

Soziotechnische Systeme als Herausforderung der Funktionalen Sicherheit

Joachim Draeger

Zusammenfassung

An die funktionale Sicherheit soziotechnischer Systeme werden hohe Anforderungen gestellt. Um trotz der zunehmenden Komplexität dieser Systeme verlässliche Aussagen über bestehende Risiken machen zu können, ist man immer stärker auf den Einsatz formaler Methoden angewiesen. In diesem Beitrag wird der bisher noch wenig untersuchte Aspekt der formalen Validierung soziotechnischer Systeme näher beleuchtet, also der Nachweis, dass sich das zur Bestimmung der Risiken genutzte Modell und das reale soziotechnische System tatsächlich genügend ähnlich sind. Es zeigt sich, dass Verfahren der formalen Validierung prinzipiell auch auf soziotechnische Systeme anwendbar sind. Allerdings müssen dabei die Besonderheiten soziotechnischer Systeme beachtet werden. Die individuellen Unterschiede bei der Perzeption von Risiken durch Menschen können jedoch unter Umständen eine Modifikation des Risikobegriffs erfordern.

1 Einleitung

Die Methoden der funktionalen Sicherheit (Spellmann & Whiting, 2009) dienen dazu, etwaige Risiken bei der Verwendung eines technischen Systems S zu identifizieren. Dabei gilt üblicherweise die Regel, dass die Beherrschbarkeit des Systems S umso größeren Herausforderungen unterliegt, je höher seine Komplexität ist (z.B. Navlaka, 1986). Der Begriff der „Komplexität“ ist in diesem Zusammenhang anschaulich im Sinne einer abnehmenden Einsicht und Durchschaubarkeit zu verstehen. Bezieht man den Menschen mit seiner enormen Komplexität in S mit ein, d.h. wird statt eines technischen ein soziotechnisches System (Mate & Silva, 2005) betrachtet, ist man hinsichtlich der Untersuchung von S aus der Perspektive der funktionalen Sicherheit mit gänzlich neuen Herausforderungen konfrontiert.

Eine dieser Herausforderungen ist sicherlich die Diskrepanz zwischen der gewaltigen Komplexität eines als Komponente von S betrachteten realen Menschen einerseits und seiner notwendigerweise stark abstrahierten Repräsentation in einem Modell M von S andererseits. Zudem resultieren aus fehlenden Kenntnissen über das spezifische Verhalten einzelner Individuen zahlreiche Unsicherheiten. Damit stellt sich die Frage, inwieweit M und S tatsäch-

lich miteinander korrespondieren. Für die funktionale Sicherheit ist eine solche Korrespondenz jedoch essentiell, da ohne sie die Übertragbarkeit des auf der Basis von M berechneten Risikos auf das reale System S fraglich ist. Man wird daher bestrebt sein, die Korrespondenz zwischen M und S im Rahmen einer Validierung (Sargent, 2011) des Modells M bezüglich S nachzuweisen.

Die Validierung von Modellen soziotechnischer Systeme mit Hilfe nichtformaler Methoden ist mehrfach untersucht worden. Dies betrifft insbesondere die Validierung basierend auf Expertenmeinungen. Dabei wurden gemäß Goerger (2004) und Harmon et al. (2002) jedoch häufig Defizite hinsichtlich der Qualität der Validierung festgestellt. Die mangelnde Eignung semiformaler Methoden oder auf reiner Anschauung basierender Argumentationen scheint klar, da beide in gewissen Sinne auf einen unmittelbaren Einblick in das System und seine interne Funktion angewiesen sind. Diese Voraussetzung ist für soziotechnische Systeme angesichts ihrer Komplexität jedoch nicht gegeben. Dementsprechend findet sich oft ein ausgeprägter Bias in den Expertenmeinungen.

Die Probleme informeller Methoden legen die Verwendung formaler Methoden bei der Behandlung soziotechnischer Systeme und ihrer Modelle nahe (Cebulla, 2002). Entsprechende Untersuchungen zur Validierung vernachlässigen aber bisher die Besonderheiten risikobasierter Betrachtungen oder beschränken sich auf spezielle Systemtypen oder Detailprobleme. Beispiele dazu sind Tsalgatiidou & Loucopoulos (1991), Moya et al. (2008) und Nechyba (1998).

Der vorliegende Artikel über die Herausforderungen bei der formalen Validierung risikobasierter Modelle soziotechnischer Systeme soll daher eine Lücke in der Literatur füllen. Er ist wie folgt gegliedert. In Abschnitt 2 werden zunächst die Begriffe des Risikos und des soziotechnischen Systems eingeführt und diskutiert. Abschnitt 3 widmet sich den Voraussetzungen, die an das zu validierende Modell zu stellen sind. Abschnitt 4 untersucht das geeignete Vorgehen bei einer formalen Validierung und Methoden zu ihrer Durchführung. Den Optionen zur Behandlung dabei eventuell auftretender Probleme ist Abschnitt 5 gewidmet. Abschnitt 6 diskutiert die Verwendung der beschriebenen Methodik in der Praxis und stellt einige dafür besonders gut geeignete Anwendungsgebiete vor. Der Artikel schließt mit einer kurzen Zusammenfassung und einem Ausblick in Abschnitt 7.

2 Begriffe und Definitionen

Wir führen zunächst die Begriffe des Risikos und des soziotechnischen Systems ein. Sie sind grundlegend für die weitere Arbeit.

2.1 Begriff des Risikos

Die Untersuchung eines Systems mit Methoden der funktionalen Sicherheit erfolgt gemäß Spellmann & Whiting (2009) häufig risikobasiert. Als Risiko $R = \sum_{e \in \mathcal{E}} L(e)P(e)$ wird dabei der Erwartungswert einer Verlustfunktion $L: \mathcal{E} \rightarrow [0, \infty[$ über einen Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) mit Resultatmenge Ω , Ereignismenge \mathcal{E} und Wahrscheinlichkeitsfunktion $P: \mathcal{E} \rightarrow [0, 1]$ verstanden

(Berger, 1985). Das konkrete Vorgehen zur Bestimmung eines Risikos hängt vom jeweils betrachteten System ab. Anmerkungen zum Einfluss der besonderen Eigenschaften soziotechnischer Systeme auf diese Methodik machen Asnar & Giorgini (2008) sowie Anderson & Felici (2009). Zio (2009) diskutiert die Risikoanalyse für komplexe Systeme im Allgemeinen, wobei soziotechnische Systeme ausdrücklich eingeschlossen werden. Da wir uns auf Fragen der Validierung konzentrieren wollen, wird hier auf die Thematik der Risikobestimmung für soziotechnische Systeme nicht näher eingegangen.

