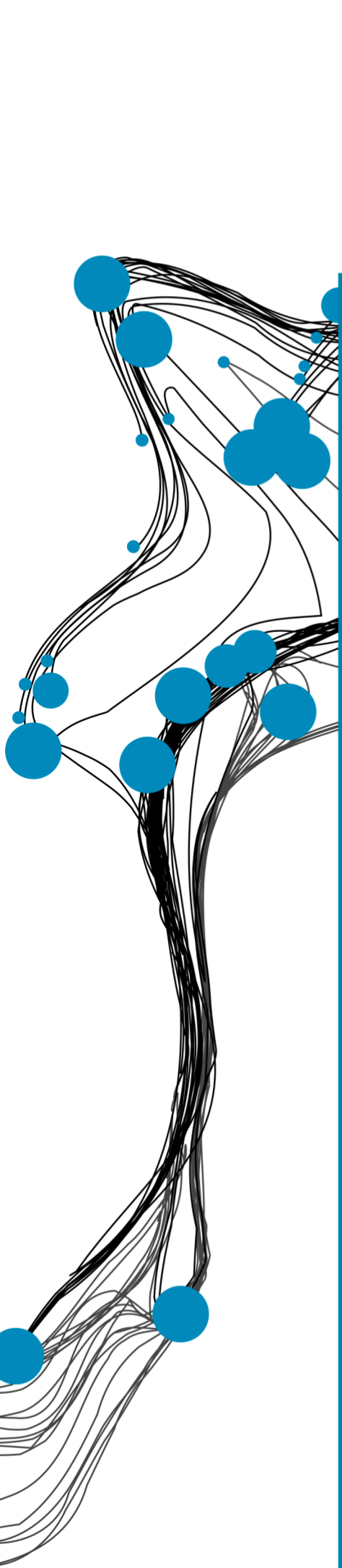

BACHELOR THESIS

The Construction of Evidence in the case of Conducted Electrical Weapons in North-Rhine Westphalia

Kilian Stahlhut (s2391473)

B.Sc. Public Governance across Borders

1st Supervisor: Guus Dix
2nd Supervisor: Shawn Donelly
Wordcount: 11,122
Ethical Approval: 210914

University of Twente, Enschede, Netherlands
Faculty for Behavioural, Management and Social Science

Westfälische Wilhelms-Universität, Münster, Germany
Institut für Politikwissenschaft

Wednesday, 30 June 2021

UNIVERSITY OF TWENTE.

Abstract

In 2020 the German federal state of North-Rhine Westphalia implemented a pilot study on Conducted Electrical Weapons (CEWs). This choice was curious as the scientific community has studied the weapon since the late 90s and already offered evidence in various directions. This thesis investigates evidence on CEWs. More in particular, the thesis examines how two groups of industry affiliated and not industry affiliated researchers produce knowledge about the technology. It utilizes concepts from evidence-based policymaking, philosophy of science and sociology of science to understand differences in scientific practice and the nature of 'appropriate evidence' for policymaking. The thesis derives three hypotheses from the theoretical literature on science and analyses research papers and interviews to test the expectations. The analysis leads to the conclusion that all three hypotheses are well-supported. Against the background of limitations in the existing evidence, the German pilot study can add significant aspects to the body of knowledge on CEWs by allowing policymakers to observe the technology in its native environment.

Abbreviations

ARDs	Arrest Related Deaths
CEWs	Conducted Electrical Weapons
EBP	Evidence Based Policymaking
NRW	North-Rhine Westphalia
RCTs	Randomized Control Trials

Table of Contents

1. Introduction

Conducted electrical weapons have been implemented into police work in over 100 countries. The leading manufacturer, AXON, reports that out of 1,201 CEW uses, 99.75 % result in no serious injuries (AXON, 2020). Evidence provides that the technology is safe for an average, healthy population (Baliatsas, et al., 2021). But research on CEWs must be examined carefully. In 2011, Azadani et al. statistically proved that AXON affiliated researchers are 75% more likely than non-affiliated researchers to prove the technology's safety (Azadni, et al., 2011, p. 535). There appear to be differences in how AXON affiliated and not AXON affiliated researchers derive knowledge about the technology. AXON affiliated research is published frequently. The manufacturer provides a research index on their website, which lists 820 CEW research papers about the technology's implications. Out of these 171 are reported as "partial or fully funded by Axon" (AXON, 2019).

Given this situation, it is important to look whether there are patterns in the evidence on CEW. Dymond and Rappert (2014) suggest pilot studies as an "ideal situation" for doing so. They argue that patterns in evidence on CEWs need to be discovered to assess the technology's practicality for a specific context. In 2020 the German federal state of North Rhine-Westphalia (NRW) has implemented a pilot study on CEWs "to thoroughly test the Tasers for their practicality in a long-term test." (Ministerium des Inneren NRW, 2020; Westdeutscher Rundfunk 2021). Societal opinions on the pilot study diverge. Police and Ministry celebrate the technology can "finally" be implemented (Gewerkschaft der Polizei, 2021). They argue it will improve officer safety. In parliament the green faction found that CEWs "do not appear to be an appropriate means for police duty" (Aachener Zeitung, 2020). After a video was published on social media which captured the use of the weapon in one of the pilot study cities, citizens criticized the "illegitimate use of the weapon" (Westdeutscher Rundfunk (b), 2021). The ministry of the interior found that the weapon was used "legally, permissible and proportionate" (Ruhr Nachrichten, 2021).

Taking up on the scientific and societal controversies around CEWs, this thesis will use the NRW pilot study as a case to examine the relation between science and policymaking. More in particular, it will be analysed how different scientific practice influences the appropriateness of evidence for policymaking. It will be examined whether the appropriateness of evidence for policymaking differs between the AXON affiliated and not AXON affiliated researchers. It will be determined how the two groups conduct research, especially which scientific choices they

make. The influence of scientific practice on the appropriateness of evidence for policymaking will be identified. Based on identified patterns of construction of evidence recommendations will be given for how the pilot study can or cannot add to the existing body of knowledge on CEWs and increase its appropriateness for policymaking.

The theoretical framework will be built upon evidence-based policymaking (EBP). EBP highlights science as the producer of evidence and establishes a concept of appropriate evidence. Appropriate evidence for policymaking, must meet certain criteria that exceed the simple provision of high chances to attain political goals. From an EBP perspective, evidence must be constructed in useful ways, apply to the local policy context, and address the political goal of interest (Parkhurst, 2017). The selected case suggests that appropriateness of CEW evidence is restricted. This might have necessitated the pilot study.

This Thesis will conceptualize appropriate evidence for policymaking to analyse the evidence on CEWs. The understanding of (appropriate) evidence will be reinforced drawing from philosophy of science. It will be presented how science is practiced, and discussed that *inductive risks* are inherent in science. Science is an uncertain activity, the choice for a research question, and methodology can reduce the quality of evidence (Douglas, 2000). It will be scrutinized which choices scientists make and how they justify them. Drawing from sociology of science it will be discussed that scientific practice can limit epistemic diversity, which reduces the appropriateness of evidence for policymaking (Dix, 2019). Three hypotheses will be derived from the theoretical framework to test the appropriateness of CEW evidence for policymaking in NRW. Interviews will be conducted and CEW research studied.

Against the background of the literature on the construction of scientific evidence, the German pilot study might improve evidence and permit evidence-based policymaking. It follows Dymond and Rappert's suggestion to test evidence in the field. It remains an open question, however, why policymakers did not directly act on the evidence that tells CEWs are safe and implemented the weapons. This thesis analyses the production of evidence on CEWs so far to get a better sense of its characteristics and its limits. After doing so, it becomes possible to define what a pilot study can and cannot provide that is different from the current evidence on CEWs. The central research question of the thesis is the following:

What can the German pilot study on conducted electrical weapons (not) learn from the available evidence as it is currently constructed by different groups of research?

The research question is complemented by six sub-questions, which are introduced below. These will be answered chronologically, provide coherence, and improve the answer to the central research question.

(1) *What is evidence-based policymaking?*

EBP and its main elements will be discussed. It will be presented how EBP conceptualizes policymaking and how it perceives the role of science for policymaking.

(2) *What is appropriate evidence for policymaking?*

Drawing from EBP it will further be presented, which characteristics define whether evidence is or is not appropriate for policymaking. This is crucial to understand whether evidence as currently constructed by the two groups of researchers is appropriate for policymaking.

(3) *How is evidence produced by science?*

EBP perceives evidence as the outcome of scientific activities. It will be presented how science is practices, where scientific practice inherits risks, and how these should be mitigated by scientists. Also, it will be discussed how epistemic diversity is lost, which reduces the appropriateness of evidence for policymaking.

