

Algorithmisch basierte Entscheidungsunterstützungssysteme für die deutsche Kinder- und Jugendhilfe? Messages from Research

Ackermann, Timo; Gillingham, Phillip

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Ackermann, T., & Gillingham, P. (2020). Algorithmisch basierte Entscheidungsunterstützungssysteme für die deutsche Kinder- und Jugendhilfe? Messages from Research. *Widersprüche : Zeitschrift für sozialistische Politik im Bildungs-, Gesundheits- und Sozialbereich*, 40(158), 63-78. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-91865-7>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Phillip Gillingham & Timo Ackermann

Algorithmisch basierte Entscheidungsunterstützungssysteme für die deutsche Kinder- und Jugendhilfe?

Messages from Research¹

1. Einführung

Im Laufe der letzten 25 Jahre haben Kinderschutzorganisationen weltweit und besonders in den englischsprachigen Ländern elektronische Informationssysteme genutzt, um Daten über Adressat*innen und Interventionen zu sammeln und auszuwerten. Die gleiche Entwicklung lässt sich für die Felder Gesundheitsfürsorge, Erziehung sowie der Strafgerichtsbarkeit aufzeigen. Neue digitale Technologien, die weithin als „Big Data“ bekannt geworden sind, ermöglichen es, aus multiplen Datenbasen große Datensets herzustellen und prognostisch auszuwerten. Daten aus den Feldern Gesundheitsförderung, Soziale Dienste und Strafgerichtsbarkeit wurden in jüngerer Vergangenheit bereits über Technologien des „machine learning“ miteinander relationiert, um algorithmisch gestützte Entscheidungsunterstützungssysteme, „decision support tools“ (DST)² zu entwickeln (vgl. z.B. Allegheny County 2017; Wijenayake/Graham/Christen 2018). Auch im deutschsprachigen Raum findet die Digitalisierung Sozialer Arbeit zunehmend Beachtung. Die Einführung von algorithmisch gestützten Entscheidungsunter-

-
- 1 Der Text versucht Argumente und Positionen – die in den letzten Jahren von Philip Gillingham und anderen vor allen Dingen im englischsprachigen Raum vertreten wurden – für weitere Diskurse in Deutschland verfügbar zu machen und an aktuelle Diskurse anzuschließen.
 - 2 Es wird hier und im Folgenden die im internationalen Kontext gebräuchliche Abkürzung „DST“ verwendet und der Begriff „Entscheidungsunterstützungssystem“ gleichbedeutend verwendet. Eine Übersetzung als „Entscheidungsunterstützungsinstrumente“ wäre ebenfalls naheliegend gewesen; wir haben uns für „Entscheidungsunterstützungssysteme“ entschieden, weil dies unseres Erachtens besser den vernetzten, relationalen Charakter der entwickelten Instrumente verdeutlicht.

stützungssystemen wird in diesem Kontext bisweilen als „langfristig (...) nicht mehr aufzuhalten“ bewertet (Schrödter/Bastian/Taylor 2020: 256).

Das Ziel von DST besteht im Feld Sozialer Dienste typischerweise darin, vulnerable Adressat*innen frühzeitig zu identifizieren, um präventive Interventionen zu ermöglichen (Sanders u.a. 2017). DST entbinden menschliche Akteur*innen von der Interpretation der Daten; die Entscheidung, ob Interventionen angesichts der Empfehlungen der DST durchgeführt werden sollten, verbleibt aber bei den zuständigen Personen. Die Behauptung der Entwickler*innen besteht zumeist darin, die entwickelten Technologien seien menschlichen Akteur*innen überlegen, weil sie vielfältigere Informationen in kürzerer Zeit verarbeiteten und daher in ihrer Risikoeinschätzung akkurater und zuverlässiger seien. Postuliert wird, DST seien der nächste Schritt in der aktuariellen Einschätzung von Risiken in Kinderschutzorganisationen (Macchione et al. 2013; Schrödter/Bastian/Taylor 2018, 2020).

Bislang gibt es allerdings unserer Kenntnis nach nur wenige Forschungen aus dem Feld des Kinderschutzes, um Entscheidungsträger*innen in ihren Entscheidungen zu unterstützen, ob sie sich in der Implementierung von DST engagieren sollten. Einige Studien und kritische Kommentierungen sind gleichwohl publiziert worden. In diesem Artikel gehen wir zunächst auf Beispiele für die versuchten Implementierungen von DST ein. Das Ziel besteht darin, einen kritischen Überblick über aktuelle Entwicklungen und Debatten zu geben. Zugleich soll ein Beitrag zur Frage geleistet werden, ob die Nutzung von algorithmisch gestützten Entscheidungsunterstützungssystemen auch in der deutschen Kinder- und Jugendhilfe vermehrt vorangetrieben werden sollte – oder nicht.

Im Rahmen eines Zeitschriftenartikels ist es notwendig, z.T. komplexe Phänomene zusammenfassend darzulegen. Wo immer möglich, werden wir jedoch Referenzen zu vollständigeren Darstellungen herstellen. Wir beziehen uns auf Forschungsberichte, Zeitungs- und Fachartikel, die bislang vor allem im englischsprachigen Raum erschienen sind sowie auf einige neuere Publikationen aus dem deutschsprachigen Raum. Zur Veranschaulichung des Problemkontextes gehen wir zunächst auf einige Beispiele für DST ein, die vor allem in den USA, in England, Neuseeland aber auch in den Niederlanden verwendet werden. Hierauf folgt eine Diskussion von Problemdimensionen, die durch die Entwicklung, Implementierung und Nutzung von DST hervorgebracht werden. Um es gleich vorwegzunehmen: Bislang gibt es unseres Wissens keine unabhängigen Publikationen, die die Effizienz von DST in Kinderschutzorganisationen nachweisen konnten. Insgesamt kommen wir daher zu einer skeptischen Einschätzung in Bezug auf die Einführung von algorithmisch gestützten Entscheidungsunterstützungssystemen.