2.2 Begriff des soziotechnischen Systems

In einem soziotechnischen System gibt es sowohl menschliche als auch technische Komponenten. Zur Vereinfachung werden für beide Arten von Komponenten wie bei Pizziol (2013) unterschiedliche Rollen im Gesamtsystem angenommen. Die technischen Komponenten realisieren dementsprechend die eigentlichen Funktionalitäten des soziotechnischen Systems und sind die unmittelbare Quelle vorhandener Risiken. Die menschlichen Komponenten dagegen dienen als Kontrollinstanzen, welche die technischen Komponenten steuern und überwachen.

Jedes soziotechnische System ist notwendigerweise auch ein komplexes System, da die Komplexität des Menschen die Komplexität des Gesamtsystems, bestehend aus menschlichen und technischen Komponenten, dominiert. Infolgedessen besitzen soziotechnische Systeme auch die besonderen Eigenschaften komplexer Systeme (Ladyman et al., 2013). Für uns sind folgende Eigenschaften von besonderem Interesse:

- Soziotechnische Systeme besitzen eine inhärente Dynamik aufgrund des adaptiven Verhaltens und der Lernfähigkeit des Menschen.
- Soziotechnische Systeme sind offen, d.h. das Verhalten des Systems wird durch externe Einflüsse mitbestimmt. Beispielsweise ist das Verhalten eines Menschen durch Haltung der Vorgesetzten, Lehrmeinung, Äußerungen von Kollegen usw. von außen geprägt.
- Soziotechnische Systeme sind stochastisch, da die Individualität der Menschen eine erhebliche Variation ihres jeweiligen Verhaltens bewirkt.

3 Voraussetzungen an das Modell

Wir wollen nun die Eigenschaften diskutieren, die das Modell M haben muss, um für eine formale Validierung geeignet zu sein. Es zeigt sich, dass M erstens eine brauchbare Abgrenzung von der Umgebung besitzen muss; Teile des Systemumfelds, die auf S entscheidenden Einfluss haben oder von S entscheidend beeinflusst werden können, sollten als Bestandteil von M aufgefasst werden. Zweitens sollte das Modell auf einer geeigneten Ebene der Auflösung formuliert sein. Drittens müssen die risikorelevanten Objekte von M und von S einander zugeordnet werden können.

3.1 Abgrenzung des Modells

Das Risiko R hängt u.a. von den Effekten des Systems S auf die Umgebung U von S ab, die eventuelle Fehlfunktionen der technischen Komponenten von S verursachen. Werden diese Effekte nicht durch das Modell M beschrieben, sind ihre Kritikalitäten – repräsentiert durch die Verlustfunktion L – nicht aus M herleitbar und können daher lediglich mit entsprechenden Unsicherheiten behaftet intuitiv geschätzt werden. Im Falle soziotechnischer Systeme kommt als komplementärer Effekt die Wirkung der Umgebung U auf die menschlichen Systemkomponenten von S hinzu. Weil die Menschen gemäß der Annahme aus Abschnitt 2.2. als Kontrollinstanzen der technischen Komponenten von S fungieren, deren Fehlfunktionen die vorhandenen Risiken bestimmen, hat U also einen Einfluss auf das Auftreten möglicher Fehlfunktionen von S . Nicht zu S selbst gehörend ist dieser Einfluss nicht genauer spezifiziert. Damit sind auch die Wahrscheinlichkeiten $P(e)$ für das Auftreten von Fehlfunktionen e mit Unsicherheiten behaftet. Unter diesen Umständen hat ein quantitativ kalkuliertes Risiko nur noch eingeschränkte Aussagekraft; von der Durchführbarkeit einer erfolgreichen Validierung kann man dann ebenfalls nicht mehr unbedingt ausgehen. Das Problem ist jedoch zumindest im Prinzip beherrschbar, wenn die sowohl in L und als auch in P vorhandenen Unsicherheiten so weitgehend wie möglich objektiviert werden. Eine geeignete Methode hierfür ist, das Modell M von S um entsprechende Teile der Umgebung U des Systems S zu ergänzen und so eine Beschreibung der wesentlichen Interaktionen zwischen U und S im Rahmen von M zu ermöglichen.

3.2 Auflösung des Modells

Die Eignung eines Modells wird durch seinen Detaillierungsgrad wesentlich mitbestimmt. Ein zu grobes Modell macht ungenaue oder sogar fehlerhafte Vorhersagen und stellt keine geeignete Repräsentation des Systems dar. Ein zu detailliertes Modell arbeitet dagegen mit teilweise schwer oder gar nicht zugänglichen Parameterwerten, wodurch eine Validierung bereits im Ansatz scheitert. Ein zu hoher Detaillierungsgrad macht auch hinsichtlich der unzureichenden Modellierbarkeit bestimmter kognitiver Aspekte des Menschen wenig Sinn. In diesem Zusammenhang gelten ähnliche Einwände (Proncheva & Makhov, 2012), wie sie bereits gegen Forrester's Welt-Modell (Meadows et al., 1972) des Club of Rome vorgebracht wurden. Worauf ist hinsichtlich Modellierbarkeit und damit auch Validierbarkeit zu achten?