The preceding three questions will establish the theoretical framework. Hypotheses will be derived and tested in the Analysis. Next to the hypotheses, three sub-questions will guide the Analysis and inform the answer to central the research question:

(4) *What are differences between the different sources of evidence on CEWs?*

(5) *What are strengths and weaknesses of the current evidence on CEWs?*

(6) *How can evidence on CEWs be improved to enhance its appropriateness for evidence-based policymaking?*

Answering the sub-questions will allow to determine, whether, if so why, the appropriateness of evidence on CEWs for policymaking is limited. Based on the findings the main research question will be answered, it will be formulated how the pilot study can and cannot address weaknesses of the current evidence, also how it could be designed.

2. Theory

This chapter introduces the main theoretical concepts. After a general discussion of EBP the concept of *appropriate evidence* will be introduced. Then, it will be discussed how science produces knowledge. It will be presented that scientific choices at different research stages inherit risks which evoke uncertainty about the results. It will be presented that researchers must justify their decision to mitigate these *inductive risks*. Third, it will be presented that *epistemic diversity* is important for EBP and that it is lost when one strand of evidence predominates in policymaking. One hypothesis will be derived from each of these three theoretical concepts to analyse the data and help to answer the central research question.

2.1 EBP: science as the producer of 'appropriate evidence' for policymaking

This section presents EBP, emphasis will be laid on how science relates to policymaking. Understanding this will be valuable to analyse how a pilot study can inform policymaking. Also, the characteristics of *appropriate evidence* for policymaking will be derived from EBP to evaluate the evidence on CEWs.

2.1.1 Evidence-based policymaking (EBP)

EBP scholars conceive policymaking as the process of deciding "which course of action to follow in specific settings" (Parkhurst & Abeysinghe, 2016). Policymaking is a constant trade-off between "multiple competing social values and concerns" (Parkhurst, 2017, p. 86). Russell et al. add that policymaking is a "constant discursive struggle over the naming and framing of problems" (Russell, et al., 2008, p. 43) rather than a goal-oriented task. EBP conceives policymaking as the process of balancing different social and political beliefs to set political goals.

Parkhurst presents EBP as an opposition to rational decision-making. Understanding his criticism on rational decision-making helps understand the role of science for EBP. Parkhurst describes that rational decision-making underestimates the role of the decision-makers' environment (Parkhurst, 2017, p. 87). Rational choice theory is argued to establish a "naïve rationalism" (Russell, et al., 2008). Meaning it assumes that evidence simply provides information on "what works" for policymaking (Parkhurst, 2017; Russell, et al., 2008). This reduces the evidence for rational policymaking to analytical deductions that are meant to reflect the effectiveness of instrument X to achieve goal Y (Shepsle, 2010, p. 20 f.). Here, science focuses only on the political goal that is achieved *by* an instrument, not the instrument *itself*.

This results in context-based selection mechanisms for evidence: scientific practice that influences the evidence is not subject of scrutiny. For rational decision-making what counts is the mere probability of achieving the desired political goal (Parkhurst & Abeysinghe, 2016, p. 668). This neglects the nature of science, that constructs knowledge according to different scientific interests, as well as the environment in which policy measures will be embedded, that diverges between different contexts.

Unique about EBP is that it acknowledges simple, causal relationships are hardly found in the political arena. Parkhurst rejects the "what counts is what works" conception of evidence for policymaking. Still, EBP utilizes evidence to determine probabilities to attain a desired policy outcome. EBP shares the assumption that assessments of effectiveness, signified in probabilities, help to accurately predict whether an instrument is suitable to achieve a desired political goal. But for EBP science does not establish political goals, it only informs policies. For EBP, the desirability of an outcome does not fall within the boundaries of what science can establish (Kohsrowi, 2019, p. 53; Parkhurst, 2017, p. 23). The scientific proof of an instrument's effectiveness does not directly translate to its implementation. Before evidence is utilized for policymaking, scientific practice must be scrutinized to evaluate whether evidence on an instrument – and therefore the instrument itself – is suitable for the question that policymakers are concerned with (Cartwright, 2011). This indicates that for EBP evidence is context based. It acknowledges that an effect found in one specific setting, might not be present in a different context. Studies establish that the effect "works somewhere" but evidence must prove effects that "work for us" (Cartwright, 2011) to inform policies. Taking up on the notion of desirable and context-based evidence, Parkhurst formulates three characteristics of appropriate evidence for EBP.

2.1.2 Appropriate Evidence

Appropriate Evidence is a central for EBP it presents that evidence must meet three key features, which refer to different scientific disciplines (Parkhurst, 2017, pp. 112-118).

Following a political science approach, evidence must *address the main policy concern*. The main questions of relevance for a political decision must be identified to determine whether an instrument achieves a desired outcome and *how* it manages to do so. Policymakers must have a mutual understanding of what goal they pursue (Russell, et al., 2008). This allows to identify

and scrutinize assumptions that influence scientific results and their appropriateness (Parkhurst & Abeysinghe, 2016).

Following philosophy of science, evidence must be *applicable to the local context* that a policy will be implemented in (Parkhurst & Abeysinghe, 2016). Evidence must not work somewhere, but "work for us" (Cartwright, 2011, p. 1400). The context of discovery must be evaluated to determine whether experiments and studies are only valid in a specific setting or can be inferred to another context. Scientific methods that influence the construction of knowledge must be uncovered. Again, it is essential to make the political goal explicit to identify which different methodological choices negatively influence the appropriateness of presented evidence (Parkhurst & Abeysinghe, 2016, p. 675).

Following sociology of science, evidence must be *constructed in a way that is meaningful* to solve the problem at hand. Scientific evidence is presented as a product of its environment. Norms and values influence the production of knowledge and context of discovery. Categories are established to describe phenomena according to the context of discovery. If evidence is to inform policy, the way science constructs knowledge must be useful to the problem which is desired to be solved (Parkhurst, 2017). Here it is to uncover crucial research limitations and identify how they can or do reduce the appropriateness of evidence for policymaking.

Concluding, even though science might prove that an instrument has a certain effect, for evidence to be appropriate it must address the central policy issue, be constructed in useful ways, and apply to the policy context. Therefore, to understand whether evidence is appropriate we must try to understand how it is produced. The first step to answering the research question will be to identify what the two groups of researchers present as appropriate evidence on CEWs. Identifying how the two groups present the results, benefits and risks of CEWs will provide an understanding of what they perceive as appropriate evidence. The first hypothesis is derived from the theory of EBP but made specific for the case of CEW's:

> H1: Two producers of evidence – i.e. AXON affiliated and non-affiliated researchers – have different ideas about what counts as 'appropriate evidence' for policymaking on CEWs.

2.2 Evidence and Science

To develop an enhanced understanding of how scientific practice influences the appropriateness of evidence this section, in accordance with EBP criteria of appropriate evidence, will discuss two further concepts from philosophy and sociology of science. First, it will be presented how scientific practice inherits risks that can create uncertainties. Following Hempel and Douglas it will be argued that science inherits *inductive risk*s that oblige researchers to adequately justify their scientific choices. Second, it will be presented that *epistemic diversity* is desirable when constructing evidence for EBP. It will be discussed how the quality of evidence might be lost in the process of constructing it.

2.2.1 Scientific Practice and Inductive Risks

Hempel (1965) introduced the concept of *inductive risks* that refers to risks that result from how science is practiced. Hempel describes that science is practiced through the test of hypothesis. The main goal is to produce knowledge, which is "represented by a system of statements which [are] sufficiently supported by available evidence to be accepted in accordance with the scientific principles of test and validation" (Hempel, 1965, p. 91). Inductive risks for Hempel arise because hypotheses derived from theory do often not allow observation. Scientists must infer statements from a hypothesis to put them to test (operationalization). These inferred statements allow to indirectly test a hypothesis. Tests aim to provide probabilities of their occurrence. These probabilities allow to verify or reject a hypothesis (Hempel, 1965, p. 82 f.) and consequently contribute to the body of knowledge. Because knowledge is derived through these indirect tests of inferred statements, Hempel finds science can never establish universal truth of a hypothesis (Hempel, 1965, p. 92). He finds that there is always a chance of wrongfully accepting or rejecting a hypothesis. Reality cannot be grasped in its entirety a chance remains that a different scientific approach provides contradicting findings.