2. Exemplarische Blickschneisen auf die Einführung algorithmisch basierter Entscheidungsunterstützungssysteme

Das „Allegheny Family Screening Tool“, Philadelphia (USA)

Die folgenden Ausführungen zum „Allegheny Family Screening Tool“ (FST) beruhen auf einer Evaluation der lokalen Behörde (Allegheny County 2018) sowie auf der Fallstudie von Chouldechova u.a. (2018). Das FST wurde in Allegheny County (einem Vorort von Philadelphia, USA) entwickelt. Basierend auf Daten von Kindern und ihren Familien, die bereits in Kontakt mit Kinderschutzorganisationen standen, berechnet das FST die Wahrscheinlichkeit, mit der ein Kind innerhalb von zwei Jahren (abhängige Variable) wieder bei den Kinderschutzbehörden als Opfer von Misshandlungen oder Vernachlässigung gemeldet wird. Das FST zielt insofern darauf ab, die vulnerabelsten, mit hoher Wahrscheinlichkeit gefährdeten Kinder in einer Gruppe relevanter Personen zu identifizieren. Der Algorithmus nimmt für sich in Anspruch, über 100 kindbezogene Faktoren zu berücksichtigen. Das DST nutzt dabei Daten gewissermaßen aus einem „Gemischtwarenladen“ von 29 verschiedenen Quellen, darunter Daten der Kinder- und Jugendhilfe, aber auch Daten aus der Hilfe für psychisch kranke und suchtkranke Menschen, sowie Daten aus der Strafgerichtsbarkeit.

Die schiere Menge der berücksichtigten Daten und Faktoren mag beeindruckend sein, gleichwohl wäre ein genauerer Blick auf die Konstruktion des Algorithmus notwendig, um das Instrument bewerten zu können. Bislang sind allerdings, was ein typisches Problem darstellt, keine Informationen darüber zugänglich, welche Gewichtung das FST den verschiedenen Faktoren zuspricht. An anderer Stelle wurde zudem darauf verwiesen, dass Algorithmen häufig nur eine geringe Zahl von Faktoren verlässlich als Prädiktoren nutzen können, wenn es darum geht, administrative Daten zu verwenden (Gillingham 2017; Wijenayake/Graham/Christen 2018).³ Diese erste exemplarische Betrachtung verdeutlicht: In der Bewertung von Entscheidungsunterstützungssystemen gilt es, sich nicht von der schieren Menge an Faktoren und Variablen beindrucken zu lassen, vielmehr gilt es die prädiktiven Fähigkeiten von Variablen und Faktoren zu hinterfragen, um die Effektivität und praktische Verwendbarkeit eines Instrumentes bewerten zu können.

3 Für Kinderschutzorganisationen: frühere Kontakte mit den Kinderschutzbehörden sowie die Dauer des Bezugs von Sozialleistungen und der Status als Alleinerziehende, wobei die letzteren beiden Indikatoren für Armutsverhältnisse sind (Gillingham 2017).

Das „Rapid Safety Feedback Programme in Chicago“ (RSF) (USA)

Zum RSF liegen nach unserem Wissensstand bislang keine wissenschaftlichen Evaluationen vor. Unsere Angaben basieren auf einem Zeitungsartikel, der am 6. Dezember 2017 im Chicago Herald veröffentlicht wurde. Die Einführung des RSF hat große mediale Aufmerksamkeit erregt, und im Internet finden sich zahlreiche Berichterstattungen zum RSF, aber alle scheinen auf einen Artikel von Gloor (2017) zurückzugehen. Der Artikel berichtet davon, wie das Illinois Department of Children and Family Services entschied, die Nutzung des RSF wegen Fehlern in seinen Prädiktoren wieder zu stoppen. Das RSF gehört zum Typus der supervised learning algorithm. Der Algorithmus nutzt Daten aus dem Feld der Kinder- und Jugendhilfe, um besonders gefährdete Kinder zu identifizieren, sobald sie in der Kinder- und Jugendhilfe auffällig werden. Dabei wird jedem Kind ein Wert auf einer Skala bis 100 Punkten zugeordnet.

Kurz zusammengefasst überschätzte das RSF die Wahrscheinlichkeit von Misshandlung für 1000 Kinder. Das führte dazu, dass die zuständige Behörde mit einer Vielzahl von zu bearbeitenden Fällen überfordert war und infolgedessen das Risiko in anderen Fällen unterschätzte, in denen Kinder starben oder schwerwiegende Schädigungen erlitten. Das RSF ermittelte z.B. eine 100%ige Wahrscheinlichkeit eines Todes oder einer schwerwiegenden Verletzung für 369 Kinder innerhalb der nächsten zwei Jahre. Eine solche Berechnung lässt natürlich die Frage offen, wie in solchen Fällen konkret vorgegangen werden soll, und überlässt den (aufwändigen und Aufmerksamkeit bindenden) Umgang mit der Prognose letztlich den zuständigen Sozialarbeiter*innen.

Das RSF wurde von einem privaten Unternehmen entwickelt, und der Algorithmus steht nicht für öffentliche Untersuchungen zur Verfügung. Es ist daher schwer zu bestimmen, warum es zur Überschätzung von Wahrscheinlichkeiten kam. Angesichts dessen, dass das RSF Daten aus dem Feld der Kinder- und Jugendhilfe verwendete und periodisch (oder sogar laufend) mit Daten aus dem gleichen Feld weiter „lernte“, dabei aber *zunehmend* Fehleinschätzungen produzierte, kann vermutet werden, dass das RSF einer Selektionsverzerrung („Sample Selection bias“) zum Opfer fiel.

