Ethics of Algorithms

Philosophy of Science, Technology, and Society
University of Twente

Thijs Slot

August 29, 2016

Graduation Committee:
Dr. David Michael Douglas
Dr. Johnny Hartz Søraker

ABSTRACT

Present day society is heavily shaped and influenced by the workings of algorithms on multiple levels. Near-ubiquitous in any data-processing operation, algorithms play an important role in diverse topics spanning virtually everything that is happening on the internet and in wider computer sciences, governmental data collection and processing, financial markets, health-care and medicine development, educational settings, etc. A large number of activities taking place in these spheres would be impossible without the use of algorithms, and parts of them would never have been developed without the increase in automation, efficiency and scalability algorithms offer. At the same time, a tension has arisen between their prevalence and the level of understanding of their functioning. This thesis aims to provide insight into the effects of algorithmic applications, and unpack the 'black box' of their operations by questioning their status as neutral tools or technologies. This approach will be taken specifically by focusing on the ethical implications of algorithms, which will be evaluated along two lines that correspond to the ethical traditions of consequentialism and deontology. The first of these two lines, consequentialism, will elucidate how current applications of algorithms — including the aforementioned technologies and fields which they propel — have negative consequences. In this section, a theoretical framework outlining the structural ways in which practical algorithmic functioning produces adverse effects will be introduced, after which a number of examples will be given to illustrate the current state of affairs and concomitant real ethical issues. After this, a number of ways to counter these dynamics will be provided, notably geared towards measures to increase awareness, accountability, and balancing power-relations. The second line of exploring the ethical dimension of algorithms, deontology, will move the analysis away from practical concerns, towards evaluating the choice for accepting the required conditions for algorithmic processing fundamentally. That is, in order for algorithms to be capable of functioning properly, a certain lens or epistemological stance needs to be adopted, which in itself can be ethically contestable. The relevant conditions will be stated, constituting the theoretical component, after which again a number of examples will be introduced to illustrate this dimension's pertinence. With this, the main aim of informing continued discussion on multiple levels to address the raised issues is achieved.

ACKNOWLEDGEMENTS

I would like to thank first and foremost my supervisor dr. David Douglas, who has helped me with his knowledge, patience, and continued support. It has been a very pleasant experience to receive this level of encouragement, confidence, and flexibility, which has allowed me to pursue the project in a pleasant way.

I would also like to thank dr. Johnny Søraker for his accommodation and trust, contributing to the completion of this thesis.

Furthermore I thank my family for their unquestioning moral support.

Finally I would like to thank all my friends, and all others who have allowed me to bore them with details. A special mentioning is reserved for my fellow PSTS-students Chirag Arora, Giannis Marinakis, and Jurjen Idskes, for the smiles and experiences we have shared.

CONTENTS

INTRODUCTION

Present day society is heavily shaped and influenced by the workings of algorithms on multiple levels. Near-ubiquitous in any data-processing operation, algorithms play an important role in diverse topics ranging from virtually everything that is happening on the internet and in wider computer sciences [9], to health-care and medicine development [71], traffic (control) (cf. [60] [104] [126] [38]), meteorological models [98], financial markets [76] [77], educational settings [105] [68] etc. A large number of activities taking place in these spheres would simply be impossible without algorithms, and parts of them would never even have been developed to begin with. By extension, many of the artefacts we interact with and through, such as mobile (smart) phones and computers, rely on the workings of aggregates of algorithms in a variety of ways, generating large amounts of data that are in turn algorithmically processed. The development of these algorithms and their applications has opened up many new possibilities through increasing efficiency, ease of use, and the creation and analysis of large datasets. Increasingly, algorithms have also come to influence and shape the way we inform ourselves [73], and in turn how to evaluate the processes that themselves are reliant on algorithmic processes [50]. Algorithmically sorted search results and complex statistical models[1] are two prime examples of how not only parts of content itself are generated algorithmically, but its status and functioning are further ordered, analyzed and evaluated by algorithms. It is not a coincidence that such widely varying fields rely on algorithms, since algorithms are a vital building block required for functioning. It is not a matter of (arbitrary) choice to use algorithms, as there is no viable alternative in terms of technology to perform the same functions in a comparably efficient manner. Algorithms, in short,

At the same, a tension has arisen between their prevalence, and the understanding of their functioning. Algorithms have developed into a sort of "myth" [11] [128], with being called a "new kind of object, intermediator, gate-keeper and more..."[48], serve as "pathways through which capitalist power works" [85], "play a critical role in producing and curating our communications and shared culture" [109], and stand for how "rules of rationality replaced the self-critical judgments of reason". The gap between these two realities arises in no small part from the fact that — their pervasiveness notwithstanding — the question of what an algorithm precisely is has no trivial

1 Which inform, e.g. "...education, markets, political campaigns, urban planning, welfare, and public safety." [128]

answer. Donald Knuth, widely regarded an expert in the field of algorithms [127] [32] [92] remarked that, when trying to offer a proper description of algorithms, "Of course if I am pinned down and asked to explain more precisely what I mean... I am forced to admit that I don't know any way to define any particular algorithm except in a programming language". Others, such as Gurevich [59], have even argued that because the notion of 'algorithm' is expanding, any hard definition will necessarily rapidly find itself outdated. Despite these difficulties, there are general characteristics both to algorithms themselves, and to the range of applications they are suitable for. In their most basic form, algorithms are self-contained sets of defined operations to be carried out. This entails that, in each form, they are involved in a relation of *translation*, where input and output differ as per the algorithm's specifics. Algorithms, however, not only serve to translate and produce data, they also *necessarily* come with a certain logic, a lens through which to view, structure and approach any given situation, because of the operations that are (to be) carried out. Whereas in some cases this results in rather unambiguous benefits such as in the case of faster medical diagnoses [71], adverse effects can creep in. For example, to stay in the medical setting, it has been shown that researchers can have solid, rational reasons for designing different algorithms for the same purpose of image analysis, depending on pre-existing values [81]. In this way, algorithms codify such pre-existing values and lend it a sheen of rationality and neutrality. Thus, we see that the mechanisms of translation of data and delegation of responsibility, coupled with the lack of clarity in what algorithms exactly are, pose serious challenges to contemporary scholars [128], notably Work on tackling these challenges has been undertaken from a number of different angles, where [50] [5] have taken a sociological approach; [22] [58] have looked at legal impacts; others (cf. [109] [94] [5] etc.) have focused specifically on the ethics of algorithmic functioning. This thesis builds on this existing body of literature, and aims to combine and deepen the analyses. Specifically, the goal is to explore two different ways of looking at algorithms, (a) first, practically, assessing its current functioning from a *consequentialist* normative ethical viewpoint, and (b) fundamentally, moving away from the practical critique of consequentialist approach and analyzing the ethical component of algorithms from a *deontological* perspective. Hence, the focus will first be on looking at algorithms as a technology and artefact: specifically, how they are currently employed and being developed, and what the drawbacks of this are. This is not to neglect or deny the clear benefits that accompany algorithmic applications, but serves to clarify the ethical dimension by unpacking algorithms as value-laden instruments with potential adverse effects. Here, the concept of dual-use technology will be instrumental in capturing both the positive and negative potential of algorithmic functioning, given

the current consequences. The second dimension this thesis will develop is to look at algorithms as a mode of thinking and structuring, or as an epistemic technology. Here basic tenets of algorithms will be presented, as well as the limits of algorithmically approaching any situation, that is, the desirability to impose the necessary conditions for an algorithmic approach in any given instance will be under scrutiny. Through the combination of these themes, this thesis aims to contribute to an understanding of how the "algorithmic turn" [94], or "algorithmic culture" [112] can be understood from an ethical perspective, and offer recommendations on both how to frame the developments practically, as well as fundamentally. In the final section, after having established that algorithms *do* in various ways produce ethically debatable issues, a short discussion of a number of promising directions for future research will be given. This includes whether or not algorithms are as a consequence *moral agents*. The answer to that question depends on the conceptualization of the relation between technology/artefacts and humans. While it is beyond the scope of this thesis to address this question completely, the given overview serves to show that the dynamics analyzed in this thesis can be applied in a discussion on such topics, regardless of which stance is adopted. Moreover, the idea of *stigmergy*, a notion from ecology and biology, will be discussed to see how recommender-algorithms can be understood by introducing methods from other disciplines. Finally, a short look at how transformation in the labor market are expected due to the increasing implementation of algorithms replacing human (cognitive) labor.

0.1 RESEARCH QUESTION AND GOALS

As clarified in the previous sections, there are different levels at which algorithms have impacts that are worthy of examination. The issues that this thesis will be dealing with can be formulated in the following main research question:

Research Question. *What are the ethical implications in practical and fundamental terms of the use of algorithms?*

The following sub-questions are important to fully answer this question:

1. What are algorithms and how do they function?
2. What are current applications of algorithms and what are the ethical *consequences* of these?
3. What are the effects of electing an algorithmic approach, regardless of practical implementation?

0.2 CHAPTER DIVISION

The thesis will begin by introducing the reader to what algorithms are in a number of complementary ways. Having established a foundation, current applications of algorithms will be assessed using a consequentialist approach. Following this, a deontological analysis of the theoretical possibilities of algorithms will be undertaken. Finally, a number of recommendations for future research will be given, flowing out of the preceding chapter. In the overview below, the numbers in parentheses indicate their relevance to the corresponding sub-questions of the previous paragraph.

CHAPTER ONE: WHAT ALGORITHMS ARE (1,2,3) This chapter will consist of a number of different ways to view algorithms, which combined are sufficient for reaching sensible answers to all the research (sub-)questions stated above. The methodological choice for introducing multiple different views over *one* (working) definition has been made because the latter would be too confining, whereas the different approaches are chosen specifically to facilitate subsequent analysis. The chapter will cover algorithms (a) historically, (b) through natural language descriptions, and (c) from the perspective of the philosophy of computing. Note that the emphasis is on algorithms themselves, and the ethical evaluation of algorithms in their relation to people will feature in the following chapters.

CHAPTER TWO: APPLICATIONS AND IMPLICATIONS (2) Having outlined what algorithms are, the focus will be on how algorithms are currently functioning in society. A number of different ways in which algorithms have an impact will be discussed, and an ethical *consequentialist* framework will be introduced to assess these impacts. It will be concluded that algorithms are a dual-use technology, with clear benefits in some areas, and problematic aspects in others. To remedy some of these problems, an *informed* reflection on the unintended consequences of algorithms is urged.

CHAPTER THREE: THE ALGORITHMIC CHOICE (3) Whereas Chapter Two focused on current application of algorithms, this chapter will focus on what kind of dynamics and translation algorithms *fundamentally* involve. Where all the ethical consequences in the previous chapter could be called a practical critique, to be mitigated with an improvement of the algorithm or the structure it is embedded in, this chapter will argue that, *deontologically*, there are arguments against the usage of algorithms regardless of how well it would be functioning. This argument results in the conclusion that certain tasks and areas are fundamentally unsuited for algorithmic approaches, inviting a discussion on possible alternatives.

CHAPTER FOUR: ALTERNATIVE AND FUTURE APPROACHES Having established that algorithms are currently having a major impact, and can be problematic on both the consequentialist and deontological front, this chapter will look at a number of possible avenues for future research. This will not be an attempt at being exhaustive, but rather to give the reader an impression of possibilities, as well as highlighting awareness of other approaches,accompanied by a short rationale for leaving out in this thesis. The examples here will be *technological mediation theory*, *stigmergy* and *transformations of the labor market*.

CONCLUDING REMARKS This section will summarize the main conclusions, and reiterate the answers to the research question and accompanying sub-questions.

0.3 JUSTIFICATION

The contribution of this thesis comprises of two main elements. First, to add to the existing and growing body of literature on the ethics of algorithms. On this front, the contribution is made by analyzing the *multiple different ways* in which algorithms can be claimed to have ethically debatable impacts, instead of focusing on one type only, as is often done in the literature to date. Moreover, whereas most current literature does explicitly indicate that certain dynamics brought about by algorithms have an ethical impact, this is rarely tied to any specific normative ethical position. The second contribution is by extending the analysis to the point that there are intrinsic properties of algorithms *theoretically* that are ethically contestable. To the knowledge of the author, an ethical analysis of algorithms has not been carried out without reference to functionalist arguments, that is, by pointing out that it is not bad coding, unintended consequences etc., but rather a fundamental property of algorithms which makes them unsuited for certain situations. This, then, calls for reflection on the extent to which algorithms inescapably bring about effects that outweigh their uses.

