Ich bin dann mal raus. Erklärbares Automationsverhalten und Vertrauen

Lorenz Prasch[1], Stefan Tretter[2]

Lehrstuhl für Ergonomie, Technische Universität München[1]
Department Psychologie, Ludwig-Maximilians-Universität München[2]

Zusammenfassung

Autonomes Fahren, die Fortbewegung von Fahrzeugen ohne Eingriff des Menschen, ist derzeit eine der meistbeachteten Entwicklungen innerhalb der Automobil-Branche. Allerdings sind die vorhandenen Technologien noch stark limitiert und keinesfalls fehlerfrei, weshalb nicht selten eine Übernahme durch den Fahrer von Nöten ist. Es wird angenommen, dass die Erklärbarkeit der Ursache für eine Übernahmeaufforderung durch das Fahrzeug Einfluss auf das Vertrauen gegenüber dem System hat. Eine derartige Aufforderung sollte das Vertrauen umso weniger verletzen, je mehr ein Fahrer in der Lage ist, sie für sich zu erklären. In einer Online-Studie beobachteten 36 Teilnehmer das Video einer Fahrsimulation mit unterschiedlich offensichtlichen Gründen für die Übernahmeaufforderung. Es zeigten sich die vermuteten Tendenzen. Der vorliegende Beitrag ergänzt bisherige Ansätze durch den Fokus auf psychologische Bedürfnisse und Prozesse, die die grundlegende Bereitschaft zur Anwendung betreffen.

1 Einleitung

Im September 2015 gerieten Interviewaussagen des damaligen Porsche-Vorstands Matthias Müller in die Schlagzeilen, in denen er das autonome Fahren als „Hype“ abtat, „der durch nichts zu rechtfertigen ist“ (Auto Motor und Sport, 2015). Damit konterkariert er Bestrebungen sämtlicher großer Automobilkonzerne und führender Technologiefirmen wie Google, Apple und Microsoft, die dem Thema größte Aufmerksamkeit widmen (CB Insights, 2016), sowie der amerikanischen Regierung, die bekannt gab 4 Milliarden Dollar investieren zu wollen (NHTSA, 2016). Weiter frage er sich, „wie ein Programmierer mit seiner Arbeit entscheiden können soll, ob ein autonom fahrendes Auto im Zweifelsfall nach rechts in den Lkw schießt oder nach links in einen Kleinwagen.“ Solche Fragen der Ethik werden zwar eine große Rolle spielen, das autonome Fahren der Gegenwart muss zunächst jedoch im reibungslosen und vor allem sicheren Zusammenspiel zwischen System und Fahrer geschehen. Die wahrgenommene Verlässlichkeit und das resultierende Vertrauen potenzieller Nutzer wird essenziell für die tatsächliche Anwendung hochautomatisierten Fahrens im Straßenverkehr der Zukunft sein.

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
S. Franken, U. Schroeder, T. Kuhlen (Hrsg.):
Mensch und Computer 2016 – Kurzbeiträge, 4. - 7. September 2016, Aachen.

http://dx.doi.org/10.18420/muc2016-mci-0270

2 Theoretischer Hintergrund

Ein Zwischenschritt auf dem Weg hin zu vollautomatisiertem Fahren ist der flexible Wechsel verschiedener Verantwortungsbereiche zwischen Fahrer und Fahrzeug. Momentan unterliegen entsprechende Systeme noch unterschiedlichen Limitationen. Im regulären Fall sind Prototypen bereits zuverlässig in der Lage, sich in den Verkehrsfluss zu integrieren. Treten allerdings Komplikationen auf, löst das System eine Übernahmeaufforderung (=Take-Over-Request; kurz: TOR) an den Fahrer aus. Verschiedenste Arten der Rückmeldung sind bereits Gegenstand der Forschung (vgl. Wickens & Xu, 2002). Die Gründe für solch eine Übernahmeaufforderung können auf System- wie auch Wahrnehmungsebene sehr unterschiedlich sein. Während beispielsweise eine nahende Baustelle oder fehlende Spurmarkierungen noch eine für den Fahrer recht offensichtliche Ursache sein dürften, sind schlechte GPS-Abdeckung oder unzureichendes Kartenmaterial schon nahezu unmöglich wahrzunehmen oder werden gar als Systemfehler interpretiert. Ziel muss es sein, dass Fortschritte in Fahrsicherheits- und Assistenzsystemen letztendlich auch vom Anwender genutzt werden. Einer der wichtigsten Einflussfaktoren auf die Bereitschaft zur Nutzung ist hierbei das Vertrauen des Menschen in die jeweilige Technik (vgl. Choi & Ji, 2015). Ausgelöste TORs stellen einen entscheidenden Moment für die Entwicklung des Vertrauens in das System dar (Gold et al., 2015; Wickens & Xu, 2002).