Douglas extends upon Hempel's conception of inductive risks. Based on Hempel's considerations about scientific practice she perceives that science is uncertain. Errors can occur at several stages. She identifies three stages of the scientific process where wrong choice can reduce the quality of results. At these stages, scientists decide how to conduct research and produce knowledge (Douglas, 2007). The chance of errors in the scientific process establishes uncertainty: Why were decisions made? Douglas finds uncertainty as a result of inductive risks arises from choices made at the following stages throughout the scientific process: (Douglas, 2000, p. 563 f.):

(1) The selection of Problems
(2) The choice for a certain methodology
(3) The use of scientific knowledge in society

Scientific choices at these stages must be identified and scrutinized for how they influence results. Douglas argues that "errors [in these categories] lead to clear non-epistemic [moral] consequences" (Douglas, 2007, p. 123). She finds that to reduce uncertainty, and enhance confidence in scientific results, scientific choices at these stages must be made explicit. Researchers must always justify why they made different scientific choices, evaluate how scientific choices influence their results, and consider consequences of errors to mitigate the chance of uncalculated application of findings in society. If researchers do not justify their scientific choices uncertainties will be translated to society and citizens if evidence informs policies.

The second hypothesis follows the first hypothesis but specifically focuses on the identification of scientific practices. Testing this hypothesis allows to identify different scientific choices made by the groups of researchers and determine their influence on what is presented as appropriate evidence. Following Douglas and Hempel, it must be analysed which choices the two groups of researchers make at the central stages of scientific practice that introduce inductive risks. It must be determined whether researchers adequately justify their choices to mitigate uncertainties about their scientific practice. The second hypothesis is derived from philosophy of science but applied specifically to the case at hand:

> H2: The different ideas of 'appropriateness' of evidence on CEWs for policymaking between the two groups of producers are related to scientific choices and evaluation of their consequence.

2.2.2 Loss of epistemic diversity

Constructivist theorists argue that reality is a construct that differs according to scientific lenses applied to perceive and describe reality. Parkhurst shares this conception as he finds that evidence must be *constructed* in meaningful ways (Parkhurst, 2017). Scientific practices themselves are a product of their environment, which reflect societal norms and relations (Jasanoff, 1990, p. 12). Therefore, knowledge and evidence are always to a certain extend constructed.

For EBP scientific practice prescribes methodological rigour, unbiasedness, and precision (Kohsrowi, 2019, p. 51 f.). Science for EBP must deploy reliable methods, only measure the effect of interest, and design studies in ways that allow replication of results. EBP Science does not evaluate whether a political goal is desirable or is scientifically relevant. The evaluation of different political goals is the task of policymakers. A main goal of Science for EBP is the provision of exact estimations of an instrument's ability to attain a political goal as reflected in average effect sizes (Kohsrowi, 2019, p. 65).

In a study of evidence-based educational policy in the Netherlands, Dix (2019) formulates that epistemic diversity and quality get lost when evidence is constructed for policymaking. Meaning, the scope of information, presented to policymakers, is narrowed down in the process of providing policy-relevant evidence. He finds, the combination of different kinds of evidence to calculate an average effect of a policy measure leads to the loss of contextual information (Dix, 2019, p. 190). The calculation of that average effect comes with the in- and exclusion of some studies and thereby reduces the scope of what is presented to policymakers (Dix, 2019, p. 197). Because epistemic diversity can be lost when average effect sizes are calculated it is necessary to evaluate science for its "ability to individually, or jointly with other kinds of evidence, provide comprehensive accounts of policy effects, including details on adverse effects on subgroups" (Kohsrowi, 2019, p. 75).

The third hypothesis builds up on the preceding two and will investigate limitations of the evidence that result from different scientific practice. It will be central to identify how limitiations prevent an implementation of CEWs. It will be scrutinized whether the two groups of researchers provide a comprehensive account of CEWs' effect for the NRW policy context. This hypothesis is derived from the sociology of science and made explicit for the case:

> H3: The predominance of one version of 'evidence' on CEWs – whether affiliated or non-affiliated – would lead to a loss of epistemic diversity and prevent a comprehensive evaluation of the technology's effect when used in the field.

2.3 Conclusion

This section answered the first three sub question:

Central EBP assumptions and the concept of *appropriate evidence* were presented to answer sub-questions (1) and (2). The concept of appropriate evidence for EBP was enhanced by introducing the concepts of *inductive risks* and *loss of epistemic diversity* from philosophy and sociology of science. Both concepts complement each other and describe how science produces knowledge and where it might have limited capacity to inform policymaking, thus answer sub-question (3).

From every presented concept one hypothesis was derived and related to the thesis' case. The hypotheses will guide the data analysis and provide results that are meaningful to answer the remaining three sub-questions as well as the central research question.

3. Methodology

This chapter will introduce the methods which were applied to select the data and will describe how the analysis will be conducted. First, it will be described why the NRW pilot study was selected. Then the dataset and data collection methods will be presented. Last it will be discussed how theoretical expectations will be operationalised for the data analysis.

3.1 Case Description

To answer the research question, the current construction of evidence on CEWs will be analysed. Qualitative research methods will be applied to trace how the two groups of researchers practice science and construct knowledge. It will be examined how different scientific practice restricts the appropriateness of evidence for policymaking. Finally, it will be formulated what the pilot study can and cannot add to the body of knowledge on CEWs.

Russel et al. (2008) and Parkhurst (2017) present policymaking as a discursive activity, where policymakers trade off different societal values. In the case of CEWs, societal values have been traded off: From a medical-technological perspective it appears safe to implement the weapon. CEW advocates praise the weapon as an appropriate supplement for police departments (Gewerkschaft der Polizei, 2021; Ministerium des Inneren NRW, 2020). From an ethical perspective, the argument is made that there remain uncertainties about the weapon. The technology is deemed to be safe only *under certain conditions for a healthy population* (Anaïs, 2011; Baliatsas, et al., 2021). It is argued direct implementation would impose uncertainties upon citizens who are most vulnerable to CEW exposure (Anaïs, 2011).

Policymakers are required to take a middle ground, evaluate what the evidence can and cannot provide, to formulate a final decision. In the case of CEWs in NRW, policymakers formulated a preliminary decision: to implement a pilot study. This decision suggests that science does not produce objective facts which allow straightforward implementation of an examined instrument (e.g. CEWs). In the case of CEWs there seem to be limitations in how science has constructed evidence, that restrict its appropriateness for policymaking.

The case therefore makes a "typical case" (Seawright & Gering, 2008, p. 299) for a qualitative study of EBP. It represents main criteria of policymaking as outlined by EBP scholars: Policymakers encounter evidence on an instrument that promises to achieve a political goal. They must comprehend whether evidence on CEWs provides the weapon can help to attain the political goal of enhanced safety. It must be identified whether the evidence is appropriate for policymaking in the local policy context. Selecting this typical case allows to test the hypotheses of appropriate evidence and scientific practice. It will be studied how evidence on CEWs is currently constructed by both AXON affiliated and not AXON affiliated researchers.

3.2 Data

Evidence on CEWs is the unit of analysis. Different kinds of evidence on the technology will be analysed. This follows Parkhurst's suggestion of a systematic literature review to identify presence and quality of evidence on a policy issue (Parkhurst, 2017). The data set will consist of two distinct groups: (a) AXON affiliated researchers and (b) Not AXON affiliated researchers. This distinction will allow to observe how the two groups produce evidence. Azadani et al. (2011) observe that researchers affiliated with AXON are more likely to prove CEWs' are safe. This suggests that there are differences between the scientific practice of the two groups which presuppose different results and influence the appropriateness of evidence for policymaking.

An extensive list of CEW research, funded and not funded by AXON, was utilized (AXON, 2019) to identify AXON affiliated research. Disclosure statements of research papers were examined to identify whether researchers revealed affiliations with the manufacturer besides research funding. Research was categorized as AXON affiliated if one (or several) of the following categories applied:

(1) The research paper is listed on AXON's research index as "partially or fully" funded
(2) The research paper discloses AXON as a source of funding

(3) One or several authors disclose affiliation with the company (i.e., stockowners, consultants, scientific advisory board members)

If none of the categories applied to a research paper, it was categorized as not AXON affiliated. Purposeful sampling methods were applied to select interviewees and research papers.

The Data set contains two distinct sources of evidence on CEWs:

(a) Research Papers

Papers were included based on their content, not quality. To allow an adequate comparison of evidence, animal-based research was not included. This decision was based on the observation that the extent to which animal study findings can be inferred to humans remains unsettled amongst CEW researchers. To increase certainty about the findings, this research was excluded. Research papers are divided into AXON affiliated and not affiliated research. Purposeful sampling methods guided the selection of research papers (Patton, 1990). The selected papers are diverse in content. It was aimed to gather a sample of research papers which represents contents and research methods applied in the field of CEW research. Papers should be representative of their subgroup (affiliated; not affiliated) to allow a comprehensive assessment of evidence on CEWs. Thus, the sample is a representative, intensity sample.