Menschliche und künstliche Intelligenz sind nicht ohne weiteres miteinander vergleichbar. Bei einigen Aufgaben ist der Rechner überlegen (z.B. Identifikation von Sonarsignalen), bei anderen der Mensch (z.B. Vorhersage von Wechselkursen). Die menschliche Lösungskompetenz ist bei gegenwärtigem Stand der Technik daher im Allgemeinen grundsätzlich nur approximierbar, aber nicht vollständig adäquat modellierbar. Selbst wenn Mensch und Maschine in einer Problemklasse insgesamt eine annähernd gleichstarke Lösungskompetenz besitzen, kann das Spektrum der erfolgreich behandelten Aufgaben sehr verschie-

den sein. Um die menschliche Art der Problembehandlung detailliert im Modell zu reproduzieren, sind sogenannte deskriptive Theorien zu verwenden wie beispielsweise in der Entscheidungstheorie. Es sind aber noch viele andere Gebiete betroffen. Man betrachte etwa zufällige Prozesse: Die Unterschiede zwischen „echten“ und von Menschen generierten Sequenzen von Zufallszahlen sind signifikant (Schulz et al., 2012). Die kognitiven Funktionen des Menschen werden darüber hinaus durch Emotionen, kulturellen Hintergrund, Motivation, Ermüdung, bewusste und unbewusste Erfahrungen, gesundheitliche Aspekte u.a. beeinflusst. Der Mensch verfügt zudem grundsätzlich über die umfangreichere und vielfältigere Wissensbasis einschließlich sogenanntes „*Tacit Knowledge*“ (Kimble, 2013), das z.B. aus vorgefassten Meinungen gegenüber bestimmten Sicherheitsprozeduren besteht und erheblichen Einfluss auf Entscheidungsprozesse haben kann (Giunipero et al., 1999). Zusammenfassend ist man also hinsichtlich der kognitiven Prozesse des Menschen ohnehin gezwungen, für das Modell M eine erhebliche Abstraktion von der Vielfalt der vorhandenen Details zu akzeptieren.

3.3 Art des Modells

Um zur Validierung des Modells M das Verhalten der Objekte des Modells M und des realen Systems S miteinander vergleichen zu können, sind die Objekte von M und von S zueinander in Beziehung zu setzen. Die Existenz einer solchen Korrespondenz ist jedoch nicht selbstverständlich und hängt u.a. von der Art des Modells M ab. Zumindest hinsichtlich einer Validierung scheint ein beschreibender Ansatz des Verhaltens (Horvath & Rudas, 2004) am besten geeignet. Die Korrespondenz zwischen Modell- und Systemobjekten ist dann in der Regel kanonisch gegeben, so dass die Wahrscheinlichkeiten für bestimmte Verhaltensweisen in M mit Beobachtungen des menschlichen Verhaltens in S unmittelbar verglichen werden können. Verhaltensbasierte Modelle genügen somit der entsprechenden Forderung aus Ale et al. (2012).

4 Durchführung einer Validierung

Zur formalen Validierung eines Modells M stehen verschiedene Optionen zur Verfügung. Nicht jede dieser Optionen ist jedoch für die konkrete Anwendung im Rahmen der funktionalen Sicherheit geeignet. So scheint eine Validierung lediglich des resultierenden Gesamtrisikos R unbefriedigend zu sein, da ein solcher Validierungsansatz die notwendigen prognostischen Fähigkeiten von M ignoriert, in der Praxis bisher noch unbeobachtete aber potentiell mögliche Fehlfunktionen zu berücksichtigen. Stattdessen ist eine Überprüfung vorzuziehen, ob zueinander korrespondierende Objekte des Modells M und des Systems S auch ein genügend ähnliches Verhalten zeigen. Da wir aus Abschnitt 2.2. bereits wissen, dass S ein stochastisches System ist, bieten sich verschiedene statistische Validierungsmethoden an wie etwa Korrelationsmaße, Cross-Validationen oder statistische Tests. Die Überprüfung des Verhaltens kann dabei zwei verschiedene Aspekte

betreffen, nämlich zum einen das Verhalten der einzelnen Objekte selbst und zum anderen die Beziehung zwischen ihnen.

4.1 Validierung der Modellobjekte

Die klassischen Verfahren der statistischen Validierung, wie sie etwa in Hills & Trucano (1999) und in Hunter et al. (1997) beschrieben werden, sind im Detail nicht ohne weiteres auf soziotechnische Systeme übertragbar. Eine Stochastik technischer Komponenten gehorcht häufig einer bekannten elementaren Wahrscheinlichkeitsverteilung P' , so dass für eine Validierung statistische Tests bezüglich der vorgegebenen Verteilung P' genügen. Dagegen sind die Wahrscheinlichkeitsverteilungen des Verhaltens von Menschen – abhängig von ihrer jeweiligen Persönlichkeit – aufgrund von Wertehäufungen und miteinander korrelierten Parametern teilweise stark strukturiert (Roberts & Stankov, 1999, sowie Good, 1962). Infolgedessen ist bei der Validierung des Verhaltens der menschlichen Komponenten von S häufig auf eine Korrespondenz zwischen Modell M und System S ohne vorgegebene Wahrscheinlichkeitsverteilung zu testen. Derartige Verfahren werden u.a. in Lopes et al. (2007) beschrieben.

4.2 Validierung der Beziehungen zwischen den Modellobjekten

Ein Systemmodell M beschreibt nicht nur die einzelnen Modellobjekte, sondern auch die Beziehungen zwischen ihnen. Insofern ist man über die Validierung des Verhaltens der Modellobjekte hinaus auch an einer Validierung der Beziehungen interessiert. Denn diese beschreiben die Struktur des Modells, welche durch Berücksichtigung von Redundanzen, Fehlerketten u.a. in die Berechnung des Risikos einfließt. Allerdings ist eine vollständige formale Validierung der Modellstruktur nicht ohne weiteres realisierbar. Statistische Tests auf eine kausale Abhängigkeit gelten beispielsweise als problematisch (Jacobs et al., 1978; Blossfeld & Rohwer, 1997), wobei die vorhandenen Schwierigkeiten sicherlich auch mit der Herausforderung zu tun haben, den Begriff der Kausalität überhaupt zu definieren. Welche Störung einer Variable in zeitkontinuierlichen Systemen „zuerst“ aufgetreten ist, kann nicht unterschieden werden; die Bedeutungen von Kausalität und Korrelation beginnen zu verschmelzen. Viele risikobasierte Verfahren zur Sicherheitsanalyse sind jedoch auf eine Unterscheidung zwischen Ursache und Wirkung angewiesen. Ihre Anwendung in solchen Situationen muss daher möglicherweise überdacht werden (Hollnagel, 2014).

