

## **Arbeiten in komplexen Mensch-Automations-Systemen: Das unheimliche und unsichere Tal der Automation, erste Skizze am Beispiel der Fahrzeugautomatisierung**

Frank FLEMISCH<sup>1,2</sup>, Eugen ALTENDORF<sup>1</sup>, Yigiterkut CANPOLAT<sup>1</sup>,  
Gina WEßEL<sup>1</sup>, Marcel BALTZER<sup>2</sup>, Claudia RUDOLPH<sup>2</sup>, Daniel LOPEZ<sup>2</sup>,  
Gudrun VOß<sup>3</sup>, Maximilian SCHWALM<sup>3</sup>

<sup>1</sup> IAW Institut für Arbeitswissenschaft, RWTH Aachen

<sup>2</sup> FKIE Fraunhofer Institut für Kommunikation, Informationsverarbeitung  
und Ergonomie, Fraunhoferstr. 20, 53343 Wachtberg

<sup>3</sup> IKA Institut für Kraftfahrzeuge, RWTH Aachen

**Kurzfassung:** Digitalisierung und Vernetzung ermöglichen immer leistungsfähigere Assistenz- und Automationssysteme z.B. in der Luftfahrt, in der Produktion oder für Automobile, die allerdings auch Nachteile z.B. im Zusammenspiel mit dem Menschen mit sich bringen können. In der Robotik ist ein sog. Uncanny Valley, ein unheimliches Tal bekannt, welches beschreibt, dass Roboter mit einer hohen, aber nicht perfekten Ähnlichkeit und Fähigkeit verglichen z.B. mit dem Menschen als unheimlich und unsicher wahrgenommen werden. Für die Fahrzeugautomation deutet sich ein vergleichbarer metaphorischer Designzusammenhang an, bei dem es zwischen teil- und hochautomatisierten Automationsgraden eine Zone gibt, bei denen es zu Fehlwahrnehmungen und Sicherheits-einbußen kommen kann. Der Beitrag skizziert den Zusammenhang, fasst erste bestätigende Studien zusammen und entwirft den Gestaltungsraum.

**Schlüsselwörter:** Automation, Assistenz, Robotik, Mensch-Maschine System, Uncanny / Unsafe Valley

### **1. Einführung: Automatisierung und Robotik**

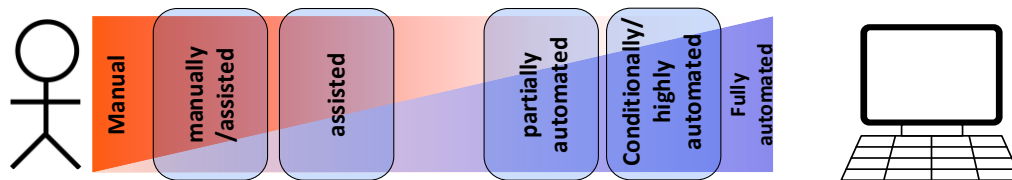
Ermöglicht durch technische Fortschritte in der Digitalisierung und Vernetzung, sowie motiviert durch Kostendruck und immer höhere Leistungsanforderungen, wuchs die Komplexität von Mensch-Technik-Systemen in den letzten Jahrzehnten stetig (z.B. Hollnagel 2007). Ein Teil dieser Komplexität kann durch Assistenzsysteme und Automation aufgefangen werden, allerdings kann es unerwünschte Nebeneffekte wie „Operator / pilot out of the loop“ oder „Mode confusion“ geben (z.B. Billings 1997). Manchmal wird etwas undifferenziert von Überautomation gesprochen, allerdings weist bereits Norman (1990) auf die Notwendigkeit einer differenzierteren Betrachtung hin, die auch die Interaktion und eine besser angepasste und kooperativere Automation in den Fokus rückt.

Robotik als eine spezifische Form von Automation kennt den Sachverhalt eines Uncanny Valley der Robotik, welches beschreibt, dass Roboter mit einer hohen, aber nicht perfekten, Ähnlichkeit und Fähigkeit verglichen z.B. mit dem Menschen als unheimlich und unsicher wahrgenommen werden (Mori et al. 1970/2012). Darüber hinaus bewegt sich Forschung von einer vollautomatisierten Robotik in Richtung einer hochautomatisierten und kooperativen Robotik, bei der Mensch und Roboter

z.T. in gleichen Arbeitsräumen zusammenarbeiten (z.B. Mayer 2012).