Appendix A lists the selected research papers for the group of a. affiliated and b. not affiliated researchers The number of affiliated research papers is slightly higher as only one AXON affiliated researcher was interviewed. Also, the affiliated research papers are shorter considering the number of pages.

(b) Interviews

Semi-structured interviews were conducted to supplement the research papers. Interviews are flexible, interactive, and in-depth (Legard, et al., 2003, p. 141) and will help to gain a deeper look into CEW research. This will emphasize the thesis' qualitative approach and pragmatically reduce the literature study workload. Interviews will help to understand and better describe the complex field of CEW research.

It was aimed to gather a sample of information rich cases, that signify what is typical for their field and group (affiliated; not affiliated) of research. Three researchers who have published on CEWs and one politician were selected. The final sample can be described as a typical, intensity sample (Patton, 1990). Interviewees were divided into a. AXON affiliated and b. not affiliated research.

Ideally, an official from the NRW ministry of the interior would have been interviewed. The ministry was unwilling to participate. Instead, a politician who sits on the parliamentary committee for the interior was selected. This does not provide insights into how the scientific community constructs evidence but gives valuable information on how policymakers perceive where the construction of CEW evidence restricts its appropriateness for policymaking. This will improve remarks on the relation between the pilot study and evidence on CEWs.

The sample of interviewed researchers includes not affiliated researchers from Great Britain (Appendix B a.) and the Netherlands (Appendix B b.). Both are currently researching medical and societal implications of CEWs.
One AXON affiliated researcher from Germany (Appendix B c.) was interviewed. This interviewee is a member of the AXON scientific advisory board
One politician from NRW parliament was interviewed (Appendix B d.). This interviewee is member of the ruling party and the parliamentary committee for the interior.
Including interviewees from different countries allows to examine how the influence of local contexts is perceived by researchers from the field.

3.3 Method of Analysis

The analysis aims to discover how evidence is constructed. It will be determined how the construction of evidence on CEWs restricts its appropriateness for EBP. Based on these findings, remarks will be given on what the pilot study can or cannot learn from the current construction of evidence to add to the body of knowledge

To answer the research question, a qualitative content analysis (QCA) will be conducted. This is an "approach of empirical, methodological controlled analysis of texts within their context" (Mayring, 2000). Categories of expected observations will be formulated to analyse the data. Expectations were derived from theoretical concepts and translated into three hypotheses. The hypotheses will be operationalized to guide the analyses. This allows to analyse whether theoretical expectations can be observed in the data. This approach to content analysis that follows theoretically derived categories is labelled "deductive category application" (Mayring, 2000, p. 6).

For deductive category application, categories are organized as “codes”. Codes are a few words, which are constructed to operationalize theoretical concepts for the analysis of a dataset (Saldana, 2016, p. 3). The construction of codes establishes a methodological tool that enables researchers to uncover patterns in the dataset. The established codes will be applied to analyse the dataset. Whenever a given passage of data refers to a code it will be labelled. Identified patterns will be translated into themes throughout and after the analysis. Therewith categories of observation can be established (Saldana, 2016, p. 14). These observations will allow to verify or reject the hypotheses and provide sound answers to the remaining sub-questions as well as the central research question.

Table 1 presents the constructed codes that will guide the analysis. The codes operationalize the theoretical expectations formulated in the hypotheses. Thus, they are organized according to the three hypotheses. Relating to the first hypothesis, the data will be analysed for what is presented as appropriate evidence. For the second hypothesis, it will be analysed which scientific choices the two groups of researchers make and how they justify their choices. According to the third hypothesis, it will be scrutinized how different scientific choices limit the appropriateness of evidence for policymaking.

Dimension	Concept	Topics
Hypothesis 1		
Appropriate evidence	appropriate evidence	Results Benefits CEWs Risks CEWs
Hypothesis 2		
Philosophy of Science	scientific choices and inductive risks	Scientific choices: - Research Question (Variables) - Research Design - Methodology - Sample and Data collection Justification of Scientific Choices Uncertainty
Hypothesis 3		
Sociology of Science	loss of epistemic diversity	Limitations: - Research Question - Methodology - Research Design Average effect sizes - Subgroups - Included Data - Sample characteristics

Table 1- Coding

The application of different codes allows to first derive what is presented as evidence, then investigate the scientific practice that constructs the presented evidence, last examine how scientific choices to construct evidence limit its appropriateness for policymaking in NRW. Based on the findings for each hypothesis it will be formulated what the pilot study can or cannot add to the existing body of knowledge.

3.4 Conclusion

This thesis takes a qualitative research approach. The case of CEWs in NRW represents a typical case for a case study that examines how different scientific practice influences the appropriateness of evidence for policymaking.

CEW research will be analysed to identify how the groups of affiliated and not affiliated researchers construct evidence. It will be examined whether presented evidence and scientific practice diverge between the two groups. Ultimately it will be examined how the current construction of evidence restricts its appropriateness for policymaking. Research papers and interviews will be analysed by applying a coding scheme, this provides a methodologically grounded approach to investigate the influence of scientific practice on the appropriateness of evidence for policymaking in the case of CEWs in NRW.

4. Analysis

This section presents findings from observing the dataset and discusses recurring patterns in the data, it will be organized according to the hypotheses. The results will be divided in affiliated and not affiliated research for every hypothesis. This encourages a clear comparison between the two producers of evidence. Following, it will be evaluated whether the hypotheses are well-supported or ill-supported by the data. The conclusion will answer the remaining sub questions.

4.1 Appropriate Evidence

Table 2 presents the codes which guided the data analysis for the first hypothesis. Results presented by both groups will be identified and differences between presented benefits and risks of CEWs emphasised. This will provide insight into which implications of CEW exposure the groups perceive as important, consequently, what evidence they find appropriate.

Dimension	Concept	Topics/ Terms
Appropriate evidence	Appropriate Evidence	Results Benefits CEWs Risks CEWs

Table 2- applied codes Hypothesis 1

4.1.1 Affiliated research

Affiliated research *results* concern health risks associated with CEW exposure. Health parameters pre-, through and post exposure are examined, laboratory Randomized Control Trials (RCTs) are a priority (e.g. Ho et al., 2012; Dawes et al., 2010; Dawes et al., 2009). Also,

affiliated researchers study single cases of Arrest Related Deaths (ARDs) related to CEW deployment. Here, autopsy and media reports are often the primary source of Data (e.g., Vanga et al., 2009). Affiliated research communicates an overall mortality rate temporal to CEW exposure of 1: 3017 (Kroll, 2019).

The main *benefit* of CEW use presented by affiliated researchers is that ARDs are very rarely the result of CEW exposure. It is concluded that CEW exposure very seldomly influences human health while the weapon is very effective in incapacitating a subject. CEWs appear to have only few risks, an "appropriate, measured and regular [CEW] use is harmless from a forensic point of view" (Kunz, 2017, 85).

An affiliated researcher perceives that there are a few, very unprobeable, *risks* associated with CEWs: "I am convinced that the weapon, if applied appropriately, is not lethal. […] But the probability is not zero." (Appendix B c.). Affiliated research presents the following effects of CEW exposure as the main causes of ARDs: CEW induced traumatic falls, which resulted in 16 fatalities in the USA (Kroll, 2019, p. 4). The risk is calculated as 2.3 per 1 million deployments. Major burns from CEW ignited fires (Brave, 2017, p. 40). Risk of CEW ignited fire fatalities is calculated as 1.9 per 1 million deployments (Kroll, 2019, p.6). The following characteristics are presented to increase *negative effects* of CEW exposure on human health: low body mass index combined with short CEW probe to heart distance (Ho et all., 2011, p. 56); high body mass index (Dawes et al., 2010, p.54); drug intoxication; agitation or confusion; age > 44; male gender; mental illness (Ho, 2009). Affiliated researchers stress the few risks must be viewed in the light of the overwhelming benefits (Ho et al., 2011, p.56; Dawes et al., 2010, p. 253).

4.1.2 Not affiliated research

Affiliated research *results* mainly concern effectiveness of CEW use (e.g. White & Ready, 2007), CEW effects on health (e.g. VanMeenen, 2010), implications of CEW use for police activities (e.g. Sousa et al. 2010), and the odds of suspect and officer injuries (MacDonald et al., 2009). Not affiliated research often compares implications of CEW use with those of other less-lethal weapons such as police dogs or OC spray[1] (e.g. Jauchem, 2005).