Eine weitere mögliche Beziehung zwischen Zufallsvariablen ist ihre (stochastische) Unabhängigkeit. Die Elementarereignisse $\omega \in \Omega$ des festgelegten Wahrscheinlichkeitsraums (Ω, \mathcal{E}, P) sind beispielsweise definitionsgemäß stochastisch unabhängig. Da wir soziotechnische und damit notwendigerweise offene Systeme betrachten, lässt sich diese Eigenschaft jedoch nicht ohne weiteres garantieren. Jedes Paar von Ereignissen, die innerhalb des Systems stochastisch unabhängig sind, kann durch außerhalb des Systems bestehende Zusammenhänge stochastisch abhängig werden. Die Konsequenzen dieser Feststellung für risikobasierte Verfahren der funktionalen Sicherheit scheinen derzeit noch offen zu sein.

5 Probleme der Validierung

Bei der formalen Validierung von Modellen soziotechnischer Systeme können verschiedene Probleme auftreten. Wir betrachten in diesem Abschnitt Probleme, die im Zusammenhang mit der funktionalen Sicherheit typisch erscheinen, und eventuelle Ansätze zu ihrer Lösung. Im Einzelnen sind dies:

- Wird die statistische Signifikanz der Validierungen insbesondere für die Wahrscheinlichkeiten der Fehlfunktionen als zu gering eingeschätzt, kann der Übergang zu einer gröberen Auflösung des Modells Abhilfe schaffen.
- Äußere (d.h. nicht zum betrachteten System gehörende) Einflüsse können bei der Kalkulation des Risikos berücksichtigt werden, indem der Wahrscheinlichkeitsraum um das sogenannte „unbekannte Ereignis“ erweitert wird.
- Bei Verwendung von Modellen, die das Objektverhalten basierend auf Objekteigenschaften kalkulieren, lässt sich unter Umständen mit Hilfe einer Modellkalibration die für die Validierung erforderliche Korrespondenz zwischen den Objekten des Modells und des realen Systems etablieren.
- Die klassische mathematische Formulierung des Risikos lässt die individuelle Perzeption des Risikos durch Menschen unberücksichtigt. Daher ist eventuell ein modifizierter Risikobegriff zu wählen.

5.1 Statistische Signifikanz und Abstraktion der Validierung

Der Versuch einer Validierung von Modellen soziotechnischer Systeme zur Untersuchung von Fragen der funktionalen Sicherheit liefert oft Ergebnisse mit geringer statistischer Signifikanz. Die Gründe dafür sind vielfältig. Die starke Streuung des individuellen Verhaltens verschiedener Personen in einer konkreten Situation spielt ebenso eine Rolle, wie die oft geringen verfügbaren Stichprobenanzahlen. Letztere ergeben sich aus der Tatsache, dass die für die Risikoberechnung betrachteten Fehlfunktionen des Systems konstruktionsbedingt vermieden werden sollen und daher üblicherweise entsprechend selten sind. Hinzu kommt gemäß Abschnitt 4.1. der notwendige Verzicht auf Zugrundelegung einer gegebenen Wahrscheinlichkeitsverteilung bei der Validierung.

Wird die erhaltene statistische Signifikanz des Validierungsergebnisses als zu gering für eine zuverlässige Aussage eingeschätzt, ist eine Vergrößerung der Betrachtungsebene eine naheliegende Option. In erster Linie betrifft diese Option die Auflösung des Modells (siehe Abschnitt 3.2.). Will man das Modell jedoch unverändert lassen – für „*Legacy Systems*“ beispielsweise kommt mangels Einblick in das verwendete Modell eine Modifikation des Modells üblicherweise nicht in Frage – lässt sich auch eine Abschwächung des herkömmlichen Begriffs der statistischen Validierung in Betracht ziehen, wie etwa die Betrachtung einer eingeschränkten Korrespondenz zwischen Modell M und realem System S . Eine derartige eingeschränkte Korrespondenz kann so aussehen, dass nicht mehr alle erfolgreich validierbaren Modelle das gewünschte Systemverhalten besitzen,

jedoch keine adäquaten Modelle abgelehnt werden. Im Vergleich mit anderen, etablierten Validierungsmethoden erscheint ein solches Vorgehen durchaus zulässig. Jede statistische Validierung kann durch ungünstige Gestalt der Stichprobe ebenfalls fälschlich ein zulässiges Modell signalisieren. Validierungen durch Betrachtung von Use-Cases (Fleisch, 1999) basieren von vornherein auf einem ähnlich eingeschränkten Anspruch. Der Vorteil der vorgeschlagenen Abschwächung des Validierungsbegriffs – im Folgenden als abstrahierte Validierung bezeichnet – ist, dass die statistische Signifikanz des Validierungsergebnisses bei geeigneter Wahl der Validierungsmethode zunehmen kann.

Ein Beispiel zum Ansatz der abstrahierten Validierung ist die Überprüfung der Einhaltung einer Ordnungsrelation anstelle einer quantitativen Korrespondenz. Ein solches Vorgehen ist ohnehin naheliegend, wenn anstelle von Experimenten mit absoluten Beobachtungen nur vergleichende Beobachtungen herangezogen werden können. Man denke in diesem Zusammenhang etwa an die Bestimmung kognitiver Fähigkeiten durch Zählung erfolgreich gelöster Aufgaben. Die beschriebene Methode validiert Wertebeziehungen, abstrahiert dabei aber von den quantitativen Werten an sich.

Steht keine Ordnungsrelation zur Verfügung, kann man stärker abstrahieren und beispielsweise die Mengen der möglichen Verhaltensweisen von Modell und System vergleichen. Damit lässt sich beispielsweise klären, ob das Modell tatsächlich das gesamte Spektrum möglicher Verhaltensweisen überdeckt oder ob das Modell eventuell ein teilweise unrealistisches Verhalten zeigt. Ein *brute-force*-Ansatz zur Konstruktion der Menge möglicher Verhaltensweisen durch Abarbeitung aller theoretisch möglichen Parameterwerte ist aufgrund des Rechenaufwands in aller Regel unpraktikabel. Man greift daher auf Verfahren wie beispielsweise ein „*Design of Experiments*“ (Hinkelmann & Kempthorne, 2008) zurück. Vereinfachend wirkt sich hierbei aus, dass wir stochastische Systeme betrachten, so dass das Auftreten singulärer Verhaltensregionen zu vernachlässigen ist. Eine weitergehende Verfeinerung des Ansatzes ist mit dem sogenannten „*Data Farming*“ möglich (Horne & Meyer, 2004).

