Delft University of Technology

Joint Interdisciplinary project
TUD4040

6.1.1 De Nationale Politie: Final Report

A formula to measure the effect of police interventions

Authors:
Raghav Shankar
Michael Glaser
Kevin Veerkamp
Laura Ottevanger

June 28, 2023

Abstract

Currently, there is not a clear and well defined metric available that can be used to measure safety and in turn justify resource allocation within the police Department. This metric will help decision makers understand the situation better prior to taking major decisions. The large amount of data available to the police is not yet used to its full potential when making these critical decisions. This project aims to translate and quantify the qualitative concept of safety by relying on measurable values found in the Netherlands. The creation of this metric will successfully allow the police to compare, over time, how police resource allocation and intervention tactics lead to a safer society in the Netherlands. The proposed final equation is put together, combining Crime-Harm Index, Utility, and Effectiveness factor of the police. Each of these individual components of the equation were studied individually and the final equation has been explained and validated with hypothetical values. This leads to a composite safety factor, which is bounded from 0 to 1. The safety factor can be later visualised, essentially displaying a hot-spot map that updates frequently. This will help the police in determining the effectiveness of their decisions and measure the impact of their interventions to a certain extent. Moving forward, we believe the Dutch National police should take such a form of measurement into serious consideration, as this equation explores a more holistic representation of safety in Dutch society through various factors.

Samenvatting

Op dit moment bestaat er geen duidelijk gedefinieerde metriek die kan worden gebruikt om veiligheid te meten. Hierdoor is het vaak lastig om de toewijzing van middelen van de politie te rechtvaardigen. De metriek voorgesteld in dit rapport zou besluitvormers kunnen helpen om de situatie beter te begrijpen voordat een grote beslissing gemaakt moet worden. De grote hoeveelheid data beschikbaar bij de politie wordt nog niet gebruikt tot haar volledige potentie bij het nemen van deze kritieke beslissingen. Het doel van dit project is het vertalen en kwantificeren van het kwalitatieve concept van veiligheid met behulp van meetbare waardes in Nederland. Het creëren van deze metriek kan de politie assisteren met het vergelijken van de toewijzing van politie middelen en tactieken over een bepaalde tijd en hoe dit kan leiden tot een veiligere samenleving. De voorgestelde oplossing is een formule waarbij de Crime-Harm Index, Utility en Effectiviteit van de politie worden samengevoegd. Ieder van deze componenten van de formule zijn individueel bestudeerd en de totale formule wordt uitgelegd en gevalideerd met hypothethische waardes. Dit leidt tot een samengestelde veiligheidsscore, die een waarde zal zijn tussen de 0 en de 1. De veiligheidsscore kan worden gevisualiseerd op een kaart die regelmatig kan worden geüpdatet. Dit kan de politie helpen in het bepalen van de effectiviteit van hun beslissingen en het meten van de impact van hun interventies. Wij geloven dat de Nederlandse Politie een dergelijke manier om veiligheid te meten in overweging moet nemen aangezien deze formule een holistische representatie van veiligheid in de Nederlandse samenleving door middel van verschillende factoren laat zien.

Contents

1 A rationale for the problem choice

1.1 Why the police proposed this problem and its importance to the Society

Currently, a lot of police work is built on best practices. Individuals gain experience while working and this experience is used to adequately make decisions. In the last couple of years, the police slowly adopted more and more types of technologies to assist them in their work. The present dependence on and usage of technology has shown to be advantageous. However, the problem at hand is that every type of technology used is distinct and there is a disconnect. Missing within these types of technologies is a central component to link everything together.

The present system for funding the police is likewise based on best practices. The issue with it is that there is no metric to assess whether or not allocating a certain amount of resources to a problem would really fix the problem. To further explain, think of the drug war in the United States. Although those involved in drug-related crimes are apprehended and convicted, it is impossible to tell if this has improved society. On the contrary, some data suggests that the drug war caused more harm than benefit. A study by Volkow showed that "Imprisonment, whether for drug or other offenses, actually leads to much higher risk of drug overdose upon release" [37].

To put it more abstractly, most police actions are currently based on experience and personal convictions on what worked in previous cases because results of performed interventions can be very difficult to measure. In addition, when a study appears that shows the effect of certain police actions, there is a lot of bureaucracy that needs to occur before a change in the current interventions can be done. However, another solution that prevents waiting on studies to happen, is measuring safety done by the police.

What must be done is an essential step in the measurement of safety, future research, and the integration of all knowledge. Simply said, decision-makers in office cannot rest their decisions on educated guesses; there must be a means to evaluate them. It is safe to say that the ability to measure something is a necessary condition for creating it. For something to be possible, a parameter must be present. Policies based on educated guesses can sometimes have unintended consequences such as rewarding people for doing the wrong thing. This is known as "Perverse Incentives" or the "Cobra Effect", named after a case that dates back decades. There were many cobras when the British empire colonized India. A reward was given to a native who caught a cobra and gave it to the British colonizers. As a result, the people began to breed cobras so they could use them and reap rewards. Simply put, one cannot create policies solely based on examples.

1.2 Individual reasons for choosing the problem

Just like the police had their reasons to propose this problem, we all had our individual reasons to choose to work on it.

1.2.1 Raghav

In terms of the potential influence this initiative might have on society, I believe it to be the most ground-breaking. I chose to pursue this project because I have a strong enthusiasm for using technology to solve issues and because it calls for a cross-domain understanding of technology, management, and policies. This project also supports a number of additional initiatives, which puts the significance of the original endeavor into perspective. I believed that I could contribute to the project by combining my IT knowledge, data analytical abilities, and management capabilities. It is inevitable that such programs at any law enforcement organization will receive attention and criticism owing to the ethics and ramifications involved. This intrigues me the most about the project, in my opinion. Another reason I was inspired to choose this topic was the chance to put what I have learned to use and to comprehend the ethics underlying this endeavor. More significantly, I thought it would be satisfying to witness the project's finished product. Growing up in a developing country, I understood the value of proper law enforcement and a trustworthy administration. I would have the chance to contribute to the same by taking part in this project.

1.2.2 Kevin

For a lot of the JIP cases, upon reading the problem descriptions, I felt that while the case sounded interesting I was not sure how I could meaningfully contribute to it. As a somewhat theoretically leaning math student, I got the feeling that most JIP cases did not need any theoretical mathematics and would be better of with another student on the case. This case stood out to me as the main exception. While most of the mathematics I learned would probably not be applicable to the quantification of police benefits, I felt like this case would really benefit from having someone who is used to working with abstract concepts. Moreover, I thought this problem would give me the opportunity to use some interesting mathematics in a practical context. Finally, even though I did not have any particular knowledge about how the police operate, having a universal way to quantify the benefits of police work intuitively seemed very useful to me. Being able to intuitively understand the usefulness of solving this problem, even without knowing much about the police, was yet another reason for me to choose this problem.

1.2.3 Laura

The problem initially jumped to my attention as my father has been a firefighter for as long as I have been alive and I am very familiar with the inner working of the firefighting department. I have heard many stories of encounters where some higher-up within the Veiligheidsregio made a decision that was not well received by firefighters and where there was no clear explanation presented to the ones lower in the hierarchy that are directly impacted by the decisions. As the police works in a similar manner, I was very eager to assist in this problem. At the initial decision, I hoped that I could be useful by analyzing data and maybe using data analysis and machine learning to solve the problem.

1.2.4 Michael

"What is safety?" If you asked this question to 100 people, you would probably get 100 different answers. To me, this leads to a very interesting project: defending a definition through quantitative analysis and reasoning - not "gut-intuition." Furthermore, this question is being tackled by one of the most powerful organisations in Dutch society. To have the opportunity as an international student to address this problem with a unique perspective is a privilege. Regardless of facts or reality, the police will always be "controversial" due to the impact that they have on so many peoples' lives. I think backing police decision-making with statistics and evidence-based reasoning is a solution to distrust, and this project is a step in the right direction.

2 Problem definition

Currently, there is not a well-defined metric to justify how the allocation of police resources will lead to a safer Dutch society. Unit-based comparison tools that measure qualitative information already exist in other fields. In medicine, for example, QALY (quality-adjusted life years) is used to quantify which interventions are the biggest benefit to society and should receive the most funds [39]. A similar "safety" metric for Dutch police is necessary.

The goal of this project is to translate the qualitative concept of safety into a precise quantitative "unit of safety" that relies on measurable values found within the Netherlands. Converting qualitative information into tangible units of measurement comes with several challenges, such as the justification of the significance of different "variables of safety," accessibility of data, and ease of adoption by the Dutch National police. This unit of measurement for safety ultimately must be the culmination of several safety factors. We believe the creation of this metric will successfully allow police to compare, over time, how police resource allocation and intervention tactics lead to a safer society in the Netherlands.

3 Sustainability goals

The sustainability goals from the United Nations Development Programme [35] that are most relevant for our problem are the following:

- 5 Gender equality
- 8 Decent work and economic growth
- 10 Reduced inequalities
- 16 Peace, justice and strong institutions

In this section, we will address these goals and how they are relevant to our project.

3.1 Gender equality

First of all, the development goal called "Gender equality" is of relevance to our solution. The main subgoal that we are targeting is: "End all forms of discrimination against all women and girls everywhere". This subgoal is relevant as we look through the perspective of the victim. There should not be any difference if the victim is a man, woman or a non-binary person.

3.2 Decent work and economic growth

A second relevant goal is "Decent work and economic growth". The relevant subgoal for our problem is: "Promote development-oriented policies that support productive activities, decent job creation, entrepreneurship, creativity and innovation, and encourage the formalization and growth of micro-, small- and medium-sized enterprises, including through access to financial services". This subgoal is relevant as the aim of the solution that is suggested in this paper is to advise the police with their job and ensure a more productive police force.

3.3 Reduced inequalities

Another relevant goal is "Reduced inequalities". The relevant subgoal for the problem is: "Ensure equal opportunity and reduce inequalities of outcome, including by eliminating discriminatory laws, policies and practices and promoting appropriate legislation, policies and action in this regard". This subgoal is relevant as the solution is regarding both the perspective of the victim as the perspective of the perpetrator. In both of these cases, there should be a regard of how well of the victim and/or perpetrator are. A victim that might have a considered amount of wealth cares less when 100 Euros are stolen while a victim that is less well of might be heavily impacted when the same amount of Euros is stolen.

3.4 Peace, justice and strong institutions

The final relevant sustainability goal is "Peace, justice and strong institutions". There are multiple goals relevant: "Develop effective, accountable and transparent institutions at all levels" and "Ensure responsive, inclusive, participatory and representative decision-making at all levels". The solution develops a transparent formula that is clear to everyone. There is no black-box algorithm indicating the impact of the police involved. The solution will make the choices made by the police a bit more transparent as it assists the decision-making process.

4 Measuring harm and mitigation as a police objective

To quantify the benefit or effectiveness of police work, we deem it necessary to first specify qualitatively the desired outcome of police work in a general sense. The police spend a large portion of their resources fighting crime, either by preventing it or arresting suspects when a crime is suspected to have occurred. For this reason, reducing the amount of crime as much as possible can be seen as the desired outcome of police work. Consequently, a reduction in the amount of crime would be viewed as a step in the right direction.

To be able to quantify the effectiveness of police work in terms of reducing crime, it is necessary to specify what counts as a reduction in crime and to quantify the extent of that reduction. A theoretically straightforward approach to quantify crime is to simply count the number of crimes that occur. Using this approach, a reduction in crime is synonymous with a reduction in the total number of crimes, and the change in the total number of crimes represents the extent of the reduction in crime.

There are (at least) two problems with the above approach; the first problem is that counting crimes is not as easy in practice, because not all crimes get reported [14]. This problem is not unique to this approach and we expect it to always be present to some extent. This problem is elaborated on in subsection 8.1. A second problem is that this approach to quantify crime treats all crimes equally. Under this approach, preventing one case of shop-lifting would be considered as beneficial as preventing one murder, even though most people would prefer the prevention of the murder.

Reducing the number of crimes that occur without considering the type of crime seems to be an inadequate objective for the police. To obtain a better objective, it seems useful to consider the underlying reasons for fighting crime. One of the main reasons why actions are considered criminal is that they cause harm [25]. As such the prevention or mitigation of harm caused by crime can be viewed as the objective of police work and we will adopt this point of view. From this point of view, the benefit of police work is equivalent to the amount of harm they prevent or mitigate. The same idea could be applied to other organizations whose goal is to prevent or mitigate harm in a specific context, such as fire fighters, hospitals or rescue workers. A brief mathematical exploration of the quantification of crime harm can be found in Appendix A.

5 Selection Process and Rationale

Traditionally, safety has been addressed from single factor indicators, such as arrest-rate, crime-rate, or number of prisoners. However, a realistic definition of safety can not be defined by only one factor; it is composed of various qualitative and quantitative indicators from a variety of stakeholders in society. Ultimately, a rational framework for safety requires multiple cognitive perspectives. These perspectives consider a perpetrator's punishment, a victim's physical harm, and the qualitative feeling of trust and peace in society.

The selection process of this project was to find key indicators that reflect the "multiple perspectives" of safety. Each solution concept for the measurement in this project expands on the previous concept, as new but equally important qualitative variables arise that expound a more realistic definition of safety. Three solution concepts have been developed:

5.1 Safety from a perpetrator's perspective: A Crime-Harm Index

The first solution concept to creating a measurement of safety is to consider safety purely in a traditional format: a perpetrator's punishment. This traditional perspective has been studied for decades in criminology; vast amounts of data on perpetrators are available. One classification of a perpetrator's punishment is average sentencing time for a crime. This quantitative measurement, often referred to as the Cambridge Crime-Harm Index [30], allows for the severity ranking of specific types of crime to be compounded with their frequency of occurrence. Therefore, the severity of different crimes, such as assault and petty theft, can be compared. Additionally, a crime-harm index can also serve as a moral indicator on the severity of certain crimes according to Dutch law and society.

5.2 Safety from a victim's perspective: Utility

While comparing types of crime in the CHI is useful, there are several flaws with this method [28]. The largest is the lack of consideration for the victim. Measuring prison sentencing time does not actually define whether a victim's safety has been jeopardized. As such, the utility (measurement of happiness or satisfaction) of a victim both before and after a crime must be studied. By finding the change in a victim's utility from a crime, the true severity of the crime can be measured from a victim's perspective. This utility, further defined in section 6.2, is composed of two factors: physical/mental health (QALY, used in medicine) and financial health (a function of a victim's wealth).

5.3 Safety from society's perspective: Trust and Effectiveness of Police

Finally, consideration of the victim and perpetrator perspective does not provide a truly holistic view of safety; the effectiveness of intervention techniques from the police must also be considered at a societal level[28]. This factor, called the effectiveness factor, should comprise two core components. The first component is a victim's measuring of police effectiveness immediately after a crime. The second component is society's measuring of police effectiveness in general, over longer periods of time.

5.4 Combining all perspectives: Safety at a macro and micro level

Through the combination of three perspectives (a perpetrator, a victim, and society) a realistic measurement of a person's safety in Dutch society can be calculated. This measurement, written in the form of an equation, can be geographically compared across the Netherlands over time with different intervention techniques implemented by the police.

6 Proposed Solution

"Interdisciplinary research is any study or set of studies conducted by experts with expertise in two or more distinct scientific fields. The research employs a technique that is not specific to any one field of knowledge and is based on a conceptual model that integrates theoretical underpinnings from other disciplines". However, as scientists, we must first engage in intradisciplinary research that involves a critical investigation of a subject matter within one or more pertinent disciplines before turning to interdisciplinary research [16]. The simplified equation for safety, shown below, is a reflection of this intra- and interdisciplinary investigation. This equation, consisting of three factors, ultimately determines a unit benefit of safety across all types of crimes, victims, perpetrators, and geographic regions:

$$\text{Factor of Safety} = \Delta\text{utility} * \text{CHI} * \text{effectiveness factor}$$

$$b = \Delta u * c * f \tag{1}$$

Where:
Δutility = Δu, where Δu is the change in utility caused by crime.
CHI = c, where c is the Dutch Crime-Harm Index.
effectiveness factor = f, where f is the effectiveness factor of the police.

Ultimately, this equation is the culmination of three solution concepts with additional new factors of safety generated from interviews with stakeholders (detectives, dutch police researchers, economists, and the general public).

6.1 Netherlands Crime-Harm Index (CHI)

The first factor to creating a measurement of safety is to consider safety as a reduction in "crime-harm". Crime-harm can be calculated through average sentencing time of a perpetrator. This quantitative measurement, often referred to as a Crime-Harm Index [30], allows for the moral severity ranking of specific types of crime to be compounded with their frequency of occurrence.

6.1.1 Background and Research

The method of assigning a weight to different categories of crime to represent moral severity has been done before. In fact, this is the idea behind the Cambridge Crime-Harm Index [31] [32]. Because the total number of crimes does not accurately reflect the harm done by crime, crimes are assigned a weight based on the type of crime. In the Cambridge CHI, these weights are based on sentencing guidelines for different types of crime [31]. The weights are expressed in number of days imprisoned. If the recommended sentence is a fine instead of a prison sentence, the amount of the fine is converted into a prison sentence by considering the number of days one needs to work at minimum wage to pay the fine.

Weighting crimes instead of simply using total number of crimes is supposed to better represent the harm caused by the committed crimes as severe crimes receive a higher weight. Using total number of crimes puts disproportionate weight on less severe crimes that occur frequently. Assigning weights per crime type (instead of per individual crime), while sometimes less accurate, allows for more efficiency in calculating CHI values. This is because crimes are already being categorized by type, and this data is readily available [31]. Furthermore, choosing the weights based on sentencing guidelines ensures that the CHI framework has a basis in the judicial system, making it more likely to be approved by the government [31]. Not taking into account the actual sentence for each crime makes sure that the offender's prior record does not influence the contribution of that crime to the CHI [31]. The idea behind this is that the harm caused by a crime does not inherently depend on the prior crimes of the offender. Typically, a CHI is applied to the method of "hot-spot policing" [21], where police concentrate resources depending on their interpretation of where the most severe crime occurs. In California, New York, Toronto, and the UK, for instance, hot-spot policing has proven to be an effective method in crime reduction when combined with a hot-spot map displaying CHI.

Finally, including a crime-harm index based on Dutch law allows for moral indicator of crime within Dutch society. Every country values the severity of crime differently, and the application of a Dutch CHI reflects the Dutch government's moral compass on safety within the Netherlands.

6.1.2 Application to Dutch Law

While a crime-harm index already is in use in English-speaking countries such as the UK, US, and Canada [31], its application in the Netherlands is still in development [28]. For proof-of-concept, an entirely unique and preliminary version of a crime-harm index for the Netherlands was created based on *Wetboek van Strafrecht* (Dutch criminal code). The initial process for developing a CHI follows that of the Cambridge CHI. First, the sentencing time for types of crime were compiled. For this case, 55 specific types of crimes were considered, and the sentencing time was based on the maximum fine for every crime, stated in the Dutch criminal code. Minimum wage was calculated at 81.06 Euros per day, and fines of categories 1-6 were based on current Dutch law. Each fine category was multiplied by the minimum-wage day to determine the sentencing time (in days) for every type of crime, shown below.