[1] Colloq.: pepper spray

The *benefits* of CEW use as presented by affiliated researchers often concern CEWs' effect on police-citizen interaction. The effects are diverse and mostly concern reduced officer and/or suspect injuries, the effectiveness of the weapon to solve violent situations, or the effect CEW use has on the use of other less-lethal weapons. White & Ready report that 80% of officers who deployed CEWs perceive the weapon's performance as satisfying (White & Ready, 438). Ross and Hazlett found that 1 out of 3500 ARDs can be associated with CEW deployments (Ross & Hazlett, 2018, p. 194). Bozeman et al. (2009) describe that of 1,201 CEW subjects, 1,198 received mild (skin irritations from CEW probes) to no injuries. Also affiliated researchers find: "We cannot say that [CEWs] cause substantial health effects" (Appendix B b.).

Risks presented by affiliated researchers often cover CEWs' effects on use of force situations in the field. Ariel et al. (2018) found that when CEWs are present, suspects are likely to become more aggressive towards officers. Also, affiliated researchers find pre-existing cardiovascular disease, agitated conditions and intoxication as further factors that increase *negative effects* of CEW exposure on health (Jauchem, 2015, p. 58). Not presented as a risk but worth noticing for police is that not affiliated research finds no association between CEW application and reduced officer injuries (MacDonald et al., 2009; Smith, 2007). An interviewed affiliated researcher finds that the technology "can be associated to psychological impacts" (Appendix B b.)

4.1.3 Conclusion

H 1: Two producers of evidence – i.e. AXON affiliated and non-affiliated researchers – have different ideas about what counts as 'appropriate evidence' for policymaking on CEWs.

Both groups of researchers find that CEW exposure does not cause severe health risks - respectively ARDs - in a healthy population. However, there is one main difference between the two groups in terms of the presented *results, benefits, and risks*.

Affiliated researchers exclusively investigate the association of CEW exposure and human health effects. Thus, affiliated researchers only present results cover *benefits and risks* of CEWs in the light of health effects. This is contrasted by not affiliated researchers which present *benefits and risks* of CEW use for human health as well as police activities. The scope of evidence presented by affiliated researchers is narrower than for the other group.

Based on the presented findings, hypothesis 1 can be confirmed. The two groups of researchers have different perceptions of appropriate evidence. The presented evidence is not contradicting but emphasises different dimensions of CEWs.

4.2 Scientific choices

Table 3 shows the codes which guided the data analysis for the second hypothesis. This section will highlight different scientific choices at central stages in scientific practice, research interests, methods, designs and samples will be scrutinized. It will be identified how the researchers reflect on their decisions to mitigate uncertainty and inductive risks.

Dimension	Concept	Topics/ Terms
Philosophy of Science	scientific choices and inductive risks	Scientific choices: • Research Questions • Methodology • Research Design • Sample and Data collection Justification of Scientific choices Uncertainty

Table 3-applied codes Hypothesis 2

4.2.1 Affiliated research

Affiliated researchers primarily utilize RCTs for their *research designs*. *Research questions* are posed to study the effect of CEW exposure on health, *variables* differ between RCTs. Exposure times vary from a "standard" 5 second CEW circle (e.g. Ho et al., 2013), to 30 second continuous exposure (e.g. Dawes, 2007). Some studies utilize resting volunteers, others deploy "rigorous exertion protocols" (e.g. Ho et al. 2017). Some affiliated researchers apply the CEW probes via hand to have an ideal probe spread when deploying the weapon (e.g. Ho et al. 2013) others shoot the probes at subjects to replicate in field situations (e.g. Dawes, 2007).

RCTs are designed to simulate situations that hypothetically increase negative effects of CEW exposure on human health. Consequently, different parameters are measured to identify effects of CEW exposure in different situations. For instance, Ho et al. (2009) investigated whether exertion increases the probability of suspect injury from CEW exposure. It was hypothesized that cases of strong physical agitation and exertion, are associated with increased risks of ARDs. Exertion was simulated by a series of push-ups and treadmill running. This illustrates how

affiliated research modulates RCTs according to different research interest. Other researchers include alcohol intoxicated subjects, subjects performing tasks, or subjects cuffed to matts to simulate different CEW exposure scenarios.

The *methods of data collection* are the same for every RCT. Consequently, sample characteristics only show minimal differences. All RCT samples consist of law enforcement personnel collected at AXON training courses. Participants are compensated with CEWs (e.g., Ho et al., 2012; Moscati et al, 2007) Samples have a mean age of 35 to 40 years, are mostly male (more than 90%) and include 30 to 60 participants. Some RCTs have significantly less participants (Dawes et al., 2010, n=12) others significantly more (Ho et al. 2017, n=115). Pregnant, mentally ill, or otherwise vulnerable participants are always excluded.

Besides RCTs, affiliated research utilizes medical-, media- and police-reports to conduct case studies. These investigate extraordinary ARDs that might be associated with CEW exposure (e.g., Kroll et al., 2019; Ho et al. 2009). The number of included cases is usually small. The dominant *research question* in these studies is whether CEW deployment can be considered as the cause of an investigated ARDs.

For affiliated research the main *research interest* is whether CEWs cause ARDs. It seems primarily concerned with improving the weapon's safety profile. It appears that this group perceives the weapon as safe when it causes no deaths.

An interviewed researcher finds that for affiliated research: "The motivation behind the selection [and] examining […] outcomes is not that clear" (Appendix A b.). The researcher finds that affiliated researchers tend to:" say 'we found some associations but those are not clinically relevant'. But in the paper, it is not clear how clinical relevance is defined." (Appendix A b.). Affiliated researchers do not give explicit *justifications of their scientific choices*, for example the decision to measure certain health parameters (e.g. catecholamine, Ho et al., 2010). For the recipients of these studies there remain *uncertainties* about why different parameters are of interest.

Also, the study protocols for affiliated researchers are often not extensive enough to allow replication of results (Appendix A b). For some affiliated research papers, it is difficult to comprehend how researchers arrive at conclusions. Extensive discussions, that relate the results

to wider contexts of use of force and policing are not included. This raises uncertainties about the results.

4.2.2 Not affiliated research

Non-Affiliated researchers pose *research questions* that trace effects of CEW exposure on human health (e.g. Vilke et al., 2007) and CEW effects on police use of force (e.g. Sousa et al., 2010).

Non-affiliated researchers *design* RCTs to investigate the effects of CEW exposure on human health. Compared to affiliated research, the RCTs are more conservative: Probes are mostly applied via hand, when exposed to CEWs subjects are often in supine position (e.g. VanMeenen et al., 2010), if exposed while standing, spotters are provided to prevent injury from falling (e.g. Vilke et al., 2009). Exposure times are limited to maximum ten, predominantly five seconds (e.g. Vilke et al., 2007). But there are also fewer not affiliated RCTs, thus the investigated health parameters are less diverse.

The *methods of data collection* and *sample characteristics* for not affiliated research resemble those of affiliated research. RCT samples contain healthy police officers gathered at police trainings. Even though monetary compensation is offered by researchers, it depends on police departments whether the participants are allowed to be compensated (e.g. VanMeenen, 2013). Samples contain 20 to 30 subjects, who are under 40 years old and dominantly male. To investigate CEW effects, not affiliated researchers publish literature reviews, content analyses and conducts case studies (e.g. Jauchem, 2015; Swerdlow et al., 2009). These also resemble not affiliated case studies, rely on media-, medical- and/or police-reports, and intend to uncover the relation between ARDs and CEW deployments.

The biggest *methodological difference* between affiliated and not affiliated research are large scale field studies. These are conducted by affiliated researchers to investigate the effect of CEWs on use of force in police-citizen encounters. Field studies utilize empirical *data* from police departments. *Research interests* for these studies are the effect of CEW implementation on police operations (Ariel et al., 2018); patterns of CEW use across officers and police departments (Ready & White, 2011; Dymond, 2020); CEW effects on probability of suspect and officer injuries (MacDonald et al., 2009; Bozeman et al., 2009); effectiveness (White & Ready 2007). Except two studies from the UK, these studies are conducted in the USA. *Sample*

characteristics diverge for field studies: MacDonald et al. (2009) cover 24,380 use of force incidents. They used data from 12 police departments to derive odds of officer and suspect injury for different non-lethal weapons. Other studies include significantly less departments accordingly the number of analysed units decreases. Ariel et al. (2018) only include 6,000 use of force incidents. CEWs were only used nine times, which included de-holstering and pointing at targets (Ariel et al., 2018, p. 292 f.).