Eine vergleichbare Entwicklung in Richtung kooperativer Assistenz und Automation läuft nun ausgehend von der Luftfahrt auch im Bereich von Bodenfahrzeugen ab. Dabei wurde, ausgelöst durch Grundlagenkonzepte wie Levels of Automation, zunehmend klar, dass Assistenz und Automationssysteme, die zusammen wirken, zusammen betrachtet werden sollten und auf einer Skala bzw. einem Spektrum abgebildet werden können (z.B. Flemisch et al. 2008). Dies ist eine Betrachtungsweise, die später in Überlegungen zur Absicherung und Normierung von Fahrzeugautomation übernommen wurde (z.B. Gasser et al. 2012 – für die BASt oder SAE 2014) und in denen z.B. zwischen teil- und hochautomatisierten Systemen unterschieden wird. So zeigt Abbildung 1 eine Verteilung der Kontrolle zwischen Mensch und Automation in den Assistenz- und Automationsgraden manuell, assistiert, teil-, hoch- und vollautomatisiert bzw. autonom.

Eine für ein mögliches, unsicheres Tal der Automation wesentliche Region befindet sich dabei in der rechten Hälfte der Skala, bei der es zwischen teil- und hochautomatisierten Modi eine Zone geben könnte, die eher unheimlich bzw. unsicher sein könnte.



**Abbildung 1:** Kontrollverteilung zwischen Mensch- und Automation als Assistenz- und Automationskala, hier mit expliziten Automationsgraden / Modi (inspiriert durch Sheridan 1980, Flemisch et al. 2003ff, Gasser et al. 2012, SAE 2014).

## 2. Erste Hinweise auf die Existenz eines unsicheren Tales

Erste systematische Erkundungen auf dem Gebiet zwischen teil- und hochautomatisiert erfolgten für Landfahrzeuge ab 2001 (NASA-H-Mode) bzw. ab 2004 im Rahmen der DFG-H(-orse)-Mode-Projekte. Diese stützen sich auf die Design-Metapher Reiter-Pferd (Flemisch et al. 2003/2015; Bengler et al. 2011; Altendorf et al. 2015) und führten, flankiert von einer Reihe von nationalen und EU-Projekten (z.B. Hoeger et al. 2011), zum teil- und hochautomatisierten Fahren. Dabei war anfangs noch unklar, ob ein oder mehrere Modi zwischen assistiert und vollautomatisiert Sinn machen und wie sie beschaffen sein sollen, insbesondere bzgl. des Involvierungsgrades des Fahrers, der Stärke des Eingriffs der Automation und der nötigen Absicherungsmaßnahmen z.B. durch eine Fahrerbeobachtung.

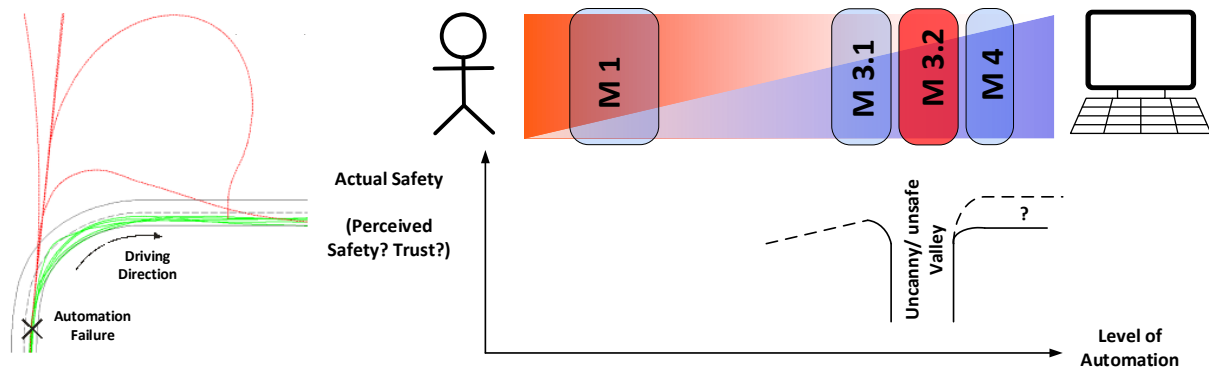
So wurde in einem Simulatorexperiment des DLR gezeigt, dass beim teilautomatisierten Fahren, bei dem die Fahrer noch Lenkmoment beisteuern müssen, Ausfälle noch aufgefangen werden können, während bei einer Variante, bei der die Fahrer, die Hände vom Lenkrad nehmen konnten, Ausfälle nicht mehr aufgefangen werden konnten (Schieben & Flemisch 2008). In einer Reihe weiterer Experimente zeigten z.B. Damböck et al. (2013) sowie Schwalm (2015) und Voß und Schwalm (2015), dass stark automatisierte Versionen zu einer Schwächung der Übernahmefähigkeiten führen können. Nachgewiesene Zusammenhänge bestehen hier z.B. in der fehlenden kompensatorischen Reaktion bei der Reduktion nicht-