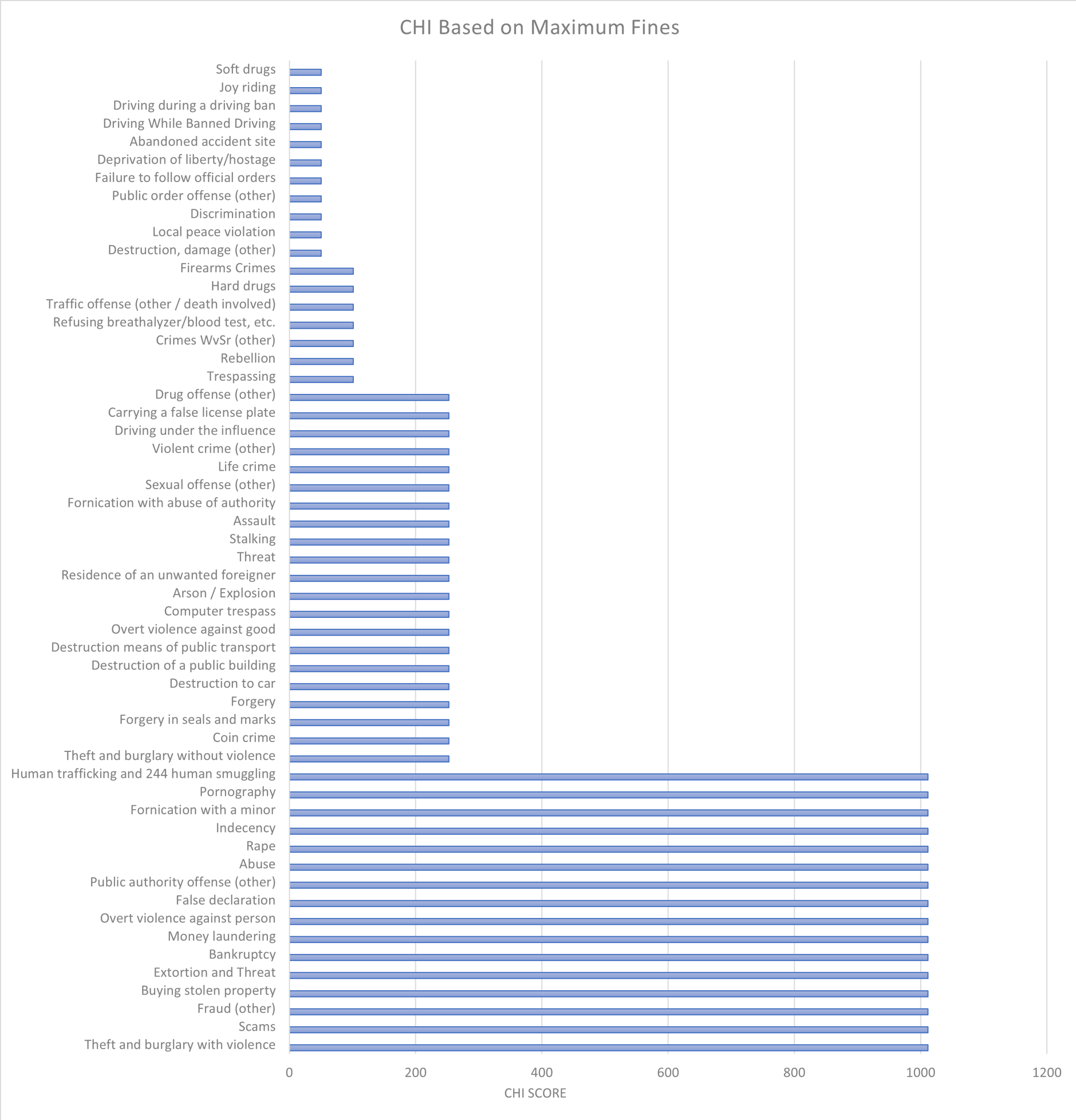


Figure 1: Ranking of all crimes based on maximum fines

Evidently, ranking the severity of crimes based on maximum fines does not lead to a diverse distribution in the severity of crime. This is because, unlike UK and US criminal law, nearly every crime according to Dutch criminal law can be addressed with a maximum fine. According to this system of measurement, a "life crime" (one involving the killing of a person) is equivalent to "destruction of a car." This mapping of maximum crime can be found in Figure 1. One option to address this flaw while maintaining the integrity of the measurement is to consider "average sentencing time". In this instance, "average sentencing time" refers to the calculated average between the maximum fine penalty and the maximum prison sentence penalty. This calculation still upholds the central purpose of the measurement: an indicator on the moral severity of every crime according to Dutch law. The created values are shown in Figure 2.

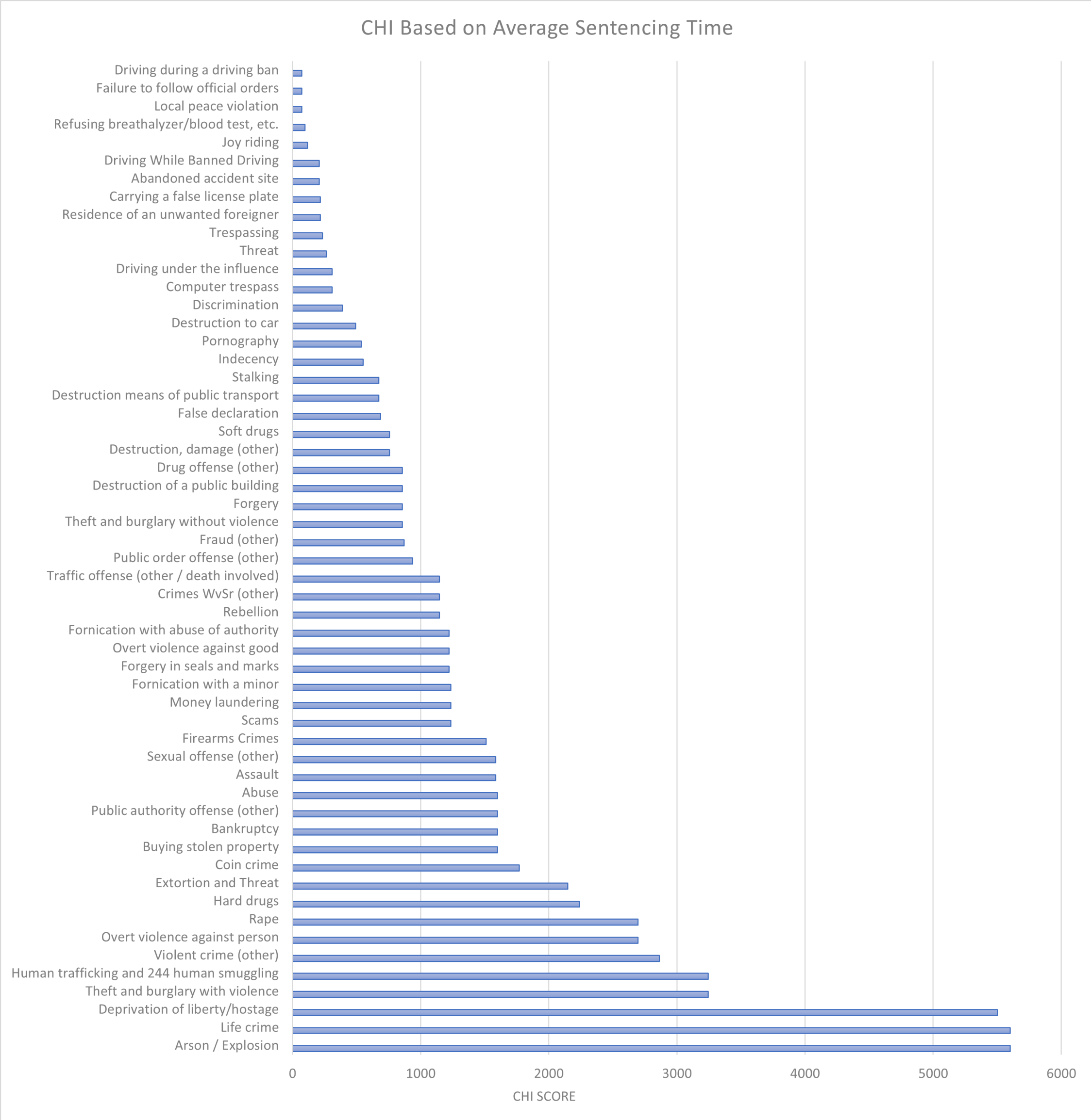


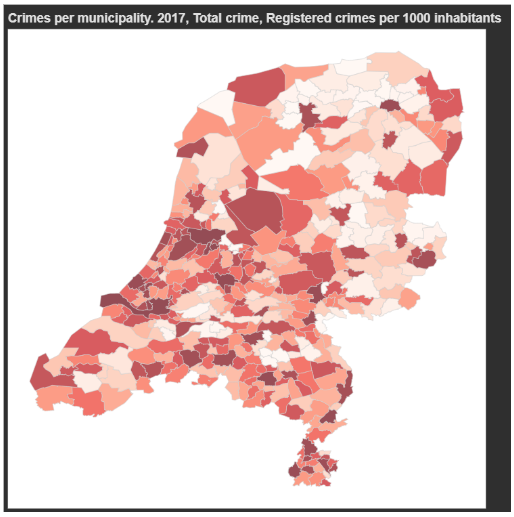
Figure 2: Ranking of all crimes based on average sentencing time

By taking an average between maximum fine and sentencing time, a more diverse reflection on the moral severity of crime can be interpreted. Now, “life crime” scores the highest for the CHI, while “destruction of a vehicle” is considered significantly less severe.

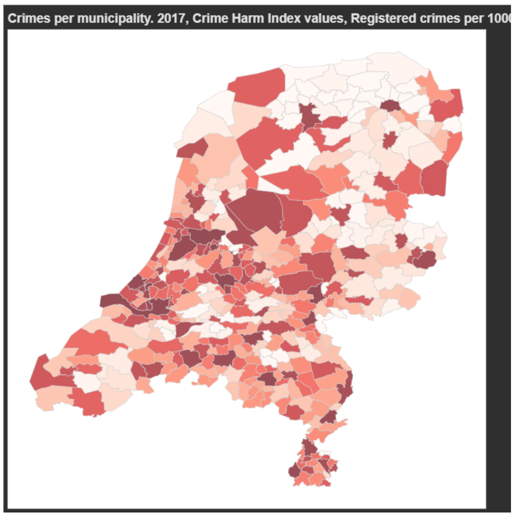
It is possible to bound this ranking system on a scale, 0-1. When considering the most severe crime, “life crime” (or murder), at a maximum CHI score of 5601, this score can be equated to 1. All other crimes are scaled around this value. For example, “destruction of a vehicle,” originally with a score of 491, now has a CHI score of 0.088. This scale becomes significant when directly comparing CHI values with other factors of safety, such as police efficiency and victim utility.

6.1.3 Results

With an accurate measurement of the moral severity of every crime in the Netherlands defined by a crime-harm index, a more accurate hot-spot map of harm can be displayed. Rather than displaying only the total crimes per municipality (per 1000 inhabitants) [8], as is shown in Figure 3a, the total number of crimes times the CHI per municipality (per 1000 inhabitants) can also be mapped, which is shown in Figure 3b. This accounts for two factors: the over-representation of petty crimes and the under-representation of more severe crimes.



(a) The amount of crimes in the Netherlands mapped per municipality per 1000 inhabitants



(b) The amount of crimes times the CHI mapped per municipality per 1000 inhabitants

As expected, there are only small changes between the hot-spots of crime-harm in the two maps. However, even small changes are still significant. The CHI map properly addresses a more accurate reflection of the amount of harm being caused by certain types of crimes in each region, based on the severity of sentencing time in Dutch criminal law.

6.1.4 CHI's Reflection of Safety

While this has been the first known application of a CHI to the Netherlands, official organisations in the Netherlands are already conducting scoping literature reviews on its applicability to Dutch society [28]. Through one-to-one interviews with researchers at the Netherlands Institute for Crime and Law Enforcement, a greater understanding of the benefits and draw-backs of a Dutch CHI were explored. In general, CHIs are an accurate measure for determining the allocation of police resources according to hot-spot policing. However, measuring crimes purely based on prison sentences has a number of downsides. To begin with, there is no method to reliably measure the rehabilitation rate of perpetrators, leading to no understanding on the status of repeat offenders. Furthermore, non-criminal offences are not considered at all. These types of offences, which often require community policing, are entirely overlooked despite having a vast aggregate effect on harm at a societal level. Finally, the sentencing time for crimes are often based on external political influences, and they may not be a true indicator on the individual harm addressed by each crime. The vast difference in criminal drug laws between countries serves as an example of this.

Despite the critiques, a more extensive version of a CHI for the Netherlands is now in development. This is because a CHI can serve as a “bottom line indicator” of whether public safety is increasing or decreasing in the Netherlands [28]. The Netherlands Institute for Crime and Law Enforcement has access to vast amounts of more data on criminal law. As a result, the CHI currently being developed will consider over 600 types of criminal codes, while the current CHI developed by the JIP team “only” considers 55 criminal offences.

Perhaps the most significant criticism of the CHI is that, for some categories, the severity of crimes within that category can vary drastically. As a result, assigning the same weight to all crimes in that category might not accurately represent the harm done [3]. In fact, the CHI offers no way to compare the harm done by crimes in the same category, unless smaller subcategories with separate sentencing guidelines are available. For example, a violent assault that results in a broken finger and a violent assault that results in permanent paralysis would be considered equally severe. This leaves the question if the punishment for a crime is necessarily proportional to the harm done. While the harm done is important in determining both sentencing guidelines and actual sentences, it is not the only factor. The intent of the perpetrator is likely also considered. For example, the punishment for murder is higher than for killing someone in a car accident while drunk driving even though the harm done to the victim might be very similar. Because the CHI is based on sentencing guidelines for categories of crime it cannot be used effectively to quantify harm on the level of an individual crime. Instead, this must be addressed through change in the amount of harm done to a victim. This is called utility.

6.2 Utility functions

We have decided to quantify the harm done to individuals as a result of crime by using utility functions. Before going into utility functions themselves we will first discuss our choice to consider the harm done by crime to individuals as well as the types of harm we will consider in subsubsection 6.2.1. Next we will explain how utility functions work and how they can be used to quantify harm done to individuals in subsubsection 6.2.2 through subsubsection 6.2.4. We will also discuss the QALY (quality-adjusted life year) as an example of the use of utility functions in medicine in subsubsection 6.2.5. Our approach is similar to the QALY, so a good understanding of the QALY should help to understand the utility functions we introduce. In subsubsection 6.2.6 we propose a general form for utility functions used to quantify harm. In subsubsection 6.2.7 and subsubsection 6.2.8 we narrow down the general form to an example utility function that can be used to calculate changes in utility as a result of crime. In subsubsection 6.2.9 we explain how we calculate example utility values. In subsubsection 6.2.10 we discuss the possibility of normalizing utility function, so harm to individuals can be measured on a 0-1 scale. In subsubsection 6.2.11 we consider some of the downsides of measuring harm through the utility functions we propose.

6.2.1 Bearers and types of harm

To quantify harm more directly, instead of using sentencing guidelines as a proxy, it is necessary to more precisely consider which types and bearers of harms to consider. In the paper by Greenfield (2013) [11] a framework is proposed that allows for the identification and limited qualitative comparison of harms. While different from the quantitative comparison we pursue, the bearers and types of harm identified in Greenfields paper [11] are still of use to us.

In the paper by Greenfield [11], four types of harm and four categories of bearers are identified. The bearers identified are individuals, private-sector entities, government entities and the environment. The types of harm identified are harms to functional integrity, material interest, reputation and privacy. Here the exact meaning of functional integrity depends on the category of bearer. For individuals this means physical and psychological integrity, for private-sector and government entities this means operational integrity and for the environment this means physical, operational and aesthetic integrity.

Greenfield [11] gives no way of comparing harms across different types or different categories of bearers. Because our intent is to quantitatively compare harms and to keep the scope of our JIP case manageable, we have decided to consider only some of these types and bearers of harm.

We have decided to only consider individuals as bearers of harm. The main reason why we have chosen to only consider individuals is that at least part of the harm done to bearers in the other categories can be captured by looking at indirect harm done to individuals. For example, if the government suffers financial harm through fraud, this could be viewed as indirect financial harm to all taxpayers. If the environment is harmed through the (illegal) dumping of chemical waste, this could result in physical harm to the people living nearby.

For types of harm we have decided to limit our attention to harms to functional integrity and material interest, or in other words to physical, mental and financial harms. We decided to limit our attention to these types because they seemed easier to quantify. Financial harm can be expressed in monetary values and physical and mental harm can be expressed through the QALY used in medicine. We will consider the QALY in more detail in subsubsection 6.2.5. We also think that the more severe cases of harms to reputation and privacy can affect the victim mentally and hence be captured in that way.

Having chosen to limit our attention to harms done to individuals, it is important to address which individuals should be considered when assessing the harm done by a crime. In some cases this is relatively straightforward. If someone's bike is stolen, there is a direct victim whose harm should be considered. In some cases, for example a murder, a crime might not just harm the direct victim but also the people that know the victim. Their harm might also have to be considered. For some crimes, for example fraud, there might not be a direct victim and identifying indirect victims and assessing their harm can be difficult. In some situations, for example violence between different criminal organizations, the distinction between victims and perpetrators might be unclear. A framework such as in Greenfield's paper [11] could be useful in identifying which harms to consider in more complex cases.

Aside from harms, there can also be benefits associated to crime. For one, in many cases the perpetrator likely benefits from the crime they commit or they would not commit the crime in the first place. But there can be others than the perpetrator who benefit from the crime [11]. If crime benefits are included in the analysis, then some crimes might not seem to cause any harm. Theft that only causes financial harm to the victim might benefit the perpetrator enough to offset the harm caused to the victim resulting in a crime with no net harm even though the victim was clearly harmed [12]. For this reason we have decided to not include illegitimate benefits as a result of crime in the quantification of harm.

6.2.2 Ordinal and cardinal utility functions

A utility function is a concept in economics that is used to represent the preferences of people. It assigns a real number to alternative options available to a person. A utility function u is said to represent a person's preferences when for any two options A and B, the person prefers the option for which u is higher. If $u(A) = u(B)$, then the person is indifferent between options A and B. If the numerical value $u(A) - u(B)$ has relevance beyond expressing which of the two options is preferred, the utility function is called cardinal; otherwise it is called ordinal [34]. For an ordinal utility function u, its values only serve to rank options by preference and nothing more.

Our approach is to quantify the harm done to a person as a result of a crime, by comparing the person's utility in case the crime does not occur with their utility if the crime does occur. Suppose A is the option that the crime in question does not occur and B is the option that the crime occurs. If $u(A) > u(B)$, the person would prefer the crime not to occur. This is the likely case if the person in question is a victim or knows a victim. If $u(A) < u(B)$, the person prefers the crime to occur, which likely means that the person in question is a perpetrator or some other beneficiary of the crime in question. If $u(A) = u(B)$, then the person is indifferent about the crime. This will likely apply to the people that are not involved in this crime. Since we intend to not include illegitimate benefits, we are mostly interested in the scenario where $u(A) > u(B)$.

If the utility function u is only ordinal, then based on u it is only possible to distinguish between people that benefit from a crime, are harmed by a crime or neither. In case a person is harmed, we would also like u to provide us with information about how much they have been harmed. Similarly, if a person benefits, it seems useful to know how much they benefit as this might explain why people commit crimes in the first place. We want $u(A) - u(B)$, the utility lost as a result of the crime, to represent the extent of the harm caused by the crime, with a larger value coinciding with more harm done. Similarly, we want the utility gained as a result of a crime to represent the benefit gained. Therefore, we need our utility functions to be cardinal to use it in the way we intend.

Utility functions in general are not unique. If the utility function u, which can be cardinal or ordinal, represents a person's preferences, certain transformations of u will also be utility functions of the same type for this person. In other words, it is possible to transform u into equivalent utility functions without loss of information. For ordinal utility functions the allowed transformations are to transform u into the composition $f \circ u$ for any strictly increasing $f : \mathbb{R} \to \mathbb{R}$ ($f(x) > f(y)$ if $x > y$), where $(f \circ u)(A) = f(u(A))$. These transformations of u represent the same ranking of preferences, which is all information contained in u. Cardinal utility functions contain more information, so there are more restrictions on the allowed transformations. For cardinal utility functions the allowed transformations are the positive affine transformations, transforming u into $au + b$ for constants $a > 0$ and $b \in \mathbb{R}$ [34]. Note that these are special cases of the allowed transformations for ordinal utility functions with $f(x) = ax + b$.

There are a number of questions about (cardinal) utility functions that we have not yet addressed. First of all, given a person's preferences does a utility function that represents that person's preferences always exist? The answer depends on the assumptions that are made about the preferences a person can have and on the type of utility function. For example if a person has seemingly contradictory preferences, where they prefer option A to option B, B to C and C to A, no utility function will be able to represent these preferences. A second question is how to construct or approximate a person's utility function assuming one exists. For an ordinal utility function only a ranking of the preferences is needed, so the person can be asked in a questionnaire to compare pairs of options. Their answers can be used to construct an (approximate) ordinal utility function u. Can u automatically be considered a cardinal utility function? u could be declared cardinal, but so could u^3 which is equivalent to u as ordinal utility functions. u and u^3 are not equivalent as cardinal utility functions, so which one, if any, is an appropriate cardinal utility function? A cardinal utility function should represent a person's preferences in such a way that differences in utility have numerical significance, but so far we have not specified what this numerical significance is exactly. So a third question that needs to be answered is what the numerical significance of utility differences even is in the first place.

6.2.3 Von-Neumann-Morgenstern utility functions

These questions can be addressed, at least to some degree, by considering preferences under uncertainty. Suppose the preferences of a person with regards to some collection of options are known and an ordinal utility function u represents these preferences. Consider a situation in which one of the options $A_1, \ldots, A_n$ from the collection is chosen at random and p_i is the probability of choosing the i-th option. Such a situation is called a lottery. If L is this particular lottery and M is a different lottery, the person might prefer one of the two lotteries or be indifferent between them. The expected value of the utility function u when lottery L is conducted is given by

$$E_L(u) = \sum_{i=1}^{n} p_i u(A_i). \tag{2}$$

Since u ranks the preferences of the person between the individual options it would be convenient if the expected value of u would rank their preferences between lotteries. In general it does not, but a utility function that represents a person's preferences in such a way that they will maximize the expected utility when choosing between lotteries is called a Von-Neumann-Morgenstern utility function [34]. The nice thing about a Von-Neumann-Morgenstern utility function is that it makes it easy to determine a person's preferences between lotteries.

We want to construct utility functions that can quantify the harm caused by crime. In most situations, people do not get to choose if they become the victim of a crime or not. However, people can make decisions that influence their probability of becoming victims of a crime. For this reason the fact that Von-Neumann-Morgenstern utility functions represent a person's preferences under uncertainty seems useful in the context of crime harm.

If u is a Von-Neumann-Morgenstern utility function for some person, are equivalent ordinal utility functions for this person also Von-Neumann-Morgenstern utility functions? In other words if $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing and L a lottery, does $L \mapsto E_L(f \circ u)$ provide the same ranking of preferences between lotteries as $L \mapsto E_L(u)$? If $E_L(f \circ u) = f(E_L(u))$ it does, because f is strictly increasing. To ensure that $E_L(f \circ u) = f(E_L(u))$, f must be an affine transformation ($f(x) = ax + b$). As a result the possible transformations that transform one Von-Neumann-Morgenstern utility function into another are exactly the positive affine transformations [34]. This shows that their behaviour under uncertainty is what gives Von-Neumann-Morgenstern utility functions their numerical significance and makes them cardinal.

