Jo Reichertz | Verena Keysers (Hrsg.)

Emotion Eskalation Gewalt

Wie kommt es zu Gewalttätigkeiten vor, während und nach Fußballspielen?

BELTZ JUVENTA

Jo Reichertz | Verena Keysers (Hrsg.)
Emotion. Eskalation. Gewalt.

Jo Reichertz | Verena Keysers (Hrsg.)

Emotion. Eskalation. Gewalt.

Wie kommt es zu Gewalttätigkeiten vor, während und nach Fußballspielen?

BELTZ JUVENTA

Dieses Buch ist erhältlich als:
ISBN 978-3-7799-3926-9 Print
ISBN 978-3-7799-5184-1 E-Book (PDF)

1. Auflage 2018

in der Verlagsgruppe Beltz · Weinheim Basel
Werderstraße 10, 69469 Weinheim

Herstellung: Hannelore Molitor
Satz: Christine Groh, Frankfurt am Main
Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza
Printed in Germany

Weitere Informationen zu unseren Autor_innen und Titeln finden Sie unter:
www.beltz.de

Inhalt

III. Theoretisierungen

Jo Reichertz

Emotion. Eskalation. Gewalt.

Probleme und Ergebnisse eines interdisziplinären DFG-Projekts

1. Ziele des Vorhabens

Wenn Menschen in großer Zahl in der Öffentlichkeit zusammenkommen, sei es bei Fußballspielen, Konzerten, Stadtfesten oder Demonstrationen, entstehen oft und leicht Gruppenemotionen (angenehme wie konfliktäre). Manchmal können konfliktäre Gruppenemotionen (scheinbar) unberechenbar eskalieren und zu teils massiven Gewalttaten von Einzelnen oder Gruppen führen (siehe hierzu die klassische sozialwissenschaftliche Literatur – Le Bon 1982; Freud 2005; Canetti 1996; Borch 2013; Tarde 2015). Ob und wann die Gruppenemotionen in einer bestimmten Situation eskalieren, ist bisher wissenschaftlich verlässlich weder vorherzusagen noch rechtzeitig zu erkennen noch sozialwissenschaftlich zu erklären.

Die zentrale *grundlagentheoretische* Fragestellung des vom 01.02.2015 bis zum 31.01.2018 laufenden, von der DFG bewilligten interdisziplinären Projekts *Emotion. Eskalation. Gewalt. Entwicklung eines video-basierten Verfahrens zur Früherkennung von Emotionsprozessen bei Großveranstaltungen*[1] lautete des-

1 An der Erarbeitung und Erstellung des DFG-Antrags waren neben mir vor allem Gregor Schöner, Daniela Horn, André Ibisch, Leif Klemm, Christine Moritz und Marc Tschentscher beteiligt. An den späteren Projektarbeiten waren zudem noch Verena Keysers, Selma Gleisberg, Joanna Meißner, Chantal Otterbein und Nils Spiekermann beteiligt. Richard Bettmann und Thomas Hoebel unterstützten die Projektarbeit mit Feldbesuchen und durch ihre Teilnahme an einer Reihe von Interpretationssitzungen. Matthias Meitzler und Caroline Plewnia unterstützten regelmäßig die Interpretationssitzungen. Eltje Gajewski und Lara Pellner absolvierten jeweils ein dreimonatiges Praktikum im Projekt und bearbeiteten eine eigene Fragestellung. Verena Keysers und ich haben im WS 2016/17 im Rahmen eines Methodenseminars an der Universität Duisburg-Essen mit einigen Studierenden das erhobene und sorgfältig anonymisierte Interviewmaterial ausgiebig hermeneutisch ausgewertet. In diesem Rahmen wurden fünf längere Hausarbeiten erstellt, in denen Interviewteile hermeneutisch analysiert wurden. Zudem hat Verena Keysers im SoSe 2017 an der Universität Duisburg-Essen mit Studierenden die aktuelle theoretische Debatte zur Entstehung von Gewalt aufgearbeitet. Zwischenergebnisse des Projekts wurden auf Tagungen in Essen, Berlin, Paderborn, Bielefeld und Amsterdam vorgestellt und diskutiert. Im Verlauf des Projekts kam es zu drei gemeinsamen Arbeitssitzungen mit dem Forschungszentrum Jülich, das an ähnlichen Fragestellungen arbeitet.

halb, ob sich Eskalationsprozesse vielleicht automatisch mittels einer beobachtenden Kamera und einer entsprechenden Auswertungssoftware automatisiert oder teilautomatisiert erkennen lassen. Zu diesem Zweck arbeiteten ein *sozialwissenschaftliches/kommunikationswissenschaftliches* Team vom Kulturwissenschaftlichen Institut Essen und ein Team vom *Institut für Neuroinformatik* der Ruhr-Universität Bochum zusammen.

Um das oben beschriebene Ziel zu erreichen, sollte das *sozialwissenschaftliche/kommunikationswissenschaftliche* Team das implizite Wissen von erfahrenen Beobachtern solcher eskalierenden Gruppenemotionen (Polizist_innen, die für die Beobachtung und ‚Steuerung' von Fußballspielen, Demonstrationen und Rockkonzerten zuständig sind, Sicherheitsbeauftragte und Ordnungsdienste) mittels *teilnehmender Beobachtung* und *Befragung* erheben und verdichten. Zum Zweiten sollte es den Prozess der Entwicklung von Eskalationen mittels *Videografie* systematisch erheben und vermessen. Die Videoanalyse positiver und negativer (kontrastierender) Datensätze sollte es ermöglichen, Eskalationsprozesse zu kategorisieren und detailliert zu beschreiben. Auf diese Weise sollte bestimmt werden, welche Merkmale von erfahrenen Beobachter_innen implizit und explizit zur (Früh-)Erkennung und Beurteilung von Eskalationsprozessen herangezogen werden und wie sich diese mit Verfahren der modernen Bildverarbeitung abbilden und echtzeitfähig implementieren lassen. Angestrebt wurde die Verschränkung und Verbindung einer hermeneutisch-wissenssoziologischen Forschung und einer grundlagentheoretischen Forschung zur automatischen Erkennung von Gruppenemotionsprozessen und deren Feststellung und Vorhersage.

Forschungspraktisches Ziel des Projektteams vom *Institut für Neuroinformatik* war die Entwicklung und praktische Erprobung eines mobilen Mehrkamerasystems nebst zugehörigem Softwareframework, das in der Lage ist, Menschengruppen von beschränkter Größe und Dynamik zu erfassen und ein Maß für die aktuelle Eskalationssituation zu schätzen. Die Auswertung sollte in Echtzeit, d. h. entsprechend der Zeit, in der die Information Relevanz besitzt, geschehen. Das Framework sollte eine Visualisierung implementieren, die den Benutzer_innen Lokalität einer als relevant klassifizierten Gruppenemotion, deren Art, die Konfidenz der Schätzung und ggf. die Merkmale, anhand derer die Klassifikation vorgenommen wurde, schnell und eindeutig erfassbar zur Verfügung stellen. Eine Schnittstelle zur vom sozialwissenschaftlichen Team genutzten Video-Annotationssoftware *Feldpartitur* sollte das Training maschineller Lernverfahren basierend auf kategorisiertem (gelabeltem) Videomaterial ermöglichen. Zusätzlich sollte in einer späteren Phase die Güte der Emotions- und Eskalationserkennung automatisiert bewertet werden. Zur Entwicklung dieses System war eine enge und permanente fachübergreifende Zusammenarbeit zwischen der Soziologie und der Neuroinformatik unabdingbar.

Der vorliegende Sammelband liefert einige der Ergebnisse des Projekts. Mit dieser Einleitung versuche ich weniger, die Projektergebnisse zusammenfassend vorzustellen, sondern mir geht es mehr darum, den Ablauf und die Schwierigkeiten der Forschungsarbeiten vorzustellen und zu diskutieren. Am Ende soll dann noch ein knapper Ausblick auf weitere Forschungsperspektiven gegeben werden.

2. Arbeitsprogramm – wie vom Plan vorgesehen

Die gesamte Forschungsarbeit war, so sah es der Plan vor, in drei sich von Beginn an einander durchdringende Arbeitsbereiche (Sozialwissenschaftliche Analyse von Feldbeobachtungen und Interviews – Videoanalyse – Entwurf und Programmierung von Software) mit bestimmten Arbeitspaketen gegliedert. Das Projekt ruhte auf drei Säulen. Sehr grob vereinfacht lässt sich der geplante Projektverlauf etwa so skizzieren:

2.1 Sozialwissenschaftliche Analyse

Aufgrund der inhaltsanalytischen und hermeneutischen Auswertung von Interviews mit Expert_innen aus verschiedenen Bereichen (Polizei, Sicherheitsdienste, Fans, Mitarbeiter_innen von Fanprojekten etc.) sollte mittels einer *sozialwissenschaftlichen Analyse* (a) deren Wissen um Eskalationsprozesse und deren Merkmale erfasst werden und (b) sollte aber auch eigenständig vor dem Hintergrund soziologischer Theorien versucht werden, den Eskalationsprozess selbst angemessen zu verstehen und zu erklären.