The rationale for exploring this issue in the manner proposed can be found along two lines: (a) philosophically, and (b) through science and technology studies. Furthermore, the choice for the two divergent ethical frameworks has been made to highlight the different ways in which algorithms can be argued to call for ethical deliberation. See more on this in the section on methodological justification below.

PHILOSOPHICAL JUSTIFICATION Given the large impact of algorithms on society, and the projected increase in these effects, it has become a serious concern for the (social) sciences to make sense of

their functioning. This thesis aims to contribute to an understanding of specifically how the dynamics of translation — making everyday phenomena into quantifiable input and output for and by algorithms — coupled with the adoption of a specific lens — a way to view the world that the algorithm produces and sustains — are playing out, with special reference to ethical implications. Through this focus, the aim is to develop insights that clarify and structure normative positions to be taken with regards to algorithms' present functioning, intrinsic qualities, and future developments. Algorithms have an effect on epistemology insofar as they delineate ways of thinking about subjects. Through this delineating, then, we see the algorithmic trap where we find a situation where algorithms are shaping parts of our lives in potentially controversial ways, but the ethical dimension through which to evaluate these effects are in turn informed by algorithms. This requires an analysis that looks specifically at the functioning of algorithms, and how to evaluate this functioning without resorting to its logic.

SCIENTIFIC AND TECHNICAL JUSTIFICATION Algorithms are extremely prevalent technologies, underpinning many of everyday's practices. The reliance on their functioning is large, and expected to grow. Given this reality, the black box of algorithms [128] [101], paired with the question of what in fact an algorithm is, makes for a case of relevance for this thesis. Critically, many developments in science and engineering, such as Autonomous Vehicles [19], as well as also quantum-experiments [82] are informed by the functioning of algorithms. Thus, more than it simply being an active technology, it is also embedded in the way that science and technology are being developed and conceptualized, and are an explicit topic *of* research, as well as a tool *in* research.

METHODOLOGICAL JUSTIFICATION An ethical analysis of algorithms as proposed in this thesis concerns a number of core issues. First of all, realizing that the choice for *which* ethical viewpoint to focus on has consequences for *what* kind of elements and dynamics will be highlighted. The explicit choice for *consequentialism* to assess current practices has been made because it represents a broad category of ethical viewpoints placing emphasis on the normative properties of actions and their consequences. Moreover, because there are many varieties of consequentialism, adopting this view allows for multiple interpretations and integration in future studies, making the analysis and its conclusions potentially more useful. The second choice of ethical theory has been *deontology*, to study what adopting an algorithmic lens or logic entails *in principle*. Deontology is often considered a complementary ethical tradition to consequentialism, and places emphasis rather on adherence to rules and duties. This has the potential

to bring out the dynamics of algorithms regardless of how they are functioning practically, looking at the theoretical ramifications of the choice for using algorithms. Finally, with regards to moral status, the choice has been made to leave the question — notably brought to the fore by *technical mediation theory* [123] [121] [103] — of where agency resides, and hence who moral agents are, open to further debate. Extreme positions in this field are either that (a) artefacts (technologies, in this case algorithms) are neutral tools, and cannot be claimed to have agency (weak view), or conversely, (b) that artefacts actively co-shape people's being and behavior and have intentionality (strong view). This is a very interesting and relevant debate, but for the goals of this thesis, it will not be addressed directly. Instead, a middle-ground position will be taken, with a small number of comments on how the analysis could change given this debate in the section on alternative approaches (chapter 4).

1

WHAT ALGORITHMS ARE

In order to be able to analyze the ethics of the functioning of algorithms both in practical terms as well as fundamentally, it is important to understand what, in this thesis, is meant by an *algorithm* to begin with, i.e. an ethical evaluation of algorithms hinges on a proper understanding of the terms involved. While algorithms are ubiquitous both in terms of their application as well as in narratives surrounding these applications, texts dealing with these topics often leave the question of *what algorithms are* unanswered [128] [127] [92]. Even in the cases where the issue is addressed, different contexts often yield different conceptualizations, suited for that specific purpose only. Moreover, the task of conceptualizing algorithms is far from trivial, because of their multifaceted nature, as well as that the notion is still developing and expanding [59]. Yet in order to develop an overview of normative positions one can take on algorithms in their current functioning, and as epistemic technologies fundamentally, it is crucial to understand *what* algorithms are, and what they can and cannot do. Therefore, instead of providing any specific, confining definition, number of different ways to view algorithms will be introduced here, that combined provide a secure basis for the analyses required for answering the research questions presented in the previous chapter. These different, complementary views will be, in turn:

1. **Historical** — Exploring its roots and seeing that its present-day usage has not emerged in a vacuum.

2. **Natural language description** — What can be said about algorithms in terms of how they function.

3. **Philosophical attributes** — As a technology and method, what principles are associated with its functioning.

With this, the explicit methodological choice has been made to avoid attempting to artificially create clarity, creating an illusion of certainty that does not exist, as well as being needlessly narrowing. It is instructive to realize that while this approach necessarily leaves out certain aspects pertaining algorithms, yet it offers sufficient information for reaching the intended goals. This method allows for the goals to be

supported by the conceptualization, and, in turn, lets this conceptualization derive its legitimacy from the intended goals as presented in the previous chapter. This choice follows [11], where it is mused that "[R]ather than decry the ambiguity around algorithms as a term of art, perhaps we should embrace it." In sum, this chapter will focus mostly on algorithms themselves, where subsequent chapters put the dynamics of algorithmic functioning in relation to people. This entails that the actual ethical evaluation will not take place in this chapter, but rather that this chapter serves to facilitate an informed assessment. Thus, of the two axes that this thesis is composed of — *possibility* and *desirability* — this chapter will be covering the first, enabling the work of interpretation and evaluation in subsequent sections. This chapter will contain very few examples, as most noteworthy examples will be featuring in the sections where their relevance will be fit into the ethical assessment.

1.1 A BRIEF HISTORY OF THE CONCEPT OF ALGORITHMS

The word 'algorithm' is, etymologically speaking, linked to *Abū Jafar Muhammad ibn Mūsā al-Khwārizmī*, a mathematician who wrote the earliest known book on algebra in the 9th century AD [24]. Although he is often absent in western history books [6], the influence of al-Khwārizmī on the development of mathematics is profound, introducing not only the words for algebra, but also developing a large number of notation-forms still in use today, as well as popularising the adoption of so-called Arabic numerals [6] [112]. The final part of his name, *al-Khwārizmī* — which translates to *from Khwarezm*, a region east of the Caspian Sea in central Asia — has undergone a "mangled transliteration" [112] to form the current Latinized word *algorithm*. Variations of the word and concept appear in the English language from circa the 18th[1] century, and it was only in the 20th century that *algorithm* became the dominant form to denote a set of procedures. The entering of 'algorithm' in the orthography used today is, however, historically far from obvious, as originally the word *algorism* denoted the same phenomenon, and was far more used. As al-Khwārizmī's work was translated in the 12th century to European languages, it introduced novel methods of approaching problems of arithmetic, gradually replacing methods relying on a counting table or abacus. The type of algorithm that al-Khwārizmī describes is for simple computations, for example for finding the value of x in equations such as $x^2 + 10x = 39$. Formally, this is called a *classical sequential algorithm*, which was "the only algorithm in use from antiquity to the $1950s$" [59], and in an elementary form part of everyday life. It is, put simply, a plan involving a number of steps, a way of getting

1 Though Striphas [112] notes that the earliest known appearance can be found in Chaucer's *Canterbury Tales*, in the form of *augrim*.

from a to b, or from input to output. Thus, while algorithms can denote simple processes having taken place long before its formulation or etymological roots, the reflection resulting in these mathematical texts explicitly describing the process in abstract form is a milestone in the development of formal mathematics.

The influence of the notion, as foreshadowed by the previous quote, has long remained constant, illustrated by the fact that Knuth [80] has remarked that "[b]y 1950, the word algorithm was most frequently associated with 'Euclid's algorithm"', a method for finding the greatest common divisor of integers. This changed rapidly, however, with a number of fundamental breakthroughs in the domain of theory of computation, and the advent of electronic computer chips, capable of performing specific tasks with immensely increased efficiency compared to humans[2]. The appliance of their combined new possibilities spurred the development of a flood of different functions, going far beyond simple sequential tasks for mechanically — that is, without variation or changing of conditions — solving a mathematical problem. Algorithms are currently properly considered in the narrowest sense a *technology* [3], with others expanding their view to "a particular form of decision-making... or an epistemology onto itself. Still others take a more expansive view, conceiving of algorithms as a particular form of rationality, symptomatic of a general mode of social ordering. And then there are those who see algorithms as a sociotechnical process." [11] This thesis will use elements of each interpretation to bolster the analysis, as each interpretation offers distinct insights. The development and embeddedness of algorithms, then, has certainly taken off since the 1950s, currently permeating all electronic technologies, as well as having revolutionized ways thinking about and structuring the world.

1.2 NATURAL LANGUAGE DESCRIPTIONS

The sheer success that algorithms have enjoyed in a wide variety of settings invites the question of what it is about them that is so useful. This section will attempt to describe algorithms, in turn (a) by giving two separate accounts characteristics of algorithms in the general sense; one classic and the other contemporary, and (b) by separat-

2 On this note, it should be clear that, following [48], it should be clear that this thesis is concerned with a subset of algorithms, namely those performed electronically by "computer chips". "Computer" is here to be taken liberally, and does not necessarily relate to PCs or similar artefacts. The underlying dynamics and methods are not different, but these same processes carried out by humans are not the scope of this thesis.

3 One simple example of how it works as a technology comes from the fact that has been shown that running an efficient algorithm on outdated hardware yields better results than running less efficient algorithms on newer hardware. That is, their technological, computational artefactual existence is felt in the way they steer processes they are a (constituent) part of.

ing the term from the similar notion of *programs*. The combination of these elements will serve to clarify what kind of processes are involved with algorithms, and how they relate to other parts of the technical processes involved. The reason for giving two separate natural language accounts of algorithms is that this highlights the development of algorithms in the 20th century, as well as that contrasting the approaches helps focus on why the more contemporary account was required.

KNUTH'S CHARACTERISTICS Regarded as groundbreaking in his algorithmic analysis in the late 1960s, Donald Knuth's seminal work on algorithms provides a good starting point for the current purposes. Algorithms are, in his words "a finite set of rules which gives a sequence of operations for solving a specific type of problem" [80]. He is quick to note, however, that algorithms differ from related words such as "*recipe, process, method, technique, procedure, routine*" [80] because of five basic characteristics. What follows here is a brief description of these five features:

1. **Finiteness** — An algorithm must terminate, i.e. stop operating, after a finite number of steps, regardless of the (accepted) input. This finiteness is a theoretical condition, as the number of steps is only bounded by finiteness in an abstract sense, meaning that the number of steps can be arbitrarily large, as long as it does not exceed finiteness. Put simply, given a theoretically long amount of time, the algorithm is required to terminate because it reached its final step. This feature is, by Knuth's admission [80] not restrictive enough for practical purposes.

2. **Definiteness** — Each step of the algorithm's functioning must be 'well-defined', i.e. not containing any possible ambiguity. This feature sets it apart from the related words mentioned above, and also implies that algorithms cannot be properly consisting of any natural language, or in fact any language without a formal, context-free grammar which prescribes and constrains all relations between the possible sets of strings of elements in it. This hearkens back to the earlier quote of Knuth, stating he knew of no way to define algorithms without resorting to formal languages. Here, we learn, that in his view this is not only not possible because he does not know any formulation, but rather that any definition given outside of formal languages is logically inconsistent.