The construction or approximation of a Von-Neumann-morgenstern utility function also seems relatively straightforward in principle. Instead of only asking a person their preferences between certain outcomes, they should be queried about their preferences between lotteries. This can be done through the use of questionnaires, assuming that a Von-Neumann-morgenstern utility function exists for this person. The existence of Von-Neumann-Morgenstern utility functions is guaranteed under certain assumptions about a person's preferences. We will state and briefly discuss these assumptions taken from the paper by Strotz (1953) [34]. A proof that a Von-Neumann-Morgenstern utility function exists under the assumptions stated below can be found in the paper by Marschak (1950) [18].

The first assumption is that a person's preferences regarding prospects are completely ordered and transitive. Here a prospect is any certain outcome or lottery. Completely ordered means that for any pair of prospects, the person will prefer one over the other or be indifferent. Transitive means that if prospect A is preferred over B and B over C, then A should also be preferred over C. As long as a person has sufficient time to think, it seems plausible that they can decide which prospect they prefer for every pair of prospects. Transitivity also seems like an intuitively plausible assumption.

The second assumption is that given prospects A, B and C such that A is preferred over B and B over C, there is a probability $p \in (0, 1)$ such that the person is indifferent between the certainty of B and the lottery where A is chosen with probability p and C with probability $1 - p$. In other words, any prospect B that is worse than A and better than C is as good as some particular lottery between A and C. For sufficiently small p the person will likely prefer the certainty of B and for sufficiently large p they will prefer the lottery. There will likely be a tipping point of sorts where their preference between the certainty of B and the lottery flips. It seems plausible that at this tipping point, they will be indifferent.

The third assumption is that for any prospect A and probability $p \in (0, 1)$, a prospect B exists such that the person will not be indifferent between the certainty of A and a lottery where A is chosen with probability p and B with probability $1 - p$. Unless the person is indifferent between all available prospects, choosing any prospect B for which the person is not indifferent to A and B should intuitively suffice. If A is preferred over B, the certainty will intuitively be preferred over the lottery and if B is preferred over A the lottery will be preferred. So this assumptions also seems plausible.

For the fourth and final assumption suppose that A, B and C are prospects such that the person is indifferent between A and B, and $p \in (0, 1)$. In this case the person should also be indifferent between the following two lotteries. In the first lottery A is chosen with probability p and C with probability $1-p$. In the second lottery B is chosen with probability p and C with probability $1-p$. We can interpret this as the person throwing a (possibly unfair) coin that comes up heads with probability p and tails with probability $1-p$. If the coin comes up heads then prospect A is chosen in the first lottery and B in the second lottery. If the coin comes up tails then C is chosen in both lotteries. Because the person is indifferent between A and B, they will be indifferent between the lotteries if the coin comes up heads. If the coin comes up tails, the same prospect is chosen in both lotteries, so the person will also be indifferent. Framed in this way, it makes sense for the person to also be indifferent to the lotteries without knowing the outcome. So this assumption also seems plausible.

The above assumptions seem intuitively plausible for a rationally behaving person, which implies that the preferences of a rationally behaving person can be represented by a Von-Neumann-Morgenstern utility function. However, the behaviour of people in practice does not have to be in line with a Von-Neumann-Morgenstern utility function. Strotz [34] gives an example where the hypothetical preferences of a person between a number of lotteries involving money are considered. Assuming the existence of a Von-Neumann-Morgenstern utility function, such a utility function is constructed. However, according to this utility function, this person prefers receiving 20 dollars over receiving 25 dollars, which seems strange assuming the person should prefer more money. Upon closer inspection the initial preferences given seem strange as well, but this is not apparent at first glance. Does this mean that the above assumptions are not in line with rational behavior? Strotz [34] suggests that the assumptions about preference can be considered rational in the following sense. A sane person might hold preferences that are not consistent with the given assumptions. However, if that person is given a clear explanation of how their preferences violate the assumptions and they understand this violation, this should cause them to adjust their preferences. We agree that while in a lot of cases people might state preferences that contradict the four assumptions, possibly because they might not have an intuitive understanding of probabilities, most people could be persuaded to adjust their preferences once they realize what does not make sense about them. For this reason we believe that Von-Neumann-Morgenstern utility functions can be used to represent people's (revised) preferences.

We briefly consider a property of Von-Neumann-Morgenstern utility functions, which we call the extension property, that will become relevant later on. Suppose that u is a utility function for some person over some collection $\mathcal{C}$ of certain outcomes. Let $\mathcal{C}'$ be the set of lotteries over outcomes in $\mathcal{C}$. Let $E.(u) : \mathcal{C}' \to \mathbb{R}$ be the function that maps a lottery to its expected utility. By identifying a lottery with a guaranteed outcome in $\mathcal{C}$ with that outcome, $E.(u)$ can be viewed as an extension of u. If u is Von-Neumann-Morgenstern utility function, then $E.(u)$ represents the person's preferences between lotteries and vice versa. It is possible to also consider preferences over lotteries with outcomes in $\mathcal{C}'$. These are lotteries over lotteries and as such have final outcomes in $\mathcal{C}$. If a person is assumed to base their preferences over lotteries with outcomes in $\mathcal{C}'$ only on the probabilities of the final outcomes in $\mathcal{C}$, then $E.(u)$ will be a Von-Neumann-Morgenstern utility function. A proof of this result can be found in Appendix B. This result allows us to extend a Von-Neumann-Morgenstern utility function u on $\mathcal{C}$ to a Von-Neumann-Morgenstern utility function on the set of lotteries $\mathcal{C}'$. We will identify u with this extension.

6.2.4 Aggregating utility functions

Given (cardinal) utility functions for individual people in society, it is important to consider how to combine these into a utility function that represents the preferences of society as whole. We will refer to it as the total utility function and denote it with a capital letter (usually U). The utility functions of individual people will be referred to as individual utility functions and denoted with a lowercase letter (usually u). For a society of n people, where u_i is a utility function for the i-th person, one way to define a total utility function U for the whole of society is to define

$$U(A) = \sum_{i=1}^{n} u_i(A). \tag{3}$$

As discussed before, the individual utility functions u_i are not unique: certain transformations of them contain the same information. If v_i are allowed transformations of the u_i, then

$$V(A) = \sum_{i=1}^{n} v_i(A) \tag{4}$$

should still represent the same preferences as U. Moreover, if we want U to be cardinal, V should be a positive affine transformation of U. If we assume the u_i to be cardinal, then the allowed transformations are of the form $v_i = a_i u_i + b_i$. In this case

$$V = \sum_{i=1}^{n} a_i u_i + \sum_{i=1}^{n} b_i. \tag{5}$$

This is not, in general a positive affine transformation of U. If we assume that $a_i = a$ for all $i \in \{1, \ldots, n\}$ then

$$V = a\sum_{i=1}^{n} u_i + \sum_{i=1}^{n} b_i = aU + \sum_{i=1}^{n} b_i \tag{6}$$

is a positive affine transformation of U. So to define a cardinal total utility function as the sum of the individual utility functions, it is necessary to further restrict the allowed transformations for the individual utility functions as done above.

Defining the total utility function as the sum of the individual utility functions has the following practical benefit: to calculate the difference in total utility, it is sufficient to only consider the change in utility of those members of society whose utility has changed. For a more general total utility function the change in total utility might depend on the utility of members of society whose utility has not changed. This would for example be the case if the total utility function would be defined as the minimum of the individual utilities. The practical simplicity of defining the total utility as the sum of the individual utilities was the main reason to choose this option, but we also did some research to support this decision.

The paper by Blackorby (1984) [4] considers the more general problem of how to translate individual utility functions $u_1, \ldots, u_n$ into a preference ordering for the whole society (consisting of n people). More precisely they consider the general form of a function F that maps the tuple of individual utility functions $\boldsymbol{u} = (u_1, \ldots, u_n)$ to a preference ordering for the whole society under certain assumptions. We will briefly mention the general assumptions made about F in the paper by Blackorby [4].

The first assumption is that there are no restrictions on the utility functions, so any $\boldsymbol{u}$ from the possible options to $\mathbb{R}^n$ is allowed.

The second assumption is that of Pareto indifference: For any possible tuple of utility functions $\boldsymbol{u}$ and pair of options A and B, if $\boldsymbol{u}(A) = \boldsymbol{u}(B)$ then society is indifferent between A and B under $F(\boldsymbol{u})$. In other words, if all members of society are indifferent between two options, then so is society as a whole. This certainly seems plausible.

The third assumption is that of binary independence of irrelevant alternatives: If $\boldsymbol{u}$ and $\boldsymbol{u}'$ are two tuples of utility functions and A and B are a pair of options such that $\boldsymbol{u}(A) = \boldsymbol{u}'(A)$ and $\boldsymbol{u}(B) = \boldsymbol{u}'(B)$, then society should have the same preference regarding options A and B under $F(\boldsymbol{u})$ and $F(\boldsymbol{u}')$. This essentially means that the preference of society regarding options A and B only depend on the values the individual utility functions take in A and B. In other words, the utility assigned to options different from A and B has no impact on society's preference between A and B. This also seems like a reasonable assumption. Blackorby [4] calls these assumptions taken together welfarism.

Two more assumptions that are assumed in some parts of the paper by Blackorby [4] are the weak Pareto assumption and anonymity. The weak Pareto assumption says that if $u_i(A) > u_i(B)$ for all $1 \leq i \leq n$, then A is preferred over B under $F(\boldsymbol{u})$, where $\boldsymbol{u} = (u_1, \ldots, u_n)$. In other words if everyone prefers option A over B, then so should society, which seems reasonable. Anonimity is the condition that if $\boldsymbol{u}(A)$ is a permutation of $\boldsymbol{u}(B)$, then society is indifferent between A and B under $F(\boldsymbol{u})$. In other words society is indifferent if the utilities of people are swapped around, which represents an equal treatment of sorts.

Blackorby [4] also considers the fact that certain transformations can be performed on $\boldsymbol{u}$ without loss of information. To ensure that F does not use more information than is actually contained in $\boldsymbol{u}$, F should be invariant under the allowed transformations of $\boldsymbol{u}$. So if $\boldsymbol{v}$ is an allowed transformation of $\boldsymbol{u}$, then $F(\boldsymbol{v}) = F(\boldsymbol{u})$. They call this property information invariance. If $\boldsymbol{u}$ is assumed to contain more information about individual preferences, less transformations without loss of information are possible. As a result, there are less restrictions on F.

Blackorby [4] considers different sets of information preserving transformations for $\boldsymbol{u}$ and the consequences this has for the allowed functions F. The assumptions we made earlier about the allowed transformations for the individual utility functions when constructing the total utility function were as follows: the allowed transformations of $\boldsymbol{u} = (u_1, \ldots, u_n)$ are $\boldsymbol{v} = (au_1 + b_1, \ldots, au_n + b_n)$, where $a > 0$ and $b_1, \ldots, b_n \in \mathbb{R}$. These allowed transformations correspond with cardinal utility functions that are unit-comparable [4]. These transformations allow for the interpersonal comparison of utility differences but not of utility levels.

Blackorby [4] considers the allowed possibilities for F when assuming welfarism, weak Pareto, anonymity and information invariance with unit-comparable cardinal utility functions. Blackorby [4] shows that, under these assumptions, the preferences of society should be represented by the sum of individual utilities. If more transformations of the utility functions are allowed, more options become available [4]. This lends some justification to our definition of total utility.

We propose the quantification of harm through Von-Neumann-Morgenstern individual utility functions. By identifying these with their extension to the collection of lotteries through the extension property, we can assume that $E_L(u_i) = u_i(L)$ for each individual utility function u_i and each lottery L by identifying L with an equivalent lottery among deterministic outcomes if necessary. From this it follows that for any lottery L we have that

$$E_L(U) = E_L\left(\sum_{i=1}^{n} u_i\right) = \sum_{i=1}^{n} E_L(u_i) = \sum_{i=1}^{n} u_i(L) = U(L). \tag{7}$$

This shows that under the identifications we have made, the total utility U corresponds with its extension to the collection of lotteries. So its extension is a utility function, hence U itself is a Von-Neumann-Morgenstern utility function. As a consequence U is a cardinal utility function on the level of society.

6.2.5 The QALY

An important example of Von-Neumann-Morgenstern individual utilities that are summed to obtain a utility function for society is the quality-adjusted life year (QALY) used in medicine. The QALY is a unit of measurement for a person's utility based on their longevity and health. Each health state h is assigned a (cardinal) health utility $q(h) \in [0, 1]$ where a 1 corresponds to perfect health and a 0 to death [39]. The health utilities are determined through questionnaires, including ones where participants choose between lotteries. The utility of living t years in a constant health state h with health utility $q(h)$ is set to $Y(h, t) = q(h)t$ QALYs. Under certain assumptions a utility function that depends on health state and number of years lived is necessarily of the given form, but if some of the assumptions are relaxed a wider range of utility functions is possible [26].

The assumptions made in the paper by Pliskin (1980) [26] to show that the QALY formula $Y(h, t) = q(h)t$ is the correct form for a utility function based on health and longevity are as follows. The first assumption is that longevity t and health state h are utility independent of each other. This means that if one of these two variables is held fixed and lotteries over the second variable are considered, the preferences of a person over these lotteries do not depend on the value at which the first variable is fixed [26]. In other words, if a person can choose between lotteries regarding remaining life-years with their health state unchanged, then their choice does not depend on their health state. The same should hold if the roles of remaining life-years and health state are reversed. Note that if health states worse than death are considered, then this assumption will not hold. This is because while for most health states a person will prefer to live longer, for health states worse than death they will prefer to live shorter. As long as health states worse than death are not considered this assumption seems plausible [26].

The second assumption made is the constant proportion trade-off assumption of life years for health status. This assumption states that if a person with t life-years remaining is willing to give up a proportion of their life-years for an improvement of health state from h_1 to h_2, then they would be willing to give up the same proportion of life-years if they had t' life-years remaining for any t' [26]. So if a person with 15 life-years remaining is willing to give up 3 of those years for a health improvement, then they would be willing to give up 4 years for the same health improvement if they had 20 life-years remaining [26]. This assumption might not hold universally, but is in line with empirical results in the paper by Pliskin [26].

The third and final assumption is that of risk neutrality on life-years. For a fixed health state, this means that when choosing between lotteries involving life-years, a person will choose the lottery that maximizes the expected life-years [26]. This implies that a Von-Neumann-Morgenstern utility function for life-years alone should be a linear function of life-years [26]. According to Pliskin [26], this assumption is the least well-supported of the three. Based on empirical results in the paper by Pliskin [26] this assumption applies to about half the people considered. It is shown in the paper by Pliskin [26] that the QALY formula implies these three assumptions and under these three assumptions the QALY formula is the only possible Von-Neumann-Morgenstern utility function based on longevity and wealth up to positive affine transformations.

The QALY formula $Y(h,t) = q(h)t$ works if a constant health state h is assumed, but in general a person will not be in the same health state their entire life. Therefore we can consider a person's health state h to be a function of time t in years. To evaluate a person's lifetime QALYs it is necessary to have a more general formula.

Based on the QALY formula for constant health, $q(h(t))$ can be viewed as the amount of QALYs gained per year by that person at time t. So if the health state is allowed to change over time the QALY formula can be generalized to

$$Y(h) = \int_{t_0}^{t_e} q(h(t))dt \tag{8}$$

for the total QALYs in a person's lifetime. Here t_0 is the time of birth and t_e is the time of death for that person. Viewed in this way, the total QALYs in a person's life can be interpreted as the area under the graph of $q(h(t))$. Note that if h is assumed to have a constant value h_0, then the above simplifies to $q(h_0)(t_e - t_0)$, which is essentially the earlier QALY formula.

Usually quality of health is not measured continuously but at discrete moments. Because of this, h is assumed to change at a finite number of times and to be constant in between those times. To make this more precise, suppose $t_0 < t_1 < t_2 < \cdots < t_N = t_e$ and the person is in health state h_n between times t_{n-1} and t_n. So $h(t) = h_n$ for $t_{n-1} \leq t < t_n$. In this case the lifetime QALYs from the integral formula reduce to

$$\sum_{n=1}^{N} q(h_n)(t_n - t_{n-1}). \tag{9}$$

QALYs are assumed to be comparable across individuals and changes in QALYs are summed to obtain a total change in QALYs [39]. The QALY can be used to assess the benefit of a treatment and to compare the benefits of different treatments. Note that since a person's remaining life-years and future health state are unknown, these will usually be estimated. The QALY has served as a source of inspiration both for our company coach in coming up with this case and ourselves in coming up with a solution.

6.2.6 Utility functions based on health, longevity and wealth

As mentioned before, we have decided to measure harm done to individuals by crime. The types of harm we consider are financial, physical and mental. For this reason we need cardinal utility functions that can measure these harms. We have decided to consider utility functions with longevity t, health h and wealth w as their variables. Financial harm results in a decrease of w, mental harm results in a decrease in h and physical harm in a decrease of h and possibly t.

The general form for a cardinal utility function $u(h,t,w)$ based on health h, longevity t and wealth w is derived in the paper by Hammitt (2013) [13]. It is assumed that there is a cardinal utility function $Y(h,t)$ for health and longevity, such as the QALY, with $Y(h,0) = 0$ and $Y(h,t) \geq 0$. Moreover, it is assumed that $Y(h,t)$ is consistent with preferences over health and longevity for any fixed level of wealth. In other words it should hold that for all fixed levels of wealth w', $Y(h',t') > Y(h'',t'')$ if and only if $u(h',t',w') > u(h'',t'',w')$. These assumptions imply that

$$u(h,t,w) = Y(h,t)a(w) + b(w) \tag{10}$$

is the general form of a utility function based on health, longevity and wealth [13]. Here a and b are functions of wealth, where $a(w) > 0$ for all w. $U(h,0,w) = b(w)$, so b represents the utility a person assigns to the inheritance they leave behind after they die [13]. It is assumed that b is non-decreasing in w and a is strictly increasing in w ($b'(w) \geq 0$ and $a'(w) > 0$) [13]. Note that it is implicitly assumed in the paper by Hammitt [13] that a and b are (continuously) differentiable.

If we assume that for any fixed level of wealth, the QALY is a cardinal utility function for health and longevity, then we obtain the following utility function if we assume constant health and wealth:

$$u(h,t,w) = a(w)q(h)t + b(w). \tag{11}$$

There are a number of questions that should be answered about the above utility function. First of all, what is meant exactly by a person's wealth? Should this be interpreted as a person's income, their net worth or something else? Based on the interpretation of $b(w)$ as the utility of an inheritance in the paper by Hammitt [13], it seems appropriate to interpret w as a person's net worth. To be comparable across different years the effects of inflation should ideally be included in the utility function. We have decided to leave the effects of inflation outside the scope of the JIP case, but it should be considered in future research.

The next question is how to generalize the above utility function in case health states and wealth are allowed to vary throughout a person's lifetime. We have decided to generalize it in a way similar to the QALY. In particular, suppose that a person's health h and wealth w are functions of time t in years. Based on the utility formula for constant health and wealth, $a(w(t))q(h(t))$ can be viewed as the amount of utility gained per year by that person at time t. Furthermore $b(w(t_e))$ can be interpreted as an additional source of utility based on the inheritance left behind. Here t_e is the person's time of death, so $w(t_e)$ is the net worth of their inheritance. So if the health state and wealth is allowed to change over time the lifetime utility formula can be generalized to

$$u(h, w) = \int_{t_0}^{t_e} a(w(t))q(h(t))dt + b(w(t_e)) \tag{12}$$

Note that if h is assumed to have a constant value h_0 and w a constant value w_0, then the above simplifies to $a(w_0)q(h_0)(t_e - t_0) + b(w_0)$, which is essentially the utility function for constant health and wealth from before.

If wealth and quality of health are measured only at discrete moments, then the utility function can be approximated by considering h and w to change at a finite number of times and to be constant in between those times. To make this more precise, suppose $t_0 < t_1 < t_2 < \cdots < t_N = t_e$ and the person is in health state h_n with wealth w_n between times t_{n-1} and t_n. So $h(t) = h_n$ and $w(t) = w_n$ for $t_{n-1} \leq t < t_n$. In this case the lifetime utility from the integral formula reduce to

$$\sum_{n=1}^{N} a(w_n)q(h_n)(t_n - t_{n-1}) + b(w_N). \tag{13}$$

In summary we obtain individual lifetime utility functions of the form

$$u_i(h_i, w_i) = \int_{t_{0,i}}^{t_{e,i}} a_i(w_i(t))q(h_i(t))dt + b_i(w_i(t_e)). \tag{14}$$

Here the underscore i is to show that the function or variable belongs to the i-th person. If utility differences can be compared inter-personally, then we can define a total utility function for a society of N people by defining

$$U(h_1, \dots, h_N, w_1, \dots, w_N) = \sum_{i=1}^{N} u_i(h_i, w_i) = \sum_{i=1}^{N} \left(\int_{t_{0,i}}^{t_{e,i}} a_i(w_i(t))q(h_i(t))dt + b_i(w_i(t_{e,i})) \right). \tag{15}$$

Here U depends on the individual health and wealth functions (or trajectories) of the members of society. Two related questions to be addressed are how to choose the functions a_i and b_i, and how to make sure that utility differences can be compared inter-personally. We recall that to allow for the interpersonal comparison of differences in utility and to justify the summation of individual utilities to obtain a total utility some assumptions need to be made on the allowed transformations of the tuple $(u_1, \dots, u_N)$ of individual utility functions. In particular, the allowed transformations of $(u_1, \dots, u_N)$ are $(cu_1 + d_1, \dots, cu_N + d_N)$, where $c > 0$ and $d_1, \dots, d_N \in \mathbb{R}$ or a subset of these transformations [4].

