Automationshilfen mit der Anzahl aufgetretener Fehlentscheidungen dieser Systeme sinkt (siehe z.B. Dzindolet et al. 2003).

Andere Arbeiten wiederum hatten die Interaktion von Vertrauen und mentaler Beanspruchung des Nutzers zum Gegenstand. Mentale Beanspruchung meint hierbei Aktivierungen verschiedenster Leistungsfunktionen des menschlichen Informationsverarbeitungssystems, welche die Grundlage des subjektiven Gefühls des Angestrengtseins bilden (Manzey 1998). Ist der Nutzer dabei stark in die Überwachung eines Systems involviert, so steigt auch dessen mentale Beanspruchung. Demnach ermöglicht der Grad an mentaler Beanspruchung einen Rückschluss auf die Überwachungsleistung; eine Komponente, die wie eingangs erwähnt maßgeblich für die Beschreibung der Rolle des Menschen in automatisierten Systemen ist. Dabei weisen erste Studien auf eine negative Korrelation der beiden Konstrukte Vertrauen und mentale Beanspruchung hin (Safar & Turner 2005). Je weniger Vertrauen ein Nutzer einem System also schenkt, desto stärker scheint er an der Überwachung des Systems beteiligt zu sein, da sich ein höherer Grad an mentaler Beanspruchung feststellen lässt.

Jian et al. (2000) stellen ein empirisch fundiertes Messinstrument zur Messung des subjektiven Grades an Vertrauen vor. Darauf basierende Folgestudien liefern Grund zur Annahme, dass es sich bei der vorgelegten *Checklist for Trust between People and Automation (Trust Checklist)* um ein reliables und valides Instrument zur Messung der subjektiven Ausprägung von Vertrauen handelt (Spain et al. 2008).

Der vorliegende Beitrag greift die Rolle von Vertrauen in automatisierten Systemen auf und will deren Zusammenhang mit der Reliabilität eines Systems belegen. Zusätzlich wird angenommen, dass eine verminderte Reliabilität eines Systems bei ebenfalls vermindertem Vertrauen zu erhöhter mentaler Belastung des Nutzers führt.

2. Methoden

2.1 Geräte und Material

Für die Messung von mentaler Beanspruchung und Vertrauen sollen in der nachfolgenden Studie Fragebogenmaße eingesetzt werden. Hier wurde eine deutsche Version des NASA-Task Load Index (NASA-TLX) von Hart und Staveland (1988) zur Messung von mentaler Beanspruchung eingesetzt, während der individuelle Grad an Vertrauen durch die bereits erwähnte Trust Checklist (Jian et al. 2000) erfasst wurde, wobei im Rahmen der Arbeit eine Übersetzung ins Deutsche vorgenommen wurde. Die Übersetzung des Fragebogens erfolgte dabei auf Basis der Empfehlungen der International Test Commission (2005) zur Wahrung der inhaltlichen Bedeutung jedes einzelnen Items.

In engem Zusammenhang mit Vertrauen steht das Persönlichkeitsmerkmal der Kontrollüberzeugung, in der vorliegenden Studie spezifiziert auf die Kontrollüberzeugung im Umgang mit Technik. Bei einer niedrigen Ausprägung dieses Konstrukts erlebt der Nutzer technische Probleme als nicht durch eigene Anstrengungen händel- und lösbar. Es ist anzunehmen, dass in diesem Fall das Vertrauen in das technische Gerät oder die Automation höher ausfallen sollte. Eine wichtige Rolle im Umgang mit Technik spielt zudem das Konstrukt der Selbstwirksamkeit, also der Überzeugung, erwünschte Resultate durch persönliche Kompetenzen erreichen zu können. Auch hier ist anzunehmen, dass eine geringe Konstruktausprägung zu vermehrtem Vertrauen in die Entscheidungen des automatisierten Systems führen sollte. Da den Konstrukten demzufolge ein direkter Einfluss auf die Einschätzung des Vertrauens in das automatisierte System zugeschrieben werden kann, sollen beide initial durch Fragebögen erhoben werden. Dies geschieht durch den Fragebogen Kontrollüberzeugung im Umgang mit Technik (KIU) von Beier (2004) sowie die Skala zur generalisierten Selbstwirksamkeitserwartung (SW) von Collani & Schyns (2014).