3. **Input** — An algorithm requires zero or more inputs, taken from specific, defined sets. That is, the algorithm functions only when it is provided a quantity (data)of a given order, prior to its first step.

The input needs to come from a defined set of acceptable inputs of a *quantifiable* degree, that is, unambiguous[4] by nature.

4. **Output** — The algorithm produces one or more outputs. These outputs are related to the input, via the (finite, well-defined) steps.

5. **Effectiveness** — This feature refers to the level of complexity involved in each step. In Knuth's words this would mean that each step should be trivial enough that a reasonably intelligent person should be able to carry out that step given pencil and paper. This also restricts the sort of steps that are allowed: Nonsensical or logically inconsistent steps are not part of a proper algorithm.

The combination of these features rules out any of the above, related words, and justifies the separate category of algorithms. A simple example of an algorithm (necessarily) satisfying these conditions is MERGESORT. This algorithm is used for internally sorting a finite (set of) arrays, roughly through the following steps [1], with n as the length of the array:

1. If n is 0 or 1, it is sorted and the algorithm can move to step 4. If not:

2. Divide the string into n sub-array each containing 1 element.

3. Repeatedly merge sub-array in ordered[5] manner until one sub-string of length n remains. This is the final sorted sub-string.

4. Terminate the algorithm.

This short example of MERGESORT exhibits *finiteness*: the number of steps is finitely dependent on the input of length n), *definiteness*: each step is well-defined and valid for each allowed *input*, and produces an *output* of equal length n. The *efficiency* of this algorithm is well-defined to be of the $O(n \log n)$ category [1]. Note, however, that this is very much a natural language approach, and not by any means a description fit for a (non-human) computer.

CONTEMPORARY CHARACTERISTICS Appealing as Knuth's list of five features may be, it is also in two ways outdated [32]. First, it is argued that it fails to outline a general definition of 'procedure' [32] [127], making it open to interpretation and unsuited for purposes of definition. In fact, [**Yanofsky**] has explicitly stated that Knuth simply replaced the ambiguous concept of *algorithm* with the ambiguous concept of *procedure*, doing little in the direction of clarifying. Secondly,

4 Unambiguous does not mean that the input cannot be open to multiple interpretations. If the latter is the case, this can still be unambiguously processed.

5 In the example depicted in Figure 1, the ordering is of positive integers from low to high. The ordering can in other arrays and contexts be set differently.

Knuth makes certain claims about algorithms that are no longer currently applicable, such as that an algorithm halts on every input[6]. Dean [32] attempts to rectify this by giving an alternative list of "common observations", which include the elements from Knuth, but are framed in a different, more contemporary-proof manner. In Dean's view, algorithms involve [32]:

1. **Mathematical procedures** — Algorithms can be thought of as procedures acting upon mathematical objects, in that it takes input of a certain quantifiable class, and returns an output. This is achieved through a set of instructions expressed as "imperative statements", which are repeatable and can yield different outcomes on different inputs. This set of instructions is carried out in a particular order (with possible feedback mechanisms), dependent on both the steps and the input itself.

2. **Mathematical problems** — Algorithms are virtually always introduced to serve a specific mathematical problem. Moreover, given that the input itself must be quantifiable and unambiguous, this restricts the type of problems amenable to algorithmic approaches.

3. **Finitiness** — Algorithms are distinguished from 'mathematical procedures' in the general sense because of a triple condition of finiteness. First, algorithms are bound by finite specifiability, or a finite number of "primitive expressions of which the canonical example is a statement in a programming language".[7] This is to ensure that each 'step' of the algorithm is not mediated by an infinite amount of other conditions. Second, the *nature* of the primitive expressions themselves must be finite, leading to a requirement of efficiency, i.e. a reduction of the resources required to terminate the algorithm. Third, it is virtually always required that the algorithm performs a finite number of steps before completion. Dean notes that the latter two conditions are not strictly speaking true for all algorithms, but so common that they belong in this description.

4. **Repetition** — Most well-known algorithms relate to methods for simplifying mathematical instances. In an algorithm, this is often accomplished through repetitions or *iterations* of step(s), the amount and extent of which is controlled by a parametric (loop) variable. This is not a strict requirement for an algorithm, but since "virtually all non-trivial examples" do have this feature, it is included in this list.

6 This is, other than the theoretical *halting problem* (discussed in the next section), a requirement by Knuth on a properly functioning algorithm. In more recent times, continuous new input and output without halting can in some systems be desirable, such as in weather modeling.

7 More on the relation between algorithms and programming languages can be found in the next section.

5. **Abstraction** — Despite algorithms being conventionally specified as "imperative-like statements over a natural or formal language", this should not be confused with the actual algorithm such a statement could be specifying. This point is subtle, and while feature again later when we will be discussing a philosophical approach computing and algorithms, but is akin to the crucial distinction between natural and formal languages. The mere expression of it in these terms cannot capture the level of abstraction at which an algorithm technically functions.

Whereas Knuth's characteristics focused heavily on the necessary conditions without which an algorithm would not function (properly), Dean has outlined a more procedural approach by looking at the workings of algorithms practically. While the description is fairly similar, more insight into the procedure itself, as well as differentiating finiteness and dropping the *strict* requirement of finiteness regardless of the input, are two major points of contribution.

ALGORITHMS AND PROGRAMS There exists some confusion about the status and difference between the notions of *algorithm* and *program*. After all, a program also exhibits, e.g. Knuth's five characteristics of (a) finiteness, (b) definiteness, (c) input, (d) output, and (e) effectiveness. A quite extreme position is to claim that algorithms and programs are properly the same thing, since they fulfill the same functions [16]. However, others, such as [74] [41], note that the difference between algorithms and programs is that algorithms can be read by humans (natural language), whereas programs are formulated in a programming language and can be implemented by (electronic) computers. These views, however, do not match the conceptualization of algorithms in this thesis, which is closer to the more moderate view that "an algorithm is a general technique for solving a problem (that is, problem-oriented), whereas a program is the concrete formulation of an algorithm as it is needed for being executed by a computer (and is therefore machine-oriented)."[72] Moreover, [72] does go on to state that "the algorithm may be viewed as the heart of the program." On top of this, algorithms and programs differ not only in *for who* or *what* it is intended and/or generated by, but also in terms of breadth of vocabulary, as programming languages are often deliberately kept limited. In fact, several clearly and distinctly different programs can implement the same algorithm [127]. In sum, while there are many similarities between the two notions, since the aim of this thesis to look at the codified elements, which allow for mechanically obtaining a result moving from input to output, *algorithms* are a more appropriate unit than *programs*.

EFFICIENCY A separate note on *efficiency* is in order here, all the more because Knuth and Dean do not address it thoroughly when

giving their characteristics. Efficiency, in the field of *analysis of algorithms* — which is devoted to assessing the efficiency of algorithms specifically — defines it as the amount of resources the algorithm requires to terminate and produce output, with relation to the size of an input n. Note that 'resources' here is a category, often evaluated using the two axes of (a) time, and (b) space [47]. *Time* is measured from the beginning of an algorithm to its termination, i.e. the time required to move from input(s) to output(s) and reaching its final step. It is often mistakenly believed that with the advances of computer hardware, evaluation runtime is less of a concern. This misconception is understandable in simple applications, but because of the requirement of scalability in complex systems, efficiency in terms of runtime remain important. In some instances, there are possible classifications of *worst-case runtime*, *best-case runtime* and *average runtime*. The choice between algorithms with different attributes is context-dependent, although the evaluation happens most commonly for worst-case runtime[8] [72]. *Space* refers to the computational resources required to terminate, and is mostly defined in terms of Random-Access Memory (RAM). There are four different elements involved in how an algorithm 'uses' RAM-space, (a) the amount of memory space the algorithm's code itself occupies, (b) the total amount of memory space the input(s) of the algorithm occupies, (c) the total amount of memory space the output of the algorithm occupies, and (d) the amount of memory space the operation *during* the execution of the algorithm requires. The total of these four elements determine its efficiency in terms of space. Giving preference to either of these evaluations of resource-dependent efficiency depends on the context. In some cases speed (appealing to the time axis) is essential, such as for the purpose of quick (medical) diagnoses or so-called 'flash trading'. This speed comes at the expense of computational power, no matter the specifics of the algorithm itself. The balance in the making of meteorological models is often more towards optimizing[9] for computational power, as runtime of the algorithm is not of direct concern. Algorithms can be altered to give preference to any given type of resource optimization.

1.3 PHILOSOPHY OF COMPUTING

The description of algorithms in the previous section has given a sufficient impression of what algorithms are, and how they roughly operate. An understanding of their functioning, however, is not complete

8 Exceptions to this can come in different forms, e.g. for algorithms with widely varying runtimes for different inputs, stating the worst-case runtime is not informative over a longer period of time, making *average runtime* a better candidate.

9 Throughout this thesis, "optimization" refers both to the more colloquial meaning of *improving* and actual optimal instances. In the latter case this will be explicitly clarified.

without at least superficially going into the *philosophy of computing*, and the status of algorithms when seen from that perspective. Following [21] this thesis will restrict the philosophy of computing to the "nature, possibilities and limits of computation". The intent with this is to move beyond the *practical* descriptions of the previous section, and to facilitate the aim of being able to assess arguments *pro* and *contra* the use of algorithms on an abstract level, regardless of a specific algorithm's functioning. For that, this section will consist of (a) a short discussion on computability, the Church-Turing Thesis and Turing Machines, and (b) tracing the limits of computational and algorithmic operations practically and theoretically. Similar to the natural language approach, the idea is not to give any clear-cut definitions, but rather to highlight several important aspects that will become relevant in the ethical analyses of the subsequent chapters.

COMPUTABILITY AND THE (CHURCH-)TURING THESIS The study of computability refers roughly to the study of what is in theory (un)computable, and the nature of computing itself [21]. The field saw a major breakthrough in the year 1936, when Alonzo Church and Alan Turing published, initially independently of each other, seminal works on fundamental computability [17], which haves come to shape the field [28]. Both Church and Turing engaged with what Hilbert [62] posited as the *Entscheidungsproblem*, which, in the words of Church [27] holds: "By the Entscheidungsproblem of a system of symbolic logic is here understood the problem to find an *effective method* by which, given any expression Q in the notation of the system, it can be determined whether or not Q is provable in the system."[10] This translates roughly to the question of whether there is an effective method to establishing whether a first-order logical expression is universally valid [21], where the words 'effective method' in this quote are an informal way of expressing that there needs to be a procedure, or algorithm to do this in a finite number of steps. What Turing, and Church from a different direction, did is take this *informal* expression and transform it into a *formal* expression, that is, in terms of logic and mathematics [119]. Since the answers of Church and Turing are equivalent, in that they replace the informal notion with the same set of mathematical functions, we will only be looking at one of the two answers. The choice here has been made in favor of Turing for reasons of simplicity, as Church's answer involves λ-calculus that requires a large section to arrive at the answer.[11] Before diving into Turing's answer to the Entscheidungsproblem, two steps are necessary: (a) looking at Turing's point of departure of "effective

10 Emphasis added.

11 For a good introduction to the relation between Turing's and Church's answers to the Entscheidungsproblem, see [17]. For a more mathematical approach to Turing Machines and the Church-Turing Thesis in general, see cf. [17] [59]. For a mathematical attempts at (partial) definition of (types of) algorithms, see cf. [92] [**Yanofsky.2010**].

method" informally, and (b) focusing on his concept of Turing Machine to move from the status quo to a formalization of the method. According to [29], an effective method M contains the following elements:

1. M is set out in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols);

2. M will, if carried out without error, produce the desired result in a finite number of steps;

3. Mcan (in practice or in principle) be carried out by a human being unaided by any machinery save paper and pencil;

4. M demands no insight or ingenuity on the part of the human being carrying it out.

This notion is fairly similar to the natural language description of an algorithm in the previous section. To overcome these ambiguities and imprecisions, especially those contained in the "no insight or ingenuity" part, Turing thought of a structure he called "logical computing machines", later to be called *Turing Machines* (TMs). These machines were an attempt to boil down the calculating process to its most basic elements, and strip away any arbitrariness, so as to get to the heart of computation itself. The TMs in Turing's conceptualization still referred to humans, albeit in a way to ignores everything about them but their 'computing' ability, and moreover in a way that makes no claim to describe a human (or biological) way of computing. *Turing Machines* consist conceptually of an abstract scanner and a limitless or boundless memory tape, that moves back and forth with respect to the scanner. The tape is divided into segmented, distinct elements, each of which is either blank or contains a single symbol; most commonly either 0 or 1, but fundamentally any symbol from a finite set or *formal* alphabet suffices [118]. The scanner is capable of examining one such element at a time, and operates accordingly by changing the direction the tape *moves*, and the quantity of different elemental spaces in either direction (calculated step-wise but giving a net-result). Furthermore, it has the capacity to *erase* and *print* symbols on the scanned element. Added to this is a connected 'device' that allows the scanner to *change its m-configuration* or "state", responsible for how it post-processes each encountered atomized element. Adapting its m-configuration serves to preserve information about what the scanner has just processed, and can thus be seen as a form of memory. Combined, the four elements of *moving*, *erasing*, *printing* and *changing state* are the basic or "atomic" [118] operations. This machine, Turing claimed, can in theory execute every computation that could be done mechanically or algorithmically. There are many

variations with varying degrees of rigor of the Turing thesis. One of the most accessible forms of Turing's thesis is [29]:

Church-Turing Thesis. *LCMs*[12] *can do anything that could be described as "rule of thumb" or "purely mechanical."*

This thesis — in its formulation as the Church-Turing Thesis modified slightly to reflect the terminology of Church's method of obtaining the same result — outlines whether there is an effective method of solving a problem, by showing whether it is possible to develop a Turing Machine for it. This formulation quickly amassed support from logicians and mathematicians, to the point where it soon became an accepted proposal. Von Neumann took this concept and engineered it into physical, electronic devices that were the predecessors of contemporary computers.

There are two notable limitations to TMs, called the *halting problem* and the *printing problem*, which are both theoretical limitations to what is possible using a TM-based (i.e. algorithmic) approach. The *halting problem* denotes the idea that it is in advance impossible to decide whether a TM will eventually halt. An example of a non-halting TM would be calculating the number π with ever-increasing accuracy. Given that π is an irrational number, the digits to be computed are endless without falling into a repeating pattern, and so without further specification, a TM calculating these digits would run forever without halting. The halting problem consists of the fact that it is impossible in some cases to decide beforehand whether or not the TM will halt or not. The *halting theorem* therefore states that: "The halting function is not computable by a Turing Machine" [28]. The *printing problem* consists of the problem that it is not determinable whether an arbitrary program on a TM will at some point print 0, or not. This problem underpins Turing's answer to Hilbert's *Entscheidungsproblem*, in the way that if a TM would be able to decide whether a statement is true in the system, it would also be able to overcome the printing problem. Since this is, in principle, impossible, the answer to the Entscheidungsproblem is that it is fundamentally unsolvable. Finally, it is important to realize that undecidability is in fact the *norm*, not the exception [21], when approaching computation in an algorithmic sense. This means that in order to be able to algorithmically approach computation, a translation of either the input, or of the problem and program may be necessary to facilitate such an approach.

While the Church-Turing Thesis is a phenomenal achievement, delineating the possible computations to be executed by TMs and all approaches reducible to it, its usefulness is often misunderstood [21] [28] [17]. The main difficulty arises when the result is interpreted in less exact ways, as is frequently the case [59], the result of which is often found in the form of the so-called "strong interpretation" [53]:

12 Logical Computing Machines: Turing's expression for Turing machines

Stong Version of Church-Turing Thesis. *A TM can do (compute) anything that a computer can do.*

This is factually untrue, and also a statement never made by either Turing or Church themselves. In fact, Turing has, in his original publications, already made mention of computational devices that would be irreducible to TMs, *choice machines* [53], a theoretical concept where an 'oracle' or black box element of the machine[13] partially influences the TM's behavior. So far, this has not been translated into any physical, operating machine. Rejecting the strong version of the Church-Turing Thesis, however, is significant because this interpretation misunderstands the foundational work of Church and Turing, and hence the practical and theoretical limitations to TM-based computing.

BEYOND TURING MACHINES The Turing Machine-model of computation is extremely influential, and rightfully regarded as foundational for electronic computing. However, as Goldin and Wegner [53] argue, it would be mistaken to equate everything executed by computers currently is equivalent to TMs. The confusion or myth that this would be the case, has to do with the influence that the so-called "mathematical worldview" [53], i.e. the adherence to the strong interpretation of the Church-Turing Thesis, has exerted, notably in the field of computer science. This worldview maintains that all computable problems are based on functions, where the input needs to be clearly defined. In contemporary applications, however, we find feedback-loops where the input is being changed during the algorithmic functioning, for example in applications concerning navigation, or in modern *neural networks* with semi-supervised learning mechanisms. This is not to say that algorithms are not the deciding factor in a fundamental sense, as its logic underpins all applications, but rather that the importance of TM-equivalence should not be overstated: they form the basis of all contemporary computation, but have been extended through the notion of *persistent Turing Machines* [53], that use *persistent stream language* involving interaction and persistence. These nuances notwithstanding, Goldin and Wegner [53] that "TMs can simulate any algorithmic device", justifying the elaboration on the idea in this section. These nuances, however, serve to highlight the contemporary extensions going beyond TMs and algorithms *alone*, while preserving their central role.

COMPUTATIONAL COMPLEXITY THEORY Beyond the theoretical limitations to computability in the algorithmic sense, there are a number of practical concerns that make computability unattainable. Whereas the field of computability theory looks at whether can be solved by a TM-equivalent computer *at all*, the field *computational complexity the-*

13 Turing is explicit in saying that this element is itself *not* a machine.

ory looks at the practical attainability of results. Beyond the restricted, i.e. finiteness of available resources — an infinite tape is a splendid concept to describe the phenomenon of computation, but hardly a practical proposal — as discussed in the natural language approach to algorithms, there is also the problem-side of practical computability. The field of computational complexity theory studies the level of complexity of problems, and classifies them in terms of the time required by an optimal algorithm to solve it. The most relevant distinction in this field is between the two classes called P and NP, where P stands for *polynomial* and NP for *non-deterministic polynomial*. Problems in P can, as the name suggests, be calculated by a (deterministic) TM in polynomial time, that is, for example n^2, where n stands for any natural number, i.e. $n \in \mathbb{N}$. Given a large enough value for n, this is clearly going to be problematic in practical terms to solve, but this complexity is rightfully of a different order than NP, where problems are, using the similar example, 2^n. To fully illustrate the difference between the two, consider the following consequence [21]: "If $n = 100$, the former amounts to 10.000 steps whereas the latter amounts to a number higher than the number of microseconds elapsed since the Big Bang." The non-determinacy of this class of problems revolves around the idea that it is non-determinable whether a solution will be achieved in any amount of time, yet when it *has* been determined, verifying the solution is possible in polynomial time. The thesis that these classes of problems are not reducible to one another, is called the $P \neq NP$ thesis. An overwhelming consensus among mathematicians, philosophers and logicians exists on the side of this thesis being true. While it makes some problems extremely difficult to calculate effectively, i.e. there is no efficient or even effective algorithm to do this, this is also the underlying principle of cryptography, where it would take extreme computational resources to 'solve' the problem (colloquially: break the code), while it is easy to verify the answer once it has been found. This places the burden of computational resources required heavily tilted towards those attempting to solve it rather than merely verifying the accuracy. The fundamental message here is that algorithms, despite their myriad uses, are inherently bounded in the extent to which certain problems can be solved.

1.4 CONCLUSION

In this chapter we have looked at how algorithms function, and the extent of possible applications in a theoretical and practical sense. This chapter has answered sub-question 1: "What are algorithms and how do they function?" In this sense, it is the necessary building block for the subsequent sections that revolve around the ethical sides to the application of these algorithms. This theoretical background is necessary as a superficial or false understanding of algorithms has the

potential to result in unrealistic evaluations, or in the consideration of unrealistic scenarios. While the description given in this chapter is by its nature not all-encompassing, it is sufficient to meet the goal of answering the research question, so that it should be regarded as a *means to an end*.

2

APPLICATIONS AND IMPLICATIONS

This chapter will focus on current applications of algorithms, and how we can develop an ethical understanding of them. For this, the attention will first be on the *consequentialist* branch of normative ethics. A brief overview and justification for adopting such a position will be provided. Next, the notion of *dual-use technology* will be introduced, to later frame algorithms as such a technology, meaning that there are both positive and negative aspects to them. This is fairly standard practice in evaluating technologies from a consequentialist perspective [61]. After this, a number of dimensions and processes of algorithms that result in ethically contestable instances will be described. This helps situate the examples, or cases, of the subsequent section, giving the reader a sense of what sort of things to look for. In turn, the examples involving algorithms featuring ethically controversial aspects will give practical meaning to the dimensions, so that the sections can be combined in a systematic way to show the *manners* in which algorithms have ethical impacts. The cases are not meant to be exhaustive, or to cover each possibility extensively, but rather serve to call attention to the fact that the objections that are raised against algorithmic applications are not merely part of a theoretical discussion, but of practical and actual concern. The final part will be devoted to suggesting a number of possible (directions to) solutions to overcome these difficulties.

2.1 CONSEQUENTIALIST NORMATIVE ETHICS

Consequentialism is the philosophical tradition that, befitting its name, holds that normative properties are dependent on the consequences of actions alone. As [111] notes, there are many types of consequentialist theories, yet without adhering to this basic statement, it falls outside of the umbrella term. Consequentialism, in a similarly broad manner, advocates that we should "base our actions on promoting good consequences and avoiding bad ones" [55], that is, by extension, to identify and pursue that which is considered preferable while minimizing any offsets. It is arbitrary what can be considered preferable, for whom, and to what extent these arbitrary evaluations should counteract one another, resulting in disagreements among consequentialists themselves [66]. These difficulties notwithstanding, negative

consequences themselves can be brought to the fore without resolving the complex and contradictory elements inherent to consequentialist analyses[1].

At its core, this chapter aims to pick up the challenge that [115] describes as requiring "critics to argue that the ... technology might cause harm to some stakeholders and thus cannot be pursued freely". That is, providing solid arguments for why algorithms cannot be framed as neutral instruments to achieve certain goals, but need to be thought of as having ethical implications, requiring continuous reflection on what they are achieving beyond the explicitly stated goals. Any shortcoming of any normative ethical tradition will inevitably influence its analytical strength, but addressing such critiques in-depth is beyond the scope of this thesis. This caveat notwithstanding, consequentialism is widely regarded as a good starting point for ethical analyses in general, and since this chapter deals with the practical, current, real-life applications, this choice of method seems amply justified. Moreover, focusing on *observable* consequences now bolsters the case for reflection on a practical level, not just to a theoretical or meta-ethical (distanced) degree. It is, after all, not the intent of this thesis to use algorithmic applications as a case study to amass points of critique towards consequentialism. Rather, it takes consequentialism as a starting point, and judges algorithmic applications on the basis of the consequences it has. Finally, the point of taking algorithms as a subject for analysis while they are embedded in larger technical and social systems, is justified because they are argued to be a "particularly prevalent and potentially significant component of our evolving infrastructures." [110] In other words, there is a particular dynamic that is brought about by algorithms, and that can be found in more than one context. Furthermore, it is argued to be the most common way of addressing the ethics of algorithms [110], because it connects with real-life events and actualities.

2.2 DUAL-USE TECHNOLOGY

A critical analysis of any subject contains a real danger of portraying that which is analyzed too much in negative terms. To counteract this, it is important to understand that by the fact that this thesis is concerned precisely with the ethically *problematic* aspects of algorithms, it will therefore focus less on the positive uses. To formalize this idea a bit further, algorithms will now be framed as a *dual-use technology*, where the focus of this paper will be mostly on the part of its 'dual' nature that gives rise to ethical issues.