Zu einer Verzerrung durch Stichprobenselektion kommt es, wenn genutzte Daten selbst auf die eine oder andere Form bereits einem Bias unterliegen (Berk 1983). In Daten aus der Kinder- und Jugendhilfe ist davon auszugehen, dass Kindesmisshandlung und Vernachlässigung überdurchschnittlich häufig vorkommen. Der Algorithmus lernt auf diese Weise alle möglichen Formen von Faktoren mit Kindesmisshandlung und Vernachlässigung zu verbinden, ohne dass in der

allgemeinen Bevölkerung solche Faktoren zu Kindesmisshandlung oder Vernachlässigung führen müssen. Im Laufe der Zeit, in den Selbst-Lern-Schleifen des Algorithmus, kann sich dieser Bias einschleifen und verschärfen, was zu zunehmend fehlerhaften Einschätzungen von Risiken führen kann. Der Leiter der Behörde erklärt das Programm jedoch insbesondere deshalb für gescheitert, weil gerade besonders schwerwiegende Fälle nicht vorhergesagt werden konnten (Gloor 2017).

Das „Hackney’s Early Help Profiling System Social“, London (GB)

Dencik u.a. (2018) beschreiben das Early Help Profiling System (EHPS), das in London Verwendung fand, wie folgt: Der Bezirk Hackney County Council ging mit privaten Firmen eine Partnerschaft ein, um das EHPS zu entwickeln und – ähnlich wie oben berichtet – gefährdete Kinder und Familien so früh wie möglich zu identifizieren. Zu der Entwicklung und Evaluation des EHPS sind keine wissenschaftlichen Studien verfügbar, allerdings eröffnet ein Zeitungsartikel von Ed Sheridan (2019) im „Hackney Citizen“, einer lokalen Zeitung, interessante Einsichten.

Sheridan (2019) berichtet, wie das EHPS genutzt wurde, um in der allgemeinen Bevölkerung, die noch nicht in Kontakt mit der Kinder- und Jugendhilfe stand, vulnerable Kinder zu identifizieren. Dazu wurden Daten aus dem Kontext der Hilfesysteme sowie Angaben über abweichendes Verhalten aus anderen Kontexten verwendet, etwa Daten über häusliche Gewalt sowie über Schulschwänzen und -ausschlüsse.

Die Nutzung des EHPS wurde zwischenzeitlich gestoppt, vor allen Dingen aus zwei Gründen: Erstens erreichte seine Leistung nicht die erwarteten Standards, und zweitens waren die Bürger*innen nicht darüber informiert worden, wie ihre Daten genutzt wurden, was im Nachhinein auch von politisch Verantwortlichen als Missachtung von Persönlichkeitsrechten anerkannt wurde. Sheridan (2019) berichtet zudem über Zweifel an der Qualität der von EHPS genutzten Daten. Die private Firma, die mit dem Bezirk Hackney County zusammenarbeitete, erhielt umgerechnet etwa 410.000 € für ein Instrument, welches nie effektiv zur Anwendung gebracht werden konnte (Sheridan 2019).

Das „Predictive Risk Modelling“ (PRM) in Neuseeland

Die folgenden Ausführungen zum PRM fassen die wichtigsten Eckpunkte von zwei Publikationen zusammen, die an anderen Orten erschienen sind (Gillingham 2016, 2017). Das PRM wurde als Teil einer größeren Reform des Kinderschut-

systems von einem Team der Universität Auckland entwickelt (CARE, 2012; Vaithianathan u.a. 2013), im deutschsprachigen Diskurs wurde es bereits als erfolgreiches Beispiel für DST behandelt (Schrodter/Bastian/Taylor 2018: 9). Die Entwickler*innen nutzen Daten aus den Feldern der Kinder- und Jugendhilfe („Child Welfare“), Gesundheitsfürsorge, Erziehung und Sozialhilfesysteme. Der mit diesen Daten lernende Algorithmus sollte dazu dienen, Kinder in Gefährdungslagen zu identifizieren, sobald ihre Eltern zum ersten Mal Sozialhilfeleistungen beantragten. Im Anschluss an die Identifizierung sollte dann interveniert werden, um Misshandlungen zu stoppen und Hilfeleistungen zu installieren (Ministry of Social Development in New Zealand 2014a, 2014b).

Die Beteiligten entwickelten einen Algorithmus, dessen Genauigkeit mit 76% angegeben wurde, und es wurden Pläne gemacht, einen Versuchslauf sowie die Implementierung anzugehen. Allerdings erzeugte der PRM vielfältige Kritik in den Massenmedien und der Wissenschaft. In Frage gestellt wurde: (a) die ethische Zulässigkeit Daten zu verwenden, die für einen anderen Einsatz vorgesehen waren und zu deren Verwendung die Betroffenen in dieser Hinsicht nicht zugestimmt haben, (b) die Stigmatisierung von Kindern und Familien, die durch das PRM identifiziert werden, (c) die moralische und rechtliche Basis von Interventionen, die auf Algorithmen beruhen und (d) der Mangel an Transparenz darüber, wie das PRM entwickelt wurde und wie es operierte (Dare 2013; Keddell 2015).

Die Pläne für eine Erprobung des Instrumentes unter Echtzeitbedingungen als reine Beobachtungsstudie, in dem die Verlässlichkeit des Instrumentes an einer Gruppe von 60.000 Neugeborenen getestet werden sollte, wurden zurückgestellt: dies vor allem angesichts politischer Bedenken. Die intervenierende Ministerin wird mit den Worten zitiert „these are children, not lab rats“ (Jones 2015). Einige weitere Schwachstellen des Instrumentes lassen sich festhalten; in Kürze: Das Datenset enthält z.B. Daten über Kinder, die Opfer von Misshandlungen wurden, aber auch Daten von Geschwistern, die Opfer von Misshandlungen wurden sowie Daten von Kindern, die Hilfen für psychisch Kranke erhielten, und von unbegleiteten, minderjährigen Geflüchteten. Infolgedessen hätte der Algorithmus wohl in der Tat zu einer Über-Identifizierung von vermeintlich von Misshandlung gefährdeten jungen Menschen geführt.