fahrbezogener Aufgaben in Vorbereitung einer Übernahmesituation (Voß & Schwalm 2015; Schwalm et al. 2015).

Im Gegensatz dazu zeigten eine Reihe weiterer Experimente z.B. Schieben et al. (2008) und Petermann und Schlag (2009), dass Varianten mit einer Unterscheidung zwischen teil- und hochautomatisiert und mit gut gestalteten Transitionen gut funktionieren können, und Rauch et al. (2010) demonstrierte, dass eine Fahrerzustandserkennung wesentliche Beiträge zur Sicherheit und Akzeptanz leisten kann. Offenbar gibt es jedoch zwischen gut funktionierenden Modi und Varianten Bereiche, die deutlich unsicherer sind.

### 3. Skizze des Effektraums: Was passiert in einem unsicheren Tal?

Was sind die wesentlichen Effekte, die ein solches Tal bestimmen? Das Beispiel in Abbildung 2 (links) deutet bereits an, dass es einen Zusammenhang zwischen der Kontrollverteilung zwischen Mensch und Automation, die sich in Modi oder Levels of Automation ausdrücken kann, und einer Leistungsgröße wie Performanz oder Sicherheit gibt. So zeigt der rechte Teil von Abbildung 2 den im linken Teil verdeutlichten „Absturz“ anhand der Kontrollverteilung und einer geschätzten Sicherheit: Während ein Modus niedrigerer Automatisierung, hier M3.1 (z.B. teilautomatisiert / partially automated) noch ausreichende Sicherheit liefert, ist eine höhere Kontrollverteilung (M3.2), offenbar unsicher, während eine noch höhere Automatisierung (M4), bei welcher der Mensch nicht mehr eingreifen muss oder nur mit einer garantierten Eingriffszeit eingreifen kann, bereits wieder sicherer sein könnte.



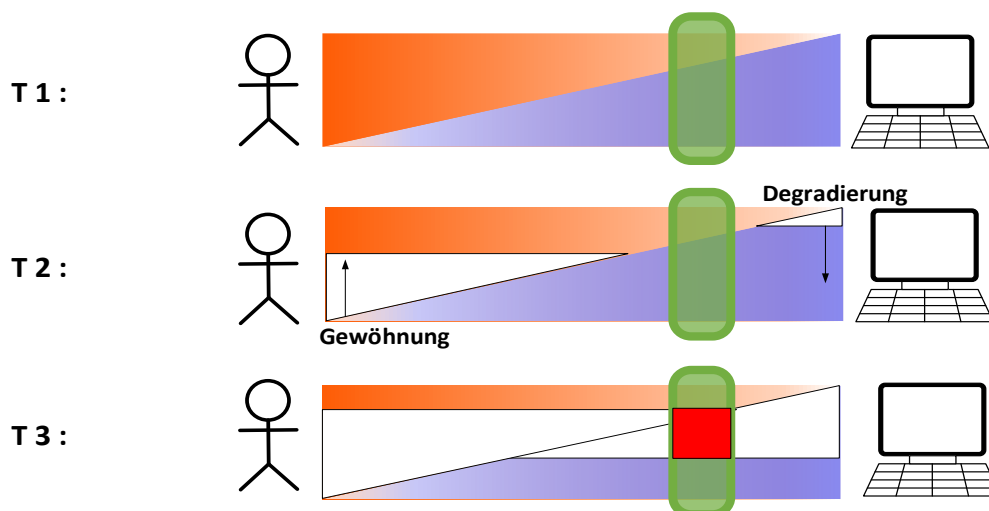
**Abbildung 2:** Links: Beispiel für einen „Absturz“ in ein Uncanny Valley: Abfangen eines Systemausfalls vor einer Kurve (grün / durchgezogen: Systemvariante teilautomatisiert; rot / gestrichelt: Systemvariante hochautomatisiert).  
Rechts: Möglicher Zusammenhang von Kontrollverteilung und Sicherheit.

