



Wertbasiertes Design von Entscheidungsunterstützungssystemen

Von der Wertetabelle zur praktischen Umsetzung

Nick Lüthi · Christian Matt · Thomas Myrach

Eingegangen: 30. September 2021 / Angenommen: 1. Februar 2022 / Online publiziert: 2. März 2022
© Der/die Autor(en) 2022

Zusammenfassung Durch fortgeschrittenere Algorithmen in Entscheidungsunterstützungssystemen (DSS) bieten sich Unternehmen neue Möglichkeiten, um kognitive Belastungen bei Entscheidungsprozessen zu senken und bessere Ergebnisse zu erreichen. Allerdings mahnen Stimmen aus Forschung und Praxis vor damit einhergehenden Gefahren. Im Raum stehen etwa die systematische Fehlbeurteilung bestimmter Personengruppen, fehlerhafte Einschätzungen von (unternehmerischen) Risiken sowie unklare Verantwortlichkeiten im Falle personeller oder materieller Schäden. Der besseren Berücksichtigung von Werten wird diesbezüglich zunehmend Beachtung geschenkt, denn Algorithmen und die darauf aufbauenden DSS sind keine neutralen Artefakte; sie werden durch aus dem Entstehungsprozess hervorgehenden Wertvorstellungen und Wertannahmen geprägt. Das sogenannte Value Sensitive Design (VSD) wird genutzt, um dieser „Wertgeladenheit“ Rechnung zu tragen, indem Wertetabellen erhoben werden, die Nutzern und Entwicklern in Bezug auf ein DSS als wichtig erscheinen. Fernab konzeptioneller Überlegungen mangelt es aktuell allerdings noch an konkreten praktischen Operationalisierungen solcher Werttabellen. Am Beispiel der algorithmenbasierten Gefahrenerkennung beim Predictive Policing zeigen wir auf, wie sich eine Wertetabelle beim Design eines DSS konkretisieren lässt und welche Schwierigkeiten dabei auftreten können. Das vorgeschlagene dreistufige Operationalisierungsverfahren bietet Praktikern klare Richtli-

Nick Lüthi (✉) · Christian Matt · Thomas Myrach
Institute of Information Systems, Department Information Management, University of Bern,
Engenhaldenstrasse 8, 3012 Bern, Schweiz
E-Mail: Nicola.luethi@iwi.unibe.ch

Christian Matt
E-Mail: Christian.matt@iwi.unibe.ch

Thomas Myrach
E-Mail: Thomas.myrach@iwi.unibe.ch

nien, um die Wertgeladenheit und dadurch hervorgehende Problematiken bereits bei der Entwicklung von DSS zu berücksichtigen.

Schlüsselwörter Value Sensitive Design · Wertgeladenheit · Operationalisierung von Werten · Ethics-by-design

Value-based Decision Support System Design

From Value-list to Practical Implementation

Abstract The implementation of complex algorithms in decision support systems (DSS) offers companies new opportunities to reduce cognitive load in decision-making processes and to achieve better outcomes. However, voices from research and practice warn of the associated dangers. These include the systematic misjudgment of certain groups, incorrect assessment of (entrepreneurial) risks, and unclear responsibilities in the event of personal or material damage. Values are an increasingly important concept in this regard because algorithms and DSS based on them are not neutral artifacts, they are rather shaped by values and value assumptions arising during the development process. Value Sensitive Design (VSD) is used to account for this “value-ladenness” by creating a list of values that is considered important by users and developers with respect to a DSS. Except conceptual considerations, however, there is currently a lack of concrete practical operationalization of such value-lists. Using predictive policing as a practical example, we detail how a value-list can be used for the design of a DSS and what difficulties can arise in this process. The employed three-step operationalization process provides clear guidelines to practitioners for considering the value-ladenness and resulting problems already during the development of DSS.

Keywords Value Sensitive Design · Value-Ladenness · Operationalization of Values · Ethics-by-design

1 Hintergrund und Einordnung

Entscheidungsunterstützungssysteme (engl. Decision Support Systems, DSS) übernehmen immer mehr Funktionen, die bisher Menschen vorbehalten waren, indem sie wichtige Entscheidungen unterstützen oder diese gar vollständig übernehmen (Newell und Marabelli, 2015). Getrieben wird diese Entwicklung durch immer leistungsfähigere Algorithmen, die zunehmend komplexere Operationen ausführen können. Dies bietet aber nicht nur Chancen, ebenso kann dies zu systematischen kognitiven Verzerrungen (Bias), Verantwortungsproblematiken und weiteren ethischen Problematiken führen (Bozdog 2013; Martin 2019b; Albayrak et al. 2018; Mittelstadt et al. 2016). Wissenschaft und Praxis sind fortan gefordert, konkrete Lösungsvorschläge zu entwickeln, um das Potenzial der Algorithmen unter Einhaltung ethischer Richtlinien bestmöglich entfalten zu können. Gerade im Bereich der Künstlichen Intelligenz ist hier in den letzten Jahren ein starker Anstieg an Literatur, insbesondere zu Designrichtlinien, festzustellen (Jobin et al. 2019). Ein Teil der Literatur

fokussiert sich auf die Wertgeladenheit von Informationssystemen und die Berücksichtigung entsprechender Prinzipien im Designprozess. Diese „ethics by design“-Ansätze (Nussbaumer et al. 2021) streben die präventive Vermeidung von ethischen Problemen und damit einhergehenden Konsequenzen an (Martin 2019a).