3. Diskussion: Probleme in der Implementierung von Entscheidungsunterstützungssystemen

Ethische Perspektiven

In den letzten Jahren wurde der Einsatz algorithmisch basierter Entscheidungsunterstützungssysteme unter Verwendung von Daten aus dem Feld der öffentlichen Sozial- und Gesundheitsbereiche vielfältig kritisiert (Kedell 2015; de Haan and Connelly 2014; Beranek u.a. 2019). Dabei wurden einige Anstrengungen unternommen, den Einsatz von DST unter ethischen Gesichtspunkten zu reflektieren (Leslie u.a. 2020; Drake/Jonson-Reid 2018; Dare 2013). Als unethisch wird vor allem kritisiert, dass die Entwicklung von DST die Privatsphäre der Bürger*innen verletzt; Daten wurden immer wieder ohne entsprechendes Einverständnis genutzt (bzw. nur mit Einverständnis für einen anderen Zweck). Für eine ethisch sensible Umsetzung von DST wird gefordert, dass alle Prozesse der Entwicklung wie auch der Nutzung von DST transparent sein müssten und auch offen für Untersuchungen (im Gegensatz zu den oben vorgestellten Beispielen). Dabei würde eine Offenlegung des Algorithmus nicht ausreichen; vielmehr müssten Datengrundlagen und soziotechnische Verwendungszusammenhänge darüber hinaus nachvollziehbar sein (Zweig 2019). Entwicklungsprozesse wären zudem unter Einbeziehung wichtiger Stakeholder, inklusive der Adressat*innen (Eltern und Kindern) zu führen (Leslie u.a. 2020; Devliegere and Gillingham, in press; Schneider/Seelmeyer 2019).

Nachvollziehbarkeit von Entscheidungsprozessen

Die Entwicklung von Entscheidungsunterstützungssystemen ist eng verbunden mit Vorstellungen von Darstellbarkeit, Verantwortlichkeit und Nachvollziehbarkeit von Entscheidungsprozessen über das Kindeswohl (Ackermann 2017, 2020b): Als Vertreter*innen des Staates und der öffentlichen Verwaltung sind Sozialarbeiter*innen Tag für Tag herausgefordert, ihre Entscheidungspraxis gegenüber multiplen Öffentlichkeiten (Kolleg*innen, Nutzer*innen, Richter*innen, andere Professionelle, ...) nachvollziehbar zu machen. Auch für die Entwicklung und den Gebrauch von Algorithmen sind bereits Richtlinien entwickelt worden, die eine verantwortliche, transparente und nachvollziehbare Nutzung von Algorithmen überhaupt (USACM 2017) sowie im Feld öffentlicher Verwaltung (Djeffal 2018) absichern sollen; in der Praxis zeigen sich allerdings einige Probleme, solche Grundsätze zu verwirklichen (Gillingham 2019b; Binns 2018).

Grundsätze zum Umgang mit Algorithmen verlangen, dass algorithmisch basierte Entscheidungen – wie auch menschliche Entscheidungen – nachvollzieh-

bar sein müssen. Ananny und Crawford (2016) verweisen auf die Limitationen algorithmischer Transparenz, z.B. darauf, dass die Offenlegung und Einsicht in einen Algorithmus keinesfalls mit Transparenz verwechselt werden darf. Die Funktionsweise eines DST mag, selbst wenn sie offengelegt ist, zumindest zu Beginn für die Sozialarbeiter*innen schwer zu erklären und für die Nutzer*innen häufig kaum zu verstehen sein. Gerade komplexe Entscheidungsunterstützungssysteme (z.B. Amrit u.a. 2017) werden vermutlich gleichermaßen das Verständnis von Sozialarbeiter*innen und Nutzer*innen überfordern. Wiederkehrend wird daher gefordert, den Einsatz von DST aus dem Grund mangelnder Nachvollziehbarkeit der Entscheidungsprozesse zu unterlassen (Binns 2018; Ananny/Crawford 2016). Die Nachvollziehbarkeit algorithmisch gestützter Entscheidungsprozesse stellt, so möchten wir festhalten, gemeinsam mit der problematischen Frage der Datennutzung, eine wesentliche Herausforderung in der Realisierung von DST in der Sozialen Arbeit dar.

Problematische Prädiktoren

Wie bereits oben angedeutet, sind die von Algorithmen verwendbaren Prädiktoren zur Bestimmung von Gefährdungen begrenzt – trotz der Behauptung einiger Entwickler von DST, die Exaktheit von Vorhersagen könne durch die Verwendung multipler Datenquellen verbessert werden (Wijenayake/Graham/Christen 2018). Das oben besprochene PRM in Neuseeland verfügt z.B. über drei Hauptprädiktoren: Frühere Kontakte mit Kinderschutzbehörden („Child welfare agencies“), die Beziehungssituation der Sorgeberechtigten (wobei Alleinerziehende für ein höheres Risiko stehen) und die Dauer von Sozialhilfebezügen (Gillingham 2017). An anderer Stelle wurde bereits darauf hingewiesen, dass die Nutzung des Familienstatus als Risikofaktor zur sich selbst verschärfenden Stigmatisierung von Alleinerziehenden beitragen kann (De Haan and Connelly 2014). Mit Blick auf das in Neuseeland entwickelte PRM lässt sich z.B. festhalten, dass das DST, über die Verwendung von Armut्सindikatoren (Sozialhilfebezug, Hilfesgeschichte, ...), eher dazu geeignet scheinen, von Armut betroffene Familien zu identifizieren als dazu, Interventionsentscheidungen zu begründen (TCC Group 2015: 5). Die Nutzung vormaliger Kontakte zu Kinderschutzorganisationen als Prädiktoren kann zudem häufig zu Ungenauigkeiten führen, weil in der Regel längst nicht alle Meldungen in der Feststellung von Misshandlungen oder der Installation von Hilfen münden (Gillingham 2020a). Einige Meldungen werden sich immer als unberechtigt herausstellen und einige mögen sogar, wie Eubanks (2018) zeigt, üble Nachrede sein. Den bisher entwickelten Algorithmen fehlt bislang die Fä-

higkeit, zwischen Kontakten zu unterscheiden, die ein höheres Risiko indizieren und solchen, die dies nicht tun.