Ein wesentlicher Zusammenhang, der zu einem Zusammenbruch der Sicherheit führen kann, besteht zwischen der Performanz der Automation und der Performanz des Fahrers im Falle einer Übernahme: Hier deutet sich ein Zusammenhang zwischen der Involviertheit des Fahrers und der Übernahmefähigkeit dahingehend an, dass bei zu niedriger Involviertheit der Fahrer nicht rechtzeitig übernehmen kann. Zu starke und gleichzeitig nicht ausreichend starke Automation kann zu einer rein überwachenden Funktion des Fahrers im Sinne einer „supervisory control“ führen, für die Menschen schlecht vorbereitet sind (z.B. Endsley 1995). Ein hieran angelegter Ansatz von Schwalm et al. (z.B. 2015; vgl. auch Voß & Schwalm 2015) postuliert,

dass Fahrer im Falle zu hoher Automation einen kontinuierlichen Regelprozess verlassen und nicht mehr in der Lage sind, regulative Maßnahmen im Sinne eines funktionalen Situationsmanagements anzuwenden. Dieser regulative Prozess würde ihnen bei voller Involviertheit in die Fahraufgabe nämlich erlauben, die Fahrsituation zu analysieren, zu antizipieren und hierauf basierend die kognitiven Ressourcen angemessen zu verteilen und bereitzustellen.

Ein weiterer wesentlicher Zusammenhang dabei könnte zwischen dem Vertrauen des Menschen auf die Leistungsfähigkeit seines maschinellen Counterparts und der Übernahmefähigkeit sein, wie dies z.B. Manzey und Bahner (2005) skizzieren. Relevant ist hier die Passung zwischen wahrgenommener Erscheinung und der Attribution von Eigenschaften und Fähigkeiten der Automation einerseits, und dem Erleben der tatsächlichen Fähigkeiten des Systems andererseits, was auch emotionale Aspekte beinhalten könnte.

Deutlich wird außerdem, dass die zeitliche Komponente dabei eine wesentliche Rolle spielt. So zeigt Abbildung 3 den zeitlichen Verlauf eines „Absturzes in das unsichere Tal“: Zunächst ist noch alles in Ordnung, da die Fähigkeiten der Automation ausreichen und der Mensch noch übernahmefähig ist (Abbildung 3, T1). Im Laufe der Zeit kann es zu Gewöhnungseffekten kommen (T2), und kommt es nun zu einem Ausfall oder einer Degradierung, muss der Mensch die Lücke schließen, hat dazu aber eine unzureichende Übernahmefähigkeit (T3). Das unsichere Tal ähnelt hier eher einer Gletscherspalte, die zunächst von einer auf den ersten Blick tragfähigen Brücke überspannt wird, die dann aber zusammenbricht und den zu vertrauensseligen Operateur verschlingt. Ein anderer Vergleich ist der mit einer Schere, deren Schenkel aus der Übernahmefähigkeit des Fahrers und der Verfügbarkeit der Automation besteht, sodass der Fahrer in einem ungünstigen Moment von der Kontrolle seines Fahrzeuges abgeschnitten wird.



**Abbildung 3:** Möglicher zeitlicher Ablauf eines Absturzes ins unsichere Tal

Dieses Phänomen wird in der Literatur auch unter den Begriffen *overtrust / overreliance* (z.B. Inagaki & Itoh 2013; Lee & See 2004) und *automation bias / complacency* (z.B. Bahner 2008; Mosier & Skita 1996) diskutiert. Erst in kritischen Situationen, die die tatsächlichen Systemgrenzen aufzeigen, findet eine Anpassung der Attributionen statt. Dies hat einen starken Vertrauensabfall zur Konsequenz, der

gemäß vorhandener Literatur in den meisten Fällen nur schwer zu kompensieren ist (Hoffman et al. 2013; Lee & See 2004).