Field studies' main *research interests* are CEW effects on health, number of injuries in the field and effects on suspect-officer relations. Not affiliated *research designs and methods* are more diverse. Applying variable research designs, establishes more *inductive risks*. There will be more *uncertainty* in not affiliated research due to stratified scientific choices. This group often compares the benefits and risks of CEWs with those of other less-lethal weapons to mitigate uncertainty and present results in the wider context of policing. Effect sizes found in field studies are expressed in likeliness of injury and compared with other police use of force equipment (baton, police dog, etc.). Affiliated researchers *justify* their choices based on extensive literature discussions. Variables are often derived from use of force theory and findings related to of use of force literature. This mitigates uncertainty, recipients can understand what informed a scientific choice, also this makes it easier to comprehend findings in the context of policing and use of force. The level of *uncertainty* is lower for this group due to more extensive justifications of scientific choices.

4.2.3 Conclusion

H2: The different ideas of 'appropriateness' of evidence on CEWs for policymaking between the two groups of producers are related to scientific choices and evaluation of their consequence.

According to Hempel science can never establish truth, knowledge is only attained through indirect hypotheses testing. Based on Hempel's consideration Douglas finds that choosing a research question, a certain methodology or how to use knowledge in society necessarily pose threats to truth content of scientific evidence. Any choice for a scientific practice can introduce inductive risks, which limit the capacity of evidence to be inferred to a wider public. To mitigate these risks, it must be justified why scientific decisions are made allowing recipients to understand why decisions were made and how these influence the presented evidence.

Hypothesis 2 can be accepted. Different research interests determine different scientific practice and choices. Consequently, different inductive risks arise for both groups. While not affiliated researchers often engage in extensive discussions about the theoretical background that informed their choices, affiliated researchers only vaguely justify their scientific choices. Comparing the two groups of researchers it can be examined that not only the perception of appropriateness but also the scientific practice and justification of scientific choices diverge.

4.3 Epistemic Diversity

Table 4 shows the codes which guided the data analysis for hypothesis three. It will be investigated how the different scientific practice limits the appropriateness of evidence for policymaking. Different epistemic practice was observed in the previous section. It will be scrutinized how scientific practice establishes limitations that reduce the quality and appropriateness of evidence for policymaking.

Dimension	Concept	Codes
Sociology of Science	loss of epistemic diversity	Limitations: - Research Question - Methodology - Research Design Average effect sizes: • Subgroups • Included Data • Sample characteristics

Table 4-applied codes hypothesis 3

4.3.1 Affiliated research

The *method of data collection* introduces a significant *limitation* for affiliated research RCTs. Every RCT includes healthy, mentally stable, police officers who participate in CEW training courses. Cases for RCTs are nested: The demographics of affiliated RCTs depend on the demographics of police officers who participate in CEW training courses. Samples contain almost exclusively middle aged, male subjects. There are further differences between lab and field CEW targets: 80% of CEW targets in the field are intoxicated (Bozeman et al., 2009), 95% of CEW targets in the field are emotionally disturbed (White et al, 2007), also it is probable that targets have significant health issues. The influence of CEWs on intoxicated, mentally ill, and vulnerable suspects has never been investigated in affiliated research RCTs. Researchers

assume these characteristics increase negative effects of CEW exposure on human health. Interviewee B explains that:

> "If somebody is healthy, relatively younger, probably, and the taser is used […] yes there won't be any major health effects. But […] those results can only be generalized for those people, but not the average population. I understand the choice [to select police officers]. [It is] Part of the limitation of the current research because it is not possible in other ways." (Appendix B b.)

Affiliated research methods of data collection, and resulting *sample characteristics*, prevent generalization. The established effect sizes do not include the most vulnerable groups, which might be considered a subgroup compared to the wider population but represent the majority of CEW subjects in the field.

Second, methods of data collection do not provide adequate *sample sizes*. Interviewee B finds: "a study including ten people or fifteen people […] then finding some associations or finding nothing. Then you wonder […] is this number enough to do this kind of research?" (Appendix B b.). Small samples prevent inference to the public at large. To detect rare cases and establish robust CEW effect sizes larger number of RCT subjects would be needed (Ho et al., 2011, p.57).

Third, the *research designs* limit affiliated research. RCTs do not accurately represent police use of force situations. In the field suspects are likely to be more violent, more agitated, and under the influence of drugs. RCTs will always be artificial and restrict the inferential power of research (e.g. Ho, 2009; Moscati et al., 2010). Yet, it is "ethically and morally impossible [to conduct RCTs] in borderline situations" that include vulnerable participants (Kunz & Adamec, 2017, p. 84; Appendix B b. & c.). Also, contextual factors of use of force, respectively use of CEWs, cannot be included in RCTs. Policing in the field is dynamic, it should not be assumed that weapons will be applied with similar levels of caution in the field as in RCTs.

Also, the dominant *research interest* in the relationship between CEW exposure and ARDs limits the research. Especially the psychological impact of CEW exposure has not been researched enough. One Interviewee finds:" We do not know enough about the psychological ramifications of the use of tasers" (Appendix B a.). Only formulating research questions that aim to investigate whether CEWs exposure is causal of ARDs, or other severe health issues excludes research that examines further dimensions associated with CEW use, thus limits the quality and appropriateness of evidence for policymaking.

Last, affiliated research exclusively *originates in the USA*. While this does not influence laboratory studies, as their environment should be clinical and artificial, this is more concerning for field data. Affiliated researchers who conduct case studies utilize only *secondary police data*. This is limiting the research in two ways.

Interviewee D finds: "I do not compare the US police to the German police, especially not the NRW police. The Police in NRW is trained very differently" (Appendix B d.). Interviewee C perceives differences in police culture restrict the applicability of US data to the NRW context:" It is a problem: The political and media presentation of the weapon has simply been adopted from the USA. They definitely have more of a cowboy mentality in the USA." (Appendix B c.).

Adding on to that, Police and media reports can be perceived as unreliable and incomprehensive (Ho et al. 2009). An in-depth analysis of ARDs should not be conducted exclusively based on *data transmitted* through intermediate instances as currently practiced by affiliated researchers conducting case studies.

4.3.2 Not affiliated research

Not affiliated RCTs have the same *limitations* as those of the other group of researchers. The sample sizes are small, samples do not represent the average CEW suspects, and lab-based research designs do not represent police use of force in the field.

Field studies have two main limitations:

First not affiliated field studies utilize *police use of force reports*. These are perceived as "inherently problematic" (Dymond, 2020, p. 406). Often officers need to assign labels to provide narratives of use of force events. These Labels are ambiguous (e.g. mental stability), it is difficult to comprehend single events in their entirety just from police reports (Appendix A a.). Correlating and causing factors that influence police use of force can never, or only with difficulty, be identified. Also, if there is no unified use of force reporting but reporting differs between included police departments it is difficult to compare data. One police department might include other variables than another, also variables might be scaled differently. Utilizing police use of force reports as the only source of data for research thus limits its quality.

Second, policing differs not only between countries but also between regions or even departments. *Sample characteristics* limit the quality of presented results. *Contextual factors* such as demographic structures of citizens, demographic structure of police officers, crime rate, workload, department structures and use of force policies are significant determinants of police activities, respectively use of force, and will influence field study results (Ariel, 2018; Ready, White, 2011). Most field studies originate in the USA, some in the UK. Interviewee B's observation that results may only be generalized for sample populations is also relevant for not affiliated field studies: The German policing context is different than that in the USA. German policing "aims to deescalate, American police is more eager to use equipment" (Appendix B d.)

Some studies try to mitigate this limitation by including departments from regions with different socio-economic contexts (Ready & White, 2011). But not only the socio-economic differences influence police work, also internal structures and policies are decisive for how police use force. Field studies have not yet determined exact effects of different contextual factors. It cannot be derived which contextual factors are most significantly influencing the use of force, thus, limitations of sample characteristics for field studies will remain significant regarding contextual factors. The conclusion that "not every agency's experience with [CEWs] will be the same" (Smith et al., 2007, p. 435) must be emphasised.

4.3.3 Conclusion

Hypothesis 3: The predominance of one version of 'evidence' on CEWs – the company's – would lead to a loss of epistemic diversity and prevent a comprehensive evaluation of the technology's effect when used in the field.

Observation of data shows that neither of the two groups provide a comprehensive account of CEWs that offers appropriate evidence for policymaking. Affiliated researchers are limited in their sole focus on health effects which are examined in RCTs. Not affiliated researchers experience difficulties as they are dependent on police use of force data. Hypothesis three is supported by the data.