Als ein führender Ansatz innerhalb dieser Designvorschläge hat sich Value Sensitive Design (VSD) etabliert (van den Hoven 2007). VSD ist ein Designansatz für technologische Artefakte, bei welchem menschliche Werte eine zentrale Rolle spielen. Der Ansatz sieht dafür vor, dass die Werte aller relevanten (d.h. von der zu entwickelnden Technologie direkt oder indirekt betroffenen) Stakeholder in einem iterativen Designprozess berücksichtigt werden. VSD kann dementsprechend sowohl auf bereits bestehende als auch auf noch zu entwickelnde Technologien angewendet werden. Gerade für Entscheidungsunterstützungssysteme und die damit einhergehenden Konsequenzen bietet sich VSD durch den Fokus auf Werte besonders an. Als Wert wird nach VSD alles erachtet, was für die entsprechenden Personen in Bezug auf Ethik von Wichtigkeit ist (Friedman et al. 2017). Diese Werte sind kontextspezifisch und können sich somit zwischen verschiedenen DSS und einzelnen Designprozessen unterscheiden, teils sogar zwischen verschiedenen Nutzungskontexten des gleichen DSS (Friedman et al. 2008). Zentral ist die Erstellung einer Wertetabelle, die die Werte für ein bestimmtes DSS in einem bestimmten Nutzungskontext umfasst. Ein VSD-Prozess umfasst notwendigerweise drei Untersuchungen (Friedman et al. 2008): 1) Eine konzeptionelle Untersuchung, bei der bestehende Theorie und Forschungsergebnisse herangezogen werden, um potenzielle Werte und Stakeholder zu ermitteln. 2) Eine empirische Untersuchung zur Befragung der zuvor identifizierten Stakeholder, um die potenziellen Werte zu verifizieren. 3) Eine technologiebezogene Untersuchung, in der die Werte in konkrete Designvorschläge umgesetzt werden sollen. In der ursprünglichen Form des VSD werden diese drei Untersuchungen in mehreren Iterationen durchgeführt, bis ein möglichst adäquates DSS entstanden ist (Friedman et al. 2017). Dementsprechend können sich konkrete VSD-Prozesse in der genauen Abfolge und Anzahl der drei Untersuchungen leicht unterscheiden.

VSD versucht, ethischen Aspekten im Kontext des verstärkten Einsatzes von Algorithmen besser Rechnung zu tragen und potenzielle, ungewollte Effekte holistisch zu erfassen, um diese möglichst frühzeitig minimieren zu können. Bei VSD stellen sich jedoch mehrere Hürden bei der praktischen Umsetzung (Robra-Bissantz und Strahinger 2020). Erstens gestaltet sich die Umsetzung erhobener Werte problematisch; es gibt zwar eine wachsende Literatur, in der Wertetabellen erhoben werden (Dadgar und Joshi 2018; Deng et al. 2016), eine Operationalisierung dieser Wertetabellen zu konkreten Systemanforderungen wird dort allerdings nicht vorgenommen. Zweitens können gerade bei der Entwicklung von neuen Systemen die Werte aller relevanten Stakeholder häufig erst während der aktiven Nutzung erhoben werden, während selbst bei der Anwendung von agilen Methoden grundlegende Veränderungen am DSS möglichst frühzeitig einfließen sollten. Drittens ist die Kontextspezifität von Wertetabellen insbesondere bei der Entwicklung von neuen DSS problematisch, da häufig noch nicht vollständig definierte Nutzungskontexte vorliegen und Wertetabellen aus anderen Nutzungskontexten nicht einfach übertragen werden können.

Ausgehend von diesen Umsetzungsproblematiken ergeben sich zwei konkrete Fragen, die wir in diesem Artikel genauer beleuchten:

- Wie kann die Wertetabelle eines VSD-Prozesses konkret in Systemanforderungen für ein DSS umgewandelt werden?
- Welche Probleme stellen sich bei der Umsetzung von VSD bei der Entwicklung von neuen DSS und wie können diese gelöst werden?

Anhand eines Forschungsprojekts aus dem Bereich Predictive Policing soll nachfolgend verdeutlicht werden, wie sich die Werte eines DSS operationalisieren lassen (Lüthi et al. 2021). Beim hier illustrierten Beispiel stellt sich früher oder später die Frage der Übertragbarkeit auf andere Nutzungskontexte. Wengleich sich die hier aufgezeigten Werte und deren Operationalisierung teils nicht ohne weiteres auf beliebige unternehmerische oder öffentliche Kontexte übertragen lassen mag, so liegt der Fokus im Folgenden nicht auf dem tatsächlichen Nutzungskontext, sondern auf dem Prozess der Transformation von Wertetabellen in konkrete Systemanforderungen. Dieser Prozess erfolgt unabhängig vom Nutzungskontext. Zu erwähnen ist allerdings, dass VSD primär in Kontexten mit schwerwiegenden ethischen Auswirkungen zum Einsatz kommt. Bei genauerer Betrachtung wird jedoch klar, dass mit zunehmender Leistungsfähigkeit von Algorithmen, diese Algorithmen auch zunehmend in Kontexten und Entscheidungsprozessen mit substantiellen ethischen Auswirkungen eingesetzt werden und somit wertbasierte Designansätze insgesamt an Relevanz gewinnen sollten.