Biases und sich selbst verschärfende Ausgrenzung

Alle Aspekte aus dem vorherigen Abschnitt führen zur Problematik von kognitiven Verzerrungen und systematischen Fehlern bzw. Messabweichungen („biases“). Im Anschluss an die Überlegungen aus dem obigen Absatz und etwas vereinfachend: Die Meldung über ein Kind aus einer relativ armen Familie wird mit höherer Wahrscheinlichkeit nachverfolgt, untersucht und bekräftigt werden als die Meldung zu einem Kind aus einem vermögenden Elternhaus. Eubanks (2018) weist ebenfalls darauf hin, dass ärmere Familien mit größerer Wahrscheinlichkeit untersucht werden, weil sie sich keinen Zugriff auf private Unterstützungssysteme leisten können, und staatliche Unterstützungsagenturen mit höherer Wahrscheinlichkeit Meldungen an Kinderschutzorganisationen übermitteln. Darauf programmiert, historische Daten zu nutzen, spiegelt sich diese Perspektivierung und – gerade im Fall von Armut und vorherigen Kontakten – verschärft das DST möglicherweise sogar diese Aufmerksamkeitsverzerrungen. Weitere Bias können bereits in den Daten eingeschrieben sein und z.B. dazu führen, dass die Marginalisierung von ethnischen Gruppen unter dem Deckmantel der Verwendung von „Big Data“ fortgeschrieben oder sogar verschärft wird (Crawford 2013, Speyer 2017).

Problematische Daten

Die in Kinderschutzorganisationen gesammelten Daten sind immer wieder als lückenhaft, wenig akkurat und widersprüchlich beschrieben worden. Unübersichtlichkeit und Un-einheitlichkeit erschweren es, Algorithmen, Muster und Korrelationen zu erkennen und Vorhersagen zu treffen (Gillingham 2020b). Mehr noch: Eine Vielzahl von Daten mag, bei großer Inkonsistenz, für eine Verwendung im Kontext von Entscheidungsunterstützungssystemen gänzlich unbrauchbar sein. Weitere Probleme entstehen, weil Messungenauigkeiten häufig nur schwer zu bestimmen sind und lokale Unterschiede darin bestehen, wie Kinderschutzorganisationen auf Meldungen und Fallkonstellationen reagieren (Gillingham 2009).

Überdies nutzen Entscheidungsunterstützungssysteme Daten aus dem Kontext der Verwaltung, die Hilfeprozesse im Kontext von Kindesmisshandlungen häufig nur ungenau abbilden bzw. eigene Realitäten erzeugen (vgl. Wolff 2004; Ackermann 2020b). Dies war z.B. bei der Entwicklung des PRM in Neuseeland der Fall: Die Effekte von Hilfen werden in den Datenbanken nicht erfasst, eher wird

ein „Fall abgeschlossen“ vermerkt. Auch in den deutschen Jugendämtern werden solche fallbezogenen Statistiken geführt, ohne die Effekte von Hilfen genauer zu erfassen. Den Abschluss eines Falls als Ersatzvariable für die Beendigung einer Kindesmisshandlung zu behandeln, muss allerdings als problematisch bewertet werden, weil beide Phänomene nicht bedeutungsgleich sind. Probleme bestehen insofern nicht nur in der Repräsentativität einzelner Faktoren, sondern auch in der Repräsentativität der in der Verwaltung erhobenen bzw. verfügbaren Daten (Gillingham 2020b).

Daten für einen anderen Zweck zu verwenden als für den, für den sie erhoben wurden, kann insofern zu weiteren Problemen führen. Dennoch wird ein solches Vorgehen immer wieder befürwortet, schon alleine, weil die Daten bereits vorliegen und daher (vermeintlich) einfach zu gebrauchen sind (Salganik 2018).

Schrödter/Bastian/Taylor (2018; 2020) berichten einerseits kritisch (dieselben 2018: 9-11; dieselben 2020: 260) und vielleicht auch mit einer gewissen Faszination andererseits von der neuen Macht der „Digitalisierungsrevolution“ (Bastian/Schrödter 2019: 40): „Wenn wir davon ausgehen, dass das Risiko zukünftiger Ereignisse, wie etwa einer Kindesmisshandlung, treffsicher und auch sinnvoll nur durch statistische Verfahren prognostiziert werden kann (...), wäre Big Data Analytics folglich die beste Wahl“ (Bastian/Schrödter 2019: 46). Daten könnten im Zuge maschinell gestützter Prozesse miteinander kombiniert und ohne größeren Aufwand im Zuge von „Big Data“ Analysen genutzt werden. Dies könne „einen dynamisch anwachsenden Datenpool abgeben“ (Schrödter/Bastian/Taylor 2018: 7).

„etwa die Anzahl und die Dauer der Besuche einer Sozialpädagogischen Familienhilfe, aufgezeichnete Interviews mit den Eltern und Kindern, bisher erhaltene Sozialleistungen der Eltern, Verspätungen oder Versäumnisse bei Untersuchungen, medizinische oder therapeutische Daten (Unfälle, Entwicklungsverzögerungen, Verhaltensstörungen (...), Impfungen, Medikamentierungsgeschichte usw.), Selbsterzeugnisse der Kinder oder Entwicklungsberichte von Fachkräften aus Kindertageseinrichtungen oder Grundschulen, Schuleingangsuntersuchungen des Gesundheitsamtes, die öffentlich zugänglichen Aktivitäten der Eltern und Kinder in sozialen Netzwerken (...). Big Data Analytics kann diesen Datenpool nutzen, um sie nach Mustern zu durchforschen, die Hinweise auf Kindeswohlgefährdung geben.“ (ebenda)

Angesichts der weiter oben ausgeführten Gründe kann einem solchen Ansatz hier nicht gefolgt werden. Der Einsatz von zu anderen Zwecken erhobenen Daten, zumal in so weitreichender Form wie hier angedeutet, muss auf ethische und rechtliche Bedenken stoßen. Es ergeben sich zudem praktische Probleme, die hier ebenfalls mit Blick auf die oben untersuchten Beispiele angedeutet wurden, kann eine solches Vorgehen doch nicht zuletzt zu ungenauen Risikobestimmungen führen, die Kinderschutzorganisationen eher be- als entlasten.