Die in der Praxis erworbene ‚Intuition' von Expert_innen für das Erkennen und Bewerten von Gruppenemotionen auf Großveranstaltungen sollte für die angestrebte Erforschung von Gruppenemotionen nutzbar gemacht werden. Im Sinne von Bourdieu soll die „einverleibte, zur Natur gewordene und damit als solche vergessene Geschichte" (Bourdieu 1987, 105) der Polizist_innen und Sicherheitskräfte erfasst werden. Die detaillierte Rekonstruktion der polizeilichen Praxis (aufgrund von teilnehmenden Beobachtungen, Interviews und Artefaktanalysen), und der darin inkorporierten Expertise im Umgang mit Gruppen und ihren Emotionen sollte dann Ausgangspunkt für die trennscharfe Bestimmung und detaillierte Beschreibung von typischen Situationen, Handlungen und Labels werden.

Die durch die Analyse der Praktiken gewonnenen Erkenntnisse sollten auf die Verbindung von Top-down- und Bottom-up-Prozessen zugespitzt und die Bedeutung von Subgruppen erfasst werden. Besonders die früh erkennbaren

Top-Down-Prozesse wie z. B. das von Polizeibeamt_innen beschriebene ‚Knistern in der Luft', galt es zu betrachten und empirisch zu erfassen.

2.2 Sozialwissenschaftliche Videoanalyse

Im Rahmen einer *sozialwissenschaftlichen Videoanalyse* sollten ebenfalls zwei Ziele erreicht werden:

a) Einerseits sollten aufgrund der detaillierten und akribischen Analyse von selbst gedrehten und im Netz gefundenen Videos, auf denen Eskalationsprozesse dokumentiert sind, die einzelnen Merkmale von Eskalationsprozessen identifiziert (Labels) und auch in einen theoretischen Zusammenhang gebracht werden.
b) Zum Zweiten sollten die Videos so gelabelt (= eindeutige Kennzeichnung von sichtbaren Elementen, die trennscharfe Merkmale von Eskalationsprozessen sind) werden, dass die Informatiker daraus letztendlich Programme und Programmteile entwickeln konnten, welche die Detektion von Eskalationsprozessen ermöglichen.

Als Instrument diente zu diesem Zweck das Videoanalyseinstrument *Feldpartitur* (Moritz 2013). Diese Software erlaubt

a) die multikodale Transkription (hier: Bewegungsmuster, Gestiken, Kodierungen),
b) die Integration unterschiedlicher Datensorten, in diesem Fall: der hermeneutisch-wissenssoziologischen Videoanalysedaten, Textmaterial aus Interviews, Feldbeobachtungen und Interpretationsgemeinschaften sowie quantitativ generierten Metadaten und
c) bietet Videomanagement-Funktionen an, die für die Handhabung des umfangreichen Datenkorpus aus forschungspraktischen Gründen notwendig sind. Die Generierung der am timecode der Videodaten fixierten Datenmatrix („grounded") ermöglicht darüber hinaus
d) die Evaluation der zunächst empirisch-qualitativ „entdeckten" Videopatterns am bestehenden umfangreichen Datenkorpus.

Die Ableitung von Merkmalen zur Kategorisierung gruppenemotionaler Zustände und deren Überführung in Merkmale der echtzeitfähigen Bildverarbeitung sollte in kleinschrittiger Zusammenarbeit zwischen Sozialwissenschaftler_innen und Techniker_innen anhand von Trainingsdaten (positive und negative Datensätze, kontrastierendes Datenmaterial) mittels technischer Deskriptoren

entwickelt und schließlich bis zur finalen Entwicklung eines Prototypsystems validiert werden.

2.3 Entwicklung von Algorithmen durch Neuroinformatiker_innen

Die Neuroinformatiker_innen sollten einerseits immer wieder die Besonderheiten der neuroinformatischen Datenaufbereitung und Datenverknüpfung an die sozialwissenschaftlichen Teams zurückspielen und somit die Findung/Erstellung guter Daten und guter Labels ermöglichen. Andererseits wurden immer wieder aus den gelabelten Daten Programmteile entwickelt und ausprobiert. Die Forschungsfrage lautete: Wie können die ermittelten und die damit verbundenen Konzepte effizient in Technik implementiert werden?

Erst wurde die Entwicklung des in den anderen Forschungsphasen zur Verwendung kommenden Kamera- und Softwaresystems begonnen. Hierzu gehörte sowohl die Anschaffung und der Test der Hardware, der Kameras und des mobilen Rechnersystems als auch erste Implementierungsarbeiten. Letztere umfassten den Aufbau einer modularen Bildverarbeitungspipeline innerhalb eines in der Arbeitsgruppe entwickelten Frameworks. Ferner konnten erste Bildverarbeitungsalgorithmen wie Personendetektion und -tracking sowie ein grundlegender optischer Fluss erstellt und anhand von öffentlich verfügbaren Datensätzen evaluiert werden. Auch die Schnittstelle zur Videoannotationssoftware *Feldpartitur* wurde in dieser Phase implementiert.

Ziel hierbei war es, die aus den Felderhebungen und der Videoanalyse gewonnenen Erkenntnisse (Merkmale zur Kategorisierung gruppenemotionaler Zustände) mit den nach dem Stand der Technik verfügbaren bildverarbeitenden Methoden zu vereinen. Die von geschulten Beobachtern verwendeten Merkmale waren zu formalisieren und in der Software abzubilden. Hierzu wurden eine Reihe weiterer Bildmerkmale implementiert und ihre Eignung mit verschiedenen modernen maschinellen Lernverfahren überprüft. Für diese Forschungsphase waren abschließende Tests und eine praktische Erprobung des fertiggestellten mobilen Erkennungssystems vorgesehen. Dieses Arbeitspaket war geprägt durch die zunehmende technische Explikation (i. e. S. Programmierung) des anfangs qualitativ konzipierten Modells.

2.4 Integration der Arbeitsbereiche

Die einzelnen Arbeitspakete waren einzelnen Mitarbeiter_innen zugeordnet. Inhaltlich wurde vor allem auf das Geschehen vor und in Fußballstadien fokussiert. Allerdings wurden auch einige Feldbeobachtungen und Interviews durchgeführt, die Demonstrationen, Tanzveranstaltungen und Rockkonzerte in den

Blick nahmen. Zudem haben wir einige Experimente durchgeführt – einerseits um die Kamerasysteme zu testen und zu verbessern, andererseits auch um erste Ideen für die Beschreibung von Eskalationsprozessen zu erhalten. So haben wir Personengruppen von 10 bis 60 Personen räumlich verdichtet und entweder durch den Experimentaufbau oder durch einen Moderator emotional aufladen lassen und sie dann kollektive Bewegungsaufgaben im Raum tätigen lassen.

Die drei Teams sollten jedoch nicht nebeneinander, sondern miteinander vernetzt arbeiten. Geplant war ein wöchentlich stattfindender Prozess des Daten- und Ergebnisabgleichs, was sich jedoch schnell als zu optimistisch erwies – allein schon aus dem banalen Grund, dass es im Projekt drei Standorte gab (Essen – Interviews und Feldbeobachtungen, Bochum – Neuroinformatik, Stuttgart – Videoanalyse) –, ein Problem, das sich nicht zufriedenstellend durch Telefon- und Videokonferenzen lösen ließ.

Von Beginn an sollte die Aufmerksamkeit aller Teams jedoch nicht darauf gerichtet sein, an den *Körpern* von *einzelnen Personen* Merkmale von Eskalationsprozessen zu identifizieren, sondern von Anfang an stand im Fokus das Interesse, sich *Konstellationen von Personen* anzusehen, deren Bewegung, deren Ausrichtung, deren Dichte und deren Geschwindigkeit. Es interessierte also nie eine einzelne Person und deren Verhalten, sondern immer nur *Gruppen* von Personen und das Verhalten dieser Gruppen.