Selbst wenn die so erzielte Überdeckung des Parameterraums nicht dicht genug ist, um eine Klassifikation der Verhaltensweisen zu erreichen, ist der beschriebene Ansatz für eine Validierung brauchbar. Denn wenn bei n Läufen des Modells kein einziger dieser Läufe ein unrealistisches Verhalten ergibt, nimmt mit steigender Anzahl n die Glaubwürdigkeit des Modells zu.

5.2 Vollständigkeit der Ereignismenge und das Unbekannte Ereignis

Soziotechnische Systeme sind notwendigerweise auch offene Systeme. Damit erhebt sich jedoch die Frage nach der Vollständigkeit der betrachteten Ereignismenge. Denn aus dem Systemumfeld können unbekannte Ereignisse auf das System S wirken und so das Gesamtrisiko beeinflussen. Der für S formulierte Wahrscheinlichkeitsraum der Fehlfunktionen lässt solche Einflüsse definitions-

gemäß unberücksichtigt. Wie kann er dann jedoch zur Kalkulation des Risikos genutzt werden?

Das in komplexen Systemen möglicherweise auftretende unbekanntes Ereignis ist eines der Argumente, mit denen Alternativen zur klassischen Wahrscheinlichkeitstheorie wie *Fuzzy Sets* oder die *Possibility Theory* propagiert werden. Man kann ein solches Ereignis aber auch im Rahmen der klassischen Wahrscheinlichkeitstheorie abbilden und so den klassischen Risikobegriff, der auf den üblichen Wahrscheinlichkeiten basiert, weiter nutzen. Der Lösungsansatz besteht darin, das „unbekannte Ereignis“ in Form zusätzlicher Elementarereignisse in den Wahrscheinlichkeitsraum aufzunehmen. Dabei wird für jeden Wert der Verlustfunktion, den man für die Berechnung des Risikos separat benötigt, ein eigenes Elementarereignis eingeführt.

Es bietet sich an, in den neu eingeführten Elementarereignissen auch sogenannte „*low probability - high consequence*“-Ereignisse zusammenzufassen, sofern der verwendete Wahrscheinlichkeitsraum dies zulässt. In der Regel sind die Wahrscheinlichkeiten dieser Ereignisgruppe ohnehin nicht validierbar, da ihre Elemente definitionsgemäß allenfalls als Einzelfälle beobachtet worden sind. Trotz ihrer geringen Wahrscheinlichkeit tragen sie aber zum Gesamtrisiko signifikant bei, da der durch sie verursachte Schaden (d.h. ihr Wert der Verlustfunktion L) definitionsgemäß groß ist (Waller, 1984). Dabei kann das Fehlverhalten von Menschen eine entscheidende Rolle spielen (Faber & Stewart, 2003).

5.3 Eigenschaftsbasierte Modelle und Modellkalibration

Die verhaltensbasierte Modellierung ist ein top-down orientiertes Verfahren zur unmittelbaren Beschreibung des Verhaltens von Modellobjekten. Interne Parameter der Objekte und der Weg zu einer Entscheidung über das Objektverhalten sind dabei nicht von Interesse.

Gibt es zu viele Verhaltensoptionen, um diese auf praktikable Weise explizit beschreiben zu können, wird häufig auf eine eigenschaftsbasierte Modellierung zurückgegriffen, die das jeweilige Verhalten im Modell bottom-up aus entsprechenden Eigenschaftsparametern ableitet. Dieser Ansatz ist weit verbreitet. Beispiele sind etwa Durupinar (2010), Erlenbruch (2002), Henscheid et al. (2006) und Lampe et al. (2007). Ein potentiell Problem eigenschaftsbasierter Modelle soziotechnischer Systeme ist, dass die Eigenschaften, die in eine Entscheidung über das weitere Verhalten einfließen, oft keiner unmittelbaren Interpretation zugänglich sind. Was soll beispielsweise konkret bedeuten, dass eine Person ein „Aggressivitätslevel“ gleich 0,25 besitzt? Erst Verfahren wie etwa die Kalibration (siehe etwa Hill, 2000) des Modells setzen Eigenschaftsparameter und Verhalten in eine wechselseitige Beziehung. Eine solche Beziehung ist essentiell hinsichtlich der Frage, welche Parameterkombination eigentlich ein bestimmtes Objektverhalten reproduziert bzw. wie gut ein bestimmtes Verhalten überhaupt approximierbar ist. Dadurch legt die Kalibration fest, für welche Objektparameter

ein Vergleich des Objektverhaltens zwischen Modell M und realem System S zu erfolgen hat. Allerdings etabliert eine solche Kalibration nicht immer eine 1:1 Beziehung zwischen den Parametern des Objekts und den einzelnen „Dimensionen“ seines Verhaltensraums. Dies führt zu Komplikationen, wenn beispielsweise zwei Parameter das Verhalten eines Objekts in ähnlicher Weise beeinflussen und sich so in ihrer Bedeutung überschneiden, oder wenn mehrere Parameterkombinationen dasselbe Verhalten reproduzieren und so Mehrdeutigkeiten einführen. Dementsprechend scheinen eigenschaftsbasierte Modelle schwieriger zu validieren zu sein als verhaltensbasierte Modelle.

5.4 Risikoperzeption und Modifikationen des Risikobegriffs

Im Rahmen dieses Artikels wird angenommen, dass Menschen, die zu einem soziotechnischen System S gehören, darin als Kontrollinstanz fungieren. Sie überwachen und steuern die technischen Komponenten von S . Die damit verbundenen Entscheidungen hängen jedoch von der Perzeption des Risikos des jeweiligen Menschen ab und basieren somit oft nicht auf dem klassischen Risikobegriff (Sjöberg, 2000; Botterill & Mazur, 2004). Wenn aber die reale Kontrolle der technischen Komponenten auf einem anderen Risikobegriff basiert als die Kalkulationen für die funktionale Sicherheit, wird der Mensch als Kontrollinstanz unter Umständen andere Entscheidungen fällen, als sie der mathematischen Risikokalkulation zugrunde liegen.

Selbst ein einzelner Mensch kann bedingt durch Erscheinungen wie kognitive Phasenübergänge (Kelso, 2010) situationsabhängig mehrere unterschiedliche Kontrollfunktionen realisieren. Man denke in diesem Zusammenhang an den plötzlichen Wechsel von Beurteilungsergebnissen bei aufkommender Unsicherheit oder Panik oder bei einem Übergang zu grob heuristischen Methoden im Falle des Gefühls der Überlastung. Damit wird die Bewertung der vorhandenen Risiken situationsabhängig.