2 Von der Wertetabelle zu konkreten Systemanforderungen

2.1 Szenario und erfasste Wertetabelle

Sogenannte „Risk-Assessment-Systeme“ werden von der Polizei für ihre präventiven Aufgaben genutzt, ebenso aber auch vor Gericht von Psychiatern für die Ist-Beurteilung in strafrechtlichen Fällen. Man spricht in diesem Zusammenhang oft von Predictive Policing. Die Entscheide, bei denen Risk-Assessment-Systeme zum Einsatz kommen, sind meist sensitiv und haben weitreichende Konsequenzen für die damit beurteilten Personen (etwa eine fehlerhafte Klassifikation als Straftäter). Dementsprechend kommen Werten bei deren Gestaltung und Nutzung eine große Bedeutung zu.

Wir haben im Rahmen einer früheren Erhebung insgesamt 16 Entwickler und Nutzer von zwei DSS interviewt, die von Behörden im Rahmen des Predictive Policing genutzt werden (Lüthi et al. 2021). Die beiden Systeme sind darauf ausgelegt, die Gefährlichkeit potenzieller Gewalttäter zu beurteilen. Aufgrund der geringen Nutzerzahl und dem erschwerten Zugang zu diesen Nutzern, verteilten sich die Interviews über einen Zeitraum von anderthalb Jahren. Die Nutzer wurden aus den Bereichen Polizei, Sozialarbeit und medizinisches Fachpersonal ausgewählt. Die Interviews wurden transkribiert und in einem dreistufigen Coding-Verfahren (Values Coding, Versus Coding, Focused Coding) wurde die vorliegende Wertetabelle erhoben (Tab. 1).

Tab. 1 Erhobene Wertetabelle für Risk-Assessment-Systeme (Lüthi et al. 2021)

Wert	Erklärung
Nachvollziehbarkeit	Nutzer wollen verstehen, wie Klassifizierungen zustande kommen
Neutralität	Nutzer wollen persönliche Vorurteile vermeiden und daher bei der Erstellung eines Risk-Assessments neutral bleiben
Zuverlässigkeit	Nutzer benötigen DSS, die sie mit zuverlässigen Informationen versehen, d. h. die bei gleichen Parametern über verschiedene Fälle hinweg gleiche Informationen liefern
Unterstützung	Nutzer wünschen sich, dass DSS ihre Arbeitsprozesse vereinfachen und beschleunigen
Autonomie	Nutzer wollen autonome Entscheide treffen, die von ihnen und nicht von Algorithmen stammen
Freiheit von Verzerrungen	Nutzer wollen Systeme nutzen, die frei sind von systematischen Verzerrungen
Compliance	Nutzer benötigen DSS, die ihnen erlauben, regulatorische Anforderungen zu erfüllen
Vertrauen	Nutzer wollen DSS, denen sie vertrauen können
Haftung	Nutzer sind verantwortlich gegenüber Drittparteien und wollen diese Verantwortung durch die Nutzung der DSS bestätigen
Privatheit	Nutzer können den untersuchten Personen durch den Gebrauch des DSS Anonymität gewähren

2.2 Der Operationalisierungsprozess

Die Erhebung einer solchen Wertetabelle kann zwangsläufig nur ein Anfang sein. Es stellt sich die Frage, wie sich die ermittelten Werte nun in konkrete Systemanforderungen umsetzen lassen. Bisher fehlen in der Forschung konkrete Prozessvorschläge, welche die Operationalisierung einer Wertetabelle erlauben. Entsprechend schlagen wir folgenden Operationalisierungsprozess vor: Ein VSD-Prozess sollte ganzheitlich aufgebaut sein und daher die mathematischen Teile eines DSS (Algorithmus), dessen Ausgestaltung (User Interface) sowie zusätzliches, nicht direkt im System implementiertes Material (Handbücher, Schulungen, Werbung) berücksichtigen. Wir verwenden hierfür drei Prozessschritte, anhand derer jeder Wert beurteilt und eingeteilt wird.

1. Beurteilung der betroffenen Systemkomponenten (Algorithmus, User Interface, zusätzliche Materialien)
2. Beurteilung des Ausmaßes der notwendigen Maßnahmen (niedrig, mittel, groß)
3. Falls große Veränderungen nötig sind, Beurteilung, ob diese noch konkretisiert werden müssen

Als erster Schritt sollte für jeden der erhobenen Werte aus Tab. 1 festgelegt werden, welche Systemkomponenten von diesem Wert betroffen sind. So zeigt sich bei wertbasierten Ansätzen häufig, dass nicht nur der Algorithmus oder das User Interface berücksichtigt werden müssen, sondern dass auch Anpassungen der Kommunikation an die Nutzer beispielsweise durch angepasste Handbücher nötig sind. Dies zeigt sich in unserer Wertetabelle am Beispiel des Wertes „Freiheit von Verzerrun-

gen“. Eine systematische Verzerrung (Bias) ist ein mathematisches Problem, betrifft also direkt den Algorithmus des Systems und muss dementsprechend dadurch gelöst werden. Ein Beispiel ist die schlechtere Einstufung des Kreditrisikos von Kunden mit einer bestimmten Postleitzahl, trotz vergleichbarer Finanzdaten. Eine Anpassung des Algorithmus könnte hier bereits mathematisch korrekte Beurteilungen liefern, es muss aber bedacht werden, dass Werte immer Bedürfnisse von Nutzern ausdrücken. Hier stellt sich die Frage, ob mit dem Wert noch weitere Nutzerbedürfnisse verknüpft sind. So könnten Nutzer etwa das Bedürfnis äußern, faire Kreditentscheide treffen zu wollen und diese auf einem verzerrungsfreien System abzustützen (ganz abgesehen von den unternehmerischen Konsequenzen für falsche Krediteinschätzungen). Dementsprechend bietet sich möglicherweise eine veränderte Kommunikation an Nutzer an, bspw. durch explizite Nennung von erforderlichen Datengrundlagen zur Erzielung von verzerrungsfreien Beurteilungen in Handbüchern oder im Rahmen von Schulungen. Eine andere Möglichkeit der Kommunikation stellt eine Warnung im User Interface dar, sobald eine verzerrungsfreie Beurteilung aufgrund ungenügender Datengrundlage nicht gewährleistet werden kann.