Combining both sources of evidence – affiliated and non-affiliated – can create a more detailed account of CEWs, but limitations concerning sample characteristics and contextual factors will not vanish. It is important to note that both groups of research do (can) not include those that

are most vulnerable to CEW exposure in their RCTs. Average effects from RCTs cannot be applied to the wide public, especially not police suspects, who are likely to show characteristics that are excluded from both groups' RCTs. It remains questionable how to arrive at conclusions about the effect of CEW exposure on vulnerable subgroups, ethical considerations prohibit to include these groups in RCT samples.

There remain certain gaps in the research around psychological and societal implications of CEW use. Additionally, the effect of contextual factors on police uses of force, especially use of CEWs, must be determined. This will allow to understand which aspects besides CEWs are capable to increase safety in police-citizen encounters.

Selecting evidence of only one group would lead to a loss of epistemic diversity. Even if both groups' evidence was combined to inform policymaking, limitations of effect sizes and uncertainty about contextual factors would remain. It can be assumed that practicality limits CEW research: It does not seem likely that in a field of research, which is considerably small, studies will be conducted in multiple, international environments to enhance the understanding of different contextual factors.

4.4 Conclusion

Analysis of the data provided that all three hypotheses are well-supported. Based on the analysis the remaining three sub-questions will be answered chronologically below:

The fourth sub-question was: *What are differences between the different sources of evidence on CEWs?* The main differences between the two groups are the research design and interest. Affiliated research mainly utilizes RCTs. This group's dominant research interest is the relationship between CEW exposure and ARDs, or health. Not affiliated research utilizes RCTs and conducts field studies in both the USA and UK. Not affiliated researchers investigate how CEWs effect human health, also how the weapon effects police-citizen encounters in the field.

The fifth sub-question was: *What are strengths and weaknesses of the current evidence on CEWs?* The main strength of affiliated research are the extensive analyses of CEW effects on different health parameters, its main weakness is that it does not investigate how the weapon performs in the field. The main strength of not affiliated researchers is the investigation of CEWs in their native environment as well as the use of empirical data to develop CEW injury

profiles based on police data. Contextual factors of field use are not researched enough as study sites as well as use of force reporting mechanisms differ.

The sixth sub-question was: *How can evidence on CEWs be improved to enhance its appropriateness for evidence-based policymaking?* To improve the appropriateness of evidence for policymaking the effect of CEW exposure on vulnerable individuals must be studied. How this is to be achieved is a matter of debate as it appears ethically impossible. CEW exposure might establish a lethal threat for vulnerable subgroups. Further, national use of force reporting systems would allow coherent data collection and reduce limitations of data quality due to differing reporting systems. Also, an independent body of reporting would be desirable to reduce the influence of police forces on reporting. Psychological consequences of CEW exposure must be studied from a subject perspective. CEWs must be studied in different environments to understand how contextual factors affect police use of force and CEW effectiveness. Empirical data should be prioritized.

5. Conclusion

This thesis investigated how two groups of researchers construct evidence for policymaking in the case of CEWs in NRW. The selected case was exploited to analyse how scientific practice influence the appropriateness of evidence. Six sub-questions were formulated to organize the thesis. Three Hypotheses were derived from EBP theory, philosophy of science and sociology of science. This allowed to test whether the evidence on CEWs, as it is currently constructed, is appropriate for policymaking in NRW. The dataset included research papers and interviews. Data was divided in AXON affiliated researchers and not AXON affiliated researchers to compare whether there were differences in the construction of evidence between the two groups. The derived hypotheses guided a qualitative content analysis and were all supported by the data.

The analysis provided three main findings: The two groups have different perceptions of appropriate evidence, presented risks and benefits diverge but are not contradicting. The two groups make different scientific choices and evaluate the associated inductive risks differently. Affiliated researchers do not always justify their scientific practice. The evidence as currently constructed by either group does not provide a comprehensive account of CEW effects. Established effect sizes do not apply to most vulnerable subgroups as these are excluded from RCTs that serve to provide CEW effect sizes. Contextual factors determine police use of force

and differ between countries and regions, as CEW research primarily originates in the USA it cannot be assumed that results can be translated to the German context.

5.1 Discussion

The central research question "*What can the German pilot study on conducted electrical weapons (not) learn from the available evidence as it is currently constructed by different groups of research?*" can be answered based on the findings and answers to the sub-question.

The pilot study in NRW can respond to limitations of the evidence on CEWs as currently constructed by the two groups of actors. It allows scientists to study CEWs in their native environment and analyse how contextual factors influence the weapon's capability to attain the political goal of enhanced security.

Informed by observation of how CEW research is practiced and the evidence on CEW is constructed, the pilot study should take into consideration the following points to mitigate limitations and enhance the appropriateness of evidence for policymaking:

- contextual factors in NRW

The pilot's greatest strength is that it can put CEWs to test in its native environment. Contextual factors are decisive for police operations, identifying which contextual factors influence the effectiveness and use of CEWs will substantially add to the body of knowledge. Having an awareness for intervening, contextual variables will help to understand in which situation the weapon can or cannot be a valuable addition to police equipment in NRW.

It must be kept in mind that NRW is a large, socio-economically diverse state. Policing takes place in different contexts which makes the pilot study prone to selection biases, which would reduce the generalisability of findings. If the selected sample does not accurately reflect different policing contexts in NRW, it would be difficult to infer findings from the four pilot study sites to every other police department. Different policing context must be included. This would allow generalization from the pilot study sites to other police departments. The interviewed politician stated that:" We want to cover the most different areas of police operations in the entire state" (Appendix B d.). The selected study sites are at least different to the extent that three departments operate in urban areas and one in a rural area.

- medical implications

The pilot study should define medical implications as broadly as possible to understand where CEWs affect suspect health and to be assured that no possible effect on health is excluded. Observing gaps in the evidence on psychological effects of CEWs, it is desirable to collect data that enables researchers to investigate psychological implications of CEW exposure. Learning from field experience, medical support should be offered to every suspect immediately after CEW exposure, respectively incapacitation. This reduces the probability of ARDs.

- Officer and citizen experience

To add on to the body of knowledge not only officers, but also suspects must provide a narrative of why and how CEWs were used. It is especially interesting to investigate how the perceptions of CEWs differ between these groups. One interviewee suggested there is a "big gap […] between public perception of when tasers should be used and police officer perceptions when to use it" (Appendix B a.). Testing this observation in NRW would allow to find out whether one strain of arguments dominated during policymaking trade-offs that established the goal of enhanced security to be attained by the implementation of CEWs.

- Ethical imperative

The pilot study must take into consideration the limited knowledge about CEW exposure on most vulnerable groups. It is alarming that mentally disturbed individuals are most often exposed to CEWs. If this also is the case in NRW, the risk of ARDs will be increased. Police officers should be advised to not deploy the weapon on mentally disturbed or heavily agitated subjects. Still, it must be acknowledged that these labels are difficult to be assigned. This emphasises the urgency to offer medical assistance to any suspect after incapacitation.

5.2 Limitations and Future Research

Reflecting the findings, three main limitations come apparent for this thesis:
First, the dataset does not represent both groups of research in its entirety. Studies were included to represent the main characteristics of the specific groups, but it remains a cut-out

Second, the sample of interviewees would have benefited from a second interview with an affiliated researcher. Out of several contacted researchers only the included interviewee responded. Also, it would have been desirable to include an interviewee from the ministry of the interior as it conducts the pilot study. But the ministry was unwilling to participate. As a

replacement a politician was included. The interviewee is part of the ruling party, thus his presentation of CEWs is more optimistic than that of opposition party politicians. It was tried to include at least one opposing politician, but out of several (12) contacted politicians only the interviewee responded.

Third, the theoretical framework only allowed to evaluate how evidence is constructed. For EBP the general political goals are not subject of scientific scrutiny. In this case the goal of enhanced safety, for which to achieve CEWs are considered an appropriate instrument, was not scrutinized. This thesis did not investigate how that goal was established, whether it is desirable to be pursued or not. Awareness of the political goal and how it was established would have allowed statements on whether the pilot study and CEWs are appropriate means for enhancing security in the sense it is politically envisaged. It remains questionable whether insecurity comes apparent in numbers of injured police officers and suspects. It is a general limitation of EBP that it perceives political goals are established outside the scientific realms (Kohsrowi, 2019).