To make sure that we do not allow too many transformations, we make the assumption that $a_i = a$ and $b_i = b$ for all i, so the functions a_i and b_i are the same for each person. This decision has some added benefits. For one, it seems easier from a practical point of view, because it is no longer necessary to determine appropriate functions a_i and b_i for each person separately. Secondly, this simplification can be considered more fair, as each person's utility is assessed equally based on their time-dependent health and wealth. A downside is that the individual utility functions might no longer represent each person's preferences accurately. Under this simplifying assumption we have that

$$U = \sum_{i=1}^{N} u(h_i, w_i) = \sum_{i=1}^{N} \left(\int_{t_{0,i}}^{t_{e,i}} a(w_i(t))q(h_i(t))dt + b(w_i(t_{e,i})) \right). \tag{16}$$

6.2.7 The choice of a and b

It remains to address the question of how to determine the functions a and b. One way to do this is to propose a general form for the functions a and b including one or more parameters. Based on empirical date, it can be assessed if the suggested general form is appropriate and what parameters are reasonable. For a person with no remaining lifespan, their remaining utility is given by $b(w)$, where w is the net worth of the inheritance they leave behind. By asking people to choose between lotteries about the inheritance they leave behind, appropriate forms for the function b can be determined. Recall that we assume our utility functions to be Von-Neumann-Morgenstern. Once b is determined, a can be determined in a similar way by asking people to choose between lotteries involving their wealth. For computational simplicity the questions asked to people can be about situations where health and wealth are constant in time.

To suggest plausible general forms for a and b we consider what properties we want a and b to have. Based on the paper by Hammitt [13], a should be positive and strictly increasing, while b should be non-decreasing. It seems reasonable to us to assume that there is a limit to how much utility can be gained by increasing a person's wealth or utility lost by increasing their debt. This means that we assume a and b to be bounded. Following Hammitt [13], we also assume that a and b are continuously differentiable. A very simple option for b is $b = 0$. This reflects that a person would be completely indifferent to the inheritance they leave behind.

In case we want b to be strictly increasing just like a, the properties of a and b are very similar. The only other difference is that a should be positive. Considering the allowed transformations of cardinal utility functions we can apply a transformation such that b is positive as well. Assuming b to be positive and strictly increasing as well, we are looking for candidate functions f that are positive, bounded, strictly increasing and continuously differentiable. Here f is a candidate function for a or b. Given the continuity and boundedness of f, we must have that $f(\mathbb{R})$ is a bounded interval. Because f is strictly increasing this interval has to be open.

It seems reasonable to assume that the utility of additional wealth decreases as w approaches $\pm\infty$. In other words at some point the utility of an additional Euro becomes smaller the richer a person becomes. Similarly at some point the value an additional Euro becomes smaller the further a person gets in debt. This can be ensured by choosing $a'(w)$ and $b'(w)$ that are decreasing for sufficiently large w and increasing for sufficiently small w. So we consider candidate functions f such that f' is decreasing for sufficiently large values of w and increasing for sufficiently small values of w. In other words f is concave for sufficiently large w and convex for sufficiently small w. A simple way of ensuring this is to choose $w_0 \in \mathbb{R}$ and construct f such that it is concave for $w > w_0$ and convex for $w < w_0$. We have found three parameterized families of candidate functions f based on arctangent functions, negative exponentials and negative power functions respectively. The functions were chosen such that $f(\mathbb{R}) = (0, 1)$ but can be re-scaled and shifted to take values in another bounded open interval instead. The three families along with their first and second derivative are given in Appendix C. Note that the second and third families of candidate functions do not have a second derivative in $w = w_0$. All these candidates were obtained by stitching together two functions at w_0 in such a way that the resulting function f is continuously differentiable. In the above examples, the functions stitched together were of the same type (i.e. both exponential functions), but it should also be possible to stitch together functions of different types.

6.2.8 Proof of concept choice for a and b

To be able to calculate utility values for example scenarios we have to make a plausible choice regarding which functions a and b to use. As mentioned before, ideally these would be based on answers to questionnaires where people state their preferences about lotteries involving wealth. Unfortunately we were not able to find appropriate data about these types of questionnaires.

Due to a lack of time and data needed to properly determine a and b, we have decided to make some rough but not unreasonable guesses to come up with functions that we can use to calculate example utility values. The sole purpose of the functions we come up with is to show how the calculation of utility values could be done. Neither the functions themselves nor the utility values calculated should be viewed as an accurate representation of reality.

We first decided to set $b = 0$, so we only have to determine a. Setting $b = 0$, corresponds to people being indifferent about the bequest they leave behind. We do not expect this to be completely accurate, but we do think that the remaining lifetime utility for relatively young people is mostly determined by a. We also think that the utility of a bequest can still be captured to some degree by the increase in utility of the people receiving an inheritance.

For a we decided to limit our attention to the first family of candidate functions f_1 in Appendix C. This family has slightly more appealing mathematical properties, as it is twice continuously differentiable everywhere. As mentioned before, we have that f_1' is increasing for $w < w_0$ and decreasing for $w > w_0$. So w_0 is the point at which the marginal utility of wealth is at its highest. The parameters c_1 and c_2 are such that the function takes the value $\frac{c_1}{c_1+c_2}$ at $w = w_0$. So if $c_2 = cc_1$, then the function take the value $\frac{1}{1+c}$ at $w = w_0$. So the ratio $c = \frac{c_2}{c_1}$ determines the value of the function at $w = w_0$. We have decided to set $w = 0$ and $c = 4$. This corresponds to the marginal utility of wealth being maximal at $w = 0$ and $a(0) = 0.2$. The assertion that the marginal utility of wealth is highest at a net worth of €0 seems intuitively plausible because this is the point at which gaining or losing wealth is the difference between having a positive or negative net worth. In practice the difference between a positive and negative net worth might not be as relevant and as a result it is entirely possible that a positive or negative value of w_0 is more in line with reality. $a(0) = 0.2$ as a consequence of choosing $c = 4$, implies that the utility difference between having unlimited wealth and zero wealth, $\lim_{w\to\infty} a(w) - a(0) = 1 - 0.2 = 0.8$, is 4 times as great as the utility difference between having zero wealth and unlimited debt, $a(0) - \lim_{w\to\infty} a(w) = 0.2 - 0 = 0.2$. In particular this means that at zero net worth, there is more utility to be gained than lost based on wealth. This last implication seems plausible, but the value of c that is in line with reality might very well differ from 0.2. In short, our choice of parameters so far, while not unreasonable, has no serious theoretical or empirical foundation and should not be treated as such.

Since we have chosen $c_2 = 0.2c_1$, it remains to choose the parameter c_1. Note that with $w_0 = 0$ fixed and $c_2 = 0.2c_1$, the parameter c_1 serves as a scaling parameter in the sense that for all w, $f_1(w)$ for $c_1 = d$ is equal to $f_1(dw)$ for $c_1 = 1$. In other words, decreasing c_1 stretches the function horizontally and increasing c_1 squeezes it horizontally. To make sure that our choice of a does not completely miss the mark when it comes to quantifying the utility of wealth we decided to choose the value of c_1 based on average willingness to pay estimates for a QALY and CBS data regarding wealth in the Netherlands.

Assuming $b = 0$, the remaining lifetime utility at time t_0 of a person whose health and wealth are given by functions of time $h(t)$ and $w(t)$ is given by

$$u(h, w) = \int_{t_0}^{t_e} a(w(t))q(h(t))dt, \tag{17}$$

where t_e is their time of death. We want our choice of a to be consistent with the average willingness to pay for a QALY given a person with average net worth. In other words the lifetime utility of the person should be the same if they lose a QALY or lose an amount of wealth equal to the willingness to pay for a QALY. To use the general integral formula for the utility, knowing that the person lost a QALY is not enough. It would be necessary to know exactly how $q(h(t))$ is affected. Similarly, we would need to know how the payment of the willingness to pay amount would impact their future wealth. To make things easier we have made several simplifying assumptions. First of all we assume that the person's wealth has a constant value w. Under these assumptions we have that

$$u(h, w) = \int_{t_0}^{t_e} a(w)q(h(t))dt = a(w)\int_{t_0}^{t_e} q(h(t))dt = a(w)Y(h), \tag{18}$$

where $Y(h)$ are this person's remaining lifetime QALYs. This form makes it much easier to determine the change in utility based on changes in lifetime QALYs and wealth. Let w be the person's original level of wealth, before they decide to pay anything. Let h be the person's original expected health as a function of time. Suppose the person can choose to either have their health function h change to h' such that $Y(h') = Y(h) - 1$ or pay an amount Δw such that their wealth changes from w to $w - \Delta w$. If they choose not to pay, their utility becomes $a(w)(Y(h)-1)$ and if they choose to pay it becomes $a(w-\Delta w)Y(h)$. We interpret the willingness to pay for a QALY as the amount Δw for which the person is indifferent between losing a QALY or paying the amount. So we have that

$$a(w)(Y(h) - 1) = a(w - \Delta w)Y(h). \tag{19}$$

We have decided to choose the parameter c_1 such that this equation holds if w is the average wealth in the Netherlands and Δw is the average willingness to pay for a QALY in the Netherlands. Obradovic (2012) [23] has determined an average willingness to pay estimate for a QALY equal to Δw = €65797 in 2012 based on Dutch EQ-5D tariffs. We have decided to use the average wealth in the Netherlands from the year 2012 as well, so it will be from the same year as the willingness to pay estimate.

According to CBS data the average net worth of a Dutch household in 2012 was equal to $w =$ €158500 [6]. To determine the average remaining lifetime QALYs, we followed the same approach as in the paper by Obradovic [23]. Obradovic [23] uses average quality of health $q(h)$ for different age groups based on literature. Using the average age among participants in their study and average age of death in the Netherlands, Obradovic [23] calculates the lifetime QALYs of a person whose current age is the average among participants in their study, age of death is the age of death in the Netherlands and health state is always equal to the average health state of the relevant age group. The result of this calculation is not given in the paper by Obradovic [23], but is used as part of the willingness to pay calculation. We have performed this calculation ourselves. The average age of participants was equal to 39.39 years and age of death used was 80.55 years. The age groups and corresponding quality of health used can be found in Appendix D. Based on this we obtained an average amount of remaining lifetime QALYs equal to $Y(h) = 33.17764$. Based on these numbers we varied the parameter c_1 until $a(w)(Y(h)-1) = a(w-\Delta w)Y(h)$ would hold approximately.

For $a = f_1$ we settled on the parameter value $c_1 = 2 \cdot 10^{-7}$ for which we have $a(w)(Y(h) - 1) = 6.95485171036$ and $a(w - \Delta w)Y(h) = 6.94877695349$. A Graph of the function f_1 can be found in Appendix C. Graphs of f_2 and f_3 with parameters chosen in a similar way can also be found in Appendix C. We want to stress that while we found a value of c_1 for which $a(w)(Y(h) - 1)$ and $a(w - \Delta w)Y(h)$ are very close, and if desired we could further tweak the value of c_1 to make them even closer, this does not imply that our choice of a is an accurate representation of the utility of wealth. Because of all the assumptions and guesses made along the way, this choice of a can be used to calculate example utility values but should not be expected to be applicable in practice. In future research the choice of appropriate utility functions can be further examined by also considering $a = f_2$, $a = f_3$ with different parameter values and possibly other forms.

6.2.9 Calculating example utility values

To illustrate the use of utility functions in the context of crime harm we will consider several scenarios in which a crime occurs and we try to assess the harm done by this crime in terms of utility loss. The general procedure we follow when considering scenarios is as follows. We consider one or more people that are affected by the crime. We assume that in case the crime does not occur, the health and wealth of the people involved may change a finite number of times and are constant in between. So we assume that $h(t)$ and $w(t)$ for each person are piecewise constant. If we consider the utility of a person from time t_0 onward, t_e is their time of death and $t_0 < t_1 < \dots t_N = t_e$ are such that h and w are equal to h_n and w_n in between t_{n-1} and t_n for all $n \in \{1, \dots, N\}$, then their remaining lifetime utility is given by

$$u(\text{before}) = \sum_{n=1}^{N} a(w_n)q(h_n)(t_n - t_{n-1}) = \sum_{n=1}^{N} a(w_n)Y_n. \tag{20}$$

Here $Y_n = q(h_n)(t_n - t_{n-1})$ is the amount of QALYs of this person associated with the period between t_{n-1} and t_n. To quantify the harm done by a crime we consider how this person's utility changes as a result of the crime. To keep thing simple we usually assume that if this person suffers financial loss, then their wealth $w(t)$ decreases by an amount Δw that is constant in time. Physical and mental harm are taken into account by considering how many QALYs ΔY_n the person loses in each time period from t_{n-1} to t_n. The person's remaining lifetime utility if the crime occurs is then given by

$$u(\text{after}) = \sum_{n=1}^{N} a(w_n - \Delta w)(Y_n - \Delta Y_n). \tag{21}$$

The harm done to this person by the crime is then given by $\Delta u = u(\text{before}) - u(\text{after})$. In case the (expected) time of death changes for the person in question this can be accounted for by changing the value of t_e. If there are multiple people affected by the crime we can consider the change in total utility U, which is defined as the sum of individual utility. So the total change in utility for a society of N people is given by

$$\Delta u = U(\text{before}) - U(\text{after}) = \sum_{n=1}^{N} u_n(\text{before}) - \sum_{n=1}^{N} u_n(\text{after}) = \sum_{n=1}^{N} \Delta u_n. \tag{22}$$

Here u_n refers to the utility of the n-th person and Δu_n to the change in this utility.

In case we want to consider an average person, we will base our $h(t)$ values on the paper by Obradovic [23] to be consistent with the willingness to pay values used to determine the function a. Our $w(t)$ values will be based on CBS data on average and median wealth in different age groups. We use CBS data from 2012 because we have chosen the function a based on data from 2012. The data we use regarding average health and quality of health by age can be found in Appendix D.

6.2.10 Boundedness and normalization

Ideally we want for individual utility values to be easily understandable and interpretable at a glance. Currently, telling someone that a person's utility decreased by 0.1 will not be of much use to the person you are telling unless they have a point of reference or know what values the utility function can take. In particular, it would be useful for the person you are telling to know if the utility function is bounded and if so, what its least upper bound, or supremum, is equal to. Moreover, if the supremum is known, then the utility function could be divided by this supremum to obtain a normalized utility function. This normalized utility function would have supremum equal to 1 and measure utility on a 0-1 scale.

The general formula for a person's utility from time t_0 to their time of death t_e with $b = 0$ is given by

$$u(h, w) = \int_{t_0}^{t_e} a(w(t))q(h(t))dt. \tag{23}$$

Here q takes values in $[0, 1]$ and a takes values in $(0, 1)$. This means that the integrand of the above integral defining the utility takes values in $[0, 1)$. If we consider a person's lifetime utility starting from their birth, so t_0 is their time of birth, then $u(h, w)$ takes values in $[0, t_e - t_0)$. So a person's lifetime utility is bounded from above by their lifespan. If we assume that there is a universal upper bound T on the possible lifespan of a person, then T is also an upper bound on the lifetime utility for any person. So $0 \leq u(h, w) < T$

where h and w are the health and wealth as a function of time for any person. This makes it possible to define a normalized utility function $\frac{u}{T}$. This normalized utility function is bounded from above by 1. The downside is that unless T is the supremum of all human lifespans, the supremum of the normalized utility $\frac{u}{T}$ will be smaller than one. So the normalized utility function might not actually measure utility on a 0-1 scale, but on a 0-x scale, where $x \in (0, 1]$ is not known exactly. On the other hand, if T is believed to be an upper bound on possible human lifespan and this believe turns out to be false, then our normalized utility function could take values exceeding 1.

No supremum for possible human lifespan is known, so choosing T to be the supremum of possible human lifespans is not possible in practice. There is no consensus on a theoretical maximum possible lifespan for humans [10]. This makes it difficult to even approximate the supremum of possible human lifespans. Because of this we have decided to instead choose T based on the current maximum observed human lifespan. Currently the oldest known person in history reached the age of 122 years in 1997 and it is believed that it might take a long time until this record is broken [10]. So we have decided to choose $T = 122$, such that the utility function will have a supremum of 1 when considering all people that have lived in the past. If longer lifespans are observed in the future and individual utilities on a 0-1 scale are desired, it might be necessary to update the value of T to the then longest observed lifespan.

6.2.11 Shortcomings of measuring harm through utility functions

The main appeal of using utility functions to measure crime harm, is that they can in principle accurately quantify the harm done to a victim by a specific instance of a crime, regardless of the perpetrator's intentions. A related downside to using utility is that it does not capture the morality behind a crime. Even if the same amount of harm was done from a utility perspective, it might be desirable to distinguish the severity of crimes based on morality. For example most people would consider a murder to be more severe than a death as a result of drunk driving, even though the harm done in terms of utility might be comparable.

Another downside of the utility approach we have taken is the need to conduct questionnaires to obtain appropriate utility functions, which results in additional costs to implement this method. Moreover, even if the appropriate utility functions are known, to accurately calculate changes in utility as a result of crime, certain information about the victim is needed. In particular the victim's (estimated) current and future financial status as well as their physical and mental health state should be known in scenarios where the crime would or would not occur. Again this requires more information to be gathered in order to assess harm through utility.

The approach we have taken might also have some ethical issues as well. As our utility functions make use of the QALY, they share some of its ethical concerns. For example, in terms of QALYs gained, it is more beneficial to save the life of a younger or more healthy person compared to an older or less healthy person. The same thing is true for our utility functions based on health, longevity and wealth. The inclusion of wealth in the utility function might add some ethical concerns on top of this. Because for our functions the marginal utility of wealth decreases for sufficiently large wealth, the loss of a fixed amount of money is considered more harmful when it occurs to a poor person compared to a rich person. In this regard our utility functions imply that preventing financial losses of similar magnitude should be prioritized for poorer people. On the other hand, because our utility function is defined as an integral over a product of the quality of health and a function increasing in wealth, financial losses are considered more harmful when they are suffered by more healthy people. Similarly physical and mental harms are considered more harmful when they occur to rich people. So there are certain cases where our utility functions could incentivize prioritizing preventing certain types of harm to people that are better of in terms of health or wealth.

6.3 Effectiveness Factor

Unlike the perspectives of a perpetrator or victim, measuring the trust and effectiveness in police performance from a societal perspective is not an individualistic calculation; all results must be aggregated to formulate some qualitative result. Afterwards, this qualitative result needs to be converted to some quantitative measurement that can be directly related to CHI and utility. As a result, questionnaires are considered the best method in measuring society's trust in police.

Questionnaires and surveys are considered the best form of measurement for society's opinion on police effectiveness for two core reasons. First, questionnaires represent the aggregation of qualitative information into a single unified standard that can be assessed over time. Because society is being measured - and not an individual - questionnaires serve as a good representation for the constantly changing qualitative opinion of this entity. Second, questionnaires attempt to address the conversion of qualitative information into comparable data. This data can be assigned a quantitative value and subsequently used in a general equation for safety.

Two core types of questionnaires should be addressed when defining the effectiveness factor. The first is a victim's measuring of police effectiveness immediately after a crime. The second is society's measuring of police effectiveness in general, over longer periods of time.

6.3.1 Questionnaire 1: A victim's response to police effectiveness

Many English-speaking countries employ questionnaires to victims immediately after a crime to assess the response-effectiveness of police. For example, Canadian provinces currently employ over 26 surveys that measure the effectiveness of police according to victims after a crime has occurred [20]. A single, unified, survey that asks similar questions to these types of surveys could be employed across the Netherlands. This would allow researchers, police, and policy analysts to geographically compare police effectiveness on a case-by-case basis over time.

The following list serves as an example of proposed survey questions posed to victims after a crime has occurred. These survey questions are based on current survey questions used in practice across provinces in Canada:

- Still thinking about your most recent phone call to the police, how satisfied were you with the way your call was handled? Were you...
  - Very satisfied
  - Somewhat satisfied

  - Somewhat dissatisfied
  - Very dissatisfied
  - (Do not read) Don't know
  - (Do not read) No response
- Between the time the call was made and the responding officer arrived on scene, would you say the wait was . . . [read]
  - Longer than you expected
  - About the amount of time you expected, or
  - Less time than you expected?
  - (Do not read) Don't know
  - (Do not read) No response
- (If applicable) Still thinking about when police were dispatched to your home or business, how satisfied were you with the way the responding officer handled the matter when they arrived? Were you. . . [read]
  - Very satisfied
  - Somewhat satisfied
  - Somewhat dissatisfied
  - Very dissatisfied
  - (Do not read) Don't know
  - (Do not read) No response
- (If applicable) Still thinking about your most recent visit to a police station, how satisfied were you with the way police handled your concern or issue? Were you. . . [read]
  - Very satisfied
  - Somewhat satisfied
  - Somewhat dissatisfied
  - Very dissatisfied
  - (Do not read) Don't know
  - (Do not read) No response
- (Use this preface ONLY if police made more than one contact with respondent: Thinking about your most recent contact,) How satisfied were you with the way the police handled the matter? Were you. . . [read]
  - Very satisfied
  - Somewhat satisfied
  - Somewhat dissatisfied
  - Very dissatisfied
  - (Do not read) Don't know
  - (Do not read) No response

The purpose of these survey questions is to gauge police effectiveness on a case-by-case basis for victims. These questions are designed to be simple, yet effective, as asking too many questions may overwhelm a victim that is in need of psychological support. What is essential for these survey questions is that they are asked consistently and routinely across the Netherlands. In doing so, an accurate measure of victim response to police effectiveness can be compared across the entirety of the country. Based on these scores, police intervention and response tactics should be compared at a departmental level, and insights into specific neighborhood reactions to police should be analysed.