Wir gehen angesichts der bisherigen Erfahrungen davon aus, dass – wenn nützliche, verlässliche und genaue DST für Kinderschutzorganisationen entwickelt werden sollen –, *dies den Aufbau eines maßgeschneiderten Datensets erfordern würde*. Im Kontext von Kinderschutzorganisationen würde dies z.B. bedeuten, Daten über Hilfeprozesse und Effekte zu produzieren – im Unterschied zu administrativen Daten, die eigene Realitäten herstellen und dabei Handlungsschritte der Verwaltung und nicht unbedingt Fall-Realitäten reflektieren (vgl. Wolff 2004; Ackermann 2017). Elektronische Dokumentationssysteme müssten aufwändig neu designt werden, Sozialarbeiter*innen wären darin auszubilden, Daten anders als bisher zu erheben. Und es würde eine gewisse Zeit brauchen, um ausreichende Mengen an Daten in einem neuen elektronischen Dokumentationssystem aufzubauen. Nicht zuletzt wären alle wichtigen Stakeholder in transparente Prozesse einzubinden. Dies wäre, so viel dürfte klar geworden sein, eine sehr kosten- und zeitintensive Aufgabe.

Verwendungsweisen von DST im Alltag

Empirische Forschungen haben in den letzten Jahren immer wieder gezeigt, dass Instrumente zur Risikoeinschätzung im Kinderschutz bei weitem nicht immer so genutzt werden, wie dies die Entwickler*innen entsprechender Instrumente intendierten (vgl. Bastian 2017; Ackermann 2020a, 2020b). Sozialarbeiter*innen werden in einen intensiven Austausch mit Risikoeinschätzungsinstrumenten verwickelt, verhalten sich aber durchaus eigensinnig. Sie beeinflussen die Instrumente so, dass favorisierte Ergebnisse und Einschätzungen erzeugt werden; oder sie ignorieren die Empfehlungen aus einer Vielzahl von Gründen, z.B. weil ihnen keine Möglichkeiten zur Verfügung stehen, entsprechende Hilfen anzubieten (Gillingham 2009; Ackermann 2017).

Wie weiter oben ausgeführt, ist ein wichtiger Aspekt in der Nutzung von DST in der Nachvollziehbarkeit der Algorithmen zu sehen; Nachvollziehbarkeit muss aber nicht notwendigerweise zum Vertrauen der Sozialarbeiter*innen in die Einschätzungen des Instrumentes führen, die z. T. von „Pseudo-Mathematik“ sprechen (Ackermann 2017, 2020b). In der Studie zur Verwendung des DST in Allegheny County gibt es Hinweise darauf, dass Sozialarbeiter*innen ein Viertel der Empfehlungen von DST zurückwiesen und außer Kraft setzten (Chouldechova u.a., 2018). Selbst wenn die Instrumente so funktionierten wie geplant und alle Probleme ausgeräumt würden, bliebe immer noch das Problem der Integration in praktische Handlungsabläufe und Entscheidungsprozesse.

4. Fazit: Algorithmisch gestützte Entscheidungsunterstützungssysteme für die deutsche Kinder- und Jugendhilfe?

Die Entwicklung und Nutzung von DST kann als noch in den Kinderschuhen steckend beschrieben werden – es mag daher voreilig sein zu urteilen, dass DST ungeeignet sind, um Sozialarbeiter*innen im Kinderschutz zu unterstützen. Festhalten lässt sich aber: Es handelt sich um eine „Kindheit“, die durch ethische, rechtliche und praktische Probleme gekennzeichnet ist.

Mit Blick auf die Zukunft – zur Verbesserung der Praktikabilität und Anwendbarkeit von DST – wären eine ganze Reihe von Aspekten zu beachten. Entwickler*innen und Politiker*innen müssten sich von der ineffektiven und unethischen Praxis verabschieden, in und zwischen staatlichen Akteur*innen Daten auszutauschen und zu kombinieren, die zu einem anderen Zweck erhoben wurden, um diese als Grundlage zur Etablierung von DST zu verwenden. Genauer gesagt müsste es darum gehen, konsistentere und komplettere Datensets aufzubauen, deren Verwendung den Bürger*innen umfänglich transparent zu machen wäre und zu deren Verwendung das Einverständnis des Betroffenen erreicht werden müsste. Dies würde aber eine massive Umorganisation elektronischer Berichtssysteme und ihrer Verwendungsweisen in Kinderschutzorganisationen verlangen und enorme organisatorische und fiskalische Anstrengungen über Jahre hinweg bedeuten. Selbst wenn diese Investitionen getätigt würden, gibt es keine Garantie, dass die DST in der alltäglichen Arbeitspraxis der Jugendämter wirklich hilfreich sein können.

Angesichts der immensen notwendigen Investitionen und der vielfältigen ethischen, rechtlichen und praktischen Probleme wäre es möglicherweise sinnvoller, vermehrt anderweitig in die Systeme der Kinder- und Jugendhilfe zu investieren, z.B. in Fort- und Weiterbildung, in die Gewinnung, Supervision und Unterstützung des Personals sowie in eine insgesamt qualitativ hochwertige Kinder- und Jugendhilfe.

Zudem würde durch die weitere Einführung von algorithmisch gestützten DST in Deutschland eine Risiko-Orientierung in der Kinder- und Jugendhilfe weiter befeuert, die sich als „Kindeswohlgefährdungsvermeidung“ verstehen lässt (Dahmen/Kläsener 2019). Eine Tendenz, mit der in fachlichen Diskursen das Vermeiden von Risiken und vermeintliche Sicherheiten in den Vordergrund gestellt werden. Prozesse der Analyse und des Verstehens von strukturbedingten Ursachen von Gewalt und möglichen Kindeswohlgefährdungen werden dabei jedoch in den Hintergrund gedrängt. Die Aufmerksamkeit liegt auf der Frage, wie wahrscheinlich die KWG (vermeintlich) ist. Nicht so sehr auf der Frage, wie es zu ihr kommt und was angesichts dessen zu tun wäre.