Nach der Beurteilung der betroffenen Systemkomponenten sollte festgestellt werden, wie umfangreich die Veränderungen sind, die durch die Berücksichtigung der jeweiligen Werte erforderlich sind. Als Faustregel kann hier angenommen werden, dass Werte, welche alle drei Systemkomponenten betreffen große Veränderungen erfordern. Werte, die zwei Systemkomponente betreffen erfordern mittlere Veränderungen, während Werte, die eine Komponente betreffen niedrige Veränderungen verlangen. Hier empfiehlt sich ein gewisser Pragmatismus; nicht jede durch einen Wert hervorgerufene Veränderung steht in proportionalem Verhältnis zur dadurch erzielbaren Vermeidung an ethischen Problemen. Oft zeigt sich, dass bereits durch kleine Veränderungen viel erreicht werden kann. So kann bspw. der Wert Vertrauen, welcher sich nur schwer auf einzelne Systemaspekte reduzieren lässt, bereits stark durch kurze Texte im User Interface verbessert werden. Diese Texte können Nutzern bspw. Aufschluss über einzelne algorithmische Grundlagen liefern und diesen kommunizieren, mit welchen Wahrscheinlichkeitswerten eine Krediteinstufung erfolgte. Dabei ist zu bedenken, dass die Wertvorstellungen von Entwicklern und Nutzern auseinanderlaufen können und bspw. ein Wert wie Vertrauen von Seiten der Entwickler ganz bewusst nicht konkretisiert werden soll, um zu vermeiden, dass Nutzer sich allein auf die Ergebnisse des Algorithmus verlassen und eine eigene systemunabhängige Beurteilung unterlassen.

Nach den ersten beiden Prozessschritten ist klar, welche Werte zentral für weitere Systemanpassungen sind. Für Werte, die entsprechend große Anpassungen erfordern, sind weitergehende Schritte erforderlich. Dabei sollte beurteilt werden, ob weitere Erhebungen notwendig sind, um genauer zu bestimmen, wie die notwendigen Veränderungen in das System übertragen werden können. Im Rahmen dieser Beurteilung entstehen zwangsläufig weitergehende Fragen, die konkretisiert werden müssen. Falls die Wertetabelle in einem qualitativen Prozess erhoben wurde und entsprechende Interviewdaten vorhanden sind, bieten diese häufig bereits gute Anhaltspunkte für weitere Konkretisierungen. Dieser dritte Schritt gleicht dem klassischen Prozess der Anforderungserhebung. Es ist jedoch zu betonen, dass Anforderungserhebung

rungen zwar erhoben, respektive konkretisiert werden, diese aber im Unterschied zu klassischen Anforderungserhebungsverfahren ausschließlich wertbasiert erfolgen.

2.3 Eine beispielhafte Operationalisierung an einem bestehenden System

Tab. 2 zeigt eine beispielhafte Operationalisierung, nachdem das beschriebene dreistufige Verfahren auf die zuvor erhobene Wertetabelle aus Tab. 1 angewendet wurde. Wichtig ist zu erwähnen, dass in unserem Fall die DSS bereits bestehen und genutzt werden, etwaige Änderungen daher an einem bereits laufenden System durchgeführt werden. Hierbei zeigte sich, dass innerhalb der eruierten Werte eine Hierarchie vorherrscht und es zentrale Werte wie Nachvollziehbarkeit gibt, die durch das entsprechende DSS erheblich beeinflusst werden. Andererseits gibt es auch Werte wie Neutralität, die eng mit dem spezifischen Nutzungskontext des Systems verknüpft sind.

Zur Beurteilung des Änderungsaufwands ist zu sagen, dass im sensitiven Kontext der von uns untersuchten DSS davon auszugehen ist, dass Neutralität, Zuverlässigkeit, Verzerrungsfreiheit, Compliance, Vertrauen, Haftung und Privatheit bereits grundlegende Eigenschaften des Systems sind, die bei der Entwicklung berücksichtigt wurden und dementsprechend auch von Nutzerseite noch einmal als Wert hervorgehoben wurden. Ohne die Sicherstellung von Neutralität, Zuverlässigkeit und Verzerrungsfreiheit könnten die Systeme kaum ihren Nutzen in einer schnelleren, besseren Beurteilung von potenziellen Gewalttättern ausspielen. Ihre Nutzung wäre dementsprechend zweifelhaft. Die Werte Compliance und Haftung basieren auf bestehenden Prozessen und Rechtsgrundlagen, die entweder antizipiert und dann auch entsprechender Zertifizierungen/Verfahren bedürfen, oder sich über die Zeit etabliert haben und somit durch die Systemnutzung akzeptiert wurden. Vertrauen ist

Tab. 2 Dreistufiges Beurteilungsverfahren

Wert	Systemkomponente	Änderungsaufwand	Fragen zur Konkretisierung
Nachvollziehbarkeit	A, UI, ZM	Groß	Was genau möchten die Nutzer verstehen? Einzelne Systementscheide? Etwaige Gesamturteile? Die zugrundeliegenden Berechnungen?
Neutralität	A	Niedrig	–
Zuverlässigkeit	A	Niedrig	–
Unterstützung	A, UI, ZM	Groß	Welche Form von Unterstützung wünschen sich die Nutzer?
Autonomie	A, UI, ZM	Groß	Inwiefern kann Nutzerautonomie gewahrt werden, wenn das System Entscheide vorschlägt oder teilweise übernimmt?
Verzerrungsfreiheit	A, ZM	Mittel	–
Compliance	A, ZM	Mittel	–
Vertrauen	UI, ZM	Mittel	–
Haftung	A, ZM	Mittel	–
Privatheit	A, UI, ZM	Groß	Wie kann das System datensparsam aufgebaut und konzipiert werden?