Hinzu kommt hier eine zusätzliche Ironie des Automatisierungsfortschrittes: Während eine Automatisierung am Anfang einer technologischen Entwicklung noch Lücken und Fehler hat, auf die der Mensch aufgrund der Häufigkeit mit einer höheren Übernahmbereitschaft reagiert, wird durch steigende Verfügbarkeit der Automation der Erstkontakt mit einer Lücke oder einem Fehler immer weiter in die Zukunft verschoben. Dadurch wird u.U. ein unangemessenes Vertrauen aufgebaut, mit der Folge, dass der gleiche Fehler dann höhere Auswirkungen hat.

#### **4. Skizze des Gestaltungsraums: Lösungsdimensionen zur Absicherung eines unsicheren Tals**

Es besteht eine durch erfolgreiche Implementierungen wie z.B. Hoeger et al. (2012) oder Altendorf et al. (2015) begründete Hoffnung, dass nicht Automation per se oder zu hohe Automatisierung schlecht ist, sondern dass es unsichere Regionen gibt, um die herum es sichere Varianten von Automation und sichere Kombinationen von unterschiedlichen Assistenz- und Automationsgraden gibt. Entscheidend ist offenbar das Wechselspiel aus den jeweiligen Automationsgraden und den möglichen Transitionen. Dabei kann eine Absicherungsmaßnahme des linken Randes des unsicheren Tales sein, dass eine unzureichende Involvierung des Menschen durch eine Überwachung der Aufmerksamkeit festgestellt wird und darauf z.B. durch Aufforderungen reagiert wird, wie dies z.B. Rauch et al. (2010) und Schwalm et al. (2015) beschreiben. Gestaltungsdimensionen gegen einen Absturz von rechts kommend in das unsichere Tal sind einerseits eine hohe technische Zuverlässigkeit und eine Technik, die über ihre eigene Verfügbarkeit Bescheid weiß, andererseits die Transition hin zum Menschen, und wenn dieser nicht übernehmen kann, eine Transition hin zu einem sicheren Zustand (z.B. Hoeger et al. 2012).

Im einfachen Fall kann das unsichere Tal gut strukturiert (klar definierter Abfall und darauf folgender Anstieg) und damit vergleichsweise gut abzusichern sein. U.u. sind die Strukturen hier auch komplexer im Sinne einer Gebirgslandschaft, die erst vernünftig kartographiert werden sollte, bevor die Ränder abgesichert werden und tragfähige Brücken darüber gebaut werden können.

#### **5. Literatur**

- Altendorf E, Baltzer M, Heesen M, Kienle M, Weißgerber T, Flemisch F (2015) H-Mode, a Haptic-Multimodal Interaction Concept for Cooperative Guidance and Control of Partially and Highly Automated Vehicles. In: Winner et al. (Eds) Handbook of Driver Assistance Systems. Springer.
- Bahner JE (2008) Übersteigertes Vertrauen in Automation: Der Einfluss von Fehlererfahrungen auf Complacency und Automation Bias. Dissertation: TU Berlin.
- Bengler K, Flemisch F (2011) Von H-Mode zur kooperativen Fahrzeugführung – Grundlegende ergonomische Fragestellungen. In: 5. Darmstädter Kolloquium: kooperativ oder autonom? 2011. Darmstadt.
- Billings, CE (1997) Aviation automation: The search for a human centered approach. Lawrence Erlbaum Associates.
- Damböck D (2013) Automationseffekte im Fahrzeug – von der Reaktion zur Übernahme; Dissertation, TU München.
- Endsley MR (1995) Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors: The Journal of the Human Factors and Ergonomics Society 37 (1):32–64.