Insgesamt lässt sich eine recht große Aufregung und auch viel wohlmeinender Enthusiasmus rund um den Komplex von „Big Data“, künstlicher Intelligenz, „machine learning“, automatisierter Entscheidungsprozesse und Entscheidungsunterstützungssysteme feststellen. Nicht selten werden die beschriebenen Entwicklungen als notwendige, nächste Schritte dargestellt. Wir täten allerdings gut daran, eine „skeptische“ (Scholz 2019) Betrachtungsweise einzunehmen, um nüchtern zu bewerten, ob DST in der Tat eine Unterstützung für Sozialarbeiter*innen darstellen und ob sie den Aufwand rechtfertigen, den ihre Implementierung verlangt.

Literatur

- Ackermann, Timo 2020a: Digitalisierung in der Kinder- und Jugendhilfe und im Kinderschutz: Von Risikoeinschätzungsbögen über Fallbearbeitungssoftware bis zu Big Data. *Soziale Passagen* 12, 171-177
- 2020b: „Einige Ambivalenzen des Entscheidens über das Kindeswohl. Zwischen ‘Fallzuständigkeit’, ‘Informiertheit’ und ‘Pseudo-Mathematik’“, in: Helga Kelle/Stephan Dahmen (Hg.), *Ambivalenzen des Kinderschutzes. Empirische und theoretische Perspektiven*, Weinheim, 20-41
- 2017: Über das Kindeswohl entscheiden. Zur Fallarbeit im Jugendamt, Bielefeld
- Allegheny County 2017: *Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening Decisions*, Pittsburgh, PA, Allegheny County Department of Human Services. Verfügbar unter: <https://www.alleghenycounty-analytics.us/index.php/2019/05/01/developing-predictive-risk-models-support-child-maltreatment-hotline-screening-decisions/>[15.09.20]
- Amrit, C., Paauw, T., Aly, R. and Lavric M. 2017: Identifying child abuse through text mining and machine learning. *Expert Systems with Applications*, 88, 402-418
- Ananny, M., and Crawford, K. 2016: Seeing Without Knowing: limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society*. doi:1461444816676645
- Bastian, P. 2017: Negotiations with a risk assessment tool: Standardized decision-making in the United States and the deprofessionalization thesis. *Transnational Social Review – A Social Work Journal*, 7(2), 206-218
- Bastian, P., & Schrödter, M. 2019: Risikodiagnostik durch „Big Data Analytics“ im Kinderschutz. In: *Archiv für Wissenschaft und Praxis der sozialen Arbeit*, (2), 40-49
- Beranek, A., Hill, B.; Sagebiel, J. B. 2019: Digitalisierung und Soziale Arbeit. Ein Diskursüberblick. *Soziale Passagen*, 11(2), 225-242
- Binns, R. 2018: Algorithmic accountability and public reason. *Philosophy & Technology*, 31(4), 543-56
- CARE 2012: ‘Vulnerable Children: Can Administrative Data Be Used to Identify Children at Risk of Adverse Outcomes?’ Centre for Applied Research in Economics, University of Auckland: Auckland, New Zealand. Verfügbar unter: <https://www>.

- msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/vulnerable-children/auckland-university-can-administrative-data-be-used-to-identify-children-at-risk-of-adverse-outcome.pdf [15.09.20]
- Chouldechova, A., Putnam-Hornstein, E., Benavides-Prado, D., Fialko, O. and Vaithianathan, R. 2018: 'A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions.' *Proceedings of Machine Learning Research* 81, 1-15
- Crawford, K. 2013: 'The hidden biases in big data. Verfügbar unter: <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> [15.09.20]
- Dare T. 2013: Predictive risk modelling and Child Maltreatment: an ethical review. University of Auckland: Auckland, New Zealand. Verfügbar unter: https://csda.aut.ac.nz/__data/assets/pdf_file/0016/11923/00-predictive-risk-modelling-and-child-maltreatment-an-ethical-review.pdf [15.09.20]
- Dahmen; Kläsener, N. 2019: Kinder- und Jugendhilfe als Kindeswohlgefährdungsvermeidungsstrategie? *Soziale Passagen*, 10(2), 197-210
- De Haan, I.; Connelly, M. 2014: Another Pandora's box? Some pros and cons of predictive risk modelling. *Children and Youth Services Review*, 47(1), 86-91
- Dencik, L., Hintz, A., Redden, J. and Warne, H. 2018: Data Scores as Governance: Investigating uses of citizen scoring in public services. Abruf unter: <https://datajusticelab.org/2018/12/06/data-scores-as-governance-final-report-published/> [15.09.20]
- Devliegere, J. and Gillingham, P. (im Druck): Transparency in social services settings: a critical exploration and reflection. *The British Journal of Social Work*
- Djeflal, C. 2018: Normative Leitlinien für künstliche Intelligenz in Regierung und öffentlicher Verwaltung. In R.Mohabbat Kar, B. E. P. Thapa, & P. Parycek (Hrsg.), (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft (S. 493-415). Berlin: Fraunhofer-Institut für Offene Kommunikationssysteme (ÖFIT). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57618-7>
- Drake, B. and Jonson-Reid, M. 2018: Administrative Data and Predictive Risk Modelling in Public Child Welfare: Ethical Issues Relating to California. Verfügbar unter: http://www.cachildlaw.org/Misc/Ethical_Review_of_Predictive_Risk_Modeling.pdf [15.09.20]
- Eubanks, V. 2018: Automating inequality. How high-tech tools profile, police and punish the poor. New York
- Gillingham, P. 2020b: The development of algorithmically-based decision-making systems in children's protective services: is administrative data good enough? *The British Journal of Social Work*, 50, 565-580
- 2020a: Algorithmically based decision support tools: skeptical thinking about the inclusion of previous involvement. *Practice: Social Work in Action*. <https://doi.org/10.1080/09503153.2020.1749584>
 - 2019b: Algorithms, social policy, social justice and social work: principles of algorithmic accountability. *Practice: Social work in action*, 31, 4,1-14
 - 2019a: Can predictive algorithms assist decision making in social work with children and families? *Child Abuse Review*. DOI: 10.1002/car.2547