Legende: A Algorithmus, UI User Interface, ZM Zusätzliche Materialien

nicht nur eine vom System getragene Eigenschaft, Vertrauen wird vom jeweiligen Nutzer an das System herangetragen (Vertrauen ins System oder nicht), weshalb dieser Wert weniger durch den Algorithmus, als durch die Ausgestaltung des Systems und zusätzliche Materialien wie Handbücher beeinflusst wird.

Große Überarbeitungen erfordern im vorliegenden Fall die Werte Nachvollziehbarkeit, Unterstützung, Autonomie und Privatheit. Je nach Vorgehen in der empirischen VSD-Erhebung und den örtlichen Gegebenheiten sind für diese Werte nun weitere Erhebungen notwendig (ganz im Sinne des iterativen VSD-Prozesses), die eine Konkretisierung des jeweiligen Wertes erlauben, damit dieser entsprechend umgesetzt werden kann. Zu Zwecken der Nachvollziehbarkeit wünschen sich die befragten Nutzer beispielsweise mehr Informationen, wie ein Urteil des Systems zustande kommt, bspw. warum eine Person vom System auf der höchsten Gefährdungsstufe klassifiziert wurde. In diesem Kontext erweist sich die Nachvollziehbarkeit hauptsächlich als User-Interface-Problem, bei welchem Nutzern klar mitgeteilt werden muss, wie die Klassifizierung zustande kommt. Je nach System und algorithmischem Aufbau kann die Erklärung einer solchen Zuteilung weitgehende Konsequenzen haben. So muss unter Umständen ein Algorithmus angepasst werden, falls dieser in der momentanen Ausgestaltung nur schwer nachvollziehbar ist. Im Idealfall wird VSD daher bereits in frühen Entwicklungsphasen eingesetzt, um den Algorithmus direkt auf diese Werte hin zu entwickeln.

Der identifizierte Wert Unterstützung durch das System zeigte sich abhängig vom jeweiligen Anwendungsbereich der Nutzer. Die sinnvolle Ausgestaltung des Systems variiert je nach Nutzungszeitpunkt innerhalb des Beurteilungsprozesses und beigemessener Wichtigkeit: zu Beginn einer Beurteilung als erster Indikator, während des Beurteilungsprozesses zur Verifikation des Bauchgefühls, sowie zur finalen Absicherung am Ende. Zu Beginn sind u.U. weniger Informationen für Nutzer nötig, da nur der unmittelbare Handlungsbedarf ermittelt wird, während am Ende des Prozesses eine umfassendere Form der Unterstützung gewünscht wird, die zusätzliche Datenabgleiche ermöglicht. Dementsprechend betrifft dieser Wert auch alle drei aufgeführten Systemkomponenten (Algorithmus, User Interface, zusätzliche Materialien) und erfordert zwingend weitere Konkretisierungen, um abzuleiten, wie dieser Wert besser in das System integriert werden kann.

In Bezug auf den Wert Autonomie war es einigen der befragten Nutzer sehr wichtig, Entscheide weiterhin autonom fällen zu können. Sie bevorzugten dementsprechend lediglich die Versorgung mit Informationen durch das System, jedoch ohne Vorschlag eines finalen Urteils. Dieser Wert ist in unserem Fall eng mit den Werten Nachvollziehbarkeit und Unterstützung verknüpft, die je nach Nutzer, Nutzungskontext und vorhandener Expertise variieren können. Die untersuchten DSS haben den Nutzern nur Informationen zur Verfügung gestellt, der finale Entscheid wurde immer außerhalb des Systems getroffen. Doch auch hier hat sich gezeigt, dass umso besser Nutzer ein Urteil des Systems nachvollziehen können und hierzu vom System mit unterstützenden Informationen versorgt werden, desto besser können sie auch autonom Urteile fällen. In der geschilderten Situation, in der ein Wert nur einzelne, spezifische Nutzeranforderungen betrifft, ist eine weitere Konkretisierung zwingend. Mithilfe dieser Konkretisierung lässt sich dann bemessen, welche Ver-

änderungen am System notwendig sind, oder ob hier eine klarere Kommunikation über das System in Handbüchern bereits Abhilfe schaffen könnte.

Der Wert Privatheit war im hiesigen Fall sehr kontextspezifisch mit dem durch die untersuchten Systeme losgetretenen Digitalisierungsprozess verknüpft. Der vorhergehende Papierprozess liess keine grundlegende Anonymität zu, durch das DSS bestand nun aber zumindest theoretisch die Möglichkeit, einen solchen anonymen Beurteilungsprozess gewähren zu können. Hier stellten sich die Fragen, wie das System datensparsam konfiguriert werden kann und ob verschiedene Masken notwendig sind, die je nach Nutzerrolle (bspw. mit einer spezifischen Nutzerrolle für anonyme Beurteilungen) Informationen bereitstellen oder nicht.