3. Tatsächlich erhobene Daten und durchgeführte Auswertungsarbeiten

Ganz entscheidend für das Gelingen des Forschungsvorhabens war der Zugang zur polizeilichen Praxis oder zu anderen relevanten Feldern, dem dort vorhandenen impliziten Wissen und den verwendeten Dokumenten, Archiven und Schulungsunterlagen. Deshalb war der Feldzugang vorab gesichert worden. So hatte von Seiten der Polizei der Leiter der Beweissicherungs- und Festnahmehundertschaft der Bundespolizei (BFHu) zugesichert, das Projekt zu unterstützen. Auch die DFL, und hier insbesondere deren Geschäftsführer, Christian Seiffert, hatten ihre Unterstützung zugesagt.

Aber: Eine Forschung zu planen, ist das eine. Sie auch wie geplant umzusetzen, ist etwas anderes. Das gilt für jede Forschung. Nie decken sich Plan und spätere Realisierung. Immer zeigt sich bei der Umsetzung von Plänen, dass man zu optimistisch geplant hat. Oder aber es tauchen bei den konkreten Arbeiten Probleme auf, mit denen man nicht gerechnet hatte oder mit denen man nicht rechnen konnte. Das gehört zum Alltag der Forschung und wird in der Regel auch schon mitgeplant. Doch in diesem Projekt traten zudem außerordentliche und gravierende Schwierigkeiten auf, welche die Phase der Datenerhebung und -auswertung enorm verzögerten (und später auch viel Improvisation erforder-

ten): Die Schwierigkeiten ergaben sich aus der im November 2014 durch den Bürgerkrieg in Syrien ausgelösten Flüchtlingsbewegung, die über Monate hinweg Polizeikräfte aus allen Bundesländern band und später durch den G20-Gipfel in Hamburg im Sommer 2017.

Die Projektarbeiten begannen nämlich im Februar 2015. Es war geplant und vorab fest mit den Landespolizeien und der DFL abgesprochen, dass die Feldbeobachtungen in den und um die Stadien sowie die Interviews mit den Polizist_innen und dem Sicherheitspersonal im Frühjahr und Sommer 2015 stattfinden sollten. Aber gerade zu dieser Zeit erreichte die Flüchtlingsbewegung ihren Höhepunkt – was auch bedeutete, dass die Landespolizeien mit allen verfügbaren Zeit- und Personalressourcen an den Landesgrenzen, in und um die Auffanglager und zum Schutz von Personen und Gebäuden im Einsatz waren und für die (im Vergleich dazu weniger drängende) Frage nach den Bedingungen eskalierender Gewalt in und um Stadien keine Zeit mehr hatten. Zwar erhielten wir von den Landespolizeien NRW, Baden-Württemberg und Niedersachsen ernst gemeinte positive Antworten, doch zogen sich die Kooperationsvereinbarungen (Gestaltung der Verträge über Anonymisierung des Videomaterials, Zusammenarbeit, Sicherheit, Datenschutz, Unfallschutz etc.) trotz des ernsthaften Bemühens der einzelnen Behörden sehr lange hin. Angesetzte Termine mussten wegen akuter Einsätze an den Landesgrenzen verschoben werden, gemeinsame Treffen mit Polizeivertreter_innen scheiterten oft an der Terminfindung. Erst nachdem die Flüchtlingsbewegung abgeebbt war, kamen die Vereinbarungen zustande – weshalb wir erst sehr viel später die geplanten Feldbeobachtungen durchführen konnten. Um die Zeit nicht nutzlos verstreichen zu lassen,

- führten wir in den ersten Monaten des Projekts Interviews mit Expert_innen (z. B. Herrn Thomas Schneider von der DFL, Fans) durch,
- brachten die Kooperation mit der DFL unter Dach und Fach,
- besuchten Bundesligafußballspiele und Demonstrationen auch ohne Begleitung und erstellen Videos mit GoPro-Kameras,
- analysierten Videos von Eskalationen in Gruppen, die über YouTube verfügbar waren (z. B. Loveparade in Duisburg, von Zuschauern oder Ultragruppen ins Netz gestellte Aufnahmen von Eskalationen),
- führten gezielte Experimente mit Studierenden durch (schnelle Bewegungen in Gruppen), um Kamerapositionen bestimmen und erste Kategorisierungen für Verhalten (Labels) entwickeln zu können,
- führten einen Workshop durch, bei dem Bürger_innen gebeten wurden, bestimmte Emotionen zu gestalten; diese wurden mit fest installierten hochauflösenden Kameras aufgezeichnet und später analysiert, um die Kategorisierungen für Verhalten (Labels) weiterzuentwickeln,
- leistete ein Mitarbeiter einige Wochen aktiven Dienst als Sicherheitsperson in der zweiten und dritten Bundesliga

- und wir erarbeiteten eine (allen Auflagen des Datenschutzes gerecht werdende) Form der Anonymisierung der visuellen Daten (= Kantenbilder[2]).

Wegen der genannten Probleme wurden im *sozialwissenschaftlichen* Arbeitsbereich innerhalb des Projektzeitraums, wenn auch mit teils erheblichen Verzögerungen, folgende Daten erhoben und teils hermeneutisch, teils inhaltsanalytisch analysiert: So wurden 42 qualitative, etwa ein- bis zweistündige, themenzentrierte Interviews geführt, größtenteils mit Polizeipersonen aus drei Bundesländern (NRW, Baden-Württemberg und Niedersachsen), die in unterschiedlichen Funktionen wiederholt vor und in Fußballstadien im Einsatz waren (Polizeiführer, Mitglieder von Hundertschaften, Beweis- und Sicherheitsdienst, szenekundige Beamte)[3]. Zudem wurden ein Fanprojektmitarbeiter, drei (Ultra-)Fans und ein Hooligan, eine Person vom Sicherheitspersonal und ein Sportfunktionär ausführlich interviewt. Die Interviews wurden alle inventarisiert (mit kurzer Inhaltsangabe), 37 komplett oder zu großen Teilen transkribiert und inhaltsanalytisch, sowie ausgewählte Stellen mit dem Verfahren der Hermeneutischen Wissenssoziologie in Bezug auf die Projektfragestellung ausgewertet.

Zur *Videoanalyse* wurden Videodaten aus Feldbegehungen, ethnografischen Aufzeichnungen im Fußballstadion (unter Polizeibegleitung), auf Demonstrationen, kirchlichen Veranstaltungen erhoben; es kam zum Datenaustausch mit dem Forschungszentrum Jülich; ebenso wurden zur Projektfragestellung passende Videosequenzen aus öffentlichen Archiven wie YouTube-/Vimeo-/LiveLeak-Videos erhoben und teils sehr ausführlich analysiert – so z. B. eine Sequenz zum Spiel Fortuna Düsseldorf gegen den MSV Duisburg und andererseits die Auseinandersetzung Stuttgarter Ultras mit dem Bielefelder Ordnerpersonal vom 17.04.17. Anhand der Interview-, Feld- und Videodaten, die im Projekt zur Verfügung standen, konnten mittels Datentriangulation und Methodentriangulation einzelne multiperspektivische Rekonstruktionen von Ereignissen vorgenommen werden.

Im Bereich der *Neuroinformatik* wurden erste Ansätze zur automatischen Erfassung von Gruppenemotionen erprobt. Der Ansatz von Choi/Shahid/Sava-

2 Kantenbilder (auch *Gradientenbilder*) sind schwarz-weiß und zeigen nur noch die Umrisse (= Kanten) der gefilmten Personen und Gruppen. Einzelne Personen und deren Gestik und Mimik sind auf Kantenbildern nicht mehr erkennbar. Das stellte für die Analyse dieses Videomaterials eine sehr große Herausforderung dar.