Taking up on this thesis and its limitations, future research should investigate how security is perceived in the political arena. What are the events and political interests behind the goal of "increased security"? Investigating CEW research it appears that security and safety is achieved by CEWs as it reduces the probability of ARDs. It must be ethically scrutinized whether the only requirement for policing equipment should be its lethality. It appears this perspective simplifies the use of force, especially for suspects.

Taking an EBP approach to the question of what safety and security mean, could ask what evidence established the goal of enhanced safety. Cases with contested decisions of high public interest impose certain patterns of evidence exploitation to underpin political goals. Certainly, the case of CEWs in NRW over the past five years would make for an interesting case to trace how different political actors utilize(d) different kinds of evidence to achieve their political goals. This would also provide the ability to trace how the pilot study was implemented and improve the understanding of how it can relate to the body of knowledge on CEWs.

Generally, recent trends of enhanced executive authority must be subject of a scientific analysis that exceeds a simple investigation of how evidence for that goal is constructed. A new assembly act, and public surveillance pilot projects are two recent developments towards

enhanced executive power that are communicated as political means to improve safety The goal itself as well as political instruments must be analysed. It would be interesting to determine how the terms security and safety, that these instruments intend to enhance, are constructed by policymakers, also whether they were formulated on an evidence basis. Finally, it should be analysed how the goals of safety and security relate to other fields and political goals such as economic disparities. It must be identified who is most likely to experience intensified executive authority to investigate whether the instruments that are intended to enhance safety and security simultaneously intensify social disparities.

6. References

Aachener Zeitung, 2020. NRW-Polizei soll TASER im Streifendienst testen. [Online]
Available at: https://www.aachener-zeitung.de/nrw-region/dienststellen-der-nrw-polizei-sollen-taser-im-streifendienst-testen_aid-50502403
[last access 29 March 2021].

Anaïs, S., 2011. Ethical interventions: Non-lethal weapons and the governance of Insecurity. *Security Dialogue*, 42(6), pp. 537-553.

AXON , 2020. How Taser Energy Weapons Protect Life And Enhance Safety. [Online]
Available at: https://www.axon.com/resources/articles/how-taser-cews-protect-life-and-enhance-safety
[last access 29 March 2021].

AXON, 2019. CEW Research Index. [Online]
Available at: https://axon.cdn.prismic.io/axon%2F662254c9-6a7f-4de4-b254-c6b46075e22b_2019--03-20+cew+index.pdf
[last access 17 May 2021].

AXON, 2020. How Taser Energy Weapons Protect Life And Enhance Safety. [Online]
Available at: https://www.axon.com/resources/articles/how-taser-cews-protect-life-and-enhance-safety
[last access 29 March 2021].

Azadni, P. et al., 2011. Funding source and author affiliation in TASER research are strongly associated with a conclusion of device safety. *American Heart Journal,* 162 (3), pp. 533-537.

Baliatsas, C., Dückers, M., Gerbecks, J. & Yzermans, J., 2021. Human Health Risks of Conducted Electrical Weapons Exposure: A Systematic Literature Review. *JAMA Network Open*, pp. 1-14.

Cartwright, N., 2011. The Art of Medicine: A philosopher's view of the long road from RCTs to effectiveness. *The Lancet*, Volume 377, pp. 1400-1401.

Dix, G., 2019. Microeconomic forecasting: Constructing commensurable futures of education reforms. *Social Studies of Science*, 49(6), pp. 180-207.

Douglas, H., 2000. Inductive Risks and Values in Science. *Philosophy of Science*, 67(4), pp. 559- 579.

Douglas, H., 2007. Rejecting The Ideal Of Value Free-Science. In: H. Kincaid, J. Dupre & A. Wylie, Hrsg. *Value-Free Science Ideals and Illusions*. Oxford: University Press, pp. 120-139.

Dymond, A. & Rappert, B., 2014. Policing Science: The lessons of the Taser. *Policing*, 8(4), pp. 330-338.

Gewerkschaft der Polizei, 2021. Endlich: DPEIG-Pilot startet in vier Behörden. *DP-Deutsche Polizei. Das Mitgliedermagazin der GdP*, 21(1), p. 4.

Hempel, C. G., 1965. Science and Human Values. In: C. G. Hempel, Eds. Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. New York: The Free Press, pp. 81-96.

Jasanoff, S., 1990. Rationalizing Politics. In: S. Jasanoff, Eds. *The Fifth Branch*. Cambridge: Harvard University Press, pp. 1-19.

Kohsrowi, D., 2019. Trade-offs Between Epistemic and Moral Values in Evidence-Based Policy. *Economics and Philosophy*, 35(1), pp. 49-78.

Legard, R., Keagan, J. & Ward, K., 2003. In-depth Interviews. In: J. Ritchie & J. Lewis, Eds.. *Qualitative Research Practice*. London: Sage Publications, pp. 138-169.

MacDonald, J., Kaminski, R. & Smith, M. R., 2009. The Effect of Less-Lethal Weapons on Injuries in Police Use-of-Force Events. *American Journal of Public Health*, 99(12), p. 2268–2274.

Mayring, P., 2000. Qualitative Content Analysis. Forum: Qualitative Social Research, 1(2), Article 20.

Ministerium des Inneren NRW, 2020. [Online]
Available at: www.nrw-im.de [last access am 03 30 2021].

Parkhurst, J., 2017. The politics of evidence: From evidence-based policy to the good governance of evidence. Abdington, Oxon (UK): Routledge.

Parkhurst, J. & Abeysinghe, S., 2016. What constitutes "good" evidence for public health and social policy-making? From hierarchies to appropriateness. *Social Epistemology*, 30(5-6), pp. 665-679.

Patton, M., 1990. Designing Qualitative Studies. In: M. Patton, Hrsg. *Qualitative Evaluation And Research Methods*. Beverly Hills, CA: Sage, pp. 169- 186.

Ruhr Nachrichten, 2021. Innenminister äußert sich zu Taser-Einsatz in der Nordstadt. Ruhr Nachrichten, 10 May 2021.

Russell, J., Greenhalgh, T., Byrne, E. & Mcdonnell, J., 2008. Recognizing Rhetoric in Health Care Policy Analysis. *Jorunal of Health Services Research & Policy*, 13(1), pp. 40-46.

Saldana, J., 2016. An Introduction to Codes and Coding. In: J. Saldana, Hrsg. *The Coding Manual for Qualitative Researchers*. London: Sage, pp. 1-43.

Seawright, J. & Gering, J., 2008. Case Selection Techniques in Case Study Research. A menu

of qualitative and quantitative options. *Political Research Quarterly*, 61(2), pp. 294-308.

Shepsle, K. A., 2010. Analyzing Politics. 2nd Hrsg. New York: W. W. Norton & Company.

Westdeutscher Rundfunk (b), 2021. Anwohner filmt TASER-Einsatz in Dortmunder Nordstadt. [Online]
Available at: https://www1.wdr.de/mediathek/av/video-anwohner-filmt-tasereinsatz-in-dortmunder-nordstadt-100.html
[last access: am 27 April 2021].

Westdeutscher Rundfunk, 2021. NRW testet TASER im Pilotversuch. [Online] Available at: https:https://www1.wdr.de/nachrichten/landespolitik/nrw-pilotversuch-taser-100.html
[last access: 03 22 2021].

7. Appendix

A. Dataset Research papers

a. AXON affiliated

1) Brave, M., 2017. Medical Examiner Collection of Comprehensive, Objective Medical Evidence for Conducted Electrical Weapons and Their Temporal Relationship to Sudden Arrest. London: 19th International Conference on Forensic Science and Crime.
2) Criscione, J. & Kroll, M., 2014. Incapacitation recovery times from a conductive electrical weapon. Forensic Science, Medicine, and Pathology, Volume 10, pp. 203-207.
3) Dawes, D., 2009. Effects of CEWs on Respiration. In: M. Kroll & J. Ho, (Eds.). TASER Conducted Electrical Weapons: Physiology, Pathology, and Law. New York: Springer Science+ Media, pp. 167-179.
4) Dawes, D. et al., 2008. Second conducted electrical weapon exposure does not cause core temperature elevation in non-environmentally stressed resting adults. Forensic Science International, Vol. 176, pp. 253-257.
5) Dawes, D., Ho, J. & Miner, J., 2009. The neuroendocrine effects of TASER X26: brief report. Forensic Science International, Vol. 103, pp. 14-19.
6) Dawes, D., Ho, J., Reardon, R. & Miner, J., 2010. The cardiovascular, respiratory, and metabolic effects of a long duration control device exposure in human volunteers. Forensic Science, Medicine, and Pathology, Vol. 6, pp. 268-274.