1 For an overview of different types of consequentialism, and disagreements along the lines of what is to be valued, as well as the relation between consequences and a proper course of action, see [66] [111] [55]

The concept of *dual-use technologies* refers broadly to the idea that there are legitimate and illegitimate aspects to a technology, where the (arbitrary) separation between the two categories is dependent on context, and can vary over time. Because the notion has been used in vague and varying ways, Resnik [106] has urged for a conceptualization that is wide enough to not exclude matters of real concern, while also being narrow enough to keep the concept applicable and manageable. To translate this into a workable definition, Forge [44] attempts to take this message seriously, and come up with a workable definition. However, where the attempt to find a good middleground in terms of in- and exclusions is praiseworthy, the final definition seems to be very focused on specific types of potentially harmful technologies: "An item (knowledge, technology, artefact) is dual use if there is a (sufficiently high) risk that it can be used to design or produce a weapon, or if there is a (sufficiently great) threat that it can be used in an improvised weapon, where in neither case is weapons development the intended or primary purpose." Another shortcoming here is that Forge [44] concludes simply that, to oppose adverse effects, regulation should be in order, yet explicitly leaves the matters of *how* this should happen, and to *what extent*, open. Others, however, have been less restrictive in their use of dual-use technologies, and point to uses that do not directly pose physical harm. For example, [7] notes that "code obfuscation is used for legitimate software protection but also by malware", highlighting its dual-use nature. This highlights the dual nature of *dual-use technology* as a concept, explained by Van Wynsberghe and Nagenborg [124] to refer both to a technology that is used by dual groups of people, notably military and civilian groups, as well as "being used for good or bad purposes". Miller et al [89] indicate that a "dual-use is an *ethical* dilemma, and an ethical dilemma for the *researcher* as well as for those (e.g. governments) who have the power or authority to impede the researcher's work." This is the interpretation that will be followed in this thesis. Finally, the question of whether *intent to* and *awareness of* using a technology for good or bad matters, i.e. if unintended consequences of the use of a technology being 'bad' classify that technology as dual-use, is not of practical concern for this thesis. The concept of dual-use technology is a heuristic for highlighting the adverse sides of algorithms, not a classification that confines the breadth of analysis. Furthermore, it could be argued that if unintended consequences make up the majority of the arguments against algorithmic approaches, this points firmly in one direction where progress could be made. As Miller et al [89] already indicate in the context of the life sciences: "... most scientists and engineers do not spend a lot of time thinking about the unintended or unexpected side-effects that can occur when their products are used. They think even less about intentional misuse. Making scientists, engineers and other designers aware of the possi-

ble misuse of their 'brainchild' is the main goal of the dual use policy that has been developed. . . during the past years." In this vein, the rest of this chapter will be precisely about bringing to the fore these elements regarding algorithms, leaving the defense of the undeniable positive sides to algorithmic operations to others.

2.3 Theoretical framework: dimensions of ethical concern regarding algorithms

Instead of looking at examples and claiming there are patterns among them that warrant ethical evaluation, it is more informative to start with a framework of how algorithms are relevant, and have the examples serve as illustrations and arguments for why an ethical view is in order. This section will build such a framework, which will be used in this chapter for a consequentialist analysis, and will also partially inform the next chapter containing a deontological view of algorithmic applications. Gillespie [50] has identified six dimensions of "public relevance algorithms", a term that roughly denotes that they are of societal and hence ethical importance. This view of a subset of algorithms is shared by a 2015 report of the Global Conference of Cyberspace [48] (GCCS) and Diakopolous [36]. The step from public relevance and societal importance to 'ethically contestable' is substantiated by the idea that if there is *no* public relevance or societal importance, it means that it has little to no bearing on people. It would be difficult to base a claim for ethical evaluation on cases that do not involve people[2]. However, this is admittedly an arbitrary step that could result in a 'grey' area of what is considered relevant, and by whom, but given the fact that Gillespie is precisely concerned with the arbitrary nature of this by focusing on how algorithms themselves inform such divisions, this concern is what underpins these dimensions to begin with, making the risk of this being a problem far less likely. This is also illustrated by his warning that "we must firmly resist putting the technology in the explanatory driver's seat" [50]. Note that this list of possible areas of concern is not exhaustive, and will in fact be supplemented at the end of this section. To reiterate: the goal is not to be exhaustive or provide any definitive framework, but rather to highlight important and pervasive dynamics involving algorithms, to pinpoint why these dynamics are ethically problematic, to start conscious deliberation with these insights in mind, and where possible to suggest practical improvements on current practices. Before looking at Gillespie's overview, we will first take a short

2 This includes, clearly, concern for animals, plants, the environment etc. However, these concerns are still experienced by humans, and an ethics of how animals might be concerned about climate change, removing the human-element completely, is far beyond the scope of this thesis.

look at other attempts to to grips with how algorithms warrant ethical deliberation.

PATTERNS OF ETHICAL SIDES TO ALGORITHMS In analyzing in which way algorithms can, from a consequentialist perspective, be argued to have an ethical side, several authors have attempted to formulate abstractions from the specific examples. While contexts vary greatly, efforts to distill mechanisms have been fruitful in identifying a number of patterns that transcend the individual instances in which they are observed. The GCCS [48], for example, notes three elements that, in general "demand ethical scrutiny: complexity and opacity, gatekeeping functions, and subjective-decision making." Burrell [23] takes up the point of opacity as key to understanding "socially consequential mechanisms of classification", executed by algorithms. Her analysis focuses on three points leading to opacity, in:

1. *Intentional obscurement*, or even concealment, by corporations or the state, so that decision procedures are not open to scrutiny.
2. *Technical illiteracy* on the part of those outside the algorithm-producing community, resulting in a lack of understanding and knowledge.
3. A *mismatch between complexity of operations and human understanding*, in the sense that mathematical optimization of the analysis of data is difficult to translate back to the human level, leaving even 'insiders' puzzled.

In a way, then, the infrastructure in which algorithms are employed appear to be a mismatch with the algorithm's functioning. Complex as this interplay may be, these sub-divisions for focus all point to the manner in which algorithms are a unique component of a larger structure, and can and *should* be singled out and studied.

GILLESPIE'S DIMENSIONS For a comprehensive overview of the ethical aspects of algorithms, we turn to Gillespie's six, interrelated dimensions [50]. Gillespie is primarily concerned with the way that algorithms have come to inform and shape us in multiple ways, and have moved beyond those tasks which could conceivably have been done "by hand". To stress how important the information-providing, -generating, and –selection mechanisms are, he illustrates it by saying: "That we are now turning to algorithms to identify what we need to know is as momentous as having relied on credentialed experts, the scientific method, common sense, or the word of God." [50] A short description of Gillespie's six dimensions along which this happens will be given, followed by a supplementary note. Following this will be a number of examples — which are few and short in Gillespie's own writing — to illustrate these points in more depth.

1. **Patterns of inclusion** — "the choices behind what makes it into an index in the first place, what is excluded, and how data is made *algorithm ready*." This involves the collection of data, in that information and real-world phenomena are being digitized. As Gitelman and Jackson contend, "raw data is an oxymoron" [51], highlighting that information must be formalized in a way that readies it for algorithmic processing. Another aspect is that *exclusion* is also possible, both inadvertently and consciously. An example of the former is YouTube's active effort to keep 'suggestive' videos off the *most watched* lists, which is a 'soft' approach to keeping content curated [50]. This is in some cases a viable alternative to outright banning material entirely, which risks controversy over censoring issues.

2. **Cycles of anticipation** — "the implications of algorithms providers' attempt to thoroughly know and predict their users, and how the conclusions they draw can matter." This dimension covers, first, privacy concerns. Providers of algorithms are very concerned with gathering as much data about users as possible, to tune and retune their algorithms for increasing insight into patterns. An example of this is the ubiquitous *like-button* of Facebook, which pops up at seemingly unrelated news-articles, information pieces, videos etc. This creates a superficial symbiosis where users feel empowered and engaged, while (unknowingly) giving up more information about themselves in the process. A second part that is covered by this dimension is that algorithms necessarily focus on what is most (readily) legible to them. Algorithmic profiling is judged — from the provider's side — on sufficiency, with regards to advertisement-use, where the rest is approximated or ignored. A dangerous side to this is that the accepted level of approximation also begins to shape the experience of users, with drop-down menus asking users to specify in which category they believe to fall. The results of this are further implemented in the development of new products, propelling a dynamic of self-fulfilling anticipation.

3. **The evaluation of relevance** — "the criteria by which algorithms determine what is relevant, how those criteria are obscured from us, and how they enact political choices about appropriate and legitimate knowledge." *Relevance* is not a fixed or objective measure, and will hence be open to critique. An interesting corollary of this is that *bias* in algorithms is not a very strong point of critique, since the alternative of unbiased is a difficult position to refer to. However, approaching this at another level does provide insights, because it is an actual choice to delegate the decision-making regarding relevance to a codified mechanism in the form of an algorithm. Moreover, this codified mechanism emerges from a (corporate or governmental) culture, opening the door to suspicions. Some work

has been done (see, cf. [70]) on analyzing structural tendencies concerning bias, though this is a changing field, with some companies changing their algorithms multiple times per month. The underlying idea of codifying biases remains the same, however. This blends in with the follow point of:

4. **The promise of algorithmic objectivity** — "the way the technical character of the algorithm is positioned as an assurance of impartiality, and how that claim is maintained in the face of controversy." This dimensions links the technical aspects of algorithms to a narrative that is created around them, carefully framed by providers and presented through terms such as 'search *results*', and '*top* stories'. Managing a discourse of neutrality, objectivity and even rationality is key to the success of services, and allows the trust they inspire to trump that placed in, e.g. journalists' explicit choices or editor's selection processes. An example of how important this trust is comes in the form of Google ceasing their operations in China, rather than being forced to censor results. Gillespie quotes Morozov [91]: "Google's spiritual deferral to 'algorithmic neutrality' betrays the company's growing unease with being the world's most important information gatekeeper. Its founders prefer to treat technology as an autonomous and fully objective force rather than spending sleepless nights worrying about inherent biases in how their systems – systems that have grown so complex that no Google engineer fully understands them – operate."

5. **Entanglement with practice** — how users reshape their practices to suit the algorithms they depend on, and how they can turn algorithms into terrains for political contest, sometimes even to interrogate the politics of the algorithm itself." Gillespie goes to some lengths to stress this is a constantly ongoing process, where 'users' (and people in general) orient themselves towards recognition, as for example the popularization of hashtags shows. This also introduces a power-asymmetry, where the workings of algorithms are obscure and *obscured* on the one hand, while being optimized and tailored to make recognition-seeking strategies more readily available, without disclosing entirely how this process takes place. In this sense, algorithmic approaches mutually shape their content, with multiple loops of negotiation being possible. This leads to the final dimension of:

6. **The production of calculated politics** — "how the algorithmic presentation of publics back to themselves shape a public's sense of itself, and who is best positioned to benefit from that knowledge." A famous example of this is the so-called *filter bubble* [100], roughly corresponding to the idea that search engines (but also other services using recommender-style algorithms) select and present data differently for individual users. This is a specific culmination of

points 1 and 3, *patterns of inclusion* and *evaluation of relevance*, respectively. More poignantly, the idea of 'Big Data' seems to rest on the premise that insights may be more valid when derived from large patterns of data that may not be obvious to any one person or organization without relying on the associated techniques.