Mit den vorstehenden Ausführungen haben wir eine Möglichkeit aufgezeigt, um eine zuvor erhobene Wertetabelle zu konkretisieren und in diesem Prozess zu beurteilen, welche Systemkomponenten in welchem Ausmaß davon betroffen sind. Damit wird die erste, eingangs angesprochene Problemstellung, die Operationalisierung einer Wertetabelle, konkretisiert und in einzelne Prozessschritte aufgelöst.

3 Hürden von VSD bei der Entwicklung von neuen Systemen

Die erste Forschungsfrage, wie die Wertetabelle eines VSD-Prozesses konkret in Systemanforderungen für DSS umgewandelt werden kann, haben wir durch den dreistufigen Operationalisierungsprozess adressiert. Die zweite Forschungsfrage, die zusätzlichen Hürden bei der Umsetzung von VSD bei neuen, noch nicht bestehenden DSS, soll in diesem Kapitel erörtert werden. Die zuvor operationalisierte Wertetabelle wurde an einem bestehenden DSS mit bestehenden Nutzern erhoben. Bei einem neu zu entwickelndem DSS ist dies aus zwei Gründen nicht in analoger Form möglich: Erstens gibt es keine aktuellen Nutzer, die Erhebung von Werten, die bei der Systementwicklung berücksichtigt werden sollen, kann daher nur auf hypothetischen Annahmen oder Erfahrungen durch erste Prototypen basieren. Zweitens sollten die Werte immer kontextspezifisch erhoben werden, eine direkte Übertragbarkeit aus anderen Kontexten ist daher schwierig, während der zukünftige Nutzungskontext des zu entwickelnden Systems sich möglicherweise noch nicht präzise abschätzen lässt. Im Folgenden greifen wir diese beiden Problematiken bei der Entwicklung von neuen DSS auf und unterbreiten konkrete Lösungsvorschläge.

3.1 Frühe Berücksichtigung von Nutzerwerten

Um das Problem der noch unbekanntem Nutzerwerte zu beheben, sollte ein sehr breiter Nutzungskontext für das neue DSS definiert und entweder verwandte Nutzungskontexte identifiziert oder der bisherige (zumeist analoge) Prozess genau durchleuchtet werden. Der große Vorteil von VSD ist dessen Systemagnostizität. Prinzipiell muss nicht einmal ein DSS vorhanden sein, da eine Wertetabelle mit entsprechender Vorsicht auch für einen papierbasierten Prozess erhoben werden kann. Eine solche Wertetabelle lässt sich nicht ganzheitlich auf das DSS übertragen, aber liefert zumindest gewisse Anhaltspunkte für erste Iterationen während der Entwicklung. Forschungsergebnisse legen hierbei nahe, dass die Werte zentraler Stakeholder mög-

lichst früh im Designprozess berücksichtigt werden sollten (Campos et al. 2020). Dementsprechend bietet es sich an, Werte aus ähnlichen oder analogen Nutzungskontexten zu erheben und zu übertragen, um bereits diese später möglicherweise systemrelevanten Werte frühzeitig berücksichtigen zu können. So kann bspw. eine Wertetabelle für ein bestehendes DSS, welches in einem bestimmten Kontext genutzt wird, als erste Grundlage für ein neues DSS im gleichen Kontext verwendet werden.

Die technologische Entwicklung zeigt aber, dass diese Vorgehensweise nicht bei allen Algorithmen gleich gut funktioniert. Lernende Algorithmen, also Algorithmen, die aus Daten lernen und ihr Verhalten anpassen, bieten ungeahnte Möglichkeiten und neue Nutzungsszenarien. Es existieren daher möglicherweise keine vergleichbaren Operationalisierungen von Wertetabellen, da die technischen Möglichkeiten bislang limitierter waren. Es zeigt sich aber, dass zahlreiche, durch lernende Algorithmen neu hinzukommende ethische Probleme vergleichbarer Natur sind wie denjenigen von herkömmlichen, regelbasierten Algorithmen, wenn auch nun in teils stärkerer Ausprägung. So sind deren Berechnungen häufig unzureichend nachvollziehbar und mehr Systemtransparenz ist vonnöten (Faraj et al. 2018). Ansätze wie Explainable Artificial Intelligence (Zednik 2019) versuchen, diese Probleme zu lösen, existieren bisher aber zumeist nur auf konzeptioneller Ebene. Daher bietet sich häufig eher ein Fokus auf den Digitalisierungskontext an (Legner et al. 2017), um prozessbasiert zu analysieren, welche Aufgaben durch das System übernommen werden sollen, damit bestehende Wertetabellen für Systeme mit ähnlichen Aufgaben als erste Grundlage benutzt werden können.

3.2 Kontextspezifität

Friedman und Kahn (2003) haben betont, dass einige Werte universell (d.h. unabhängig vom Kontext), gelten, sodass eine kontextbezogene Anpassung der Werte im Idealfall nicht mehr vonnöten wäre. Die Forschung zeigte jedoch, dass universelle Werte (und entsprechende Werttabellen) zwar eine gute Grundlage bilden können, Werte aber dennoch besser kontextspezifisch erhoben werden sollten (Dadgar und Joshi 2018; Deng et al. 2016). Diese Erfahrung lässt sich auch durch die hier vorgenommene Operationalisierung bestätigen. Die spezifischen ethischen Probleme, die durch die Nutzung eines DSS auftreten können, werden durch die Berücksichtigung universeller Werte während der Entwicklung vermindert, tragen den spezifischen Nutzungskontexte jedoch nicht ausreichend Rechnung. Gerade bei sensitiven Entscheidungsprozessen sind kontextspezifische Anforderungen jedoch häufig kritisch. Auch in der von uns illustrierten Operationalisierung zeigte sich, dass es vorgängig schwierig abzuschätzen ist, welche der Werte für das spezifische System zentral sind und möglicherweise grundlegende Veränderungen an DSS und Algorithmus verlangen. Gleichzeitig ist es auch aus ökonomischen Gründen häufig nicht möglich, alle Werte gleichermaßen zu adressieren. Im Zweifel ist hier die tiefere Abdeckung von zentralen, systemkritischen Werten der breiteren, aber dafür oberflächlichen Abdeckung von Werten vorzuziehen.