Der Mensch verspürt ein grundlegendes Bedürfnis nach Sicherheit (z.B. Sheldon et al., 2001). Um dieses zu erfüllen, tendiert er dazu, unangenehme Ereignisse nachträglich erklärbar zu machen. Dies wird auch als *retrospektive Kontrolle* bezeichnet (vgl. Thompson, 1981). Die unerwartete Aufforderung zur Übernahme während des hochautomatisierten Fahrens stellt eine Gefährdung für das Sicherheitsbedürfnis dar, weshalb dementsprechend im Nachtrag ihres Auftretens versucht wird, dem Ereignis einen Sinn zuzuschreiben – und durch die Erklärbarkeit Kontrolle wahrzunehmen. Nun können die Ursachen für Übernahmeaufforderungen unterschiedlich offensichtlich für den Fahrer sein (z.B. fehlende Fahrbahnmarkierung vs. GPS-Daten), bzw. können Erwartungen an das System und welche Situationen es meistern sollte von Person zu Person verschieden sein. Im Kontext der Kontrolltheorie bedeutet das, dass die Kritikalität der TOR – und damit ihr Einfluss auf das Systemvertrauen - vom Nutzer je nach Ursache unterschiedlich wahrgenommen wird. Folglich sind die Offensichtlichkeit der Situation sowie die Erwartungen an die Systemfähigkeiten für das Kontrollerleben entscheidend. Daher sollte nicht nur die Art der Rückmeldung, sondern auch der Grund für den Ausfall des Systems berücksichtigt werden. Je nach Ursache sollte eine explizite Rückmeldung für den Grund der Übernahmeaufforderung unterschiedlich wichtig sein und damit unterschiedlich große Unterstützung bei der Zuschreibung benötigen. In einer Online-Studie soll diesem Zusammenhang und dem vermuteten psychologischen Prozess auf den Grund gegangen werden.

3 Methode & Ergebnisse

In einer Online-Umfrage wurden Probanden Videos drei verschiedener Fahrsituationen mit einer Übernahmeaufforderung aus der Ego-Perspektive präsentiert. TOR A war eine Übernahmeaufforderung wegen fehlenden Kartenmaterials, bei TOR B fehlten die Spurmarkie-

rungen ab einem gewissen Zeitpunkt, bei TOR C wird die Spur aufgrund einer Baustelle verengt und auf die Gegenfahrbahn geleitet. Die Videos mit einer Gesamtlänge von 14-29 Sekunden wurden in die Umfrage eingebettet und in einer Auflösung von 680 × 400 Pixeln abgespielt. Während jedes Videos wurde zum Zeitpunkt einer theoretischen Übernahme (10 Sekunden vor der Unregelmäßigkeit im Fahrbahnablauf) ein scharfer doppelter Sinuston abgespielt und ein blinkendes Hands-On Icon in der unteren Bildmitte dargestellt. Im Anschluss an das erste Video nahmen die Probanden ein Single-Item Vertrauensrating sowie ein Akzeptanzrating nach Van der Laan (1997) vor. Erfasst wurden Vertrauen. Akzeptanz, wahrgenommene Offensichtlichkeit, der Wunsch nach Erklärung sowie vermutete Ursachen der TOR.

Insgesamt liegen zum Zeitpunkt der Einreichung 36 vollständige Datensätze (16 weiblich) von Probanden im Alter von 18-51 Jahren (M=25,6; SD=6,3) vor. Die offene Frage nach dem Grund der TOR zeigt deutliche Unterschiede zwischen den drei Situationen. Während bei TOR A kein einziger Proband selbstständig den korrekten Grund für die Systemgrenze erkannte, fanden bei TOR B bereits 35% der Probanden den korrekten Grund heraus. Bei TOR C gelang es 78% der Probanden, den korrekten Grund (oder einen Teilgrund) für den Systemausfall zu benennen. Die Fragen nach wahrgenommener Offensichtlichkeit lassen sich in der Frage, ob die TOR als Systemfehler verstanden wurden, leicht veranschaulichen. Wie in Abbildung 1 (links) dargestellt, gibt es erhebliche Differenzen bei unterschiedlicher Offensichtlichkeit.