Als automatisiertes System diente in der vorliegenden Studie der Autopilot der Flugsimulationssoftware Microsoft Flight Simulator X – Deluxe Edition, wobei zwei Versuchsgruppen gebildet wurden: Die erste Gruppe erlebte den Autopiloten als ein vollständig reliables System, während der Autopilot der zweiten Versuchsbedingung eine ca. 40%-ige Fehlerquote aufwies. Diese Fehlerquote wurde als Deaktivierung des Systems implementiert, wobei die festgelegte Quote 6 Minuten der insgesamt 15 Minuten Gesamtzeit entsprach und randomisiert in diese integriert wurde. Die Ausführung der einzelnen Zeitpunkte der Deaktivierung erfolgte durch den Versuchsleiter als Wizard-of-Oz-Schnittstelle.

2.2 Versuchsablauf

Vor Beginn der eigentlichen Untersuchung wurden die Versuchspersonen darüber informiert, dass es in der vorliegenden Studie um die Interaktion unerfahrener Nutzer mit einem Flugsimulationsprogramm geht und gebeten, ihr Einverständnis für die Analyse ihrer Daten durch befugte Personen zu erklären. Eine vollständige Aufklärung über Ziel und Gegenstand der Untersuchung fand nach Ende des Versuches statt. Im Anschluss daran folgte eine Erläuterung des Versuchsaufbaus sowie der Funktionsweise des Autopiloten. Die Vorlage der beiden Fragebögen zur Messung der Kontrollüberzeugung (KIU) und der Selbstwirksamkeit (SW) sowie eines demographischen Fragebogens erfolgte ebenfalls vor Beginn der eigentlichen Untersuchung, um das Bilden einer geschichteten Zufallsstichprobe mit den Testergebnissen des KIU und des SW sowie den demographischen Merkmalen Alter und Geschlecht zu ermöglichen. Diese Vorgehensweise wurde gewählt, um in beiden Versuchsbedingungen eine Vergleichbarkeit der Einschätzung des Vertrauens in das automatisierte System zu gewährleisten, da allen genannten Schichtungsfaktoren Relevanz im Hinblick auf den Einfluss auf Vertrauen beigemessen wurde.

Nach einem ca. 10-minütigen Testflug der Probanden, der mit Erklärungen des Systems durch den Versuchsleiter einherging, absolvierten die Testpersonen einen 15-minütigen Versuchsflug. Die Aufgabe der Versuchspersonen bestand darin, einen vorgegebenen Flugkurs von 68° beizubehalten. Die Steuerung der Flughöhe sowie der -geschwindigkeit erfolgte über den Autopiloten, wobei die Versuchspersonen darüber informiert wurden, dass die Möglichkeit eines Ausfallens des Autopiloten

besteht und dieser somit manuell von den Versuchspersonen wieder aktiviert werden muss. Die Überwachung des Autopiloten auf Funktionsausfälle wurde als eingebettete Nebenaufgabe realisiert, da die Flugführung in einer Simulationsumgebung für Anfänger eine komplexe Aufgabe darstellt. Demnach ist sie der Einführung einer künstlichen Aufgabe als Sekundäraufgabe überlegen. Dieser Flug wurde als Between-Subjects-Design realisiert, die Probanden flogen somit entweder mit einem vollkommen reliablen oder einem nicht zuverlässig funktionierenden Autopiloten. Nach Absolvierung des Versuchsfluges füllten die Testpersonen die Trust Checklist sowie den NASA-TLX aus. Die Probanden erhielten eine finanzielle Aufwandsentschädigung im Anschluss des Versuches.