Nehmen wir an, dass der mathematische Risikobegriff das theoretisch minimale Risiko bei optimaler Kontrolle liefert, und dass die nichtideale Perzeption des Risikos durch die menschlichen Kontrollinstanzen im realen soziotechnischen System aufgrund ihrer inhärenten Unvollkommenheit und Fehlerbehaftung nur zu einer Vergrößerung dieses Risikos führen kann. Dann liefert der rein mathematische Risikobegriff eine Abschätzung des tatsächlichen Risikos nach unten. Will man anstelle einer solchen einseitigen Abschätzung den tatsächlichen Wert des Risikos bestimmen, ist bei der Kalkulation des Risikos derselbe Risikobegriff zu verwenden, wie ihn auch der Mensch im Rahmen seiner Kontrollfunktion benutzt. Die Konsequenzen dieser Schlussfolgerung sind weitreichend. Möglicherweise müssen, abhängig von den zu erwartenden Entscheidungen der individuellen menschlichen Kontrollinstanz, die Wahrscheinlichkeiten für Fehlfunktionen individualisiert werden; denn personenabhängig sind natürlich teilweise unterschiedliche Entscheidungen zu erwarten, die über spezifische Aktionen die technischen Komponenten unterschiedlich beeinflussen und so auch

zu unterschiedlichen Risikowerten führen. Darüber hinaus kann es aber unter Umständen sogar erforderlich sein, anstelle von P den gesamten der Risikoberechnung zu Grunde liegenden Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) zu individualisieren. Denn die Wahrscheinlichkeiten für die einzelnen Verhaltensoptionen der menschlichen Kontrollinstanz sind aufgrund von dessen allgemeiner Persönlichkeitsstruktur teilweise miteinander korreliert, wodurch für unterschiedliche Individuen als Kontrollinstanzen beispielsweise eventuell auch unterschiedliche Mengen Ω von Elementarereignissen gewählt werden müssen.

6 Anwendung in der Praxis

In diesem Abschnitt wird zunächst diskutiert, wie man die propagierten Methoden zur Behandlung der funktionalen Sicherheit soziotechnischer Systeme in die Praxis umsetzen kann. Anschließend sind mit dem Automotiv-Bereich und den autonomen Systemen zwei Anwendungen beschrieben, für die die hier gemachten Ausführungen sowohl leicht anwendbar als auch von besonderem Interesse scheinen.

6.1 Umsetzung in die Praxis

Das Themengebiet der formalen Validierung soziotechnischer Systeme ist insbesondere aus der Perspektive der funktionalen Sicherheit selbst auf rein theoretischer Ebene bisher wenig untersucht. Angesichts der zahlreichen Unfälle aufgrund menschlichen Versagens scheint es umgekehrt jedoch einen erheblichen Bedarf an ausgereiften und praxistauglichen Verfahren für die Anwendung zu geben. Wie können die bisher erzielten Erkenntnisse trotz der immer noch unvollständigen Beherrschung der Thematik in die Praxis umgesetzt werden?

Für eine Etablierung in der Praxis bieten sich Themengebiete an, die eine einfache Anwendung der Methodik gestatten und die gleichzeitig erheblich von den dabei erzielten Resultaten profitieren. Eine einfache Anwendbarkeit scheint gegeben, wenn die technischen Komponenten der betrachteten Systeme gut verstanden sind und eine vergleichsweise geringe Komplexität besitzen. Auch die Einflussmöglichkeiten der menschlichen Komponenten auf das Gesamtsystem sollten sowohl hinsichtlich ihrer Vielfalt als auch hinsichtlich ihrer Wirkung überschaubar sein. Umgekehrt ist bei Einbeziehung des Menschen ein signifikanter Erkenntnisgewinn zu erwarten, wenn die erzielten Resultate tatsächlich verlässlich sind – etwa infolge begrenzter Interaktionen zwischen System und Umwelt sowie geringer Unsicherheiten hinsichtlich des menschlichen Verhaltensspektrums – und zudem von den Ergebnissen einer Betrachtung auf rein technischer Ebene abweichen. Letzteres ist denkbar, wenn die Aktionen, die der menschlichen Kontrollinstanz zur Steuerung der technischen Komponenten zur Verfügung stehen, Risiken nennenswert beeinflussen können und das Gesamtrisiko zudem stark von den getroffenen Entscheidungen der menschlichen Kontrollinstanz abhängt.

6.2 Automotive als Anwendungsgebiet

Der Fahrer eines Autos übt eine wichtige Kontrollfunktion über das Fahrzeug aus. Er kann somit die Funktion und Wirksamkeit technischer Sicherheitsmaßnahmen beeinflussen (Mather, 2009; Vivoli et al., 2006). Die Untersuchung vorhandener Risiken beschränkt sich dabei üblicherweise auf Situationen, in denen dem Fahrer infolge von technischen Systemfehlern die Kontrolle über das Fahrzeug zumindest teilweise entzogen wird. Diese Sichtweise vernachlässigt jedoch Effekte wie nachlassende Aufmerksamkeit oder Ablenkung des Fahrers. Eine solche Ablenkung kann auch durch das Fahrzeug selbst verursacht werden. Beispiele dafür sind das Ablesen von Fahrzeuginstrumenten, Bedienung von Assistenzsystemen, Einblendung von Warnsignalen etc. Die auf technischer Basis kalkulierten Risiken entsprechen daher möglicherweise nicht den tatsächlichen Risiken, welche sich unter Einbeziehung der Interaktion zwischen Fahrer und Fahrzeug ergeben (van Elslande et al., 2008). Fundierte Betrachtungen der Fahrsicherheit sollten sich daher nicht auf das Auto als rein technisches System beschränken.