Erschwerend kommt hinzu, dass sowohl der Nutzerkreis als auch die Anwendungsbereiche eines DSS höchst heterogen sein können. Bspw. nutzt Firma A ein

System nur dazu, um eigens kalkulierte Kreditrisiken durch ein weiteres System abzustützen, während Firma B dieses System als zentrale Entscheidungsgrundlage nutzt, und Firma C mit ebendiesem System Finanzgutachten erstellt. Selbst nutzerzentrierte Ansätze wie VSD stoßen hier an Grenzen (Grønsund und Aanestad 2020). Für DSS mit weitreichenden Nutzerkreisen erscheint es praktisch unmöglich, alle Kontexte spezifisch abzudecken. Hier erweist sich die bereits zuvor angesprochene Idee universeller Werte von Friedman und Kahn (2003) hingegen als wertvoll. Wir erachten Vertrauen und Rechenschaft als die beiden zentralen Werte, deren Berücksichtigung gerade innerhalb von heterogenen Nutzungskontexten sinnvoll erscheint. Anders formuliert: Wenn ein DSS so gestaltet ist, dass es Nutzervertrauen befördert, indem beispielsweise den Nutzern adäquate Informationen zur Verfügung gestellt werden, damit diese guten Gewissens eine Entscheidung treffen und Rechenschaft darüber ablegen können, dann sollten die Spezifika des Nutzungskontexts zunehmend in den Hintergrund rücken. So hat auch die einschlägige Forschung aufgezeigt, dass für Vertrauen als Wert ein mittleres (weder zu hoch noch zu niedrig) Maß an Information nötig ist (Kizilcec 2016). Berücksichtigen sollte man hierbei jedoch bekannte menschliche Verhaltensweisen, um zu vermeiden, dass Nutzer sich ausschließlich auf das System verlassen oder Ängste verspüren, entgegen dessen Systemvorschlägen zu handeln (Aversa et al. 2018). Die Kontextspezifität für ein DSS kann abgeschwächt werden, wenn man sich während der Entwicklung auf die zwei Werte Vertrauen und Rechenschaft beruft. Ein idealer Start für die Entwicklung eines DSS würde demnach vier bis fünf Werte aus anderen Nutzungskontexte berücksichtigen und diese um die zwei universellen Werte ergänzen, um eine Grundlage für die ersten Entwicklungsschritte zu erhalten. Basierend auf den ersten Iterationen würden dann – ganz im Sinne von VSD – tatsächliche Werte der relevanten Stakeholder erhoben, um die Wertetabelle entsprechend anzupassen und in einem iterativen Prozess weiter zu verfeinern.

4 Fazit und Limitationen

Wir haben mit dieser Arbeit ein dreistufiges Beurteilungsverfahren vorgeschlagen, mit dessen Hilfe Wertetabellen operationalisiert und in Systemanforderungen umgewandelt werden können. Das Problem der noch unbekanntenen Werte, welches sich bei neu zu entwickelnden Systemen stellt, kann minimiert werden, wenn ein ähnlicher Nutzungskontext berücksichtigt wird und daraus 4–5 zentrale Werte abgeleitet werden. Das Problem der Kontextspezifität wird durch die Berücksichtigung der beiden Werte Vertrauen und Rechenschaft (zusätzlich zu den 4–5 zentralen, kontextspezifischen Werten) verkleinert. Der Lohn für diesen aufwändigen Entwicklungsprozess ist die frühzeitige Vermeidung von technologiegetriebenen ethischen Problemen, bevor größerer Schaden angerichtet wird. Gleichzeitig kann damit auch das Nutzererlebnis stark verbessert werden, da durch VSD ganzheitlich und wertbasiert verschiedenste ethische Problematiken im Vorhinein verhindert werden können.

Gestützt auf empirischer Arbeit zeigt das von uns vorgeschlagene Beurteilungsverfahren, wie Wertetabellen bei der Entwicklung von IS berücksichtigt werden können. Je nach Anwendungskontext ist es denkbar und möglich, dass sich weitere

wichtige Aspekte herauskristallisieren, oder dass nicht alle drei Stufen der Beurteilung gleich wichtig und umsetzbar sind. Zu bedenken ist dabei auch, dass durch die Auswahl der Personen, die die Operationalisierung durchführen, wiederum Verzerrungen entstehen können. Wir sehen hier hauptsächlich die Firma, die ein DSS entwickelt, in der Pflicht, um solchen Verzerrungen bei der Teamzusammenstellung Rechnung zu tragen. Insbesondere im Falle eines sehr homogenen Operationalisierungsteams können – selbst bei einer gewissenhaften Umsetzung des vorgeschlagenen Verfahrens – möglicherweise nicht alle ethischen Probleme beseitigt werden. Die in Abschn. 2.3 geschilderte Operationalisierung erfolgte aus einer konkreten Forschungsarbeit. Dementsprechend ist auch das von uns angebrachte Beispiel gewissen Limitationen unterworfen. Im Besonderen ist hier die fehlende Evaluation der Operationalisierung mit den befragten Nutzern zu nennen. Idealerweise wird auch bei diesem Operationalisierungsverfahren dem Grundgedanken von VSD Rechnung getragen und das Verfahren wird iterativ benutzt. Unser Verfahren ist dafür gut geeignet und ein Stück weit wird der iterative Gedanke auch im dritten Schritt bereits mitgedacht, das konkrete Beispiel wurde aber an einem bereits bestehenden System vorgenommen, weshalb der Prozess auch nicht direkt in die weitere Entwicklung mit einfließen konnte.