2.3 Beschreibung der Stichprobe

An der Untersuchung nahmen insgesamt 30 Personen zwischen 18 und 34 Jahren teil, wobei sich jede der beiden Versuchsgruppen aus sechs Frauen und neun Männern zusammensetzte. In der ersten Bedingung, in dem das vollständig reliable System implementiert wurde, waren die Probanden durchschnittlich $M_{Alter_1} = 23.9$ Jahre ($SD_{Alter_1} = 5.31$ Jahre) alt, während das Durchschnittsalter in der zweiten Bedingung mit dem unzuverlässigen System $M_{Alter_2} = 24.2$ Jahre ($SD_{Alter_2} = 4.02$ Jahre) betrug. Ebenso ist es gelungen, die beiden Stichproben hinsichtlich der Variablen Kontrollüberzeugung im Umgang mit Technik ($M_{KIU_1} = 4.88$, $SD_{KIU_1} = 0.69$ und $M_{KIU_2} = 4.87$, $SD_{KIU_2} = 0.6$) und der erlebten Selbstwirksamkeit ($M_{SW_1} = 4.63$, $SD_{SW_1} = 0.48$ und $M_{SW_2} = 4.69$, $SD_{SW_2} = 0.67$) auszubalancieren.

Nahezu alle getesteten Probanden (96,7 %) verfügten über ein abgeschlossenes Studium oder über die allgemeine Hochschulreife, wobei die Mehrheit (90 %) dem technischen Bereich, entweder als Studenten oder Berufstätige des Maschinenbaus oder Ingenieurwesens, zuzuordnen sind. Lediglich rund ein Viertel aller Testpersonen verfügte über Erfahrung im Umgang mit Simulationsumgebungen, wobei keiner der Probanden fundierte Kenntnisse im Bereich der Flugsimulation besaß. Im demographischen Fragebogen wurde ebenfalls das Nutzungsverhalten automatisierter Systeme adressiert. Analysiert wurden die Bereiche Haushalt und Versorgung, Automobil, Kommunikation und Multimedia, Gesundheit und Pflege sowie Sicherheit und Privatsphäre. Die Probanden schätzten ihre Nutzungshäufigkeit auf einer 5-stufigen Skala, von nie bzw. nicht vorhanden (1) bis täglich oder fast täglich (5) ein. Die Stichprobe zeichnete sich über eine insgesamt geringe Nutzung automatisierter Systeme aus, da sowohl Modus als auch Median aller Kategorien bei „nie bzw. nicht vorhanden“ und „seltener als einmal pro Woche“ lagen.

3. Ergebnisse

Zunächst kann festgehalten werden, dass alle 15 Probanden der zweiten Versuchsbedingung den Ausfall des Autopiloten erkannten und das System ordnungsgemäß reaktivierten. Die Befragung im Anschluss des Versuches, die eine Frage nach der vermuteten Fehlerhaftigkeit des Systems beinhaltete, stützte diese Beobachtung. Hiernach bewerteten 93,3 % der Testpersonen, welche die erste Bedingung absolvierten, das Verhalten des Autopiloten als nicht fehlerhaft, während die restlichen 6,7 % angaben, sie „wüssten nicht“ ob der Autopilot fehlerhaft gearbeitet hätte. Die Probanden der Bedingung mit fehlerhafter Funktionsweise des Autopiloten gaben mehrheitlich (60 %) an, eine Fehlerhaftigkeit bemerkt zu haben,

während die restlichen Probanden angaben, es nicht zu wissen (20 %) bzw. keine fehlerhafte Funktionsweise festgestellt zu haben (20 %). Die Probanden, die angaben, eine Fehlerhaftigkeit bemerkt zu haben, waren der Meinung, dass das System in durchschnittlich 18,3 % der Gesamtzeit fehlerhaft arbeitete.