The associated offsets of this are generally not of interest to the providers and designers of such influential algorithms, and therefore need to be scrutinized from outside of its own socio-technical culture or ecosystem. Gillespie concludes, therefore, that "we might see algorithms not just as codes with consequences, but as the latest, socially constructed and institutionally managed mechanism for assuring public acumen: a new knowledge logic. We might consider the algorithmic as posed against, and perhaps supplanting, the editorial as a competing logic." [50]

2.4 CASES OF ETHICAL SIDES TO ALGORITHMS

It is important to make explicit at the start that most of the examples that serve to illustrate the ethical dimension of current algorithmic applications are taken mostly from western societies. This is the case due to the fact that the literature available in English largely focuses on cases drawn from English-speaking countries. However, despite this caveat, the dynamics that the examples highlight are thought to (a) indicate issues that derive from algorithmic applications, rendering the specific context less important, and (b) affect enough people to be of general interest when developing an ethical analysis. Another important aspect of the examples that will be featuring here is that the algorithm needs to be a (relatively) *unique* component, which rules out very simple applications that see a virtual one-to-one translation from humans performing a task to an algorithm doing the same thing. If there is indeed nothing 'novel' or different about the way an algorithm functions — generally only the case for extremely straightforward, small-scale efforts — there is little interesting to be said about it, since there is no actual change. In short, the algorithm-component in each example is bringing to the fore *observable* new dynamics that allow for an inquiry into what this means. Following Sandvig [109] the question will be "about a certain kind of *process* or *strategy* and not about the goal itself." The focus will first be on commercial applications of algorithms, after which public and governmental applications will be considered. The final set of examples will deal with specific technologies and trends that can be found in both sectors. The numbers in parentheses in the title of each paragraph denotes the correspondence to the dimensions of the previous section.

(1, 2, 3, 4, 5, 6) In this set of examples the focus will be on the way that *search engines* and *social networks* have come to "govern" information on the internet, and perform a type of "reality construction" [73]. In order to properly do this, the focus will be on the two biggest examples of each category, Google and Facebook, respectively. Beyond being the biggest in their category, they are also, according to Just and Latzer [73], the biggest two websites employing algorithmic selection. Together, they will give insights into the workings of ubiquitous commercial algorithmic applications.

Search engines are a gateway to information on the internet. Given their importance, it is no wonder that the work on the effects of search engines in general, beyond their technical accomplishments, has a history stretching back to the seminal work of Introna and Nissenbaum in the late 1990s [70]. In their seminal paper, they explore, among other things, the political and social elements involved in the way that the algorithms employed by search engines approximate "a complex human value (relevancy)" [70]. One of the major conclusions of this work is that search engines systematically exclude certain types of sites and information, with ramifications for both content-providers and content-searchers. While in the late 1990s there were still a plethora of search engines, this multitude in options has gradually been replaced by Google, now accounting for approximately 72% of all organic desktop searches, and 95% of all mobile searches globally[3] [97]. Therefore, while other search engines may differ in their approach, an examination of the way Google handles search results and user-experiences is justified as a specific case. In a simplified manner, Google is using a *search query* as input, upon which algorithms use over 200 factors to determine which results to display, and in which order[4]. The exact balancing of factors differs per input and per user[5], although the exact mechanisms behind this fall under trade secret law, further protected by "nondisclosure agreements and noncompete clauses" [112]. In fact, attempting to pin down the algorithms too precisely would prove a futile effort, as [109] explains, *re-programmability* of algorithms is used to make "a key feature of their system modular and subject to continual revision... This has focused attention on the plasticity of algorithms... and suggests a kind of *permanent destabilization* for some algorithms (or at least long-term destabilization)." This effect is further enhanced by the fact that Google is rapidly changing its search algorithms, with 538 changes being made in the year 2011 alone [48], a pace that has only increased since. Therefore, it has be-

3 This includes traffic from China, where Google famously pulled out, but has now taken steps to move back into the market.

4 For the user-interface in which the algorithms are embedded, see the patent Google Inc. filed in 2011 [69]

5 These categories could be broken down further, but the overarching message of tailoring search results to individual users remains the same.

come unwieldy to document each change and bring it to the attention of its users. Despite these difficulties, a number of key observations about the way Google operates can be made.

A famous example is the work of Pariser [100], who analyzed search engines, and Google in particular, finding that it presents personalized search results to its users, based on their knowledge of this user. That is, through the usage of their services — in the case of Google often including more than just the search engine, e.g. Gmail, YouTube — they are able to create a profile, which an (aggregate of) algorithm(s) then uses to present certain search results, potentially leading to large differences between users. An important element in this is the functioning of *confirmation bias* [100] [65], where people feel good about being confirmed in what they already know or think. Through this, it is theorized that people will, firstly, be looking for information that confirms them in their opinions, secondly experience it as pleasant when search engines offer such results, and thirdly thus spend more 'pleasant' time using these search engines which leads to a higher revenue due to advertisement incomes. The commercial benefits, then, are clear, as "[R]ecommenders can have a positive effect on sales and web impressions"[67], and will hence be used in competitive markets, barring regulatory checks. These mechanisms thus, in a limited fashion, serve both the user and the search engine's company. Another example of how Google tailors its results, is the conflicted *Kashmir* region, which is shown to be a part of Pakistan or India, depending on one's location [48]. Although this may be seen as an attempt to be 'politically correct', or neutral, it should be noted that there is no such option, and giving people different information depending on their location inevitably influences their worldview in a value-laden way. While some of the conclusions of Pariser's book "The Filter Bubble" have been nuanced through following research, a 2015 article by Epstein and Robertson [39] takes the effect of algorithmically moderated search results quite seriously, and specifically warns about the possible influence of Google's internet search rankings on the outcome of the 2016 U.S. presidential election. Due to the *search engine manipulation effect* (SEME), which has been shown to seriously alter people's views without them noticing it, exposure to certain material could heavily sway public opinion: "(i) biased search rankings can shift the voting preferences of undecided voters by 20% or more, (ii) the shift can be much higher in some demographic groups, and (iii) such rankings can be masked so that people show no awareness of the manipulation." [39] The danger of this is exacerbated by the fact that many people are unaware [42] [39] of the fact that search engines, of which Google here serves as a prime example, in fact work algorithmically and through this develop a 'profile' which selects search results on relevancy for individual users. The danger in this, then, is that there is no longer such

a thing as a *neutral search*: no matter the phrasing, the results will be biased as a consequence of the algorithmic structure. Moreover, as Epstein and Robertson found, it is of concern that when people are unaware of manipulation effects, there is a tendency to think they have formed their opinions independently. These dynamics are also important for content-providers, who strive to maximize the traffic to- and exposure of their content. In order to achieve this, it is important that the content is both legible, as well as considered relevant to a significant group of people according to Google's algorithms. This, in effect, places constraints on the type of content that will be created, and the way it is marketed. Especially in a technological ecosystem where there is one dominant party, such as Google for search engines, this puts further responsibility on the algorithm's functioning, and the choices that go into it. Examples of how this can have adverse effects come in the 'downranking' of specific content, such as websites showing app-promoting banners [13], but more generally there is a friction between arbitrary choices between optimization of revenue-generating strategies and uncensored search results. In sum, the consequences of the specific algorithmic approach of search engines and Google in particular are that users get presented a set of pre-selected results, are unaware of this selection, and in turn base their knowledge and behavior on this dynamic, with positive reinforcement in the form of confirmation bias making it more difficult to break out of this. For content-providers, it becomes a straitjacket that requires 'fitting in', adaptation to the norm, while it remains unclear how the processes actually work [101]. This leaves the matter of evaluation even more problematic, as "[W]ithout knowing what Google actually *does* when it ranks sites, we cannot assess when it is acting in good faith to help users, and when it is biasing results to favor its own commercial interests." [101]

Similar dynamics play a role on large so-called 'social networks', of which *Facebook* will be taken as an example here. Facebook has been extensively studied from a variety of perspectives, out of which a number of aspects relevant to the topic will be discussed here. Similar to search engines, social networks, of which Facebook serves as an exemplar here, have as a part of their functionality the possibility for users to stay and become informed. However, instead of working with search queries as a basis, it works with social connections and 'subscriptions', both which require voluntary and conscious acts of expressing interest. This has proved to be a very successful model that extends beyond the original objective of social contact. Studies have shown that, already in 2014, Facebook was a source for political news for 48% of Americans [63], while 61% of millennials gets their news from the so-called *News Feed* [35], and that Facebook is now sending more traffic to top-publishers than Google, for the first time since July 2015 [120]. However, similar to Google, the algorithms responsible for

selecting content based on relevance are personalized in ways that are not only building on themselves (i.e. a positive feedback-loop, or self-fulfilling prophecy where people get strengthened in beliefs and opinions they already hold), but are also opaque in their functioning, both because studies have shown that, for example, "62% of students in a high level school were not even aware of the existence of this algorithm", but also because, like Google, the algorithm's specifics are a trade secret. A key difference with search engines is that a large number of the factors going into content-selection[6] is based on the type and depth of relationship with other users. This concept has been liked to an "echo chamber" [107] where associations formed are algorithmically interpreted, curated and fed back. To put these effects in perspective, Facebook itself has conducted a study, published in the journal *Science*, on the size of possible determining factors of what kind of results get prioritized in the News Feed [8], which contained the conclusion that social connections on the network were in fact a slightly bigger contributor than the algorithms, casting doubt on the *filter bubble* phenomenon. However, Pariser in a response has said that "Certainly, who your friends are matters a lot in social media. But the fact that the algorithm's narrowing effect is nearly as strong as our own avoidance of views we disagree with suggests that it's actually a pretty big deal." [99] Additionally, it seems intuitive people are more aware of who their friends are and what they believe in, than the demonstrably low awareness of algorithmic selection.

Traditionally, news outlets such as newspapers and television stations have, clearly, also selected which events to cover, and from which angle. That is, selection and bias are not a new phenomenon introduced by algorithms per se, and any worldview will be shaped by partial and selective information. The mechanisms described in the previous paragraphs, however, *are* novel, in that the decision-making processes are automated, and responsibility of what is included, excluded, prioritized etc. is delegated to the algorithm. Moreover, the fact that this is the case is unclear to people, and little to no efforts are being undertaken to increase awareness of the fact that algorithms are responsible for this, nor to increase transparency with regards to how these algorithms operate. Another key point to notice here is the way that news and information pieces are, through algorithmic selection, consumed on a demand-basis, with individual articles coming to the fore. This is another point of departure with more traditional media outlets, as newspaper or news broadcast on television or radio generally contain information on a range of topics, where the selection procedure is centralized, not individualized, and hence can be said to work on a supply-basis. As Just and Latzer summarize, "com-

6 As in the Google-case, outlining the entire process in details will be futile, as the structure is changing very rapidly. One source indicated that Facebook was using over 100.000 possible weight factors when selecting content, in 2013, a number expected grow [88].

pared to reality construction by traditional mass media, algorithmic reality construction tends to increase individualization, commercialization, inequalities and deterritorialization, and to decrease transparency, controllability and predictability." [73] These are all trends that require further examination and reflection, and should not be considered as value-free or neutral.

ALGORITHMIC CITIZENSHIP In the previous section the focus has been on two prime examples of commercial applications of algorithms, and the dynamics they produce and are a part of, and why this has an ethical side viewed from the consequentialist perspective. Algorithms, however, are also widely applied by governmental institutions for a variety of purposes. This section will look at the case of "algorithmic citizenship", a notion coined by Cheney-Lippold [25] to describe the role of algorithms in the construction of citizenship in the United States, focusing mostly on practices of the National Security Agency (NSA). Citizenship is traditionally defined in the terms *jus sanguinis* and *jus soli*, referring to respectively citizenship derived from 'blood', i.e. being born to parents holding a certain citizenship, and place of birth. The potential difference between the two arises from inter-state migration by either parents or during an individual's lifetime. However, online these two categories are not at all naturally apparent, since documents or other signifiers testifying to either source of citizenship are absent. Moreover, identification of individuals is mediated by technologies and infrastructures that obscure identities. *Internet protocol (IP) addresses* and *network hardware identifiers*, two common ways of identifying devices accessing the internet, can belong to machines operated by multiple users, or be essentially public as in the case of computers in libraries, schools, and other places with relatively easy access to groups of people. Consequently, when the NSA attempted to collect as much data on people as possible, while ensuring that US citizens would be exempt from being monitored — a legal requirement for their operations — it faced the problem of not being able to identify US citizens. To overcome this difficulty, it introduced a new type of citizenship, based on algorithmic scoring on a number of identifiers, to establish how foreign a user is. Cheney-Lippold describes the abrupt consequences of this scoring as follows: "If an individual's foreignness is found to be at or above "51 percent confidence" . . . then that individual legally becomes a foreigner and thus loses the right to privacy." [25]. He calls this new categorization of citizenship *jus algoritmi*, "or citizenship rights according to algorithmic logic." [25] This citizenship is based on technical requirements, and cannot be practically claimed or appealed to. Due to the constraints on investigating the specific procedure by which the '51%' is established, it is not possible to justify or describe how it resolves the problems of obscurity due to technical mediation as

described above. The conjecture of Cheney-Lippold is that there are specific instants in which an algorithm analyzes the available data of a certain connection, quantifies this and produces the output of whether this person is, according to the *jus algoritmi* classification a US citizen or not. This would also open up the possibility for the same person to move between legal citizenship by simply switching a device or producing different (meta-)data by, e.g. entering a specific search query, or communicating with someone algorithmically decided to be a non-jus algoritmi-citizen. As an illustration, the documents that Edward Snowden has famously disclosed have shown that, among others, one identifier in this algorithm is the use of language in e-mails and phone calls, which are being scanned actively. Hence, the application of an algorithmically 'scored' degree of citizenship, with clear cut-off points, produces arbitrary and ambiguous consequences through a process that is actively obscured, and which is based on metrics for which there is no demonstrated democratic or popular foundation.