6.3 Autonome Kampfroboter als Anwendungsgebiet

Für militärische Anwendungsgebiete sind immer stärker autonome robotische Systeme im Gespräch (Singh & Thayer, 2001). Bei Systemen mit Kampfauftrag stellen sich dabei angesichts der oftmals chaotischen Verhältnisse auf einem Schlachtfeld grundsätzliche Fragen hinsichtlich der funktionalen Sicherheit der Mensch-Maschine-Interaktionen. Als besondere Komplikation des Anwendungsgebiets ist in diesem Zusammenhang festzuhalten, dass die technischen Komponenten nicht fortwährend menschlicher Kontrolle unterliegen und teilweise eigene Entscheidungen treffen, welche nicht immer im Sinne der menschlichen Kontrolle sind. Zudem muss das technische System bei seinen Handlungen ein eventuelles Fehlverhalten der menschlichen Komponenten (Lin et al., 2008) berücksichtigen. Was passiert beispielsweise, wenn im Rahmen eines friendly-fire-Vorfalles ein Soldat versehentlich eine eigene Kampfdrohne unter Beschuss nimmt? Nimmt die Drohne dann ihrerseits den Soldaten unter Feuer?

7 Zusammenfassung und Ausblick

Die Arbeit diskutiert die formale Validierung von Modellen soziotechnischer Systeme. Fragen der risikobasierten funktionalen Sicherheit stehen dabei in zentralem Interesse. Es ist festzustellen, dass die beabsichtigte Validierung bereits bei der Konstruktion des Modells berücksichtigt werden sollte. Dies gilt sowohl für den verwendeten Typ des Modells als auch für dessen Ebene der Auflösung. Insofern wird eines der wesentlichen Ergebnisse von Knauf et al. (2001) bestätigt und ausgebaut. Mögliche Probleme sind eine geringe statistische Signifikanz der Validierung und das mögliche Auftreten unbekannter, bei der Modellierung nicht explizit berücksichtigter Ereignisse. Sie lassen sich vermutlich durch eine Verallgemeinerung des klassischen Begriffs der statistischen Validierung und/oder durch eine Modifikation des zugrunde liegenden Wahrscheinlichkeitsraums beherrschen. Unter Umständen ist der klassische Risikobegriff anzupassen.

Betrachtungen zu konkreten Anwendungsgebieten zeigen, dass der Mensch in Sicherheitsbetrachtungen unmittelbar einbezogen werden sollte. Auf diese Weise können meist zuverlässigere und genauere Abschätzungen des Risikos gewonnen werden als bei Beschränkung auf die rein technischen Systemkomponenten. Der Übergang von technischen zu soziotechnischen Systemen lässt einen Erkenntnisgewinn erwarten, der den damit verbundenen erforderlichen Zusatzaufwand aufwiegt. Diese Arbeit leistet hierzu einen anwendungsorientierten Beitrag.

8 Danksagung

Ich danke Herrn G. Götz und Herrn C. Hauck sehr herzlich für ihre Anregungen und Kommentare zu diesem Artikel. Fernerhin danke ich Frau O. Grigorincu für die Unterstützung bei der Literaturrecherche.

Literatur

- Ale, B.J.M., Hanea, D.M., Sillem, S., Lin, P.H., Van Gulijk, C. & Hudson, P.T.W. (2012). Modelling risk in high hazard operations: integrating technical, organisational and cultural factors. In *Proceedings of the 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012 (PSAM11 and ESREL 2012)*, June 25-29, 2012, Helsinki, Finland.
- Anderson, S. & Felici, M. (2009). Classes of socio-technical hazards: Microscopic and macroscopic scales of risk analysis. *Risk Management*, 11 (11) 208-240.
- Asnar, Y. & Giorgini, P. (2008). Ensuring Dependability in Socio-Technical System by Risk Analysis. In *Proceedings of the 6th European Dependable Computing Conference*, October 18-20, 2006, Coimbra, Portugal.
- Berger, O. (1985). *Statistical decision theory and Bayesian Analysis*. Berlin Heidelberg: Springer.
- Blossfeld, H.-P. & Rohwer, G. (1997). Causal inference, time and observation plans in the social sciences. *Quality and Quantity*, 31 (4), 361-384,
- Botterill, L. & Mazur, N. (2004). *Risk & risk perception* (RIRDC Publication No 04/043, RIRDC Project No BRR 8A). Barton, Australia: Rural Industries Research and Development Corporation.
- Cebulla, M. (2002). Using advanced formal concepts in interdisciplinary analysis and design of safety-critical sociotechnical systems: In *Proceedings of the 26th Annual Computer Software and Applications Conference*, August 26-29, 2002, Oxford, England.
- Durupinar, F. (2010). *From Audiences to Mobs: Crowd Simulation with Psychological Factors*. Dissertation, Bilkent Universität, Ankara.
- Erlenbruch, T. (2002). *Agent-based simulation of German peacekeeping operations for units up to platoon level*. PhD Thesis, Naval Postgraduate School, Monterey, California.
- Faber, M.H. & Stewart, M.G. (2003). Risk assessment for civil engineering facilities: critical overview and discussion. *Reliability Engineering and System Safety*, 80, 173–184.
- Fleisch, W. (1999). Applying Use Cases for the Requirements Validation of Component-Based Real-Time Software. In *Proceedings of 2nd IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC'99)*, May 2-5, 1999, Saint-Malo, France.