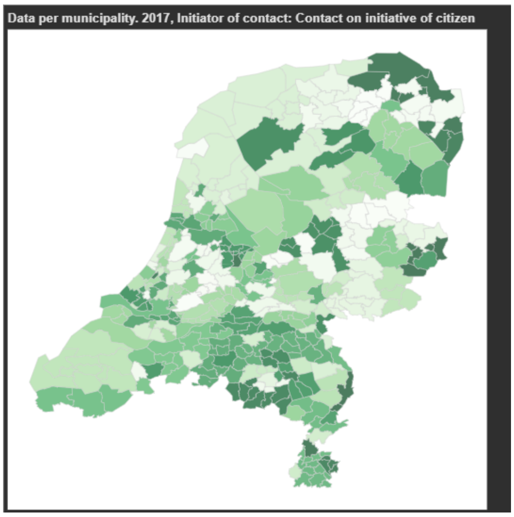
If implemented, the qualitative response given by these surveys should also be converted into a quantitative score and subsequently averaged. To stay consistent with the CHI and utility, these survey results are applied to a scale, 0-1. In this scale, a dissatisfied victim results in a police effectiveness score of between 0 and 0.3 while a satisfied victim results in a score of between 0.7 and 1. Scores in-between, such as "somewhat satisfied" and "somewhat dissatisfied", result in scores between 0.31 and 0.69, respectively.

6.3.2 Satisfaction rate

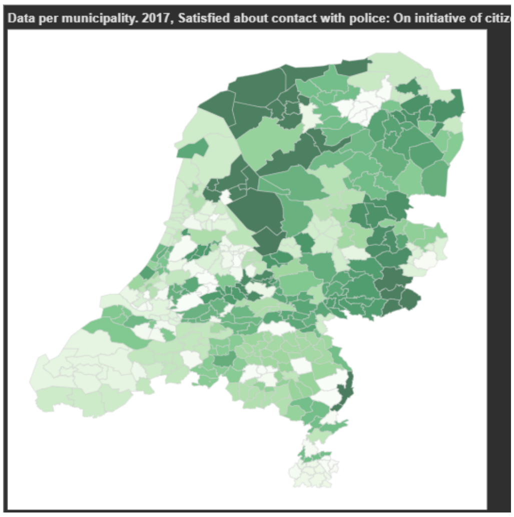
Another variable that is very relevant to the effectiveness factor is the "pakkans" (chance of a perpetrator getting caught). Unfortunately, when this factor is used in this formula, it might directly result in a self-fulfilling prophecy. When the pakkans is high, the result of the formula will be that it will be very useful to go put resources in solving this certain crime. When the pakkans is low, it will not be useful to put resources into solving this crime. When this occurs in a year, the lower pakkans will be even lower as no resources were put into this crime while the higher values will be even higher as a lot of resources will be put into these types of crimes. This is why we will not be using this "pakkans" but we will do an additional questionnaire.

To replace this pakkans, another variable would be needed. The variable that we considered is the general satisfaction with the police. We have found a survey by CBS [7] where the general opinion of the public on the police is measured. This survey is done and reported in the Veiligheidsmonitor [38] every two years. The factor of satisfaction with the police is currently measured in: Satisfied, Not Satisfied, Neutral and No opinion. We are able to convert these to a gradescore to apply it as a general satisfaction with the police variable.

This survey by CBS was done per region which is why we are able to display the survey on a map. We have plotted this item on a map to indicate which regions have more satisfaction with the police compared to other regions. This map can display the reason of contact with the police, the initiator of the contact, the way of contact, the satisfaction about certain aspects of the police (both locally and nationally) and the opinion about the availability of the police. Displayed in Figure 4a is "Initiator of contact: Contact on initiative of the citizen". The darker area represent around 90 percent while the lighter areas represent approximately 70 percent. In Figure 4b we display the "Satisfied about contact with police: Contact on initiative of citizen". The darker area indicates that approximately 77 percent is satisfied with the contact initiated by the citizen while the lighter areas represent approximately 50 percent. By comparing these figures, it becomes clear that the areas where people have initiated contact have a lesser satisfaction rate while the people that contacted the police less, have a better satisfaction rate.



(a) The percentage of people that had contact with the police that initiated the contact



(b) The percentage of citizens that initiated contact with the police that are satisfied with the police

Finally, we displayed the general gradescore about the faith in the police of the society. The map can be found in Figure 5. We have decided to display this map both in a red/green hue to show the worse and better grades and a version where we have replaced red by purple to better portray this map for the colorblind. This alternative version of the figure can be found in Appendix F. Red (or Purple in Figure 11), like shown in South Limburg is a score of 6.0 which is the worst grade on the map while bright green, like in North Groningen, is the best grade representing a score of 7.0. White represents a score in the middle.

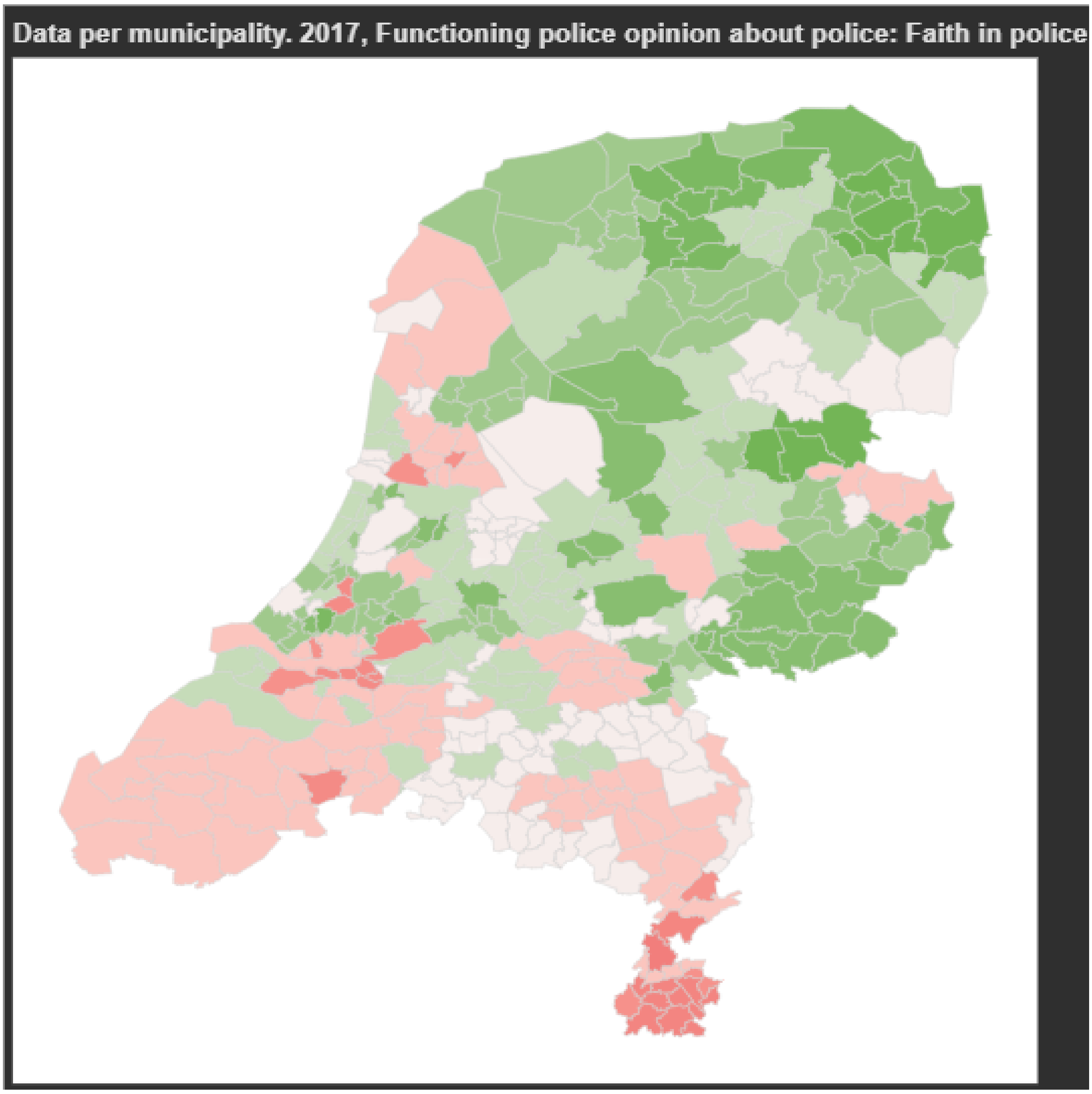


Figure 5: The score of the faith in the police per area

Using this data, we can assign a satisfaction rate of the society in a certain area. This satisfaction rate will also be on a scale of 0-1. A low satisfaction would result in a value between 0 and 0.3 while a high satisfaction would be between 0.7 and 1. Any value between 0.31 and 0.69 would be considered a medium satisfaction rate.

6.3.3 Combining into effectiveness factor

Because both surveys result in quantitative results bounded from 0 to 1, the scores of each can be subsequently averaged to create the effectiveness factor. Take the following example:

A victim goes to the police after an armed robbery. After assessing the crime, police present the victim with a questionnaire on the police effectiveness. After compiling and averaging the score of all questions, the victim yields a result of "somewhat satisfied." This corresponds to a score of 0.8. At the same time, society in the victim's municipality yields a low satisfaction rate for the police (updated every 2 years). This corresponds to a score of 0.2. When taking the average of these two scores, the overall police effectiveness for this specific crime is 0.5.

6.4 Combining all factors

As stated before, all factors can be combined to create the final equation of safety:

$$\text{Factor of Safety} = \Delta\text{utility * CHI * effectiveness factor}$$
$$b = \Delta u * c * f \tag{24}$$

Here Δu and c are the normalized versions of utility loss and crime harm score respectively. The multiplication of these factors will ultimately result in a safety factor somewhere between 0 and 1. By bounding this score, the level of safety for every type of crime can be directly compared - both individually between crimes and wholly across regions in the Netherlands.

The equation above serves to quantify the contribution of the police to the safety of a single victim. If a crime with multiple victims is considered, the total benefit B of the police can be calculated by summing the benefit to each victim. Here the effectiveness factor and change in utility can be different for each victim, but the CHI score is the same:

$$B = \sum_i \Delta u_i * c * f_i. \tag{25}$$

6.4.1 Safety from a monetary perspective

This safety unit of measurement can also be analysed through a financial perspective by the police. This equation would be seen as:

$$P = M/B, \tag{26}$$

where B is the total benefit of the police handling this crime, M the total cost and P the cost per benefit. In this form, M represents the amount of money, converted from hours of police work, needed to address a crime that has occurred. This leads to the representation of P as a cost-benefit factor of safety. Larger values of P represent a higher strain on police budget and allocation of resources. Intelligence analysts at the Dutch National Police headquarters already have values on the average number of hours needed to solve every type of crime. This information needs only to be converted to a monetary value, based on an average police officer's wage.

6.4.2 Factor Exponents

The current equation does not consider the weighting of each safety factor. This can be addressed by proposing a more generalized formula

$$b = (\Delta u)^x * c^y * f^z, \tag{27}$$

adding exponents to each factor. In our project all exponents are set to 1 as proof of concept, which corresponds to weighting every factor equally. In reality, there is a likely chance that victims, police, and society will value each factor in terms of importance differently.

At the moment, the exponents for factors of safety - x, y, and z - are not known. Overtime, research by professional analysts of the Dutch National Police with access to secure data should properly address these exponents.

6.4.3 Individual Factors

A core strength of our equation is its ability to be deconstructed; every component can be analysed individually. As a result, the factors from this project can be used to technically analyse and compare safety from multiple perspectives. For example, analysts may want to conduct studies on how varying levels of income in neighborhoods affect average prison-sentencing time in that region. The CHI would be a useful tool for this study. In another case, analysts may want to compare the level of violence against victims based on their wealth. The factor for utility would be useful in this situation.

Finally, each individual factor could even be compared directly with each other. A hot-spot map for CHI has already been created. The same hot-spot map for utility and police effectiveness could also be created based on their measured values. By comparing the geographic hot-spots of each type of factor on a map of the Netherlands, interesting conclusions may be drawn. Is it possible that a perpetrator's sentencing time (CHI) does not necessarily lead to more harm to a victim (utility)? Or are the two heavily correlated? If they are correlated, this would result in two hot-spot maps that can be directly overlaid. If not, the moral question of whether prison sentencing time is truly based on a victim's suffering must be ethically addressed.

7 Validation of the system

With the three factors of safety properly defined, the system can be validated through hypothetical cases. There are four hypothetical cases that we will consider in this section. First we will define the cases, then we will give hypothetical values for the cases and finally we will give a conclusion about the test results.

7.1 Example cases

The four hypothetical cases that we are considering are all combinations of minor and major crime and minor and major injury. The utility values were calculated following the general procedure in subsubsection 6.2.9 using the utility function chosen in subsubsection 6.2.8 and normalizing it as in subsubsection 6.2.10. Unless mentioned otherwise, all victims were considered to have median income and average quality of health based on their age. The values used for median wealth and average quality of health can be found in Appendix D. Some of the quality of health values after suffering physical or mental harm were based on the N3 model in [17]. In this paper quality of health values are assigned to different health states. The health states are based on 5 different dimensions of health and each dimension gets assigned a value 1, 2 or 3. 1 corresponds to little or no health problems in this domain, 2 corresponds to moderate problems and 3 to severe problems. The health states are written in the form *abcde*, with 11111 being the best state.

7.1.1 Minor Crime Minor Injury

A Dutch male of 65 years old with a typical life expectancy of 81 years suffer from a public disturbance (local peace violation), resulting in increased anxiety and stress over the course of 2 days. We assume that this reduces their quality of health by 0.1 for these 2 days. Their new life expectancy is unchanged. They suffer no financial harm. The victim is unsatisfied with how the police handled the situation; however, the local municipality has a high satisfaction rate with the police.

- Crime-Harm Index factor: 0.0127
- Effectiveness factor: 0.6
  - Victim: 0.3
  - Society: 0.9
- Utility loss: 0.000000957

$$\text{Safety Factor: } 0.000000957 * 0.0127 * 0.6 = 0.0000000073$$

7.1.2 Minor Crime Major Injury

A Dutch female of 25 years old with a typical life expectancy of 81 years suffers from stalking over the course of 4 years. This results in a decrease of mental health due to paranoia and anxiety for a decade (from 25 to 35 years). We assume that for these 10 years, the victim suffers from severe anxiety and depression, but her health is otherwise fine. According to the N3 model in [17], this corresponds to a quality of health value of 0.37 (the health state 11113 was used). We assume that the victim's wealth decreases by 1 % permanently as a result of this paranoia. We also assume that the victim's life expectancy is reduced to 77 years. The victim is unsatisfied with how the police handled the situation; however, the local municipality has a high satisfaction rate with the police.

- Crime-Harm Index factor: 0.120
- Effectiveness factor: 0.6
  - Victim: 0.3
  - Society: 0.9
- Utility loss: 0.0138

$$\text{Safety Factor: } 0.0138 * 0.120 * 0.6 = 0.00099$$

7.1.3 Major Crime Minor Injury

A Dutch person of 25 years old with a typical life expectancy of 81 years suffers from theft with violence. They lost 20 Euros and suffered from a broken toe. We assume that the broken toe takes 5 weeks to heal. During this time the victim suffers from moderate pain and discomfort. This corresponds to a quality of health of 0.843 according to the N3 model in [17] (the health state 11121 was used). The victim is partly satisfied with how the police handled the situation; the local municipality has a medium satisfaction rate with the police.

- Crime-Harm Index factor: 0.579
- Effectiveness factor: 0.5
  - Victim: 0.5

    - Society: 0.5
- Utility loss: 0.0000112

$$\text{Safety Factor: } 0.0000112 * 0.579 * 0.5 = 0.00000324$$

7.1.4 Major Crime Major Injury

A Dutch teenager of 18 years old with a typical life expectancy of 81 years old is murdered. The victim's relatives are partly satisfied with how the police handled the situation; the local municipality has a medium satisfaction rate with the police.

- Crime-Harm Index factor: 1
- Effectiveness factor: 0.5
    - Victim: 0.5
    - Society: 0.5
- Utility loss: 0.0891

$$\text{Safety Factor: } 0.0891 * 1 * 0.5 = 0.04455$$

7.2 Conclusion of Test Results

Based on the Safety Factor calculated for the above four examples, they can be ranked as below:

- Minor Crime Minor Injury (0.0000000073)
- Major Crime Minor Injury (0.00000324)
- Minor Crime Major Injury (0.00099)
- Major Crime Major Injury (0.04455)

As expected, example 7.1.4 has the highest safety factor value as that is the example with major crime, and major injury and the example illustrated in 7.1.1 has the lowest value as the crime does not constitute to be major and the injury to the victim is minor as well. Based on these results, the order of magnitude is almost entirely determined by utility. In this case, utility makes larger distinctions between crimes when compared to CHI and effectiveness. Therefore, the current equation may cause police to over value crimes that cause the most harm. Should the police want to prioritize the CHI and police effectiveness more, the exponent for utility can be less than one (subsubsection 6.4.2).

7.3 Edge cases

The solution has a couple of edge cases where it might be difficult to properly use the solution. In this section we will describe these edge cases and the suggested solution if there is any.

The first case are crimes where there is no human victim, also called victimless crimes. An example of a crime like this is financial fraud. When a person or company are not paying the correct amount of taxes. There is no direct victim to this crime, the tax payers who are paying the correct amount are all victims unknowingly. To use the formula in these kinds of situations, all indirect victims could be used to define the change in utility.

A second edge case is when a victim is unable to answer the survey. If a victim gets incapacitated in a manner that they are not able to adequately answer the questionnaire anymore, a solution to this problem would be to let the closest relative or friend answer the questions of the survey.

The third edge case is where people are not physically or financially harmed by a crime. For example, someone threatens your life; you are not financially or physically harmed but might have mental damage. In this cases, the mental harm should be quantified and used in the formula. This will lead into our fourth edge case; quantifying mental harm. In QALY, a survey is done to the patient to identify their current physical state before an intervention is done. Unfortunately, we can not identify the mental state of a certain person before a crime as one in most cases does not know who will be targeted by a crime. This will lead to someone attempting to quantify the mental state of a victim before the crime happened. This can be done by talking to both the victim and the close friends and family of the victim.

8 Risks, Challenges and Issues

We identified issues and problems linked with installing high-tech systems such as predictive algorithms, artificial intelligence, statistical judgments, and so on based on existing literature. We identified and listed potential challenges that the Dutch National Police may encounter by combining the conclusions of the 2019 White paper [9] and Peter Asaro's study [2] of the Chicago Police Department.

There are several organizational difficulties associated with implementing high-tech systems in the police. In order to guarantee the continuity of the police processes as more organizational operations are automated, it is crucial to maintain system compatibility. A workforce with diverse skills that might not easily fit into the traditional police employment structure may be needed for the development, acquisition, maintenance, and usage of these technologies. AI has the potential to alter existing employment and, in some situations, even replace them[9].

One should look for indications that political, economic, or social pressures may have affected historical datasets, as well as how those pressures may be influencing the way data is currently collected. One should be aware of how new data practices may change social norms and how those changes might affect the communities and individuals that a system seeks to serve. But if the creators of such algorithms are unaware of these problems, as well as the possibility of biased data or erroneous data, they may be generating garbage-in-garbage-out systems while mistakenly thinking they are developing high-quality systems (as measured by their available data). The lesson is to never assume that data is accurate; instead, one should be skeptical[2].

By freeing police officers from tedious or repetitive (bureaucratic) labor, making them more effective, and giving them more time to devote to other duties, certain specialized systems can enhance their well-being and job satisfaction (e.g. social tasks or fieldwork). However, caution must be exercised to ensure that these systems do not adversely affect workers' feeling of agency, purpose, or safety. If necessary, suitable possibilities for retraining or transfer to other positions within the business should be investigated (effects on the police personnel and police organization)[9].

By improving the likelihood that police investigations will be successful, AI may potentially have an influence on the nation's overall security. Additionally, it could facilitate the beginning of a prosecution by giving the prosecutor stronger data and legal support. AI and algorithmic systems might, however, speed up the initiation of new inquiries. The capability of the police should stay appropriate to the cases in question so that their resources are not unnecessarily dedicated to online fraud and so that such a system does not unintentionally have a detrimental effect on the nation's overall security system (effects on the general security system of a country)[9].