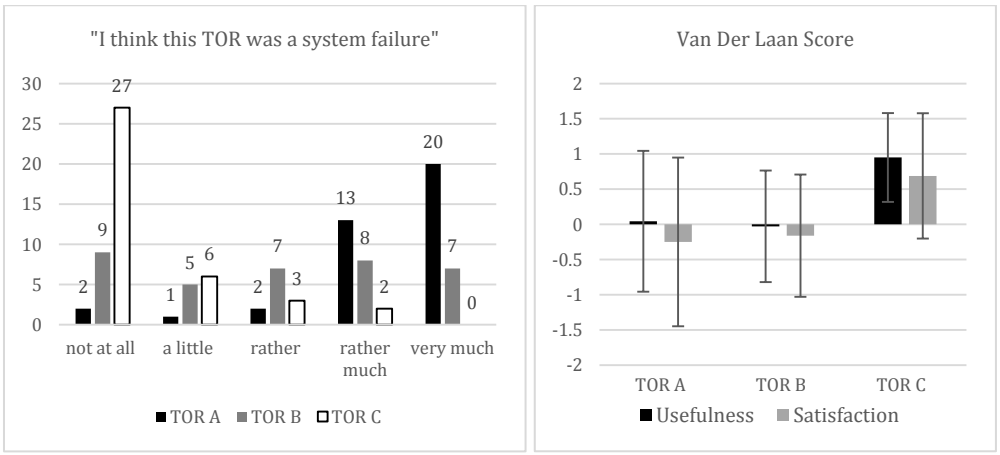

Abbildung 1: Anzahl der jeweiligen Antworten auf die Frage, ob die vorangegangene Übernahme eine Fehlfunktion des Systems gewesen sei (links) sowie wahrgenommene Nützlichkeit und Zufriedenheit der Nutzer mit dem präsentierten System (rechts).

Die Ergebnisse des Vertrauensratings sowie des Van der Laan Tests zeigen deutliche Tendenzen, zur Bestätigung der Hypothesen. Das Vertrauen in das System auf einer Skala von 0–100 ist bei TOR C (M_C=60; SD_C=25.07) deutlich höher, als bei den beiden anderen Varianten (M_A=36.36, SD_A=27.05; M_B=38.14, SD_B=22.89). Ein statistischer Mittelwert-Vergleich zwischen den aggregierten Daten von TOR A+B und TOR C (Gruppeneinteilung nach Van der Laan Score) hinsichtlich der Vertrauensratings zeigt einen signifikanten Unterschied (M_{A+B}=37.25, SD_{A+B}=24.61; M_C=60, SD_C=25.07; t=-2.3, p<0.05). Ebenso zeigen sich deutli-

che Unterschiede im Van der Laan Score der drei Situationen. Auch hier hebt sich TOR C deutlich von den beiden Alternativen ab (vgl. Abbildung 1 rechts).

4 Diskussion & Ausblick

Mit Blick auf die derzeitigen Ergebnisse kann festgehalten werden, dass die Tendenzen klar in Richtung des vermuteten Zusammenhangs gehen. Im Anschluss an eine Übernahmeaufforderung vertrauen diejenigen Fahrer dem Assistenzsystem mehr, die sich einen plausiblen Reim auf die Ursache machen konnten. Wir argumentieren, dass die Offensichtlichkeit dieses Übernahmegrundes das Ausmaß bestimmt, in dem der Fahrer sich aufgrund der Erklärbarkeit retrospektiv Kontrolle über die Situation verschaffen kann. Diese Tendenz ist Ausdruck psychologischer Bedürfnisse die es zu adressieren gilt, will man die Fortschritte im hochautomatisierten Fahren eines Tages flächendeckend auf den Straßen umgesetzt sehen. Der vorliegende Ansatz stellt durch den Einbezug psychologischer Bedürfnisse und deren Erfüllung ein Novum im Kontext hochautomatisierten Fahrens dar. Insgesamt verdeutlicht dieser Beitrag, dass es nicht nur objektive, sondern auch subjektive Kriterien zu beachten gilt, will man einen nachhaltigen Erfolgszug des hochautomatisierten Fahrens gewährleisten und verhindern, dass es letztendlich nur ein „Hype“ bleibt.

Literaturverzeichnis

Auto Motor und Sport (2015). Porsche-Chef Müller im Interview: Hybrid - die neue Porsche-Strategie? Verfügbar unter http://www.auto-motor-und-sport.de/news/porsche-chef-mueller-interview-hybrid-strategie-9968044.html

CB Insights (2016). *30 Corporations Working On Autonomous Vehicles.* Verfügbar unter https://www.cbinsights.com/blog/autonomous-driverless-vehicles-corporations-list/

Choi, J. K. & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction, 31*(10), 692–702.

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation–Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing, 3*, 3025-3032.

NHTSA National Highway Traffic Safety Administration (2016). Verfügbar unter http://www.nhtsa.gov/About+NHTSA/Press+Releases/dot-initiatives-accelerating-vehicle-safety-innovations-01142016

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology, 80*(2), 325-339.

Thompson, S.C. (1981). Will it hurt less if I can control it? A complete answer to a simple question. *Psychological Bulletin, 90*(1), 89-101.

Van Der Laan, Jinke D., Adriaan Heino, and Dick De Waard. "A simple procedure for the assessment of acceptance of advanced transport telematics." *Transportation Research Part C: Emerging Technologies 5.1* (1997): 1-10.

Wickens, C. & Xu, X. (2002). *Automation trust, reliability and attention HMI 02-03 (Technical Report No. AHFD-02-14/MAAD-02-2)*. Savoy, IL: University of Illinois, Aviation Research Lab.