- Giunipero, L., Dawley, D. & Anthony, W.P. (1999). The impact of tacit knowledge on purchasing decisions. *Journal of Supply Chain Management*, 35 (1), 42-49.
- Goerger, S. (2004). *Validating Computational Human Behavior Models: Consistency and Accuracy Issues*. PhD Thesis, Naval Postgraduate School, Monterey, California.
- Good, I.J. (1962). A Classification of Fallacious Arguments and Interpretations. *Technometrics*, 4 (1), 125-132
- Harmon, S., Hoffman, D., Gonzalez, A., Knauf, R. & Barr, V. (2002). Validation of human behavior representation. In *Proceedings of the Workshop on Foundations for Modeling and Simulation (M&S), Verification and Validation (V&V) in the 21st Century*, October 22-24, 2002, Laurel, Maryland (USA).
- Henscheid, Z., Middleton, D. & Bitinas, E. (2006). Pythagoras: An Agent-Based Simulation Environment. In *Scythe: Proceedings and Bulletin of the International Data Farming Community*, Issue 1 Workshop 13.
- Hill, M. (2000). Methods and Guidelines for Effective Model Calibration. In R.H. Hotchkiss & M. Glade (Eds.), *Building Partnerships: Proceedings of 2000 Joint Conference on Water Resource Engineering and Water Resources Planning and Management*, July 30 - August 2, 2000, Minneapolis. Reston: American Society of Civil Engineers.
- Hills, R.G. & Trucano, T.G. (1999). *Statistical Validation of Engineering and Scientific Models: Background* (Report SAND99-1256). Albuquerque, New Mexico / Livermore, California: Sandia National Laboratories.
- Hinkelmann, K. & Kempthorne, O. (2008). *Design and Analysis of Experiments*, Vol. I and II. Hoboken, NJ: Wiley.
- Hollnagel, E. (2014). *Safety-I and Safety-II*. Burlington, VA / London: Ashgate.
- Horne, G. & Meyer, T. (2004). Data Farming: Discovering Surprise. In R.G. Ingalls, M.D. Rossetti, J.S. Smith & B.A. Peters (Eds.), *Proceedings of the 2004 Winter Simulation Conference*, December 5-8, 2004, Washington, D.C., USA.
- Horvath, L. & Rudas, I. (2004). Behavior Based Modeling as an Advanced Tool for the Engineering Practice. In *Proceedings of the 2nd Serbian-Hungarian Joint Symposium on Intelligent Systems*, October 1-2, 2004, Subotica, Serbia and Montenegro.
- Hunter, N., Barney, P., Ferregut, C., Perrez, L. & Paez, T. (1997). Statistical Validation of Stochastic Models. In *Proceedings of the 15th International Modal Analysis Conference - Current Horizon for Structural Damage Detection (IMAC XV)*, February 3-6, 1997, Orlando.
- Jacobs, R., Leamer, E. & Ward, M. (1978). *Difficulties with Testing for Causation*. Los Angeles: Department of Economics, UCLA.
- Kelso, S. (2010). Instabilities and Phase Transitions in Human Brain and Behavior. *Frontiers in Human Neuroscience*, 4, Article 23.
- Kimble, C. (2013). Knowledge management, codification and tacit knowledge. *Information Research*, 18 (2) paper 577.
- Knauf, R., Philippow, I., Gonzalez, A. & Jantke, K. (2001). The Character of Human Behavior Representation and Its Impact on the Validation Issue. In *Proceedings of the Fourteenth International FLAIRS Conference*, AAAI.
- Ladyman, J., Lambert, J. & Wiesner, K. (2013). What is a Complex System?. *European Journal for Philosophy of Science*, 3 (1), 33-67.
- Lampe, T., Schwarz, G. & Wagner, G. (2007). PAX: Designed for Peace Support Operations. In *Scythe: Proceedings and Bulletin of the International Data Farming Community*, Issue 2 Workshop 14.

- Lin, P., Bekey, G. & Abney, K. (2008). *Autonomous Military Robotics: Risk, Ethics, and Design*. California Polytechnic State University.
- Lopes, R.H.C., Reid, I. & Hobson, P.R. (2007). The two-dimensional Kolmogorov-Smirnov test. In *Proceedings of the XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, April 23-27, 2007 Amsterdam, Netherlands.
- Mate, J.L. & Silva, A. (2005). *Requirements Engineering for Sociotechnical Systems*. Hershey / London u.a.: Informations Science Pub.
- Mather, R. (2009). Social Cognitive Human Factors of Automobile Driving. In S.E. Paterson & L.K. Allan (Eds.), *Road Safety: Safety, Modeling, and Impacts*. New York: Nova Science Publishers.
- Meadows, D., Meadows, D., Randers, J. & Behrens III, W. (1972). *The Limits to Growth*. New York: Universe Books.
- Moya, L.J., McKenzie, F. & Nguyen Q.-A. (2008). Visualization and rule validation in human-behavior representation. *Simulation & Gaming*, 39, Article 101.
- Navlakha, J.K. (1986). A Survey of System Complexity Metrics. *The Computer Journal*, 30 (3), 233-238.
- Nechyba, M. (1998). *Learning and Validation of Human Control Strategies*. PhD Thesis, Carnegie Mellon University, Pittsburgh.
- Pizziol, S. (2013). *Conflict prediction in human-machine systems*. Doktorarbeit, Universität Toulouse.
- Proncheva, O. & Makhov, S., (2012). J. Forrester's Model of World Dynamics and its Development (Review). In G. Setlak, M. Alexandrov & K. Markov (Eds.), *Artificial Intelligence Methods and Techniques for Business and Engineering Applications*. Rzeszow, Poland / Sofia, Bulgaria: Ithea.
- Roberts, R., Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning and Individual Differences*, 11 (1), 1-120.
- Sargent, R. (2011). Verification and Validation of Simulation Models. In S. Jain, R.R. Creasey, J. Himmelspach, K.P. White & M. Fu, (Eds.), *Proceedings of the 2011 Winter Simulation Conference*, December 11-14, 2011, Phoenix, AZ, USA.
- Schulz, M.-A., Schmalbach, B., Brugger, P. & Witt, K. (2012). Analysing Humanly Generated Random Number Sequences: A Pattern-Based Approach. *PLoS ONE* 7(7) e41531.
- Singh, S. & Thayer, S. (2001). *ARMS: Autonomous Robots for Military Systems* (Report CMU-RI-TR-01-16). Pittsburgh: Carnegie Mellon University.
- Sjöberg, L. (2000). Factors in Risk Perception. *Risk Analysis*, 20 (1) 1-11.
- Spellman, F. & Whiting, N. (2009). *The Handbook of Safety Engineering: Principles and Applications*. Government Institutes.
- Tsalgatiidou, A. & Loucopoulos, P. (1991). Rule-based behaviour modelling: specification and validation of information systems dynamics. *Information and Software Technology*, 33 (6) 425-432.
- van Elslande, P., Naing, C. & Engel, R. (2008). Analyzing Human Factors in road accidents. TRACE WP5 Summary Report. Deliverable D5.5.
- Vivoli, R., Bergomi, M., Rovesti, S., Busseti, P. & Guaitoli, G.M. (2006). Biological and behavioral factors affecting driving safety. *Journal of Preventive Medicine and Hygiene*, 47, 69-75.

- Waller, R. (1984). Low-Probability High-Consequence Risk Analysis. Berlin Heidelberg: Springer.
- Zio, E. (2009). Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety*, 94 (2) 125-141.

Autor

Dr. J. Draeger

IABG mbH
Ottobrunn

Kontakt: draeger@iabg.de