Bureaucratic labor makes up a sizable portion of police activity, such as crucial steps in the investigative process and the input of crime reports, etc. A gap between police officers and people can be seen as a result of the increasing quantity of paperwork, which leaves less resources available for fieldwork and communications with citizens. Such a discrepancy may raise questions about the productivity and job happiness of police officers, which might have a detrimental effect on the organization's overall effectiveness[9].

There are also various ways in which such systems might "go awry", often as a result of a well-meaning but mathematically unsophisticated understanding of how such systems function.

> *This includes a failure to comprehend how statistical outliers may be handled or misrepresented, as well as how historical data patterns can be self-reinforcing. One example of this is the systematic denial of opportunities for poverty eradication to poorer people and communities by denying them credit and charging them higher interest rates[2].*

Predictive policing has been criticized because, although it seeks to prevent crime, it also raises concerns about discrimination and precrime, as well as essentially treating individuals and groups as guilty of (future) crimes for activities they have not yet performed and may never commit. In addition to the obvious implications for racial, gendered, ethnic, religious, class, age, disability, and other forms of discriminatory policing, more widespread worries about the implicit biases present in historical data sets have the effect of amplifying and exacerbating the central controversy of prejudice and precrime [2].

From a utility perspective, in some cases the best approach might be to protect someone who is well off from a crime in contrast to protecting someone who is already worse off, as more utility is at stake in the first case. This means that the indicator inherently puts people with a lower age expectancy at a lower priority. This discrimination is unacceptable when it comes to the society, at least in the applications that our project is aimed at. Also from a utilitarian perspective, this makes sense and holds true. However, in other ethical theories, it does not.

8.1 Unreported Crimes

Perhaps one of the largest holes in the "equation for safety" is the reliance on crimes to be reported to police. In reality, the number of reported thefts has dropped steeply in many European countries over the last 15 years [14]. Without reporting of crime by a victim, all three factors of safety fail to compute a "safety factor". In economics, indifference curves are used to display the preference of a consumer between two goods. This same concept can be applied to a victim's response to a crime.

Below in Figure 6, is an indifference curve that displays the value of goods stolen and the probability of goods being returned [14]. In summary, if police are more likely to return stolen goods, and if the goods are of high enough value, then a victim is more likely to report a crime in the first place. This is a self-fulfilling prophecy; more effective police work will result in more victims reporting crimes.

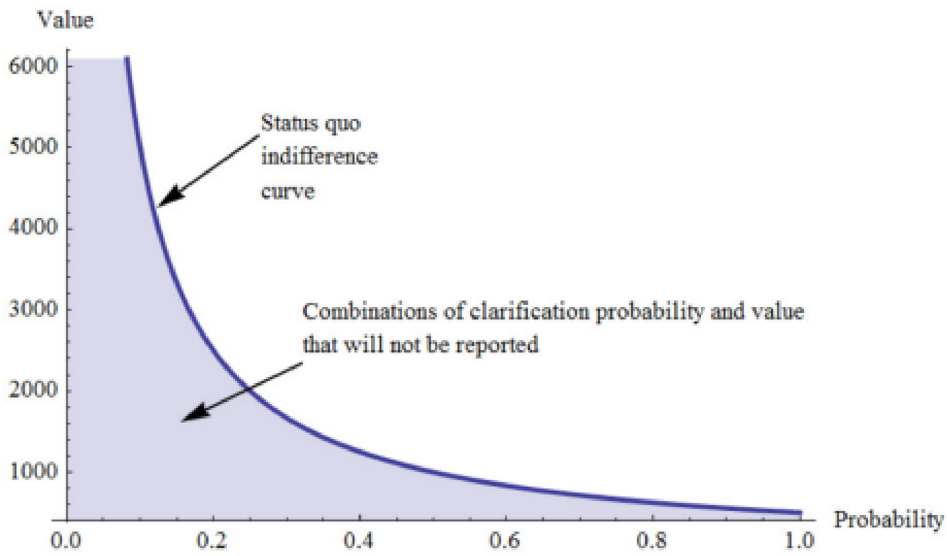


Figure 6: Indifference curve for unreported crime [14]

More effective police work can stem from a number of strategies. Consider a hypothetical scenario of initial reporting cost, y. If police increase their effectiveness, this lowers the overall reporting cost to x. This results in the indifference curve in Figure 7.

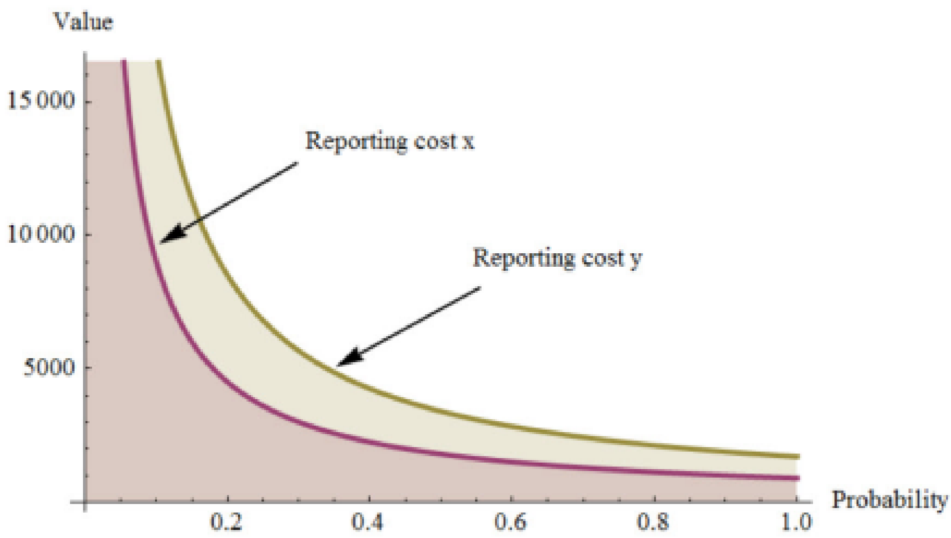


Figure 7: Indifference curve for unreported crime after lower reporting costs [14]

Various factors lead to police effectiveness. For example, more police stations or faster response rates may result in less time required for victims to report a crime, making them more likely to report a crime in the first place.

Simply put, the issue of unreported crimes can not be addressed by the current equation for safety. However, there are known strategies [14] to improving the reporting rate of crime. Should the police want to address unreported crimes, these strategies should be discussed. Further consideration of unreported crimes falls outside the scope of this project.

9 Ethics

9.1 Introduction

Our lives are secretly structured by algorithms. Algorithms may choose which political advertisements and news stories people see, as well as whether someone is recruited, promoted, awarded a loan, or given housing. However, it is unclear what role algorithms play in these crucial choices[19].

Over the past two decades, there have been significant changes in crime prevention, partly as a result of an increased dependence on systems that analyze crime data. These systems, designed expressly for use by law enforcement, judges handing down sentences, and prison administration, do both predictive and retrospective analysis to assist decision-making within the criminal justice system. Furthermore, beneath user-friendly interfaces, these software systems often include geographical informatics packages and complex statistical features. Recent research has shown issues with the algorithms in these systems' poor logic as well as the underlying biases in the data[5].

When algorithmic police monitoring receives a data protection stamp of approval or does not (appear to) process personal data, it tends to disappear into the background. Murakami Wood reported in 2009 that *"most governments do not utilize any robust procedures by which they would be able to analyze proposed surveillance technologies, either in themselves or in relation to other technologies or indeed to non-technological alternatives[22]."*

In order to ensure that algorithmic police surveillance is used in the public interest, democratic oversight of the practice should not only focus on legal compliance but also on socio-ethical issues that involve civil society and those who are subjected to it. To this end, it should use a variety of policy instruments, including standards and regular evaluations. A change in legislation and policy from a narrow emphasis on technology to socio-technical practices and the power relations that characterize them is required to address the socio-technical downsides of algorithmic police monitoring[36].

To maintain a competitive edge in contemporary culture, the police must make an effort to keep up with rapid technological advancements. This is crucial in order to stay up with criminals who make use of new technologies and to fight the emergence of new types of crime. In addition, the police are morally obligated to do their duties of serving and protecting society to the best of their ability. Technologies like artificial intelligence (AI), predictive algorithms, Big Data Analytics etc. can be used to empower citizens, such as by opening up more channels of communication with the police, as well as to increase the effectiveness and efficiency of law enforcement processes (such as investigations)[9].

All facets of the information economy are quickly adopting data-driven organizational management, such as big data, machine learning, and artificial intelligence (AI). There is little question that the gathering, sharing, analysis, and use of data in developing policies, making decisions, and carrying out everyday tasks may enhance how well the government runs and how well the public is served. Law enforcement is no different from other government services in this regard[2].

It is possible to build a society of well-coordinated and cooperative agents acting with human welfare as their top priority by empowering individual agents to act ethically and assess the ethics of other agents' actions. The author of *Even Angels Need the Rules: AI, Roboethics, and the Law*, Ugo Pallago thinks that this is insufficient. The author supports the need for primary rules guiding social standards and the ability for secondary rules to replace, modify, or supersede main rules when circumstances change[24].

9.2 Recent Excitement

There are rising worries about the social and ethical ramifications of this change as data-driven organizational management, powered by big data, machine learning, and AI approaches, continues to speed up and more procedures are automated. The employment of data-driven algorithms in police, law enforcement, and legal procedures has, however, given rise to serious issues. This involves the use of predictive policing, which identifies people or places with a high chance of committing crimes in the future and targets them for additional police. For a variety of reasons, including issues with bias and precrime and effectively labeling people as guilty of (future) crimes for activities they have not yet performed and may never commit, predictive policing has proven contentious. To quote the author's observations:

> *"This central debate over prejudice and precrime is heightened and made worse by worries about the implicit biases present in historical data sets, the obvious ramifications for racial, gendered, ethnic, religious, class, age, disability, and other forms of discriminatory policing, as well as how the use of predictive information systems affects police officers' psychology and behavior."*[2]

As more administrative procedures are automated, questions about the fairness, accountability, and openness of the algorithms used to make important decisions that affect people's rights and chances in life are becoming more and more prevalent. Less frequently discussed are the ways in which the adoption of data-centric processes and data-driven management have a significant impact on the techno-social and spatio-temporal structure of organizations, as well as on the management objectives of the organization, the nature of work, and the standard of results. That is the character of modern technocratic rule[2].

9.3 Persuasion Agents and their Perception

Persuasion agents are one of the application areas where AI tries to change people's behaviors as mentioned by Kang et al. (2015) [15] and Rosenfeld and Kraus (2016)[27]. A thorough investigation into how people view the morality of being persuaded by an AI agent was done by the authors of [33]. In order to save the lives of five people, a participant in the trolley scenario is made to purposefully hurt an innocent bystander by shoving him onto the railway track. A sacred value must be violated in order to achieve this consequentialist ethical consequence (i.e. one shall not kill).

Three different approaches to persuasion were explored by the authors: lying, appealing to participants' emotions, and utilitarian reasoning. Some participants are given the three tactics by a persuasion agent and a person posing as an authority figure (the station master of the train station). *The findings indicated that participants have a substantial preconceived negative bias towards the persuasion agent and that argumentation- and lying-based persuasion tactics are more effective than emotional persuasion strategies.* The results did not indicate a substantial difference between genders or civilizations. According to the study, the use of persuasion tactics should take individual personality, ethical stance, and domain-specific expertise into account.

Not to mention, the dynamics of human interaction will change when ethical considerations are included into such systems. Some individuals may alter their behavior in order to prevent the systems from achieving their intended goals. For instance, a youngster who is commonly seen as a non-combatant and in need of protection could deactivate an ethical autonomous gun system (assuming such a thing is possible) by covering the sensor system of the gun with spray paint, which is typically not recognized as a fatal weapon[40].

9.4 Responsible use of AI

A cooperative project involving TU Delft, University of Leiden, and the National Police resulted in the publication of the White paper in 2019[9]. This study discussed a few of the difficulties that an organization would run into if it opted to implement AI-based technologies in its daily operations. Numerous issues in our analysis, their suitable remedies, and suggestions are founded on the same principles but have been modified to fit the objectives of our suggested solution.

The Netherlands' police force is dedicated to upholding the rule of law and safeguarding fundamental human rights. The national constitution, EU Charter, particular national legislative acts, and EU directives and regulations like the General Data Protection Regulation (GDPR) or Law Enforcement Directive are just a few examples of domestic and international legal instruments that directly obligate the police to uphold this commitment. Fundamental human rights can serve as a foundation for new regulatory measures and serve as a justification for the creation, application, and usage of Hi-Tech systems.

Compliance with holding rules and regulations in a democratic nation like the Netherlands must be assumed to be standard in any application. Therefore, the legal gray areas should be the main focus of any practical concerns of ethical AI application. For the ethical use of such systems, conformity with the law is necessary yet insufficient. Naturally, the current legal frameworks should also be regularly reviewed to determine if improvements are conceivable. Finding the ethical foundations that supported the development of the current fundamental rights framework is necessary for this re-evaluation. These moral foundations are clear and understandable thanks to decades of universally embraced fundamental rights in the EU[9]

9.5 Principles for Responsible Use

These key ideas are drawn from the White paper, as was already indicated. The publication provided a comprehensive number of criteria and concepts, and the ones we believe are pertinent and crucial for our suggested approach are listed below. The elaborate list can be found in Appendix G.

- Accountability: From an ethical and legal perspective, responsibility/liability must always be assigned to a moral agent or legal person: systems are neither. However, the use of complex technology like artificial intelligence can lead to "attribution confusion", where it is not clear who, if anyone, should be held responsible [9].
- Transparency: Transparency in terms of the people involved, the rationale, operations, data and transparency to the users of the system is absolutely crucial.
- Fairness and Inclusivity: Everyone should be treated equally and it should be ensured that pre-existing biases do not contribute to such projects.
- Socio-Technical Robustness and Safety: The performance, its effectiveness and accuracy are of significance. The system also has to be robust and resilient to potential issues.

9.6 Statistics and Ethics

There is the practical worry that law enforcement cannot choose to disregard intelligence about probable crimes without abdicating their responsibility to stop crime. This is a strong intuition even when the extent and character of that responsibility are under dispute. Indeed, a lot of data-driven management is driven by the same idea. That is, if we can forecast future patterns and occurrences using previous data, we should do so in order to better spend scarce resources toward achieving a task or objective. While such intuitions are not incorrect—clearly, greater use of information may enhance policing in many ways—pursuing them without careful study, prudence, and attention to their numerous ramifications and specific implementations can swiftly result in issues[2].

> *Just think about the historical development of current statistics—horse racing betting. The oddsmakers go to considerable efforts to offer precise statistical forecasts of each horse's chances in a race. The outcome of the race itself is what counts, not whatever horse is considered the favorite to win. Only roughly one-third of the time does the favorite win. Even while it may be argued that placing bets at all is mathematically nonsensical in many games of chance, gambling would not make sense if this were not the case[2].*

Statistics are notorious for not distinguishing correlations in data from causal reasons, and it would be unjust to treat people with suspicion for coincidental correlations when the underlying causal mechanisms for criminal behavior are absent. This kind of profiling becomes deeply problematic when it becomes prejudicial, and the correlation is taken as itself constitutive of guilt, or warranting a presumption of guilt, rather than a presumption of innocence[2].

It would be unfair to suspect somebody of acting criminally because of coincidental correlations when there are no underlying causal processes for such behavior as statistics are infamous for failing to discern between correlations in data and causal explanations. When this type of profiling becomes prejudiced and the correlation is interpreted as constituting guilt in and of itself or as warranting a presumption of guilt rather than an assumption of innocence, it becomes extremely problematic[2].

9.7 The Ethics of Care Approach

According to the moral philosophy known as "the ethics of care," the moral relevance of relationships and dependence is inherent in human life's essential components. Normalizing and enhancing the wellbeing of caregivers and care recipients within a web of social relationships is the goal of care ethics, which tries to uphold relationships. Care entails looking after our own and other people's needs in addition to the needs of the environment around us. It is most frequently characterized as a practice or virtue rather than a philosophy as such[1].

As clearly explained by Peter Asaro in *AI Ethics in Predictive Policing* [2], It would be beneficial for society and themselves if the Police Department followed a "Ethics of Care" approach while taking such initiatives. In terms of law enforcement, this approach can *look for creative solutions that decrease the likelihood of crimes, necessitating less money to enforce the law.* This is in contrast to focusing on the reasons why people break the law. Any potential implications of any new data tool would be considered, and the relationship between law enforcement and the community would also be recognized as being of the utmost significance.

9.8 Ageism

Ageism is a visible but implicit ethical and moral problem that affects practically every aspect of life. Although the justification for it varies based on the conditions and the precise situation, the health economist would recommend an action to save more life-years. The public's focus is on fairness, while a doctor will work to maximize benefit. Rationing may be unnecessary if efficiency is achieved. No nation, however, is able to give all of its residents' access to cutting-edge services. There will always be situations where resources are limited, and allocation decisions must be made [29].

It is possible to counter-argue this and claim that the older population deserves better care and attention purely because of their contribution to the society until that point. The criminal justice system treats a murder convict the same regardless of the age of the victim. This is accepted by the society. The goal is to be unbiased and distribute benefits to everyone in the society from a government's perspective. Ageism makes no assumption concerning the value of anyone. It measures years lived and left to live.

It can be argued that someone who has had a long and miserable life has had less of a fair innings than one who has had a short and happy life. Yet it is just as arrogant to judge the value of the life someone has had as to judge the value of that which they will have. The healthy and wealthy may live in misery due to phobias and family traumas and the poor and disabled may be contented [29].

Decisions that appear impossible can be decided by chance or probability since chance is a necessary inevitable component of human existence. Another potential solution is to use a queue, however this might provide an unjust distribution. The wealth gap is yet another important element at play in this situation. Should law enforcement resources be accessible in the same manner that a wealthy individual may obtain pricey private medical care? Now, some could counter that this is already a reality in modern society. If it is ethical for someone to obtain a higher quality of service as a result of their own actions, should the same for law enforcement be considered ethical? Maximizing effectiveness seems to be the obvious approach. That is, maximizing the benefit from the limited amount of available resources. What policymakers should keep in mind is that this should be decided based on a rationale and made transparent to the society.

9.9 Case: The Heat List, Chicago Police Department

This instance in particular emphasizes how difficult it is to integrate cutting-edge technologies into Law Enforcement procedures. In 2012, CPD began using the Strategic Subject List (SSL) algorithm, which was created as an experiment by a researcher at the Illinois Institute of Technology. Numerous police personnel said that they had incomplete knowledge of the process used to construct the list. The SSL ratings merged those at risk of being victims with those at risk of being offenders in a single metric of "being involved in violence," while they assumed—or were persuaded to believe—that everyone on the list was a violent offender and was likely to conduct additional violent acts.

This observation teaches us a deeper lesson: the value of different types of data is highly influenced by the possibilities of action and the kind of actions that are envisioned. There is still value in thinking about what actions are available to address an issue, even though the current trend is to collect any and every accessible data in the hope that something valuable can be inferred from it. This entails doing more than just improving the effectiveness of the way a problem is currently being addressed; it also entails using data to discover new ways to act, intervene, and better understand the issue[2].

10 Recommendations

The topic of input, or how data is organized and gathered, has dominated current literature and reasearch thus far. But it's also important to consider how the data analysis is presented and how that could affect how people behave both within and outside of institutions. The way the Chicago police have been using the SSL is an excellent illustration of this problem. The decision to utilize the list for "custom notice" and for "hot lists" after crimes by precincts and police is not unrelated to the way a system like SSL is designed[2].

The following suggestions have been compiled for the police's consideration. These are clearly not all-inclusive, nor can they be. The environment of this entire project is fresh, thus it is crucial that things be dynamic in order to adapt to conditions as they develop. These are also based on the findings of the 2019 White paper [9] that has been referred to throughout the previous sections.

Without exponents in the equation for safety, utility is currently the main determining factor of safety. For further research, the police should aim to determine exponents (subsubsection 6.4.2) based on their desired values of utility, CHI, and effectiveness. Besides that, it is crucial that the future researchers account for the age and wealth discrimination. Since the utilitarianism approach to ethics aims at maximizing the utility under the circumstances, this can lead to issues while implementing the proposed idea. Therefore, it is important for the police to take this into account and act accordingly.

10.1 Organisation

1. To ensure impartial critical assessment of the use of technology inside the organisation, establish a review board within the business and consider about hiring a "High-Tech Ombudsperson."
2. Revisions should be made to the organization's "Code of Ethics" to reflect matters that are crucial for scientists who work on projects such as this.
3. Encourage and provide incentives for such research initiatives to include ethical, legal, and social aspects.
4. Continually educate scientists about ethical issues, and keep them informed about new discoveries and ethical implications.
5. Establish transparent and equitable procedures for determining who is accountable and responsible for a system's output. Instead of subjecting employees to a lot of liability, think about assuming responsibility as a company.
6. Standardize procedures and initiatives within and between different agencies. Standardization across law enforcement services may point new initiatives in the proper directions and facilitate their effective execution.

10.2 Responsible Technological Progress

1. Observe people's right to privacy. Don't collect more data than is necessary, keep it safely, and be aware that anonymization is a limited kind of security.
2. Make sure that all systems continue to have meaningful human control. There must always be one or more accountable people, thus they must be knowledgeable of and capable of interfering with the functioning of these systems.