Technical details aside, the legal, political and *ethical* implications of these practices are severe. This algorithmic approach to citizenship overrides the constitutionally protected rights of citizens, embedded in a system that actively pursues obscurity, resulting in a complete asymmetry of power favoring governmental institutions. Thus, citizens are notably unaware of this practice to begin with, users of the internet are notably not informed of the procedures, do not have a right to know or appeal against the conclusions, and are more fundamentally not made aware of the practice to begin with. Furthermore, the conclusions — and consequently this type of citizenship and protection of constitutionally protected rights — are time-dependent, malleable[7], and ephemeral. To highlight the role of algorithms in this process, it is important to look at the inherently quantified nature of the process, as in the following quote [25]: "Users produce the datafied fodder that can algorithmically modulate their perceived citizenship and foreignness at each HTTP request, with each friend they talk to, and through each access point they log into. But although each datafied action has the possibility to realign the percentage confidence that a user is more foreigner than citizen or vice versa, we do not know how that percentage confidence is determined. A user's actions are weighed by an NSA computer and analyst, while the quantitative definitions of her algorithmic citizenship remain hidden."

This analysis glosses over the many governmental practices that result in patterns of exclusion on the part of governmental institutions, as users use the internet in very different ways, for different reasons, and some may not use the internet in any way at all. Some of these concerns on the skewed relationship of governmental organizations

7 In the sense that the algorithm's scoring-mechanism may be subject to change without any public involvement or knowledge.

are beign recognised, and have recently been addressed by the White House directly, in a report called "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" [43]. However, in this report there is no mentioning of the specific case of algorithmic citizenship, and measures aimed at increasing transparency and accountability of other uses of algorithmically generated metrics remain vague.

FACIAL RECOGNITION (1, 3, 4) Facial recognition is the process where an image or video is being processed by algorithms in a way that the pixels (input) are recognized as making up a person's face. Face recognition through algorithms can be performed in multiple different ways, with each having benefits and challenges in terms of accuracy and efficiency in certain conditions [109]. Roughly, face recognition systems work through the following five steps [125]: (a) *face detection*, an object-detection or tracking problem, (b) *pose estimation*, identifying keypoints (eyes, nose, etc.) and being able to integrate those with angles and lighting, (c) *frontalization/normalization*, a step that maps the image to a normalized picture of a face, i.e. canonicalization, (d) *feature extraction*, once the previous step of canonicalization has been completed, it is represented as a vector-image that can be compared to other images, resulting in (e) *recognition*, comparing the obtained image-data to other such data in the existing dataset. The technical details of these different approaches notwithstanding[8], the implementation of face recognition can lead to significant unintended effects. An example of this is the algorithm for face detection in cameras for computers by Hewlett-Packard initially failed to recognize and track the faces of black subjects in common lighting conditions. This leads to the uneasy result that their faces were not validated as such [109], with the algorithm being the judge of this. This mistake was, due to its readily availability in the commercial sector, quickly discovered by users and addressed by the company through software updates. What is important about this case, however, is that it is precisely the selection-criteria that initially went into the design of the algorithm and hardware of the camera that produced this result. While rectification was possible and swiftly executed, this was possible primarily due to users pointing out the mistake, which requires insight into the results these algorithms produce, a specific circumstance which is often not the case for non-commercial uses. "Considering the consequences of this algorithm would lead us to conclude that in some situations, the algorithm could produce racist results." [109] Moreover, the negative consequences of the algorithm's functioning was only noticed *after* implementation, indicating that Hewlett-Packard had not considered the possibility, nor done extensive testing of their product. While this may appear to be a simple unintended effect of the algorithm, and could be phrased in a very

8 See [2] [125] for a comparison of different approaches to face recognition.

'technical' way where certain conditions result in the failure of the algorithm to perform its intended function, such an approach ignores the social reality of feelings of exclusion or racism.

Another, more recent, instance of algorithmic processing of faces involves not simply the detection, but the *recognition* of faces. While intelligence and security agencies globally use these techniques extensively, a notable case — primarily due to the available information — is that of the Federal Bureau of Investigation (FBI) in the United States. In an investigation [56] performed by the United States Government Accountability Office (GAO) it was found that the error rates of recognition varied between "a few percent up to beyond fifty percent, depending on the technology". This is largely due to outdated and under-tested versions, with updates having to be purchased, which places constraints on the possibilities to solve the issue. The underlying point, however, is that most of these error rates have not been properly tested *prior* to purchase[9] and subsequent reviews contained the same flaws, including false positives. This shows how the functioning of the algorithm underpinning the facial recognition techniques was taken for granted, became embedded in a larger (socio-cultural) system as a *neutral and rational instrument*, and has played an active role in the investigation and conviction of people, while its validity remains questionable. This interplay of serving as a seemingly neutral instrument, used to scan over 400 million different pictures, and its inherently limited functioning, has severe consequences for the labeling of people, as well as the obstruction or complication of (criminal) investigations due to the necessity to follow up on 'leads' that were in essence mislabels by the algorithm. Another point of interest here is that, while accuracy rates are generally very high in training sets, according to a 2016 study [15] humans still outperform the best automated systems in face verification when it comes to unconstrained, i.e. *random*, image sets. Relying on an algorithm hence seems to favor speed and ease over precisions and accuracy, which in some cases seems like a defensible choice, but should not go uncontested as standard procedure given the current parameters.

Recent studies have indicated that facial recognition, beyond the problems with the algorithms themselves, is beginning to pose a more pervasive risk regarding privacy issues. Despite the shortcomings mentioned in the previous two examples, the accuracy rates have steadily been going up, partially through new techniques[10], to the point that it is becoming very hard to escape recognition from images because of the quantity of exposure, even when using conventional techniques to do so [125]. The ramifications for privacy issues are

9 To be precise, "tests were limited because they did not include all possible candidate list sizes and did not specify how often incorrect matches were returned." [56]

10 An interesting new study suggest that instead of 'harvesting' more faces and feeding to an algorithm to learn from, it would be more fruitful to synthesize them [87]

severe and understudied[11] when any camera stream or picture could (automatically) be used to identify people, in both the governmental as well as the commercial sector. The debate between privacy and security is clearly ongoing, with the advances of algorithmically sifting through (live) data exacerbating some of the existing fields of tension.

As Sandvig and others [48] contend, while sometimes the results of faulty inclusion, exclusion, selection or processing of data is obvious, this may not always be the case. The guise of rationality and neutral automation, especially when embedded in complex systems, results in negative consequences that are not always obvious, and furthermore unexpected. Taking this one step further, it is quite imaginable for patterns and systems to come to rely, and *build* on the results that are obtained in this sense. *Profiling* — which is defined by Goodman and Flaxman [22] as: "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person" — on the basis irrelevant metrics in general, but exclusion, racism, and wasting time and money in practical senses, are drawbacks to this approach that need to be addressed fundamentally, and at least be taken into account when evaluating facial recognition practices.

2.5 COUNTERING CONSEQUENCES

In the previous section we have looked at a variety of examples exhibiting the dimensions of Gillespie and others that create negative consequences. Identifying the problems associated with algorithmic applications is the main focus of this thesis, but it would be remiss to leave out a number of countervailing options. Whereas the efforts to look at algorithms from an ethical perspective have only begun to develop rather recently, a couple of promising alternatives and additions have been proposed. This section will therefore deal with a number of possibilities, with the aim of reducing the adverse consequences described in the previous section.

A 'low-level' approach to addressing the problems associated with algorithms comes in the form of transparency through knowledge dissemination that begins by educating journalists and policy makers, raising their insight and sensibility to the operations. The reasoning here is that once awareness on the part of these parties is increased, this will translate into wider society and democratic processes, so that an informed discussion will take place. Some work has already been undertaken [35] on outlining how journalists, specifically computational journalists, can identify and understand power-relations mediated through algorithms. While this is a good way to tackle a number of the issues of opacity, this has its limits. First, while increased awareness on the part of 'opinion-makers' will result in more

11 For some early ethical analysis focusing on CCTV cameras, see [20].

insightful coverage and hence understanding, algorithmic applications by their very nature obscure, and that furthermore commercial and security considerations put limitations on what can and may be analyzed. Second, if the approaches are limited to methodological approaches on the part of journalists and other dedicated professionals, the functioning of algorithms hinders dissemaniation because news is acquired more on a *demand-basis*, that is, individualized so that one needs to be looking for this type of articles to begin with. Algorithmic selection works to highlight these insights for people looking for them, while reducing the chances of people being unaware *becoming* aware and informed. Furthermore, this approach is geared towards changing the structure algorithms operate in and through, rather than changing anything about the algorithm as a technical artefact, or the logic underpinning it. That is, the social construction *of* and *by* the algorithm would be subjected to change, instead of the characteristics that have contributed to this construction to begin with. However, from a consequentialist perspective, mitigating the adverse effects of the opacity of algorithms, even in a limited sense, is a worthy effort.

A more rigorous approach comes in the form of implementing policies that directly counter some of the negative consequences of algorithmic functioning. This skips the step of first informing a wider number of people, and being passive about what the outcome of an informed discussion would be. Recently, the Council of the European Union (EU) has affirmed the proposed *General Data Protection Regulation*, due to take effect in 2018 [22]. The most notable aim is to inform those affected by algorithms and concomitatant profiling, literally calling for a "right to explanation" [22]. This presents challenges for current commercial and governmental operations in the EU that need to be overhauled due to these requirements. Despite these difficulties, the authors are cautiously optimistic, calling this an opportunity precisely because it calls attention to discriminatory practices, and can result in a new paradigm for algorithmic design. This follows similar lines of reasoning as Diakopoulos [36] who has advocated a culture in which designers attempt to create a culture of transparency at all levels of algorithmic development. In a similar vein, Ziewitz [128] has argued that the study of algorithms and their effects may in fact be a way to become aware of biases, as machine-learning applications find precisely such patterns, as well as giving a codified version of how certain cultural, social or personal values are important in decision processes. This approach, at least, is capable of detecting biases that predate the algorithm's functioning itself, and dedicated scholars and journalists could use this method for such purposes. Finally, Crawford [30] invites a widening of perspective, asking us to look "beyond algorithms as fetishized objects". For this, she draws on the idea of "agonistic pluralism", referring to the po-

litical theory[12] that finds benefits in struggle and conflict. This helps illuminate how the decision-making of algorithms takes place in contested spaces. "Theories of agonism prompt us to consider in greater depth the many spaces of dispute where humans and algorithms engage." That is, there are no clear dichotomies at work here, and rather than attempting to find consensus, or a status quo vis-á-vis algorithmic applications, it should be a constant re-balancing, informed by the aforementioned methods of inquiry and disclosure.