3 Bei diesen Interviews wurde sichtbar, dass nicht alle Landespolizeien die gleichen Einsatz- und Kommunikationsstrategien im Umgang mit Fußballfans haben, sondern dass teils deutliche Unterschiede bestehen. Da es jedoch nicht das Ziel des Projekts war, Unterschiede der landespolizeilichen Behandlung von Eskalationsprozessen zu identifizieren, haben wir in der Auswertung diese Unterschiede zwar zur Kenntnis genommen, jedoch nicht herausgearbeitet, wie sie sich im Einzelnen unterscheiden und welche Folgen dies für das Eskalationsgeschehen hat. Dafür bedürfte es einer gesonderten Studie.

rese (2011) zur *Bestimmung der zeitlichen Entwicklung* der relativen Position und Körperausrichtung einzelner Personen in kleineren Menschengruppen wurde angepasst und anhand zwei verschiedener Datensätze getestet. Aufbauend auf den Arbeiten von Yao/Gall/Van Gool (2010) wurde ein System zur *raumzeitlichen Klassifikation von Handlungen einzelner Personen* in kurzen Videoausschnitten entwickelt und an den o. g. inszenierten Situationen sowie dem öffentlich zugänglichen UCF Sports Action Data Set (Soomro/Zamir 2014; Spata 2016) erprobt. Zur Erkennung anormalen Verhaltens von Menschengruppen wurde der optische Fluss geschätzt. Dabei wurden überall im Bild, wo der lokale Kontrast ausreicht, Bewegungsvektoren geschätzt, ohne zuvor eine Segmentation von Personen oder Objekten zu versuchen. Der gewählte Ansatz baute auf Mehran/Oyama/Shah (2009) auf und leistete eine computational schnelle Schätzung mithilfe einer Particle Advection (Ali/Shah 2007). Das Social Force Model (Helbing/Molnár 1995), das Bewegungstrajektorien von Fußgängern soziale Kräfte innerhalb von Gruppen zuschreibt, wurde dabei unmittelbar auf die Partikel des Schätzalgorithmus angewandt, also ohne diese unbedingt einer Person zuzuschreiben. Schließlich wurde mittels Latent Dirichlet Allocation (Blei/Ng/Jordan 2003) eine Klassifizierung in normales und abnormales Gruppenverhalten vorgenommen (Hegenbarth 2016). Mit der vorliegenden Methodik konnte der plötzliche Umschlag von Verhaltens-/Bewegungsmustern detektiert werden. Um die Dichte von Menschenmengen zu schätzen, wurde für Aufnahmen aus der Totalen ein computational schnelles Verfahren zur Detektion von Köpfen oder Kopf-Schulter-Partien aus dem Zählverfahren von Idrees et al. (2013) abgeleitet. Dieses wurde anhand des öffentlich zugänglichen INRIA Datensatzes (Dalal/Triggs 2005) evaluiert.

4. Probleme der interdisziplinären Projektarbeit[4]

In dem Projekt arbeiteten Soziolog_innen, Kommunikationswissenschaftler_innen, Informatiker_innen und Videoanalytiker_innen zusammen. Die damit einhergehenden interdisziplinären Herausforderungen haben alle Beteiligten im Rückblick klar unterschätzt, trotz einschlägiger Erfahrungen. Neben praktischen Problemen gab es eine Reihe von inhaltlichen und theoretischen Missverständnissen, welche die Arbeit immer wieder erschwerten. Manchmal erkannten wir erst nach Wochen, dass wir uns missverstanden hatten. Dies wurde im Verlauf der Projektarbeit deutlich und äußerte sich in unterschiedlichen Schwerpunktsetzungen und Planungen des Zeitverlaufs der Projektarbeit sowie auch in un-

4 Den im folgenden Kapitel benannten Problemen liegt ein Papier von Gregor Schöner zugrunde, das ich hier ausführlich wiedergebe und ergänze.

erkannten Unterschieden bei der Nutzung von Begriffen wie Varianz und bei den der Forschung zugrunde gelegten Begriffe wie *Emotion*, *Eskalation*, *Gewalt*. Aber auch die Begriffe *Gruppe*, *Masse* und *Ansammlung* waren nicht hinreichend geklärt, um für die Analyse nützlich zu sein.

Gleiches stellte sich auch bei dem Phänomen der ‚Übertragung von Emotionen' heraus. Die hierzu in der Literatur vorliegenden Theorien benutzen vor allem die Metapher der Ansteckung, der Resonanz, des Schwarmverhaltens oder der unbewussten Kommunikation.

Der Ansatz des technischen Vorgehens war es, Algorithmen des maschinellen Lernens zu nutzen, um die Detektion und Klassifikation von Ereignissen, die verschiedene Gruppenemotionen signalisieren, zu automatisieren. Solche Algorithmen lernen aus Beispielen, hier also aus Videomaterial, in dem menschliche Beobachter_innen Ereignisse gekennzeichnet haben (‚labeln'). Je größer die Vielfalt der visuellen Erscheinung der gleichen Klasse, umso mehr Beispiele sind notwendig. Durch Vorverarbeitung der Bilder kann ein Teil der Bildvarianz eliminiert werden, was die Lernaufgabe erleichtert.

Die Aufgabenverteilung im interdisziplinären Team schien offensichtlich. Die Techniker_innen beschäftigen sich mit Algorithmen der Vorverarbeitung und dem Entwurf der spezifischen Algorithmen aus dem Bereich maschinelles Lernen. Die Soziolog_innen liefern Beispieldaten als gelabelte Videoabschnitte und damit implizit die grundlegenden Begriffe zur Beschreibung von Gruppenemotionen. Im Verlaufe des Projekts zeigte sich, dass diese Schnittstelle zwischen Technik und menschlichem Beobachten bei weitem nicht genügend spezifiziert war. Im Grunde hatten beide Seiten eine ganze Reihe von impliziten Annahmen und Randbedingungen nicht kommuniziert und waren sich dieser Unterlassung zunächst gar nicht bewusst.

Eine Frage ist, ob gelabelte Ereignisse lokalisiert im Bild oder global der Gesamtsituation zugeordnet sind. Die Verfahren des maschinellen Lernens profitieren stark von lokalisierten Labeln, denn diese erlauben, die gesamte Varianz des Bildmaterials außerhalb der lokalisierten ‚region of interest' (ROI) zu eliminieren. Die Soziolog_innen gingen davon aus, dass Labels lokal in der Zeit sind, aber global die gesamte visuelle Szene bezeichnen. Das Software-Instrument der Soziolog_innen sah auch gar keine Lokalisierung der zuzuweisenden Label vor.

Die Methoden des maschinellen Lernens zielen zunächst auf eine einfache ‚Ein-Klassen-Klassifikation' ab, in dem Bilder oder ROI also als zugehörig oder nicht zugehörig zu einer bestimmten Klasse erkannt werden (‚binäre Klassifikation'). Mehrklassen-Klassifikation ist möglich, verlangt aber entsprechend größere Mengen von Beispielen. Die implizite Erwartung der maschinellen Lernern war folglich, dass die Soziolog_innen einige wenige, grundlegende Klassen vorschlagen, für die sie dann viele Beispiele liefern.

Für die Soziolog_innen war dagegen die Differenzierung der beschreibenden Begriffe ein wesentliches Ziel. Sie versuchten für jedes einzelne Ereignis

eine Charakterisierung zu erreichen, die nicht nur dem oberflächlich sichtbaren, sondern auch den mitempfundenen, vom Kontext abhängigen Dimensionen des Geschehens Rechnung trug. Eine Vielzahl beschreibender Begriffe wurde erarbeitet, und es war gerade diese Erarbeitung, die einen wichtigen Teil des Erkenntnisprozesses ausmachte. Viele Beispiele für die genau gleiche kategorielle Beschreibung zu liefern, lief diesem Bestreben entgegen.

Erschwerend für die interdisziplinäre Zusammenarbeit war in diesem Kontext auch die unterschiedliche Auffassung vom eigentlichen Erkenntnisprozess. Für die Soziolog_innen war die schrittweise Aufdeckung von möglichen Ereignissen, deren Charakterisierung und Differenzierung eine natürliche Form des Vorgehens. Für das maschinelle Lernen konnte die Arbeit erst beginnen, wenn konkrete Hypothesen in Form von gelabelten Beispielen in einem Umfang vorlagen, der ermöglichte, auf einer Teilmenge maschinell zu lernen, um die übrigen Daten zum Test der Generalisierung zu nutzen. Folglich war auch die zeitliche Koordination der Zusammenarbeit nicht einfach zu bewerkstelligen.

Neben diesen recht unterschiedlichen grundsätzlichen Perspektiven gab es kleinere Unterschiede. So war etwa für die Soziolog_innen eine bewegte Kamera, die ins Geschehen integriert ist, unter Umständen ausdrucksstärker als eine statische Kamera, die das Geschehen von außen registriert. Für die maschinellen Lerner ist dagegen eine Eigenbewegung der Kamera eine Herausforderung, da Bewegung im Bild ein wirksames Mittel zur Vorauswahl von Bildregionen ist, in denen Ereignisse auftreten mögen. Bei bewegter Kamera verliert dieser Salienzkanal seine Wirkung.

Auch die Frage der multisensoriellen Basis mancher Begriffe, insbesondere zur Hinzunahme von auditorischen Hinweisen, wurde lange diskutiert. Für die Soziolog_innen sind dies natürliche Dimensionen des zu verstehenden Geschehens. Für die Bildverarbeitung entsteht bei der Miteinbeziehung multisensorieller Hinweise die Notwendigkeit, zusätzliche Disziplinen, wie die automatische Analyse akustischer Kanäle, zu beachten. Dabei würde gleichzeitig die potenzielle Varianz des Datenmaterials weiter erhöht und entsprechend würden sich die Anforderungen an die gelabelten Beispieldaten weiter erhöhen.