10.3 Societal Context

1. Enhance collaboration with the ministry and other connected organizations by contributing to a comprehensive strategy for ethical usage and development within the Dutch Justice and Security area (Interpol, UN, municipalities).
2. The police should also pay attention to citizens, professionals, and other groups. The police can benefit from a social discourse to keep current with and aware of society perspectives around what constitutes legitimate and ethical usage of such systems. When impacted parties have a right to input on how the police behave, democratic values should be followed. On the other hand, it's crucial for the police to participate in public debates in order to advance their own opinions and goals in this regard.
3. Adoption, trust, and responsible usage may all be aided by awareness and education about both technological and ethical issues. Although practitioners are becoming more conscious of these problems in recent years, it is still usual for them to be more focused on performance metrics than ethical ramifications, such as biased decision-making. Educating all parties concerned on these problems and introducing them to alternative solutions is a first step in resolving this.
4. Ageism and wealth disparity are two moral yet objective issues that flourish secretly in the society. Acknowledging and addressing it would be the first step towards resolving issues caused by it. Open discussion, dialogue and further sociological research needs to be done to equip ourselves better to handle discrimination and dilemmas caused by the same.

11 Conclusion

The goal of this project was to translate the qualitative concept of safety into a precise quantitative "unit of safety" that relies on measurable values found within the Netherlands. Converting qualitative information into tangible units of measurement comes with several challenges, such as the justification of the significance of different "variables of safety," accessibility of data, and ease of adoption by the Dutch National Police. This unit of measurement for safety ultimately must be the culmination of several safety factors. The creation of this metric successfully allows police to compare, over time, how police resource allocation and intervention tactics lead to a safer society in the Netherlands.

The base equation for safety is:

$$b = \Delta u * c * f \tag{28}$$

The equation leads to a composite factor of safety, b, which is bounded from 0 to 1. A score of 1 would correspond to a maximally severe and harmful crime that was handled perfectly by the police, while a score close to 0 means the crime is either very minor or was not effectively handled. As such b represents how much the police managed to mitigate the harm done to the victim.

u represents the change in a victim's utility before and after a crime has occurred. A victim's utility is a function based on their decline in QALY (a medical unit that measures physical health) and their decline in financial well-being.

c represents the moral severity of a crime that has occurred, based on a crime-harm index of the Netherlands. A crime harm index scales crimes according to prison sentencing times in the Netherlands.

f represents the overall efficiency of the police in handling a crime that has occurred. police efficiency is measured through two types of questionnaires. The first questionnaire records a victim's opinion on the handling of crime by police immediately after a crime, and the second records Dutch society's opinion on overall safety in the Netherlands (updated every 2 years).

All three main variables are scaled from 0 to 1. Based on future research by police, constants can be adjusted for each variable to scale to a more "realistic" measure of safety. Each component of the equation can be individually analysed. For example, researchers can purely evaluate crime from a victim's perspective, through a victim's change in utility (u), or they purely evaluate crime from a perpetrator's perspective, through a criminal's sentencing time (c).

The implementation of a safety factor for every victim of a crime can be aggregated to create a continually updating hot-spot map of safety across the Netherlands. Over time, police can analyse this map to see if their intervention methods in certain regions actually lead to an increase in safety. We believe this safety factor considers the core perspectives of harm reduction and safety: consideration for a victim's utility, moral ranking of crime according to dutch society and law, and the public's opinion and trust on the actual effectiveness of the national police. Moving forward, we believe the Dutch

National Police should take such a form of measurement into serious consideration, as this equation explores a more holistic representation of safety in Dutch society through various factors. Currently, utility is the main determining component for the safety factor. This can be adjusted through the use of exponents, highlighted in subsubsection 6.4.2. Should this project continue, the scaling of such factors of safety through exponents would be the central focus point for research.

12 Team reflection

In the first week of the project we started by meeting in person on Monday. On Monday we would define our goals for the week and work together on parts of the project. At the end of each Monday, everyone had a clear goal in mind for the rest of the week. We would continue the week by doing a daily stand-up online everyday at 10:00. In this daily stand-up we discussed what we have done in the previous day and what we would do that day. We would indicate problems that we have encountered or expect to encounter and ask advice from the rest of the team. If it was necessary, a new Teams meeting would be planned by some members of the team to work together for a couple of hours while online.

When we were planning the week on Monday, we would always identify who would be best suited to do a certain task that week. Kevin would usually get the very difficult mathematical tasks while Laura would usually get the programming tasks. As Raghav and Michael have a lot of knowledge about management, ethics and technology in general, they would usually take up tasks that are connected to these topics or to research topics. Michael and Raghav are both very familiar and fan of doing presentations, this is why they would usually take the lead in structuring the presentation. Laura is quite familiar with LateX which is why Laura would usually take charge of outlining the report and fixing technology issues either with the report or the technical solutions that we thought of.

We have also spend quite some time travelling around the Netherlands to meet with experts in the field, we have met both people in- and outside of the police department. In these meetings, Michael would usually take the lead in explaining our project and asking the expert about their expertise while Kevin would usually explain the mathematical aspects and Laura would take notes. After every meeting, we would usually lunch together in the city that we met in, this was to increase the team-bonding.

References

(n.d.). https://iep.utm.edu/care-ethics/

Asaro, P. M. (2019). Ai ethics in predictive policing: From models of threat to an ethics of care. *IEEE Technology and Society Magazine*, *38*(2), 40–53. https://doi.org/10.1109/mts.2019.2915154

Ashby, M. P. (2018). Comparing methods for measuring crime harm/severity. *Policing: A Journal of Policy and Practice*, *12*(4), 439–454.

Blackorby, C., Donaldson, D., & Weymark, J. A. (1984). Social choice with interpersonal utility comparisons: A diagrammatic introduction. *International Economic Review*, 327–356.

Bullock, B. B., Nascimento, F. L., & Doore, S. A. (2021). Computing ethics narratives: Teaching computing ethics and the impact of predictive algorithms. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 1020–1026. https://doi.org/10.1145/3408877.3432468

CBS. (2013). *Vermogen van huishoudens; huishoudenskenmerken, vermogensklassen.* https : / / opendata . cbs . nl / # / CBS / nl / dataset / 83835NED / table ? ts = 1666875479057 (accessed: 10-27-2022)

CBS. (2020). *Burgers en politie; regio (2012-2019).* https://opendata.cbs.nl/#/CBS/nl/dataset/81928NED/table (accessed: 10-25-2022)

CBS. (2022). *Geregistreerde criminaliteit; soort misdrijf, regio.* https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83648NED/table?fromstatweb (accessed: 9-23-2022)

Dechesne, F., Dignum, V., Zardiashvili, L., & Bieger, J. (2019). Ai & ethics at the police: Towards responsible use of artificial intelligence in the dutch police. *AI & Ethics at the Police: Towards Responsible use of Artificial Intelligence in the Dutch Police.*

Eisenstein, M. (2022). Does the human lifespan have a limit? *Nature*, *601*, S2–S4. https://doi.org/https://doi.org/10.1038/d41586-022-00070-1

Greenfield, V. A., & Paoli, L. (2013). A framework to assess the harms of crimes. *British journal of criminology*, *53*(5), 864–885.

Groot, C., De Hoop, T., Houkes, A., Sikkel, D., et al. (2007). De kosten van criminaliteit: Een onderzoek naar de kosten van criminaliteit voor tien verschillende delicttypen.

Hammitt, J. K. (2013). Admissible utility functions for health, longevity, and wealth: Integrating monetary and life-year measures. *Journal of risk and uncertainty*, *47*(3), 311–325.

Idsø, J., & Årethun, T. (2018). Crime statistics: Modeling theft in favour of victims' choices. *Economies*, *6*(2), 25.

Kang, Y., Tan, A.-H., & Miao, C. (2015). An adaptive computational model for personalized persuasion. *Twenty-Fourth International Joint Conference on Artificial Intelligence.*

Kravchenko, L., Troyan, I., & Gindes, E. (1970). Interdisciplinary synthesis in economic research in the context of digitalization. https://doi.org/10.2991/assehr.k.200509.138

Lamers, L. M., McDonnell, J., Stalmeier, P. F., Krabbe, P. F., & Busschbach, J. J. (2006). The dutch tariff: Results and arguments for an effective design for national eq-5d valuation studies. *Health economics*, *15*(10), 1121–1132.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica: Journal of the Econometric Society*, 111–141.

Martin, K. (2018). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Maslov, A. (2016). Measuring the performance of the police: The perspective of the public. *Public Safety Canada*.

Mitchell, R. J. (2017). The usefulness of a crime harm index: Analyzing the sacramento hot spot experiment using the california crime harm index (ca-chi). *J Exp Criminol*, *15*.

Murakami Wood, D. (2009). A new 'baroque arsenal'? surveillance in a global recession. *Surveillance & Society*, *6*(1), 1–2. https://doi.org/10.24908/ss.v6i1.3398

Obradovic, M. (2012). Searching for the social value of a qaly in the netherlands: The willingness to pay for a qaly.

Pagallo, U. (2016). Even angels need the rules: Ai, roboethics, and the law. *ECAI*, *258*, 209–215.

Paoli, L., & Greenfield, V. A. (2013). Harm: A neglected concept in criminology, a necessary benchmark for crime-control policy. *European Journal of Crime, Criminal Law and Criminal Justice*, *21*(3-4), 359–377.

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations research*, *28*(1), 206–224.

Rosenfeld, A., & Kraus, S. (2016). Strategical argumentative agent for human persuasion. In *Ecai 2016* (pp. 320–328). IOS Press.

Ruitenburg, v. T., & Ruiter, S. (2022). The adoption of a crime harm index: A scoping literature review. *Police Practice and Research: an International Journal*.

Shaw, A. (1994). In defence of ageism. *Journal of medical ethics*, *20*(3), 188–194.

Sherman, L., Neyroud, P. W., & Eleanor, N. (2016). The cambridge crime harm index: Measuring total harm from crime based on sentencing guidelines. *Policing*, *10*(3), 171–183.

Sherman, L., Neyroud, P. W., & Neyroud, E. (2016). The cambridge crime harm index: Measuring total harm from crime based on sentencing guidelines. *Policing: a journal of policy and practice*, *10*(3), 171–183.

Sherman, L. W. (2020). How to count crime: The cambridge harm index consensus. *Cambridge Journal of Evidence-Based Policing*, *4*(1), 1–14.

Stock, O., Guerini, M., & Pianesi, F. (2016). Ethical dilemmas for adaptive persuasion systems. *Thirtieth AAAI Conference on Artificial Intelligence*.

Strotz, R. H. (1953). Cardinal utility. *The American Economic Review*, *43*(2), 384–397.

Sustainable development goals: United nations development programme. (n.d.). https://www.undp.org/sustainable-development-goals

Van Brakel, R. (2021). How to watch the watchers? democratic oversight of algorithmic police surveillance in belgium. *Surveillance & Society*, *19*(2), 228–240. https://doi.org/10.24908/ss.v19i2.14325

Volkow, N. D. (2021). Addiction should be treated, not penalized. *Neuropsychopharmacology*, *46*(12), 2048–2050. https://doi.org/10.1038/s41386-021-01087-2

voor de Statistiek, C. B. (2021). Veiligheidsmonitor. https://www.cbs.nl/nl-nl/maatschappij/veiligheid-en-recht/veiligheidsmonitor

Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: The qaly and utilities. *British medical bulletin*, *96*(1), 5–21.

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., & Yang, Q. (2018). Building ethics into artificial intelligence. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. https://doi.org/10.24963/ijcai.2018/779

A The quantification of crime harm

To quantify the harm the police prevent or mitigate by fighting crime, we first quantify the harm caused by crime. In this section, we will start by considering the possibility of quantifying harm caused by crime in a very general way and we will make our approach increasingly more specific. We assume that we can decompose crime (past, present, future or even hypothetical) into "criminal events", or simply crimes, denoted by c. Such a criminal event could for example be a specific instance of a burglary including all details. In case of a complex criminal operation the events c can also be specific actions in that operation. For example the transportation of a batch of illegal substances as part of a larger criminal operation.

We can represent a collection of crimes using a set C. Quantifying harm would now come down to finding a function H, whose input is a collection of crimes and whose output is a real number representing the associated harm. Here a harm of 0 is to be interpreted as no harm being inflicted and a negative harm is to be interpreted as a benefit.

It is important to consider that the harm associated to a collection of crimes C could in principle depend on crimes that occur or have occurred but are not included in C. For example, the harm associated with one or more instances of money laundering might depend on how that money is used in the future. So to properly define H, we will assume that $H(C)$ represents the total harm caused by crime in the (hypothetical) situation that C consists of all criminal events that have taken place and will take place.

Next we consider what properties we want H to possess. If there are no criminal events, it makes sense to assume no harm is done by these criminal events. Therefore we should have that $H(\emptyset) = 0$. Assuming that crime is never beneficial and that more crime never causes less harm than less crime, we should also have that $H(C_1) \geq H(C_2)$ whenever $C_2 \subseteq C_1$. In particular this implies that $H(C) \geq 0$ for any crime collection C.

Given a collection of crimes C and a criminal event $c \notin C$ we can define the function h by $h(C, c) = H(C \cup \{c\}) - H(C)$. The interpretation of this is that $h(C, c)$ is the additional harm caused by the crime c if it occurs on top of a collection of crimes C. In other words, if the crime c is prevented and C is all remaining crime, then $h(C, c)$ is the amount of harm that has been prevented. If c_1 and c_2 are two distinct criminal events and C is a collection of crimes distinct from c_1 and c_2, we can consider the harm prevented by preventing both c_1 and c_2. This is given by

$$\begin{aligned} &H(C \cup \{c_1, c_2\}) - H(C) = \\ &H(C \cup \{c_1, c_2\}) - H(C \cup \{c_1\}) + H(C \cup \{c_1\}) - H(C) = \\ &h(C \cup \{c_1\}, c_2) + h(C, c_1). \end{aligned} \tag{29}$$

Note that this is not necessarily equal to $h(C, c_2) + h(C, c_2)$. A similar calculation also shows that

$$H(C \cup \{c_1, c_2\}) - H(C) = h(C \cup \{c_2\}, c_1) + h(C, c_2). \tag{30}$$

In this situation, how much reduction in harm should be attributed to preventing c_1 and how much to preventing c_2? Unless we assume that $h(C, c_1) = h(C \cup \{c_2\}, c_1)$ (or equivalently $h(C \cup \{c_1\}, c_2) = h(C, c_2)$), it turns out that if multiple crimes are prevented it does not make sense to consider the harm prevented by preventing each individual crime. Moreover, when considering the harm prevented by preventing a single crime it would be necessary to specify all other crimes that have occurred.

Acts of crime do not exist in isolation, so these properties might capture some of the nuances of crime harm. Unfortunately, this level of generality has some issues. First of all, without additional assumptions, the task of constructing the function H is very daunting. Moreover, using a function H with this level of generality in practice seems cumbersome for two reasons. Firstly, assessing the benefit of preventing a single crime might require knowing the complete state of crime. Secondly when preventing multiple crimes, it might not be possible to assign a benefit to the prevention of each individual crime.

For these reasons we make the simplifying assumption that $h(C, c)$ does not depend on C (as long as $c \notin C$) and from now on we will simply write $h(c)$. Because $H(\emptyset) = 0$ this implies that

$$h(c) = h(\emptyset, c) = H(\{c\}) - H(\emptyset) = H(\{c\}) \tag{31}$$

for all c. This means we can now interpret $h(c)$ as the harm caused by criminal event c when considered in isolation. Furthermore, we have that

$$H(C \cup \{c\}) = H(C) + H(C \cup \{c\}) - H(C) = H(C) + h(c) \tag{32}$$

for all $c \notin C$. If we assume that the collections of criminal events C are always finite, which seems reasonable, it follows by induction that

$$H(C) = \sum_{c \in C} h(c) = \sum_{c \in C} H(\{c\}). \tag{33}$$

So under the assumptions we made, the total harm of a collection of crimes can be obtained by summing the harms of each crime considered in isolation. This is the approach we have decided to follow.

It is worth briefly discussing if the consideration of crimes in isolation makes sense in practice. For a given criminal event c, it seems plausible that c is not directly related to most other criminal events. For events that are directly related to c it may or may not be possible to attribute the total harm of these events to the individual events. If this is possible, there is no issue with our assumption. In case it is not possible, it might be possible to group several events together and view them as a single larger event in such a way that attribution of harm to individual events is possible now. For this reason, we believe that the consideration of crimes in isolation, while perhaps not fully accurate, could still be applied in practice.

In practice, assessing the harm caused by each individual criminal event can be a difficult task, especially when a large number of events should be considered. In this case, it might be desirable to not consider the harm $h(c)$ of individual events c, but instead the average or typical harm caused by events similar to c.

Suppose we have several categories of crimes $C_1, \ldots, C_N$ such that each crime falls in exactly one category. If h_i is the typical harm of a crime in category C_i and C contains n_i crimes in category C_i, then $H(C)$ can be approximated as follows:

$$H(C) \approx \sum_{i=1}^{N} n_i h_i. \tag{34}$$

The idea is that each C_i is a collection of similar crimes. Ideally the crimes in each C_i are similar both in nature and the amount of harm they cause. By "similar in nature" we mean here that given a particular crime, it should be easy to assign it to a category C_i and (almost) all people will assign it to the same category.

In practice, the categories can be based on existing categories used in law enforcement or the judicial system. The upside to using these categories is that people are already somewhat familiar with them and it is clear for most crimes in what category they belong. For some categories there might be a relatively large discrepancy in harm between different crimes in that category. Another downside is that these categories are likely not exhaustive. Crimes that do not fit into any existing category can be included in a category for "other crimes", but this category might contain crimes with a wide range of harms.

B Proof of extension of Von-Neumann-Morgenstern utility functions

Let $\mathcal{C}$ be any set of outcomes. Let $\mathcal{C}'$ be the set of lotteries over outcomes in $\mathcal{C}$. We can identify with each element $L \in \mathcal{C}'$, a function $p_L : \mathcal{C} \to [0,1]$ that maps each element of $\mathcal{C}$ to its probability of occurring in lottery L. Note that since L is a lottery between a finite number of outcomes, we have that $p_L(c) = 0$ for all but finitely many $c \in \mathcal{C}$, even though $\mathcal{C}$ itself could be infinite. Also note that $\sum_{c \in \mathcal{C}} p_L(c) = 1$. This sum can be calculated as a finite sum because $p_L(c) = 0$ for all but finitely many c. Let $u : \mathcal{C} \to \mathbb{R}$ be a utility function that represents the preferences of some person regarding the outcomes in $\mathcal{C}$. We consider the function $E.(u) : \mathcal{C}' \to \mathbb{R}$ defined by

$$E_L(u) = \sum_{c \in \mathcal{C}} p_L(c) u(c). \tag{35}$$

So $E_L(u)$ is the expected value of u for the lottery L. Note that because $p_L(c) = 0$ for all but finitely many c, the sum defining $E_L(u)$ can be calculated as a finite sum. For $c_0 \in C$, let L_{c_0} be the lottery with guaranteed outcome c_0. So $p_{L_{c_0}}(c_0) = 1$ and $p_{L_{c_0}}(c) = 0$ for $c \neq c_0$. Note that

$$E_{L_{c_0}}(u) = \sum_{c \in \mathcal{C}} p_{L_{c_0}}(c) u(c) = u(c_0). \tag{36}$$

If we identify $c_0 \in \mathcal{C}$ with the lottery L_{c_0}, we can view $\mathcal{C}$ as a subset of $\mathcal{C}'$. The above equation then shows that $E.(u)$ is an extension of u to $\mathcal{C}'$. By definition of Von-Neumann-Morgenstern utility functions, u is a Von-Neumann-Morgenstern utility function if and only if $E.(u)$ is a utility function representing the person's preferences over outcomes in $\mathcal{C}'$. We will prove that, under reasonable assumptions, $E.(u)$ is also a Von-Neumann-Morgenstern utility function for this person.

We consider $\mathcal{C}''$, the set of lotteries with outcomes in $\mathcal{C}'$. As before with elements of $\mathcal{C}'$, an element $M \in \mathcal{C}''$ can be identified with a function $q_M : \mathcal{C}' \to [0,1]$ that maps each element of $\mathcal{C}'$ to its probability of occurring in lottery M. Similar as before, $q_M(L) = 0$ for all but finitely many $L \in \mathcal{C}'$ and $\sum_{L \in \mathcal{C}'} q_M(L) = 1$. We also consider the function $E.(E.(u)) : \mathcal{C}'' \to \mathbb{R}$ defined by

$$E_M(E.(u)) = \sum_{L \in \mathcal{C}'} q_M(L) E_L(u), \tag{37}$$

which is the expected value of $E.(u)$ for the lottery M. As before, we identify $L_0 \in \mathcal{C}'$ with the lottery M_{L_0}, that has L_0 as its guaranteed outcome. This way, we can view $\mathcal{C}'$ as a subset of $\mathcal{C}''$ and $E.(E.(u))$ as an extension of $E.(u)$. To show that $E.(u)$ is a Von-Neumann-Morgenstern utility function it suffices to show that $E.(E.(u))$ is a utility function that represents the person's preferences over outcomes in $\mathcal{C}''$. We will show this to be true under mild assumptions.