Die Berechnung des Zusammenhangs zwischen Vertrauen und mentaler Beanspruchung, abhängig von der Reliabilitätsbedingung des Autopiloten, erfolgte durch eine einseitige Produkt-Moment-Korrelation nach Pearson. Zusätzlich wurde mittels eines t-Tests für unabhängige Stichproben untersucht, ob sich die Mittelwerte der Vertrauenseinschätzung in den beiden Versuchsbedingungen signifikant voneinander unterscheiden. Die teststatistischen Voraussetzungen einer bivariaten Korrelation bzw. eines t-Tests, namentlich Normalverteilung der Variablen und Homoskedastizität, wurden überprüft und können als erfüllt betrachtet werden.

Der Mittelwert des Vertrauensfragebogens ($1 \leq M \leq 7$) der ersten Gruppe lag bei $M_{\text{Vertrauen}_1} = 5.48$ ($SD_{\text{Vertrauen}_1} = 0.87$), während die zweite Versuchsgruppe einen Mittelwert von $M_{\text{Vertrauen}_2} = 4.49$ ($SD_{\text{Vertrauen}_2} = 1.21$) erreichte. Der t-Test ergab einen signifikanten Vertrauensunterschied in beiden Reliabilitätsausprägungen des Autopiloten ($t(28) = 2.595$, $p = .015$). Die Analyse des Gesamtmittelwertes des Vertrauensfragebogens mit dem Mittelwert aller sechs Subskalen des NASA-TLX, die mit gleicher Gewichtung in den Gesamtwert eingingen, führte zu einem Korrelationskoeffizienten von $r = -.002$ ($p = .497$) in der ersten Versuchsgruppe und einem Wert von $r = -.135$ ($p = .315$) in der zweiten Bedingung. Daraufhin erfolgte eine erneute Analyse des Vertrauensmittelwertes mit den einzelnen Subskalen des NASA-TLX, dessen Ergebnisse in Tabelle 1 dargestellt sind.

Tabelle 1: Korrelation des Vertrauensfragebogens mit den einzelnen Subskalen des NASA-Task Load Index zur Messung der subjektiven Beanspruchung der Probanden.

Subskala des NASA-Task Load Index		Geistige Anforderungen	Körperliche Anforderungen	Zeitliche Anforderungen	Leistung	Anstrengung	Frustration
Trust-Fragebogen	Gruppe 1	$r = .272$ ($p = .163$)	$r = -.120$ ($p = .335$)	$r = .199$ ($p = .238$)	$r = -.425$ ($p = .057$)	$r = .036$ ($p = .449$)	$r = .018$ ($p = .475$)
	Gruppe 2	$r = .279$ ($p = .157$)	$r = -.386$ ($p = .077$)	$r = -.161$ ($p = .283$)	$r = -.269$ ($p = .166$)	$r = -.181$ ($p = .259$)	$r = -.243$ ($p = .191$)

4. Diskussion

Auf Basis der Ergebnisse kann festgehalten werden, dass die Probanden dem automatisierten System, das durchgängig zu 100 % zuverlässig funktionierte, ein signifikant höheres Ausmaß an Vertrauen schenkten als dem Autopiloten, der Funktionsausfälle aufwies. Dies ist insofern bemerkenswert, da die Versuchspersonen keinerlei Erfahrung mit dem zu interagierenden System hatten; eine erstmalige Nutzung des Systems für 15 Minuten erwies sich also für die Ausbildung von Vertrauensunterschieden als hinreichend. Ein signifikanter Zusammenhang zwischen mentaler Beanspruchung und Vertrauen wurde für keine der beiden Versuchsbedingungen konstatiert. Dennoch soll an dieser Stelle angeführt werden, dass zumindest fünf der sechs Subskalen der zweiten Bedingung einen negativen Zusammenhang mit der subjektiven Einschätzung von Vertrauen aufwiesen. Dies befürwortet die anfangs aufgestellte Annahme, dass ein erhöhtes

Vertrauen mit einer verminderten mentalen Beanspruchung einhergeht, auch wenn für diese Auslegung keine statistische Signifikanz nachgewiesen werden konnte.