Diese Varianz innerhalb des Videomaterials, das der automatischen Detektion von Gruppenemotionen dienen sollte, war ein Thema, dessen Bedeutung wir erst im Laufe des Projektes für beide Disziplinen klar artikulieren konnten. Aus Sicht der maschinellen Lerner geht es um die Varianz auf Bildebene, also die Variabilität der visuellen Erscheinung der relevanten Ereignisse, die in verschiedenen Beispielen auftritt. Wird, beispielsweise, die Aktivität, die den Fokus einer aufkeimenden Störung darstellt, im Bild innerhalb einer kleinen Region erfasst, weil das Geschehen entsprechend weit entfernt ist, so lernt ein Algorithmus diese visuelle Erscheinungsform mit. Will man das entsprechende Ereignis unter anderen Umständen, wenn der Fokus näher an der Kamera liegt, erkennen (und somit generalisieren), so muss das Lernmaterial diese Varianz

der ‚region of interest' in ihrer Größe im Bild enthalten, also Beispiele von visuell kleinen, mittleren, und großen Abbildungen der Erscheinung ‚Fokus' liefern. Viele andere Dimensionen der visuellen Erscheinung von Ereignissen sind nicht so einfach zu umschreiben wie die Größe im Bild, sodass es auch keine einfache Art gibt, diese Form von Varianz durch theoretische Überlegungen zu eliminieren. Diese Form der Varianz ist also zu unterscheiden von der Varianz, welche die Soziolog_innen durchaus interessiert und die ja gerade der Tendenz zugrunde liegt, Ereignisse durch eine Vielzahl von Begriffen differenziert zu beschreiben.

Die zum maschinellen Lernen nutzbaren Beispielvideos waren jedoch radikal eingeschränkt durch Fragen der Filmqualität (bewegte Kamera, wechselnder Zoom, wechselnde Bildausschnitte) und durch die darin enthaltene visuelle Varianz (Hintergrund, Dichte, Abstand von der Szene, Szenenelemente). Die selbst produzierten Videos von einfach strukturierten Eskalationen konnten einerseits die Komplexität des Geschehens nicht erfassen, lieferten aber andererseits immer noch viel zu wenig Lernbeispiele um maschinelle Lernmethoden einzusetzen.

In der Begrifflichkeit der maschinellen Lerner litt das Projekt letztlich also an einer Knappheit von Daten im Sinne von einer hinreichend großen Anzahl von visuellen Beispielen für eine kleine Anzahl von Labeln. Für die Soziolog_innen war die Aufgabe, die vorhandenen Daten in ihrer Begrifflichkeit zu analysieren, schon sehr umfangreich und die Forderung nach immer mehr ‚Daten' schwer verständlich. Neben den geschilderten begrifflichen Schwierigkeiten war auch der in Deutschland sehr rigoros praktizierte Datenschutz ein Begrenzungsfaktor. Die Zusammenarbeit mit Veranstaltern wurde dadurch ebenso stark beeinträchtigt wie die Zusammenarbeit mit den im Grunde sehr interessierten polizeilichen Stellen.

Der Dialog über alle diese Probleme zwischen den Forscher_innen der unterschiedlichen Disziplinen nahm verschiedene Formen an. Die maschinellen Lerner implementierten beispielhafte Algorithmen und stützten sich dabei auf öffentlichen Datenbasen (meist aus dem amerikanischen Raum). Diese sollte zeigen, was man ‚im Prinzip' tun könnte. Die Soziolog_innen haben die maschinellen Lerner zu Workshops mitgenommen, bei denen die soziologische Videoanalyse trainiert wurde.

Ein Ergebnis dieses intensiven Austausches war die Kristallisierung der Frage, ob Phänomene im Bereich der Gruppenemotionen notwendig zuerst die Segmentierung im Bild der Einzelperson und die Entdeckung ihrer Handlungsintention erfordert. Aus soziologischer Sicht hat diese Annahme tiefe Gründe in der Handlungstheorie. Aus Sicht des maschinellen Lernens löst man bei diesem Vorgehen sehr schwierige Probleme des Computersehens, Segmentation und Intentionserkennung als ersten Schritt in einem Prozess, der dann auf recht einfache Klassifikationen der Gruppenemotion hinausläuft. So kam die

Idee auf, direkt auf der Ebene von kollektiven Variablen eine abgebildete Gruppe von Menschen zu klassifizieren. Dies wurde beispielhaft umgesetzt, indem Aufnahmen von Zuschauern bei einem Fußballspiel durch einen Algorithmus bearbeitet wurden, der den optischen Fluss berechnet. Der optische Fluss bestimmt für jeden Bildpunkt eine Geschwindigkeit im Bild, ohne vorher zu entscheiden, welchem Objekt der Bildpunkt zuzuordnen ist. So gelang es, Zustände der Synchronizität von Bewegungen in einer Gruppe von Menschen sichtbar zu machen, ohne dabei die einzelnen Handelnden visuell zu segmentieren und ihre Handlung als Einzelne zu charakterisieren. Aus diesem Dialog entstand trotz der genannten Probleme ein Begriffsapparat, der im Weiteren eine stärker hypothesengestützte Arbeit direkt an Gruppenemotionen und deren visuellen Erscheinungsformen ermöglichen kann.

5. Kurzer Überblick über die Ergebnisse des Projekts

Neben den Problemen gab es natürlich auch eine Reihe von produktiven Ergebnissen, die zu großen Teilen in diesem Band vorgestellt werden. Das wohl wichtigste Ergebnis des Projektes war die (im Grunde nicht überraschende) Erkenntnis, dass die ursprünglichen Ziele zu hoch gesteckt und im Rahmen des Projekts nicht zu erreichen waren. Das lag auch daran, dass wir im Laufe des Projektes neue Erkenntnisse über die Schnittstellen zwischen den Disziplinen erworben haben. Die ursprüngliche Einschätzung war, dass die Kennzeichnung von Gruppenemotionen und deren zeitlichen Verläufen mit den Methoden der qualitativen Soziologie Daten liefern würde, die dann mit modernen Methoden des maschinellen Lernens in automatische Detektoren und Klassifikatoren von Gruppenemotionen umgesetzt werden könnten. Diese Einschätzung war falsch. Die Daten sind in ihrer Dimensionalität, ihrem Umfang und im Hinblick auf die Varianz des zugrunde liegenden Filmmaterials weit entfernt von den informatischen Anforderungen.

Dass die Projektziele nicht erreicht werden konnten, lag (wie weiter unten in den verschiedenen Beiträgen dieses Bandes ausgeführt werden wird) auch daran, dass der Prozess der Eskalation sehr viel vielfältiger und komplexer ist als angenommen, dass der Übertragungsprozess von Emotionen bislang nur metaphorisch beschreibbar ist, ohne ihn in seiner Spezifik beschreiben und analysieren zu können. Hier ist noch sehr viel Arbeit vonnöten.

Das gilt auch für die Begriffsarbeit. Es stellte sich als grundlegend heraus, unterschiedliche Untersuchungsbereiche und damit unterschiedliche Ebenen des empirischen Zugriffs zu kennzeichnen. Im Anschluss an die soziologische Literatur haben wir erst die Begriffe *Makro-*, *Meso-* und *Mikrobereich* ausgearbeitet. Später waren wir genötigt, auch einen *Nanobereich* zu berücksichtigen. Damit gemeint ist die Ausdrucksebene sozialer Interaktion, bei der die einzelnen be-

deutungstragenden Einheiten entweder von so kurzer Dauer sind oder aber sich in minimalen Veränderungen zeigen, die entweder für die normale, wissenschaftliche Beobachtung (Augen, Ohren, Gespür) nicht wahrnehmbar, aber auf jeden Fall kaum erinnerbar und damit auch nicht für die Analyse verfügbar sind (ausführlich zum Nanobereich siehe Reichertz 2017).

Für die Rekonstruktion der Eskalationsprozesse erwies es sich als sehr fruchtbar, die Begriffe *Setting*, *Frame* und *Script* zu verwenden. Unter *Setting* wird dabei ein typisches soziales Großereignis verstanden. Innerhalb von Settings finden sich verschiedene typische Kleinformen, die zusammen das Großereignis bilden (*Frames*). Innerhalb von Frames finden sich typische Handlungsformen, die von den Beteiligten in einem bestimmten Frame erwartet werden (*Scripts*).