The elements of $\mathcal{C}''$ are lotteries with outcomes in $\mathcal{C}'$. These outcomes are in turn lotteries with outcomes in $\mathcal{C}$. As such, elements of $\mathcal{C}''$ could also be viewed as lotteries over outcomes in $\mathcal{C}$. We will now formalize this intuitive idea. Note that for $M \in \mathcal{C}''$, the probability of getting lottery $L \in \mathcal{C}'$ as an outcome in M and then outcome $c \in \mathcal{C}$ in lottery L is equal to $q_M(L)p_L(c)$. The total probability to obtain c as a final outcome is therefore equal to

$$\sum_{L\in\mathcal{C}'} q_M(L)p_L(c). \tag{38}$$

So it makes sense to consider the "composite" lottery M equivalent to the lottery in $\mathcal{C}'$ where outcome c occurs with probability $\sum_{L\in\mathcal{C}'} q_M(L)p_L(c)$. We will call this lottery $\phi(M)$. So $\phi(M) \in \mathcal{C}'$ is defined by $p_{\phi(M)}(c) = \sum_{L\in\mathcal{C}'} q_M(L)p_L(c)$. Note that $q_M(L) = 0$ for all but finitely many $L \in \mathcal{C}'$ and $p_L(c) = 0$ for all but finitely many $c \in \mathcal{C}$ for each $L \in \mathcal{C}'$. Therefore, there are only finitely many $c \in \mathcal{C}$ such that $p_L(c) \neq 0$ for at least one L such that $q_M(L) \neq 0$. As a result $p_{\phi(M)}(c) = 0$ for all but finitely many $c \in C$. Also note that

$$\sum_{c\in\mathcal{C}} p_{\phi(M)}(c) = \sum_{c\in\mathcal{C}}\sum_{L\in\mathcal{C}'} q_M(L)p_L(c) = \sum_{L\in\mathcal{C}'} q_M(L)\sum_{c\in\mathcal{C}} p_L(c) = \sum_{L\in\mathcal{C}'} q_M(L) = 1. \tag{39}$$

Here we were allowed to change the order of summation because all sums have finitely many non-zero terms. This shows that $\phi(M) \in \mathcal{C}'$ is well-defined. This defines a function $\phi : \mathcal{C}'' \to \mathcal{C}'$. We assume that the person's preferences over outcomes in $\mathcal{C}''$ are completely decided by considering the probabilities of the eventual outcomes in $\mathcal{C}$. In other words M is preferred over M' if and only if $\phi(M)$ is preferred over $\phi(M')$. In particular, if $\phi(M) = \phi(M')$, then the person is indifferent between M and M'. This assumption is made implicitly in [4]. Note that

$$\begin{aligned} E_M(E_\cdot(u)) &= \sum_{L\in\mathcal{C}'} q_M(L)E_L(u) \\ &= \sum_{L\in\mathcal{C}'} q_M(L)\sum_{c\in\mathcal{C}} p_L(c)u(c) \\ &= \sum_{c\in\mathcal{C}}\left(\sum_{L\in\mathcal{C}'} q_M(L)p_L(c)\right)u(c) \\ &= \sum_{c\in\mathcal{C}} p_{\phi(M)}(c)u(c) \\ &= E_{\phi(M)}(u) \end{aligned} \tag{40}$$

This shows that $E_M(E_\cdot(u)) > E_{M'}(E_\cdot(u))$ if and only if $E_{\phi(M)}(u) > E_{\phi(M')}(u)$ if and only if $\phi(M)$ is preferred over $\phi(M')$ if and only if M is preferred over M'. This shows that $E_\cdot(E_\cdot(u))$ is a utility function for this person, which completes the proof.

C Candidate Functions

The families of candidate functions for a and b are given below along with their first and second derivatives. The first candidate uses arctan functions:

$$f_1(w) = \begin{cases} \frac{c_2 \frac{2}{\pi} \arctan(c_1(w-w_0))+c_1}{c_1+c_2}, & w \geq w_0 \\ \frac{c_1 \frac{2}{\pi} \arctan(c_2(w-w_0))+c_1}{c_1+c_2}, & w \leq w_0 \end{cases} \tag{41}$$

with

$$f_1'(w) = \begin{cases} \frac{2}{\pi} \frac{c_2}{c_1+c_2} \frac{c_1}{1+c_1^2(w-w_0)^2}, & w \geq w_0 \\ \frac{2}{\pi} \frac{c_1}{c_1+c_2} \frac{c_2}{1+c_2^2(w-w_0)^2}, & w \leq w_0 \end{cases} \tag{42}$$

and

$$f_1''(w) = \begin{cases} -\frac{2}{\pi} \frac{c_2}{c_1+c_2} \frac{2c_1^3(w-w_0)}{(1+c_1^2(w-w_0)^2)^2}, & w \geq w_0 \\ -\frac{2}{\pi} \frac{c_1}{c_1+c_2} \frac{2c_2^3(w-w_0)}{(1+c_2^2(w-w_0)^2)^2}, & w \leq w_0 \end{cases}. \tag{43}$$

The second candidate uses exponential functions:

$$f_2(w) = \begin{cases} \frac{c_1+c_2(1-e^{-c_1(w-w_0)})}{c_1+c_2}, & w \geq w_0 \\ \frac{c_1 e^{c_2(w-w_0)}}{c_1+c_2}, & w \leq w_0 \end{cases} \tag{44}$$

with

$$f_2'(w) = \begin{cases} \frac{c_1 c_2}{c_1+c_2} e^{-c_1(w-w_0)}, & w \geq w_0 \\ \frac{c_1 c_2}{c_1+c_2} e^{c_2(w-w_0)}, & w \leq w_0 \end{cases} \tag{45}$$

and

$$f_2''(w) = \begin{cases} -\frac{c_1^2 c_2}{c_1+c_2} e^{-c_1(w-w_0)}, & w > w_0 \\ \frac{c_1 c_2^2}{c_1+c_2} e^{c_2(w-w_0)}, & w < w_0 \end{cases}. \tag{46}$$

The third candidate uses power functions with a negative exponent ($r > 0$):

$$f_3(w) = \begin{cases} \frac{c_1+c_2-c_2(1+c_1(w-w_0))^{-r}}{c_1+c_2}, & w \geq w_0 \\ \frac{c_1(1-c_2(w-w_0))^{-r}}{c_1+c_2}, & w \leq w_0 \end{cases} \tag{47}$$

with

$$f_3'(w) = \begin{cases} \frac{c_1 c_2}{c_1+c_2} r(1+c_1(w-w_0))^{-r-1}, & w \geq w_0 \\ \frac{c_1 c_2}{c_1+c_2} r(1-c_2(w-w_0))^{-r-1}, & w \leq w_0 \end{cases} \tag{48}$$

and

$$f_3''(w) = \begin{cases} -\frac{c_1^2 c_2}{c_1+c_2} r(r+1)(1+c_1(w-w_0))^{-r-2}, & w > w_0 \\ \frac{c_1 c_2^2}{c_1+c_2} r(r+1)(1-c_2(w-w_0))^{-r-2}, & w < w_0 \end{cases}. \tag{49}$$

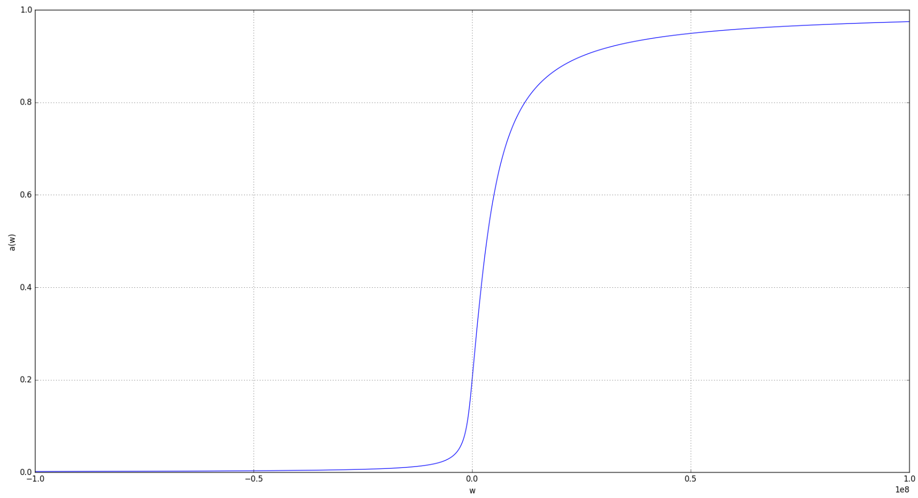


Figure 8: Example function for $a(w)$, where w is wealth in Euros. The plotted function is $f_1(w) = \begin{cases} \frac{c_2 \frac{2}{\pi} \arctan(c_1(w-w_0))+c_1}{c_1+c_2}, & w \geq w_0 \\ \frac{c_1 \frac{2}{\pi} \arctan(c_2(w-w_0))+c_1}{c_1+c_2}, & w \leq w_0 \end{cases}$ with $w_0 = 0$, $c_1 = 2 \cdot 10^{-7}$, $c_2 = 8 \cdot 10^{-7}$.

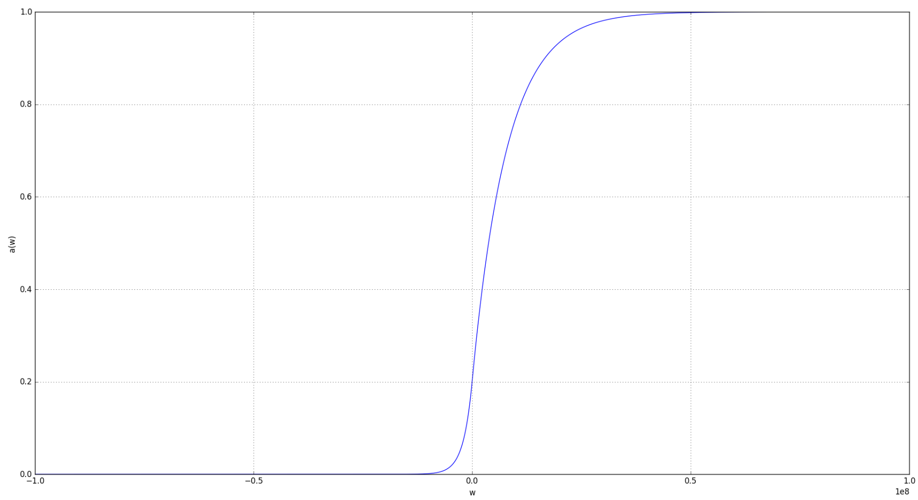


Figure 9: Example function for $a(w)$, where w is wealth in euros. The plotted function is $f_2(w) = \begin{cases} \frac{c_1+c_2(1-e^{-c_1(w-w_0)})}{c_1+c_2}, & w \geq w_0 \\ \frac{c_1 e^{c_2(w-w_0)}}{c_1+c_2}, & w \leq w_0 \end{cases}$ with $w_0 = 0$, $c_1 = 1.25 \cdot 10^{-7}$, $c_2 = 5 \cdot 10^{-7}$.

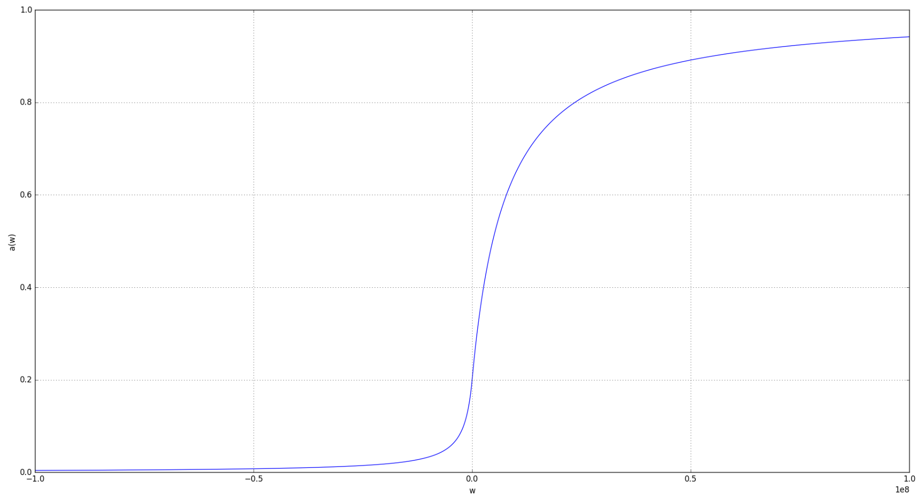


Figure 10: Example function for $a(w)$, where w is wealth in Euros. The plotted function is $f_3(w) = \begin{cases} \frac{c_1+c_2-c_2(1+c_1(w-w_0))^{-r}}{c_1+c_2}, & w \geq w_0 \\ \frac{c_1(1-c_2(w-w_0))^{-r}}{c_1+c_2}, & w \leq w_0 \end{cases}$ with $w_0 = 0$, $r = 1$, $c_1 = 1.275 \cdot 10^{-7}$, $c_2 = 5.1 \cdot 10^{-7}$.

D Average health and wealth data

Age	18-39	40-59	60-69	70+
Average quality of health	0.909	0.849	0.791	0.733

Table 1: Average quality of health values based on age. The values are taken from [23].

Age	25-	25-35	35-45	45-55	55-65	65-75	75-85	85+
Average net worth in €1000	4.3	12.5	93.4	184.9	242.8	272.4	221.9	197.7
Median net worth in €1000	0.2	-0.4	11.4	59.2	107	128.5	73.6	41.2

Table 2: Average and median net worth of households in 2012 by age of the primary breadwinner. The values are taken from [6]

E CHI Values

Type of Crime	**CCHI Score (Average Sentence)**
Abandoned accident site	208
Abuse	1601
Arson / Explosion	5601
Assault	1586
Bankruptcy	1601
Buying stolen property	1601
Carrying a false license plate	218
Coin crime	1769
Computer trespass	309
Crimes WvSr (other)	1146
Deprivation of liberty/hostage	5500
Destruction means of public transport	674
Destruction of a public building	856
Destruction to car	491
Destruction, damage (other)	755
Discrimination	390
Driving during a driving ban	71
Driving under the influence	309
Driving While Banned Driving	208
Drug offense (other)	856
Extortion and Threat	2148
Failure to follow official orders	71
False declaration	688
Firearms Crimes	1511
Forgery	856
Forgery in seals and marks	1221
Fornication with a minor	1236
Fornication with abuse of authority	1221
Fraud (other)	871
Hard drugs	2241
Human trafficking and 244 human smuggling	3243
Indecency	551
Joy riding	117
Life crime	5601
Local peace violation	71
Money laundering	1236
Overt violence against good	1221
Overt violence against person	2696
Pornography	536
Public authority offense (other)	1601
Public order offense (other)	938
Rape	2696
Rebellion	1146
Refusing breathalyzer/blood test, etc.	96

Residence of an unwanted foreigner	218
Scams	1236
Sexual offense (other)	1586
Soft drugs	755
Stalking	674
Theft and burglary with violence	3243
Theft and burglary without violence	856
Threat	263
Traffic offense (other / death involved)	1146
Trespassing	233
Violent crime (other)	2864

Table 3: CHI Values

F Alternative Color Figure

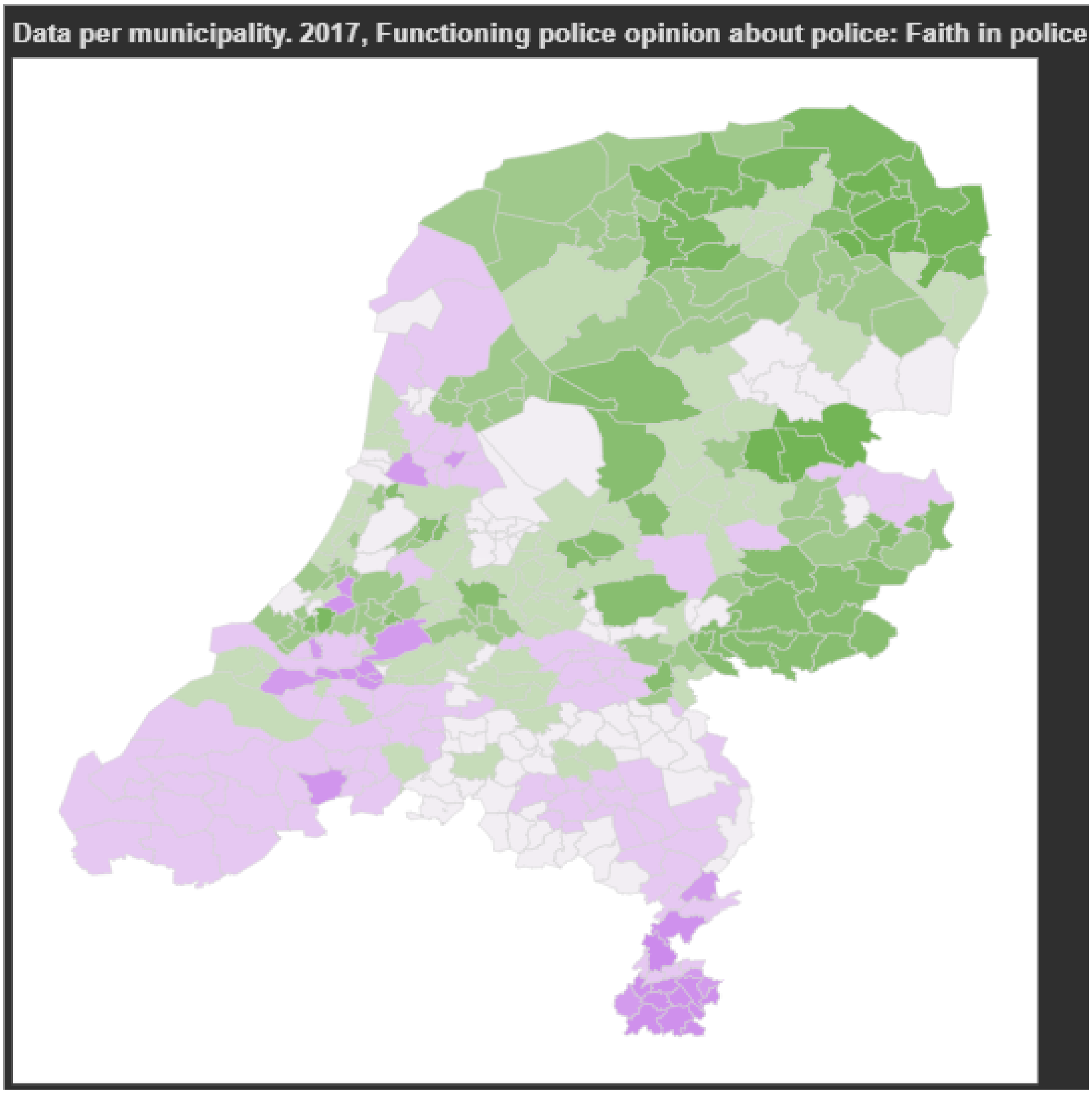


Figure 11: The score of the faith in the police per area in purple/green

G Principles for responsible use of AI

G.0.1 Accountability

In some respects, determining moral responsibility might be more difficult than determining legal accountability. No matter what technology (including AI) are engaged, it is (naturally) important that the police operate within the bounds of current legal regulations. Incentives and issues with work-ability must be taken into consideration if responsibility must be distributed within the police organization (for instance, if the program was created internally) [9].

G.0.2 Transparency

- People: Being open and honest about who is responsible for the financing, management, development, operation, and maintenance of the systems can help determine whether or not effective protocols were followed, who is responsible for what, and where biases or conflicts of interest may have arisen.
- Rationale: Using a high-tech system or process may be justified at the highest level of abstraction by being open and honest about it.
- Operation: In what ways does the system actually work? What must police officers do for the system to function properly? How does the use of the same affect other aspects of the police force or the work that they do?
- Data: One of the most crucial elements that affects how AI applications behave is data. This holds true for both types of data utilized to create these systems. What steps were taken in gathering, storing, cleaning, de-biasing, anonymizing, pre-processing, and/or presenting the data.
- Courts: In the courtroom, judges must have full access to all judgments and decision-making that are pertinent to the prosecution of criminals.
- Police: The police organization itself needs to have knowledge of the systems it is utilizing and creating. For most sorts of transparency, documentation ought to be available.
- Government: Additionally, other political parties may require more transparency as partners or watchdogs. The police force should be able to cooperate with other governmental organizations while simultaneously being responsive to the general population.
- Users: A system's users primarily need to understand how to use it, but they may also be interested in why, how, and how well it functions. Users may be police officers, but in systems that are intended for the general public, they may also be civilians.

G.0.3 Fairness and Inclusivity

The treatment of persons fairly and equally by the police is crucial in the pursuit of justice. For example, ethnic profiling is a major concern that needs to be avoided. However, bias is difficult to overcome and is present in all human judgment. A young guy is statistically more capable of committing bodily injury than an elderly woman, for example, making some bias and differential treatment acceptable.

G.0.4 Socio-Technical Robustness and Safety

- Performance – Depending on the particular application, effectiveness can take many different forms. A system's accuracy refers to its capacity to render accurate decisions, judgments, or forecasts.
- Robustness – Testing for robustness is a little trickier than accuracy testing, even though accuracy testing is a standard component of the development process. Robustness refers to how well a system can handle unique circumstances and deviations from the presumptions upon which it was based.
- Safety – The operators' and the impacted parties' safety should always come first. When a system is safe, it will function as intended without endangering people, property, or the environment. Higher precision and resilience, as well as avoiding unanticipated and unintended repercussions and faults in the system's operation, all contribute to this.