- Flemisch F, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J (2012) Towards a dynamic balance between humans and automation: Authority, Ability, Responsibility and Control in Shared and Cooperative Control Situations. *International Journal Cognition, Technology & Work* 14(1):3-18.
- Flemisch F, Kelsch J, Löper C, Schieben A, Schindler J (2008) Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. *Human Factors for Assistance and Automation* 1-16.
- Flemisch F, Schwalm M, Deml B (2015) Systemergonomie kooperativ interagierende Fahrzeuge. Projektantrag an die DFG.
- Flemisch F, Winner H, Bruder R, Bengler K (2015) Cooperative guidance, control and automation. In: Winner et al. (Eds) *Handbook of Driver Assistance Systems*. Springer.
- Flemisch FO, Adams CA, Conway SR, Goodrich KH, Palmer MT, Schutte PC (2003) The H-Metaphor as a Guideline for Vehicle Automation and Interaction (No. NASA/TM—2003-212672). Hampton: NASA, Langley Research Center.
- Gasser TM, et al. (2012) Rechtsfolgen zunehmender Fahrzeugautomatisierung - Gemeinsamer Schlussbericht der Projektgruppe Bundesanstalt für Straßenwesen (BASt), (F 83).
- Hoeger R, et al. (2011) Deliverable D61.1, Final Report. [http://haveit-eu.org/LH2Uploads/ItemsContent/24/HAVEit\\_212154\\_D61.1\\_Final\\_Report\\_Published.pdf](http://haveit-eu.org/LH2Uploads/ItemsContent/24/HAVEit_212154_D61.1_Final_Report_Published.pdf)
- Hoeger R, Wiethof M, Rheker T (2012) Complexity measures of traffic scenarios: psychological aspects and practical applications. *International Conference on Driver Behaviour and Training 2011*. Paris.
- Hoffman RR, Johnson M, Bradshaw JM, Underbrink A (2013) Trust in automation. *Intelligent Systems, IEEE* 28(1):84-88.
- Hollnagel E (2007) Keynote zur 7. Berliner Werkstatt „Prospektive Gestaltung von Mensch-Technik-Interaktion“ 2007. Berlin.
- Inagaki T, Itoh M (2013) Human's overtrust in and overreliance on Advanced Driver Assistance Systems: a theoretical framework. *International Journal of Vehicular Technology* 2013.
- Lee JD, See KA (2004) Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46(1):50-80.
- Manzey D, Bahner JE (2005) Vertrauen in Automation als Aspekt der Verlässlichkeit von Mensch-Maschine-Systemen. *Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis. Festschrift für Klaus-Peter Timpe* 93-109.
- Mayer MP (2012) Entwicklung eines kognitionsergonomischen Konzepts und eines Simulationssystems für die robotergestützte Montage. Dissertation, Aachen: Shaker Verlag.
- Mori M (1970) The uncanny valley. *Energy* 7(4):33–35 (in Japanese).
- Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE* 19(2):98-100.
- Mosier KL, Skitka LJ (1996) Human decision-makers and automated decision aids: Made for each other? In: R. Parasuraman & M. Mouloua (Hrsg.), *Automation and Human Performance: Theory and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates, 201-220.
- Norman DA (1990) *The Problem With Automation*. *Philosophical Transactions of the Royal Society of London*.
- Petermann I, Schlag B (2009) Auswirkungen der Synthese von Assistenz und Automation auf das Fahrer-Fahrzeug System. Paper presented at the 11. Braunschweiger Symposium Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET) 2011. Braunschweig.
- Rauch N, Kaussner A, Krueger H-P, Boverie S, Flemisch F (2010) Measures and Countermeasures for impaired driver's state within highly automated driving. Brussels: Transport Research Arena.
- SAE J 3016:2014 Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Society of Automotive Engineers.
- Schieben A, Flemisch F (2008) Who is in control? Exploration of transitions of control between driver and an eLane vehicle automation. *VDI/VW Tagung Fahrer im 21. Jahrhundert 2008*. Wolfsburg.
- Schieben A, Damböck D, Kelsch J, Rausch H, Flemisch F (2008) Haptisches feedback im spektrum von fahrerassistenz und automation. In *Proceedings of* 3:7-8.
- Schwalm M, Ladwig S (2015) How do we solve demanding situations – a discussion on driver skills and abilities. 57th Conference of Experimental Psychologists 2015. Hildesheim.
- Schwalm M, Voß GMI, Ladwig S (2015) Inverting Traditional Views on Human Task-Processing Behavior by Focusing on Abilities Instead of Disabilities—A Discussion on the Functional Situation Management of Drivers to Solve Demanding Situations. In: *Engineering Psychology and Cognitive Ergonomics*. Springer International Publishing, 286-296.
- Sheridan TB (1980) Computer Control and Human Alienation. *Technology review* 83(1):65–73.

Sheridan TB (1992) Telerobotics, Automation, and Human Supervisory Control. MIT press.  
Voß GMI, Schwalm M (2015) 1. Kongress der Fachgruppe Verkehrspsychologie 2015. Braunschweig.