Ein weiteres Ergebnis des zurückliegenden Projekts ist, dass Polizist_innen keine *Experten zur Entdeckung oder Früherkennung von Eskalationsprozessen* sind, wie dies teilweise zu Beginn der Arbeit angenommen wurde. Polizist_innen sind dies alleine schon deshalb nicht, weil sie *Mitakteure* in den Eskalationsprozessen sind, weil sie zudem die Ereignisse vor ihren Augen und Ohren nach erworbenen *polizeilichen Relevanzen* scannen und weil sie sich aufgrund ihrer Handlungslogik, nämlich Personen haftbar zu machen, vor allem auf *Personen konzentrieren* und weniger auf Prozesse. Die erworbenen polizeilichen Relevanzen ergeben sich aus der Aufgabenstellung der Polizei: der Gefahrenerkennung, der Gefahrenabwehr und der Beweissicherung und des Dingfest-Machens möglicher Täter. Es geht Polizist_innen in ihrem Einsatz nicht um die soziologische Analyse der Ereignisse, in die sie verstrickt sind. Ebenso sind die Ultras keine Experten der Früherkennung von Eskalationsprozessen, auch wenn sie erfahren und geübt darin sind, Eskalationsimpulse gezielt zu setzen oder dem Zufall eine gute Chance zu geben. Deshalb werden durch die sozialwissenschaftliche Analyse von Interviews mit Polizisten und Fans nur deren jeweilige Relevanzen sichtbar, was ein weiterer Hinweis dafür ist, dass Interviews, auch Interviews mit Experten, vor allem dabei helfen, die Relevanzen eben dieser Experten zutage zu fördern.

Alle Akteure handeln zwar auch aufgrund ihrer Wahrnehmungen vor Ort, aber alle handeln auch vor dem Hintergrund ihrer Erfahrungen mit solchen Ereignissen, ihren damit verbundenen Hoffnungen und Befürchtungen, kurz: Alle handeln pfadabhängig. Alle sind in unterschiedliche Kontexte und Relevanzen eingebunden, und was die Sache besonders schwierig macht: Es gibt nicht eine Gruppe, sondern mehrere, vielfältig untergliederte Gruppen, die in unterschiedlichen Kontexten und mit unterschiedlichen Medien handeln. An den Grenzen dieser Gruppen gibt es vielfältige Reibungspunkte, die sich situativ vor Ort manchmal klären lassen, manchmal jedoch nicht. Wann, weshalb und unter welchen Bedingungen ein solcher Grenzkonflikt aus dem Ruder läuft und

zu Eskalationen führt, hängt von kontingenten Faktoren ab und kann selbst später nicht mit Sicherheit rekonstruiert werden.

Bezogen auf die Frage nach dem besonderen Ablauf von Eskalationsprozessen konnten wir *fünf* Typen von *grundlegenden* und *allgemeinen* Prozessen identifizieren, die in *allen* konkreten Prozessen auftauchen, die sich über eine meist *kurze Zeit* erstrecken, die sich an dem Körperausdruck der Beteiligten zeigen, somit erkennbar sind und die räumliche Grenzen aufweisen. Diese fünf basalen Prozesse sind:

- *Fokussierung* der Aufmerksamkeit einer Vielzahl von Personen auf einen Punkt,
- *Synchronisierung* der Verhalten vieler,
- *Bildung von Rändern und Kernen,*
- *plötzlicher Umschlag* des synchronisierten Verhaltens,
- *allmähliche Erosion/Zerstreuung* der Synchronisierung und Fokussierung.[5]

Zentrales Ergebnis der Forschungsarbeiten ist jedoch, dass sich die Eskalationsprozesse nicht allein als die Folge *linearer* und *situativer* Interaktionsprozesse vor Ort erklären lassen; sie sind sehr viel vielfältiger und komplexer als angenommen: Der Eskalationsprozess ist nicht nur bestimmt durch vielfältige situative Faktoren, sondern maßgeblich durch die jeweiligen Kulturen der einzelnen Gruppen und deren Interaktion schon im Vorfeld. Mögliche Merkmale dieses Prozesses, deren Ausprägungen und deren Relationen sind nicht ganz eindeutig (Singen, Arme hochreißen, Stampfen, Vermummen etc.) sondern strukturell mehrdeutig; sie können verschiedenen Scripts und Kontexten zugeordnet werden – weshalb sich keine Merkmalsausprägungen identifizieren ließen, trennscharf zwischen Scripts zu unterscheiden. Die Merkmale, deren Ausprägungsgrade und deren Relationen entfalten sich in einer Verlaufskurve (trajectory) und ihre Bedeutung ergibt sich aus der Stellung in diesem reversiblen Entfaltungsprozess. Diese Prozesshaftigkeit erhöhte die Komplexität der Scripts enorm und überforderte die Möglichkeiten der Detektierbarkeit.

Immer sind auch übersituative Faktoren, Kontexte und Akteure für das Geschehen vor Ort verantwortlich. Das Geschehen vor und im Fußballstadion lässt sich nicht allein aus *einem* Handlungskontext erklären und verstehen, sondern man muss die verschiedenen sozialen Kontexte (aus den verschiedenen Bereichen, also aus den Makro-, Meso-, Mikro- und Nanobereichen), die auf die Eskalationsprozesse einwirken, mit in die Analyse aufnehmen. Analysen von Eskalationsprozessen müssen also *polykontextural* sein.

5 Siehe ausführlich zu diesen Prozessen den Beitrag von Reichertz in diesem Band.

Ein visuelles und automatisches Detektionssystem, das nur alle Faktoren in der Situation und vor Ort in den Blick nimmt und verarbeitet, kann deshalb nur begrenzt Aussagen über die Wahrscheinlichkeit von Eskalationsprozessen machen. Zweifellos könnten solche Systeme bei der Einschätzung bestimmter Lagen vor Ort beraten und assistieren, aber um darüber hinaus die ablaufenden Prozesse und deren Eskalationspotenzial, die in vielfältigen Kontexten aus dem Meso- und Makrobereich resultieren, zu erfassen, bedarf es grundsätzlich immer auch erfahrener Akteure, die diese Prozesse kennen und dieses Wissen in die Lageeinschätzung einbringen können. Visuelle Expertensysteme werden deshalb wohl auf absehbare Zeit vor allem Assistenzsysteme bleiben.

Für *zukünftige* Forschungen bedeutet dies, dass die jeweilige Meso-, Mikro- und Nanoebene mit geeigneten Verfahren in den Blick genommen werden muss und dass die jeweiligen Ergebnisse mit geeigneten Methoden zusammengeführt werden müssen. Im Einzelnen bedeutet dies, dass in der Forschung von sozialwissenschaftlicher Seite einerseits mittels beobachtender Teilnahme und teilnehmender Beobachtung das interaktive Geschehen und Kommunikationsprozesse beobachtet und (mit)erlebt (Mikroebene) und mittels Feldbeobachtungen und Interviews mit Beteiligten das Wissen um Frames und Scripts (Mesoebene) erhoben und analysiert werden muss. Andererseits müssen die Abstimmungsprozesse mittels hochauflösender Videoaufzeichnung und detaillierter Videoanalyse auf der Nanoebene erhoben und analysiert werden.

Für zukünftige Forschungen bedeuten die vorliegenden Ergebnisse aber auch, dass erst einmal weniger komplexe soziale Prozesse in Menschenansammlungen untersucht werden müssen: Eskalationsprozesse sind vorerst zu komplex, um sie angemessen erfassen und analysieren zu können, weshalb es sinnvoll ist, in einer ersten Phase in experimentell kontrollierten und kleinen Gruppen die kommunikativen Koordinationsprozesse auf den unterschiedlichen Ebenen zu erfassen und im Hinblick auf die Frage, wie genau und auf welchen Interaktionsebenen die Abstimmungsprozesse stattfinden, auszuwerten. In einer zweiten Phase sollte man dann die gewonnenen Ergebnisse und Erfahrungen anhand von Ereignissen in situ (Fußball, Konzert, Demonstration) überprüfen und anreichern (ausführlich zu den letzten Punkten siehe den Beitrag von Reichertz in diesem Band).

Der vorliegende Band kann nicht alle Ergebnisse des Projektes vorstellen, sondern konzentriert sich auf die Darstellung der *sozialwissenschaftlichen* und *kommunikationswissenschaftlichen* Auswertung der *Interviews* (Keysers, Spiekermann), einzelner *Videoanalysen* (Meißner, Pellner) und der *Feldbeobachtungen, Videos und Interviews* (Reichertz). Weitere *Videoanalysen* werden an anderer Stelle vorgelegt werden – so in Keysers et al. 2018 und Moritz/Corsten 2018. Zentrale Ergebnisse aus dem Bereich der *Neuroinformatik* finden sich in dem Beitrag von Daniela Horn, Sebastian Houben und Gregor Schöner. An anderer Stelle wird aus diesem Bereich die automatische Videoanalyse reflek-

tiert (Horn/Ibisch/Tschentscher 2018). In den Band aufgenommen wurde noch ein (gekürztes) Feldmemo von Richard Bettmann. Es soll einerseits demonstrieren, wie Feldberichte im Projekt angelegt wurden, andererseits zeigt es aber auch schon viel über das Feld und den Prozess des wissenschaftlichen Beobachtens.