- 2017: Predictive risk modelling to prevent child maltreatment: insights and implications from Aoteaora/New Zealand. *Journal of Public Child Welfare*, 11, 2, 150-165
- 2016: Predictive risk modelling to prevent child maltreatment and other adverse outcomes for service users: inside the „black box“ of machine learning. *The British Journal of Social Work*, 46, 1044-1058
- 2009: The use of assessment tools in child protection: an ethnomethodological study. University of Melbourne. <http://repository.unimelb.edu.au/10187/4337>
- Gloor, L. 2017: Data mining program designed to predict child abuse proves unreliable, DCFS says, *Chicago Herald Tribune*, 6 December 2017. Abruf unter: <https://amp.flipboard.com/@chicagotribune/data-mining-program-designed-to-predict-child-abuse-proves-unreliable-dcfs-says/f-4ac95e9127%2Fchicagotribune.com> [15.09.20]
- Jones, N. 2015: Anne Tolley scraps 'lab rat' study on children, *New Zealand Herald* [30. Julki 2015. Abruf unter: https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11489293 [02.09.20]
- Keddell, E. 2015: 'The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool?' *Critical Social Policy* 35(1), 69-88
- Kutscher, N., Ley, T.; Seelmeyer, U. 2020: *Handbuch Soziale Arbeit und Digitalisierung*. Wiesbaden
- Leslie, D., Holmes, L., Hitrova, C. and Ott, E. 2020: Ethics review of machine learning in children's social care. *What Works for Children's Social Care/Turing Institute/Rees Centre*, University of Oxford. Abruf unter: <https://whatworks-csc.org.uk/research-report/ethics-review-of-machine-learning-in-childrens-social-care/> [15.09.20]
- Macchione, N., Wooten, W., Yphantides, N.; Howell, J. R. 2013: Integrated health and human services information systems to enhance population-based and person-centered service', *American Journal of Preventive Medicine*, 45(3), 373-4
- Martini, M. 2017: Algorithmen als Herausforderung für die Rechtsordnung. *JuristenZeitung*, 72(21), 1017-1025
- Ministry of Social Development 2014a: Final report on feasibility of using predictive risk modelling. Ministry of Social Development: Wellington, New Zealand. Abruf unter: <https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/00-feasibility-study-report.pdf> [15.09.20]
- Ministry of Social Development 2014b: The feasibility of using predictive risk modelling to identify new-born children who are high priority for preventive services – companion technical report. Ministry of Social Development: Wellington, New Zealand. Abruf unter: <https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/00-feasibility-study-report.pdf> [15.09.20]
- Salganik, M. J. 2018: *Bit by Bit: Social Research in the Digital Age*, Princeton
- Sanders, M., Lawrence, Gibbons, D. and Calcrafft, P. 2017: *Using Data Science in Policy: A Report by the Behavioural Insights Team*, London, Behavioural Insights Team. Abruf unter: <https://www.bi.team/publications/using-data-science-in-policy/> [15.09.20]

- Schneider, D.; Seelmeyer, U. 2019: Challenges in Using Big Data to Develop Decision Support Systems for Social Work in Germany, *Journal of Technology in Human Services*, DOI: 10.1080/15228835.2019.1614513
- Scholz, T.M. 2019: Big data and human resource management. In J.S. Pedersen; Wilkinson A. (Hg.), *Big data: promise, applications and pitfalls*. Northampton, 69-89
- Schrödter, M., Bastian, P., Taylor, B. 2020: Risikodiagnostik und Big Data Analytics in der Sozialen Arbeit. In: N. Kutscher, T. Ley, U. Seelmeyer, et al. (Hrsg.) *Handbuch Soziale Arbeit und Digitalisierung*. Wiesbaden: Beltz. 255-264
- Schroedter, M., Bastian, P.; Taylor, B. 2018: Risikodiagnostik in der Sozialen Arbeit an der Schwelle zum „digitalen Zeitalter“ von Big Data Analytics. Preprint.. doi:10.13140/RG.2.2.22119.14240
- Sheridan, E. 2019: ‘Town Hall drops pilot programme profiling families without their knowledge.’ *Hackney Citizen* (Oktober 30, 2019)
- Stoltenborgh, M., et al. 2015: The prevalence of child maltreatment across the globe: Review of a series of meta-analyses, *Child Abuse Review*, 24(1), 37-50
- TCC Group 2013: Peer Review Report 1. Abruf unter: <http://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/feasibility-study-schwartz-tcc-interim-review.pdf> [15.09.20]
- Vaithianathan, R., Maloney, T., Putnam-Hornstein, E. and Jiang N. 2013: Children in the public benefit system at risk of maltreatment: Identification via predictive modelling. *American Journal of Preventative Medicine* 45(3), 354-359
- USACM 2017: Statement on Algorithmic Transparency and Accountability. US Association for Computing Machinery Public Policy Council (USACM). Verfügbar unter: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf [15.09.20]
- Wijenayake, Graham, T. and Christen, P. 2018: A decision tree approach to predicting recidivism in domestic violence, Presented at the Big Data Analytics for Social Computing (BDASC) workshop held at the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’18), Melbourne, Australia, June 2018. Verfügbar unter: <https://arxiv.org/abs/1803.09862> [15.09.20]
- Wolff 2004: Dokumenten- und Aktenanalyse. In U. Flick u.a. (Hrsg.), *Handbuch qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen* (2. Aufl.). Weinheim. 503-514
- Zweig, K. A. 2019: Algorithmische Entscheidungen: Transparenz und Kontrolle. *Analysen und Argumente*, 1-16

*Philip Gillingham, 18 Mukine Street, Jindalee, Queensland 4074, Australia
E-Mail: p.gillingham@uq.edu.au*

*Timo Ackermann, Mariannenstr. 48, 10997 Berlin
E-Mail: ackermann@ash-berlin.eu*