Wie Hoff & Bashir (2014) ausführten, bildet die Neuheit einer Situation eine wichtige Determinante des Ausmaßes an Vertrauen. Dieser Befund erklärt die verhaltene Einschätzung von lediglich 60 % Fehlerhaftigkeit bei dem System in der zweiten Bedingung, da die Stichprobe überwiegend aus unerfahrenen Nutzern im Bereich Flugführung und -simulation bestand. Die Legalität der Einschätzung des Zeitanteils der Autopiloten-Fehlerhaftigkeit, den die Probanden im Abschlussfragebogen angeben sollten, kann nicht final geklärt werden. Die Untersuchung weist darauf hin, dass die Probanden das Ausmaß der Reliabilität unterschätzten, da angegeben wurde, dass der Autopilot in rund 18 % der Gesamtzeit fehlerhaft gearbeitet hat. Ursächlich wurde eine Fehlerhaftigkeit von 40 % angestrebt, da die Probanden den Autopiloten jedoch nach kurzer Zeit wieder aktivierten und er somit nicht den geplanten längeren Zeitraum ausfiel, ist ein Rückschluss auf eine Unterschätzung des Wertes nicht möglich.

Die vorliegende Studie beschäftigte sich vor allem mit subjektiven Erhebungsmethoden von Vertrauen und mentaler Beanspruchung. Zukünftige Studien sollen den Zusammenhang zwischen Vertrauen und mentaler Beanspruchung durch die Nutzung psychophysiologischer Maße adressieren.

5. Literatur

- Beier G (2004) Kontrollüberzeugungen im Umgang mit Technik: Ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme. Band 871. Dissertation.de, Berlin.
- Collani G, Schyns B (2014) Generalisierte Selbstwirksamkeitserwartung. In: Danner D, Glöckner-Rist A (Hrsg) Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS), S 1–9.
- Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP (2003) The Role of Trust in Automation Reliance. *International Journal of Human-Computer Studies* 58:697–718.
- Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock PA, Meshkati N (Hrsg) *Human Mental Workload*. North Holland, Amsterdam, S 139–183.
- Hoff KA, Bashir M (2014) Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*:2–28.
- International Test Commission (2005) *International Guidelines on Test Adaptation*. www.intestcom.org.
- Jian J, Bisantz A, Drury C (2000) Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4(1):53–71.
- Lee JD, See KA (2004) Trust in Automation: Designing for Appropriate Reliance. *Human Factors*(46):50–80.
- Manzey D (1998) Psychophysiologie mentaler Beanspruchung. In: Rösler F (Hrsg) *Ergebnisse & Anwendungen der Psychophysiologie: Enzyklopädie d. Psychologie*. Hogrefe, Göttingen, S 799–864.
- Manzey D (2012) Systemgestaltung und Automatisierung. In: Badke-Schaub P, Hofinger G, Lauche K (Hrsg) *Human Factors: Psych. sicheren Handelns in Risikobranchen*. Springer, Berlin, S 333–352.
- Moray N, Inagaki T, Itoh M (2000) Adaptive Automation, Trust and Self-Confidence in Fault Management of Time-Critical Tasks. *Journal of Experimental Psychology: Applied* 6(1):44–58.
- Parasuraman R, Riley V (1997) Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39(2):230–253.
- Safar JA, Turner CW (2005) Validation of a Two Factor Structure for System Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 49(3):497–501.
- Sarter NB, Woods DD, Billings CE (1997) Automation Surprises. In: Salvendy G (Hrsg) *Handbook of Human Factors & Ergonomics*. Wiley, New York, S 1926–1943.
- Sheridan TB, Parasuraman R (2006) Human-Automation Interaction. *Reviews of Human Factors and Ergonomics*(1):89–129.
- Spain RD, Bustamante EA, Bliss JP (2008) Towards an Empirically Developed Scale for System Trust: Take Two. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52(19):1335–1339.