Danken möchte ich abschließend an dieser Stelle allen, die das Projekt ermöglicht und unterstützt haben: Das ist erst einmal und vor allem die DFG, die das Projekt großzügig förderte. Dann sind es die vielen Fußballfans, die mit uns gesprochen haben und die wir begleiten konnten. Besonderer Dank gilt den Polizeien aus NRW, Niedersachsen und Baden-Württemberg, die uns trotz großer Arbeitsbelastung für lange Interviews bereitwillig zur Verfügung standen und mit denen wir diverse Bundesligaspiele und das Geschehen vor und nach den Spielen beobachten durften. Auch Christian Seiffert und Thomas Schneider von der DFL möchte ich hier danken, die viele Verbindungen ermöglichten und uns mit Rat und Tat unterstützten. Schlussendlich möchte ich Christine Groh für ihre sorgfältige Durchsicht und Korrektur des Buchmanuskripts danken.

Literatur

Ali, Saad/Shah, Mubarak (2007): A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, S. 1–6.

Blei, David/Ng, Andrew/Jordan, Michael I. (2003): Latent Dirichlet Allocation. In: Journal of Machine Learning Research 3, S. 993–1022.

Borch, Christian (2013): The Politics of Crowds. An Alternative History of Sociology. Cambridge: University Press.

Bourdieu, Pierre (1987): Sozialer Sinn. Kritik der theoretischen Vernunft. Frankfurt am Main: Suhrkamp.

Canetti, Elias (1996) [1960]: Masse und Macht. München: Grin.

Choi, Wongun/Shahid, Khuram/Savarese, Silvio (2011): Learning Context for Collective Activity Recognition. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, S. 3273–3280.

Dalal, Navnett/Triggs, Bill (2005): Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, S. 886–893.

Freud, Sigmund (2005) [1921]: Massenpsychologie und Ich-Analyse. Frankfurt: Fischer.

Hegenbarth, Jens (2016): Robuste Detektion von abnormalem Gruppenverhalten mittels Particle Advection und Social Force Model. BA-Arbeit. Ruhr-Universität Bochum, Institut für Neuroinformatik.

Helbing, Dirk/Molnár, Peter (1995): Social force model for pedestrian dynamics. In: Physical Review 51(5), S. 4282–4287.

Horn, Daniela/Ibisch, André/Tschentscher, Marc (2018): Automatisierte Videoanalyse. In: Moritz, Christine/Corsten, Michael (Hrsg.): Handbuch Videoanalyse. Wiesbaden: Springer. S. 445–456.

Idrees, Haroon/Saleemi, Imran/Seibert, Cody/Shah, Mubarak (2013): Multi-source Multi-scale Counting in Extremely Dense Crowd Images. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, S. 2547–2554.

Keysers, Verena/Meißner, Joanna/Reichertz, Jo/Spiekermann, Nils (2018): Situative und übersituative Praktiken des Problematisierens beim Fußball. In: Negnal, Dörte (Hrsg.) (2018): Problem-

gruppen in Staat und Gesellschaft. Praktiken und Prozesse der Problematisierung sozialer Kollektive. Wiesbaden: Springer VS (im Druck).
Le Bon, Gustave (1982): Psychologie der Massen. Stuttgart: Kröner.
Mehran, Ramin/Oyama, Alexis/Shah, Mubarak (2009): Abnormal Crowd Behavior Detection using Social Force Model. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, S. 935–942.
Moritz, Christine (Hrsg.) (2013): Videotranskription in der Qualitativen Sozialforschung. Multidisziplinäre Annäherungen an einen komplexen Datentypus. Wiesbaden: Springer VS.
Moritz, Christine/Corsten, Michael (Hrsg.) (2018): Handbuch Videoanalyse. Wiesbaden: Springer.
Reichertz, Jo (2017): Neues in der qualitativen und interpretativen Sozialforschung? In: ZQF 1-2017, S. 71–89.
Soomro, Khurram/Zamir Amir Roshan (2014): Action Recognition in Realistic Sports Videos. In: Moeslund, Thomas B./Thomas, Graham/Hilton, Adrian (Hrsg.): Computer Vision in Sports. Cham: Springer International Publishing. S. 181–208.
Spata, Dominic (2016): Action Classification Using a Combination of Hough Voting and Random Forest. BA-Arbeit. Ruhr-Universität Bochum, Institut für Neuroinformatik.
Tarde, Gabriel (2015): Masse und Meinung. Konstanz: Konstanz University Press.
Yao, Angela/Gall, Jürgen/Van Gool, Luc (2010): A Hough Transform-Based Voting Framework for Action Recognition. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, S. 2061–2068.

I. Methode

Verena Keysers

Daten. Deuten. Verstehen.

Zu Methode und Methodologie einer hermeneutisch-wissenssoziologischen Eskalationsforschung

„Soziologie (im hier verstandenen Sinne dieses sehr vielfältig gebrauchten Wortes) soll heißen: eine Wissenschaft, welche soziales Handeln deutend verstehen und dadurch in seinen Abläufen und Wirkungen ursächlich erklären will."
(Weber 1973 <1922>)

1. Auf der Suche nach Eskalation

Die erkenntnistheoretische Einsicht, dass die Beobachtung eines Gegenstandes oder Phänomens von den Praktiken und Instrumenten der Beobachtung nicht unabhängig ist, ist für die sozialwissenschaftliche Beteiligung an dem interdisziplinären Forschungsprojekt und folglich auch für diesen Forschungsbericht „Emotion. Eskalation. Gewalt." in mehreren Hinsichten von spezifischer Bedeutung.

Einerseits ist diese Einsicht von Bedeutung, da im Zuge dieses Forschungsprojektes mittels interdisziplinärer Zusammenarbeit die Möglichkeiten und Bedingungen einer videobasierten Eskalationsfrüherkennungs-Software erkundet werden sollten. Was man für ein solches Unterfangen grundsätzlich benötigt, ist ein adäquates Modell, das Eskalationen mit hinreichender Genauigkeit beschreibt – und zwar in einer Weise, die eine kamerabasierte Früherkennung möglich macht.[1] Das heißt, das Forschungsziel der interdisziplinären Zusammenarbeit legte bereits einen speziellen Rahmen fest, innerhalb dessen sich ein sozialwissenschaftlich zu entwickelndes Modell von Eskalation zu orientieren hat.

Andererseits ist diese Abhängigkeit des Beobachteten von Praktiken und Instrumenten seiner Erfassung grundsätzlich Anlass, die Methoden der Datenge-

1 Ein Modell des zu Erkennenden benötigen selbstlernende Algorithmen zu Beginn. Schließlich kann ihr Lernprozess nicht aus dem Nichts angetrieben werden (ausführlicher dazu: Horn et al. in diesem Band).

winnung und Dateninterpretation zu erwägen und in Bezug auf das Erkenntnisvorhaben zu begründen. So ist einerseits zu prüfen, welches Konstruktionsniveau die einzelnen Datensorten aufweisen. In der Sprache des Konstruktivismus gesprochen bedeutet das, dass wissenschaftlich erhobene Daten über die Wirklichkeit eine eigene Wirklichkeit bilden und produzieren. Und zwar eine Wirklichkeit, die sowohl an Relevanz- und Ausdruckssystemen des Forschungsfeldes interessiert ist, aber zugleich Relevanz- und Ausdruckssystemen der Feldforschung folgt und notwendigerweise unterliegt. Dies bedeutet, dass schon die eigene Interessiertheit am Forschungsgegenstand eine Konstruktion beinhaltet.

Aus dem Vorangegangenen folgt, dass sich zunächst immer kritisch mit dem eigenen Wissen (Annahmen, Informationen) und Nichtwissen (Fragen, Interessen) auseinandergesetzt werden muss. Man muss fragen: Was repräsentieren die Daten und was kann ihnen abgewonnen werden?