Wir sehen das von uns vorgeschlagene Operationalisierungsverfahren als weiteren Schritt, um die Praxistauglichkeit wertbasierter Designansätze weiter zu erhöhen. Dieser Ansatz kann sowohl von weiteren Forschungsarbeiten als auch von praktischen Umsetzungen profitieren, um weiter verfeinert und validiert werden zu können.

Funding Open access funding provided by University of Bern

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Albayrak CA, Renn O, Teille K (2018) Leitlinien für das menschliche Handeln in einer digitalisierten Welt. *HMD* 55:1048–1064
- Aversa P, Cabantous L, Haefliger S (2018) When decision support systems fail: Insights for strategic information systems from Formula 1. *J Strateg Inf Syst* 27:221–236. <https://doi.org/10.1016/j.jsis.2018.03.002>
- Bozdag E (2013) Bias in algorithmic filtering and personalization. *Ethics Inf Technol* 15:209–227. <https://doi.org/10.1007/s10676-013-9321-6>

- Campos JC, Fayollas C, Harrison MD, Martinie C, Masci P, Palanque P (2020) Supporting the analysis of safety critical user interfaces: an exploration of three formal tools. *ACM Trans Comput Interact* 27:1–48. <https://doi.org/10.1145/3404199>
- Dadgar M, Joshi KD (2018) The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. *J Assoc Inf Syst* 19:86–112. <https://doi.org/10.17705/1jais.00485>
- Deng X, Joshi KD, Galliers RD (2016) The duality of empowerment and marginalization in microtask crowdsourcing: giving voice to the less powerful through value sensitive design. *MIS Q* 40:279–302. <https://doi.org/10.25300/MISQ/2016/40.2.01>
- Faraj S, Pachidi S, Sayegh K (2018) Working and organizing in the age of the learning algorithm. *Inf Organ* 28:62–70. <https://doi.org/10.1016/j.infoandorg.2018.02.005>
- Friedman B, Kahn PH (2003) Human values, ethics and design. In: Jacko JA, Sears A (Hrsg) *The human: Fundamentals, evolving technologies and emerging applications*. Lawrence Erlbaum, Mahwah, S 1177–1201
- Friedman B, Kahn PH, Borning A (2008) Value sensitive design and information systems. In: Himma KE, Tavani HT (Hrsg) *The handbook of information and computer ethics*. Wiley, Hoboken, S 69–101
- Friedman B, Hendry DG, Borning A (2017) A survey of value sensitive design methods. *Found Trends Hum Comput Interact* 11:63–125. <https://doi.org/10.1561/11000000015>
- Grønsvund T, Aanestad M (2020) Augmenting the algorithm: Emerging human-in-the-loop work configurations. *J Strateg Inf Syst* 29:101614. <https://doi.org/10.1016/j.jsis.2020.101614>
- van den Hoven J (2007) ICT and value sensitive design. In: Goujon P, Lavelle S, Duquenoy P, Kimppa K, Laurent V (Hrsg) *The information society: Innovations, legitimacy, ethics and democracy*, Bd. 233. Springer, Boston, S 67–72
- Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. *Nat Mach Intell* 1:389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kizilcec RF (2016) How much information?: Effects of transparency on trust in an algorithmic interface. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, S 2390–2395 <https://doi.org/10.1145/2858036.2858402>
- Legner C, Eymann T, Hess T, Matt C, Böhmman T, Drews P, Mädche A, Urbach N, Ahlemann F (2017) Digitalization: Opportunity and challenge for the business and information systems engineering community. *Bus Inf Syst Eng* 59:301–308. <https://doi.org/10.1007/s12599-017-0484-2>
- Lüthi N, Matt C, Myrach T (2021) A value-sensitive design approach to minimize value tensions in software-based risk-assessment instruments. *J Decis Syst* 30:194–214. <https://doi.org/10.1080/12460125.2020.1859744>
- Martin K (2019a) Designing ethical algorithms. *MIS Quarterly Executive* 18: Article 5. <https://aisel.aisnet.org/misqe/vol18/iss2/5>
- Martin K (2019b) Ethical implications and accountability of algorithms. *J Bus Ethics* 160:835–850. <https://doi.org/10.1007/s10551-018-3921-3>
- Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: Mapping the debate. *Big Data Soc* 3:1–21. <https://doi.org/10.1177/2053951716679679>
- Newell S, Marabelli M (2015) Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ‘datification’. *J Strateg Inf Syst* 24:3–14. <https://doi.org/10.1016/j.jsis.2015.02.001>
- Nussbaumer A, Pope A, Neville K (2021) A framework for applying ethics-by-design to decision support systems for emergency management. *Inf Syst J*. <https://doi.org/10.1111/isj.12350>
- Robra-Bissantz S, Strahinger S (2020) *Wirtschaftsinformatikforschung für die Praxis*. HMD 57:162–188. <https://doi.org/10.1365/s40702-020-00603-0>
- Zednik C (2019) Solving the black box problem: A normative framework for explainable artificial intelligence. *Philos Technol* 34:265–288. <https://doi.org/10.1007/s13347-019-00382-7>