Unser Interesse an der Forschungsfrage dieser interdisziplinären Kollaboration gründete auf der Frage, welche Arten von Verhaltensabstimmung stattfinden, wenn es zu emotionalen Prozessen in Menschenansammlungen kommt. Sind es neurologische Reaktionen, psychische oder sozialpsychologische Mechanismen, gar Pheromone oder andere Botenstoffe, die das Steuer übernehmen, wenn Menschen massenhaft zusammenkommen, um etwas zu erleben und diese in einen Zustand „kollektiver Efferveszenz“ (Durkheim 1984, 296 ff.) versetzen? Oder sind es doch semiotisierte Deutungsleistungen einzelner Aktivitätszentren, die jene Zustände der Intensivierung von Emotion und (Er-)Regung bewirken, die manchmal dazu führen, dass der Funke überspringt und euphorische Gemeinschaftserfahrungen geschehen, manchmal aber auch nicht in kollektiver Euphorie und Ausgelassenheit, sondern in Gewalt und Zerstörung enden? Letzten Endes ist vernünftigerweise die Frage zu stellen: In welchem Verhältnis stehen Phänomene verschiedener Art? Das heißt also die übergeordnete Forschungsfrage schlüsselte sich selbstverständlich auf in innerdisziplinäre Fragen von grundlagentheoretischem Interesse, in unserem Fall also die Bedeutung von Kommunikation in Eskalationsprozessen bzw. Prozessen „kollektiver“ emotionaler Erregung.

In diesem Artikel wende ich mich zunächst der Frage zu, welche Implikationen die hermeneutisch-wissenssoziologische Methodologie für den Gegenstand, also das *Wie* und das *Was* der Untersuchung des Gegenstandes „Eskalation auf Großveranstaltungen“ hat. Andererseits wird auch die Frage beleuchtet, welche Implikationen sich aus dem Was und Wie des interdisziplinären Forschungsgegenstandes „Eskalation auf Großveranstaltungen“ für die Kalibrierung der interpretativen Sozialforschung ergeben.

2. Deutungsdaten konstruieren

Für unser Forschungsvorhaben benötigten wir Wissen über konkrete Prozesse und Szenarien, da unserem Projekt die Frage zugrunde lag, woran man frühzeitig erkennen kann, dass ein Eskalationsprozess im Gange ist und Gewalt droht. Dieses Wissen muss folglich bestimmte Eigenschaften aufweisen. So muss dieses Wissen über Eskalationsprozesse auf Großveranstaltungen *umfangreich* sein, das bedeutet, wir benötigten nicht nur genaue Angaben zu einem historischen Ereignis einer sicherheitsrelevanten Eskalation, sondern idealerweise zu vielen verschiedenen. Dies hängt damit zusammen, dass es nicht um die Rekonstruktion eines Handelns in einer konkreten Situation geht, sondern darum, ein komplexes, wiederkehrendes Handlungsgefüge zu rekonstruieren (Reichertz/Schröer 1994, 61). Außerdem sollte sich – im Ansinnen der interdisziplinären Zusammenarbeit – dieses Wissen auf *wahrnehmbare Prozesse* beziehen und *prozessual* zu durchdringen sein. Was wir suchten sind Darstellungen oder Beobachtungen, die es zulassen, Eskalationen auf Großveranstaltungen als eine *spezifische Abfolge von Verhaltenseinheiten* zu konzeptionalisieren, die zu Gewalt führt, wenn sich diese Abfolge komplett entfaltet. So ein Modell von Eskalationen zu erstellen, das soziale Prozesse und Eigenschaften von Eskalationen definiert oder mindestens beschreibt, war also die Konstruktionsleistung, die hier gefragt war. Was uns Forschenden für dieses Modell fehlte, war Wissen. Und zwar fehlte uns nicht nur *Wissen über Eigenschaften von Eskalationen* auf Großveranstaltungen, sondern uns fehlte auch Wissen darüber, welche *empirischen Phänomene* dem Begriff „Eskalation" im Kontext von Großveranstaltungen überhaupt zukommen. Daher haben wir uns auf die Suche nach Eskalation gemacht, um zu verstehen, was für Phänomene Eskalationen sind, um analysieren zu können, welche Merkmale Eskalationen auf Großveranstaltungen haben und welche Prozesse sie charakterisieren.

Nun stellte sich die Frage: Wo in der Welt gibt es so ein Wissen, wer hat es und wo und wie können wir Forschenden etwas davon abbekommen? Aus unserer Sicht bestanden vier Möglichkeiten, wie man zu diesem Wissen gelangen kann: Die erste Möglichkeit – und klassischerweise der erste Gang jeder wissenschaftlichen Forschung – ist es, die bestehende Literatur zu sichten und zu prüfen, ob es bereits Forschung gegeben hat, die enthält, was wir für solch ein Modell der Eskalation benötigen würden. Dies ist – zu unserer Überraschung – nicht der Fall. Zwar gibt es bereits aus dem Lager situationistischer Gewaltforschung sehr erkenntnisreiche Beiträge, die für uns bereits wertvolle Hinweise enthielten, welche *Phänomene, Aspekte* und *Faktoren* in Situationen von Ausschreitungen und Gewalt eine Rolle spielen. Besonders instruktiv wa-

ren für uns die Arbeiten von und im Anschluss an[2] Randall Collins und seiner Theorie von Mikrodynamiken der Gewalt (Collins 2009; 2011; 2016) sowie seine Überlegungen zu etwas, das er als „Emotional Energy" bezeichnet (Collins 1990; 1993). All diese Konzepte dienten uns als Anhaltspunkte, also als Sensitizing Concepts[3]. Jedoch handelte es sich dabei nicht durchgängig um Konzeptionalisierungen auf Basis jener Art sinneswahrnehmbarer Merkmale, um die es sich im Zuge unserer Forschungsarbeit dreht. Eigene Datenerhebungen jenseits der Aufnahme des Forschungsstandes waren also angezeigt.

Die zweite Möglichkeit besteht darin, selbst an Großveranstaltungen teilzunehmen, um dort eigene Wahrnehmungen, Beobachtungen und Feststellungen zu machen und zu *erleben*, was dort vor sich geht. Dies ist in Bezug auf unser Interesse an wahrnehmbaren Phänomenen von eskalativen Prozessen (von denen wir im Übrigen noch nicht sagen konnten, wann diese beginnen und wie weit sie sich ausbreiten) auf Großveranstaltungen natürlich eine Datensorte erster Güte. So wurden wir[4] – sofern wir es nicht schon waren – leidenschaftliche Stadiongänger_innen, engagierte Demonstrant_innen, ausgelassene Volksfest- und mitgerissene Festivalbesucher_innen. Doch es ist das Malheur beobachtender Forschung, dass ihre Ethnograf_innen von Beobachtungsgelegenheiten abhängig sind. Was ich im Feld nicht beobachte, aufsammle oder erfahre, kann auch nicht Gegenstand meiner ethnografischen Protokolle werden (Breidenstein et al. 2015, 71).[5] So kam es, dass die Daten, die wir aus dem Feld bezüglich selbst

2 Arbeiten im Anschluss an Randall Collins, die für uns von besonderem, sensibilisierendem Nutzen waren, sind der an Turning Points (Abbott 1997) und Driving Forces ausgerichtete Modus, den jüngere Ansätze prozessualen Erklärens vorschlagen (Hoebel 2015; Aljets/Hoebel 2015; 2017), sowie die situationistisch-mechanistischen Konzepte, die Anne Nassauer im Kontext ihrer Gewaltstudien vorstellt (Nassauer 2012; 2015a; 2015b; 2015c; 2016a; 2016b).

3 „(…) The concepts of our discipline are fundamentally sensitizing instruments. Hence, I call them "sensitizing concepts" and put them in contrast with definitive concepts (…) A definitive concept refers precisely to what is common to a class of objects, by the aid of a clear definition in terms of attributes or fixed bench marks. (…) A sensitizing concept lacks such specification of attributes or bench marks and consequently it does not enable the user to move directly to the instance and its relevant content. Instead, it gives the user a general sense of reference and guidance in approaching empirical instances. Whereas definitive concepts provide prescriptions of what to see, sensitizing concepts merely suggest directions along which to look" (Blumer 1954, 7).

4 Es waren neben den drei hauptberuflich damit befassten Jo Reichertz, Christine Moritz und Verena Keysers im Laufe der drei Jahre weitere Teammitglieder des Forschungsbereiches Kommunikationskultur in die Feldforschung involviert. So ehemalige und assoziierte Mitarbeiter Leif Klemm und Richard Bettmann, die studentischen Hilfskräfte Selma Gleißberg, Chantal Otterbein, Joanna Meißner und Nils Spiekermann sowie die Forschungspraktikantinnen Lara Pellner und Eltje Gajewski.

5 Ausnahmen stellen Introspektionen und Autoethnografien dar, da dort das Tatsächliche zum Erwarteten bzw. zum Innenleben der Forschungsperson in Bezug gesetzt wird.