Commendable as all these approaches — if only for the attention they draw to the existing problems — the suggestions remain vague, consisting more of outlines and guiding principles than actual implementable frameworks, and often lack the power to transcend an non-committal nature. Even in the case of EU regulations, it is restrictive (what is *not* permissible), leaving the solutions up to commercial parties that are partially responsible for the current situation. These difficulties notwithstanding, the increased attention to the ethical sides of algorithmic applications, shedding the veneer of neutrality, is reason for cautious optimism. While the negative consequences of algorithmic functioning may shift and change, the fact that the problems are receiving attention is a sign that the veneer of unquestioned neutrality has been shed.

2.6 CONCLUSION

One of the main conclusions of this chapter is that while algorithms indeed can produce metrics and insights into patterns, they also (partially) create those patterns, shape behavior, and look at the world in a specific, quantified, processed manner, creating adverse consequences in the process. That is, they are not telling us something about the world as external observers in a vacuum, but are actively co-shaping and intertwined with that which they are describing in a multitude of ways. This realization unravels their supposed neutrality and rationality, as they can not be considered an objective tool, nor be taken as operating 'outside' of the phenomena they are intended to provide insights in. Given the negative consequences we have observed, this should invite a shift in perspective that, both in the design-phase as well as in terms of status, fundamentally analyzes algorithms as important artefacts involved in worldview-building, categorization etc. In this light, the second sub-question in the form of "What are current applications of algorithms and what are the ethical consequences of these?" has been answered sufficiently, by having looked at a variety of applications and the attached ethical issues. The relevance of such a question is underscored by recent work that attempts specifically to develop and advocate methodologies "to produce insights into the nature and work of algorithms" [78]. This

12 For an overview of the main tenets of agonistic pluralism, see [93].

chapter has contributed to such endeavors by coupling abstract dimensions to clear, relevant cases, and offering assessments that future work can build on. It is not a likely scenario to stem the development of algorithms and their applications as such, but gaining insight into the dynamics of how they produce unwanted results could result in policy, or a wider algorithmic 'culture' that urges or forces producers to pay more attention before implementing them. In this way, the ethics of algorithms should not be an afterthought, or solely be kept to the realm of studying consequences, but rather be a consideration from the start. In this way, a consequentialist analysis can contribute in a practical sense, by improving awareness and helping to draw lessons in order to find ways to counter the negative consequences. The sub-question that was introduced as supporting the overarching research question, "What are current applications of algorithms and what are the ethical *consequences* of these?" has thus been answered.

3

THE ALGORITHMIC CHOICE

In the previous chapter we have looked at the consequences of algorithmic applications presently. The arguments *pro* and *contra* were centered around whether these consequences were desirable or not, and what the trade-off consists of in various cases. In a general sense, all these arguments can be said to belong to the spectrum of evaluating specific instances or practices, that is, analyzing the present state, and offering insights into the reasons why this can be said to be ethically problematic from a consequentialist perspective. This analysis has highlighted dynamics that were uniquely brought about by algorithms, warranting a closer inspection in ethical terms. The theoretical framework of dimensions in which algorithms bring about these dynamics, as well as the examples illustrating them in a more practical sense, focused on two elements: (a) what sort of unintended consequences algorithms themselves produce, and (b) how a mismatch between the functioning of algorithms and the socio-cultural structures they are embedded or used in results in adverse results. This has been a worthwhile exercise, yielding a critique that attempts to disrupt the veneer of 'rationality', 'neutrality' or strict functionality that often accompanies algorithms. However, there is a dimension to algorithms that is not captured by such a consequentialist analysis, due to *intrinsic properties* of algorithms. In the previous chapter the issues that were raised could be addressed in practical terms, i.e. by making sure the algorithm does not produce the unintended and adverse effects, or informing stakeholders to a sufficient degree so that opacity is no longer a problematic aspect, these ethical issues would be resolved. If developers of algorithms, or opinion-makers, legislators etc. would be able to address these problems sufficiently, it would be ethically justified to use or keep using algorithms for such ends. These problems and associated countermeasures or strategies fall short, however, in addressing the question of whether the *choice for adopting an algorithmic approach* to any situation or phenomenon can itself be ethically contestable. This is not merely a theoretical dimension to algorithms, as the requirements for algorithmic functioning are specific, and inherently constricting, as well as that delegating or automating responsibility can be controversial. This chapter will look at how selecting, creating, and accepting the required conditions

for an algorithmic approach can be problematic regardless of the consequences, adopting a *deontological* ethical approach. The ethical tradition of deontology is chosen because it explicitly does not rely on evaluating the consequences or circumstances of actions, but rather on the principle on which it is undertaken. This corresponds with opting into the required conditions for algorithmic functioning and justifying its application in a broader sense, highlighting the questions that were left unaddressed in the previous chapter. In short, which ethically problematic aspects of algorithmic functioning cannot be ameliorated or resolved without questioning the use of algorithms to begin with.

First a short introduction to deontology will be given, introducing the reader to the most salient elements for the current undertaking. Following this, a short reiteration of the most relevant tenets of algorithms will feature, emphasizing that while algorithms are a very powerful and practical technology, there are limits to their use. Combining the previous two sections, a number of specific, intrinsic aspects of the choice for adopting an algorithmic approach will be given as a theoretical framework, paving the way for examples of ethically contestable instances from a deontological, *a priori* perspective. These examples will justify the subject matter and methods by showing that this is not a theoretical discussion, but addresses practical, present-day issues. The examples, however, will be substantially shorter than in the previous chapter, for the reason that what is relevant is not the specifics of actual cases — as was the case when looking at consequences — but rather the existence and urgency of moral questions preceding the choice for algorithms.

3.1 DEONTOLOGY

Deontology is the normative ethical tradition that derived from the ancient Greek word for duty, δεον. In contrast to consequentialism, deontology focuses on evaluating actions on the basis of adhering to *principles*, *obeying duties*, and *respecting rights* [122] [115] [64] [110], rather than the results. Broadly speaking, it values the intention or motivation of actions over the results [79], as well as an adherence to principles and the respect for rights, which constitute duties. Deontological evaluations hence focus on "deeply felt convictions and existential interests" [115] that guide these principles, duties and rights. In this sense, the moral question is asked before any consideration of practical nature, a clear point of difference with consequentialism, for which the underlying principles do not take center stage. As with all broad, general normative theories, there are many nuances and differences between different versions of deontology, though the characteristics described above delineate the field within such differ-

entiation takes place.[1] As was the case with consequentialism, the aim here is not to use this chapter as an exercise in finding or analyzing specific aspects of deontological traditions, but rather to use it as a heuristic [113] to explore how using algorithms is not a neutral act, but broaches important human values.

The value of adopting a deontological approach in this chapter derives from the fact that, in contrast to consequentialism, it allows for fundamentally questioning the choice to use algorithms, irrespective of the outcomes subsequent to it. Since deontology is concerned with whether certain principles are adhered to, rights respected and duties upheld, the focus will be on *which values* are at stake, and whether it is morally permissible to accept encroachment of them. Following Swierstra and Rip [115], the position that deontological concerns can have 'right-of-way' over consequentialist considerations is adopted, because of the aforementioned connection with deeply held values. The tension that can arise from consequentialist and deontological arguments stems from issues of contradicting answers to a specific case. A deontological approach is especially enlightening when consequences are *uncertain*, as well as when the difficulty in measuring and assessing *to what extent* and *for whom* the results are beneficial or positive become relevant. At a basic level, the difference between consequentialism and deontology revolves around which aspects of a given situation can justify or discount actions. In this sense, it serves as a counterbalance to the idea that the end justifies the means, in this case whether positive outcomes of electing an algorithmic approach — strictly those outcomes that arise *after* the algorithmic approach has already been adopted — justify the aspects of the choice that are in themselves ethically contestable. To clarify this point, the next section will restate a number of tenets of algorithms, which will be framed as approaching a phenomenon in a selective, limiting sense that is not always justifiable. Through this, the neutrality of the use of algorithms will be challenged by referring to (categories of) morally important values that play a role beyond the outcome. This approach also entails another difference with the previous chapter, in that instead of ending with a number of possible strategies to avoid the negative consequences, the aim here is to spark and inform discussion and reflection on whether the values in question are important enough to forego an algorithmic approach, without taking any definitive stance on the matter.

1 For an overview of different versions of deontology, as well as the contrast with consequentialism, see [3] [117].

3.2 CHARACTERISTICS OF ALGORITHMS OF DEONTOLOGICAL CONCERN

As we have seen in the first chapter, which aimed to describe algorithms in a general sense without being overly confining, there are fundamental characteristics that are *not context-dependent*. This foundation will be used in this section, and coupled with a number of dimensions of the theoretical framework of the previous chapter, which highlighted patterns of ethical concern. This coupling will show that the characteristics of algorithms *necessarily* lead to those patterns, and thus cannot be remedied by improving the algorithm or the infrastructures surrounding it, resulting in deontological concerns.

The three properties, or requirements, of algorithms that are relevant for the current analysis, are: (a) well-defined input from a formal language, (b) the codification and automation of tasks, and (c) the assumption of positivist logic. These properties all derive from the groundwork that was laid earlier, but need to be specified, because these properties give rise, when viewed from a deontological perspective, to specific issues that were not addressed in the consequentialist chapter.

WELL-DEFINED INPUT The requirement of algorithms that the *input* comes from a formal language, i.e. is not open to multiple interpretations but instead clearly defined, places constraints on the type of input that is allowed, and hence on the possible applications of algorithms *directly*. Directly is emphasized here, because while a large number of everyday phenomena do not belong to a class of formal languages, these phenomena can be transformed in such a way that they become legible or usable for algorithms. Before showing how this transformation can be problematic, it is important to explain that the reason for insisting on the usage of non-formal, rather than quantitative or quantifiable is that algorithms 'can' process input that is strictly speaking *qualitative*. That is, as a matter of perspective, the level of description can be the determining factor in calling a certain input or step quantitative or qualitative. As an example, consider an algorithm which for each input from a string of integers determines only the qualitative property of being positive, negative or zero. At a first level, this is a qualitative property of the input, yet it is straightforward to see that it can also be framed in terms of three possible categorizations, which is a quantitative and discrete process. However, taking that conceptualization as the norm is arbitrary, which is why in this thesis the choice has been to adhere to the more rigorous phrasing of non-formal, referring back to the description in chapter 1 of formal languages.

Returning to the mismatch between phenomena that are not directly suited for an algorithmic approach, and the desire to process

them in precisely such a manner, it should be noted that the transformation required for meeting this desire results in a different from of the matter. This is analogous to the more general idea that a 'model' or other type of representation is always a specific interpretation of that which it is connected, captured also in the phrase "the map is not the territory" [4]. This idea refers to the confusion or conflation of the representation and the represented, in this case an algorithmically obtained metric. While a critique of this practice easily falls into the consequentialist sphere, a deontological objection comes in the form that algorithms *necessarily* have this consequence, meaning that accepting the transition into representation should be that which is justified, not the outcome of such a transformation. The dynamic here then is towards the need for 'optimization' and 'efficiency', at the cost of understanding that which arguably makes these phenomena worthwhile to begin with. Whereas this practice of transformation is not novel — one can think of, for example, IQ tests transforming the abstract notion of cognitive abilities into a single number embedded in a model with a standardized mean and standard deviations — the fact that algorithms are increasingly employed in society, as shown in the previous chapter, emphasizes the need for the justification of subscribing to the transformation of phenomena to well-defined input, in the face of the duty to protect that which is 'lost in translation'.

CODIFICATION AND AUTOMATION Algorithms are by their very nature codified procedures for transforming input into output, albeit with varying degrees of complexity and possibilities for feedback mechanisms, e.g. an output serving as further input, etc. However, even in the case of such dynamic instances of algorithms, the dynamic nature itself is codified, so that on a derivative level the exact and fixed principles are still present.

A further point of friction on the automated nature of algorithmic operations is that the evaluation of the result is only possible *after* implementation, which is especially salient in complex situations that are either unpredictable, or themselves shaped in relation to the algorithms functioning. For example, even the best meteorological models have uncertainty and error-margins built into them, which is accepted as 'best practice'. However, the inherent complexity of other systems requires a different standard, such as when modeling the behavior of (semi-)autonomous systems, e.g. cars or software with an impact on financial markets. These arguments contain a consequentialist streak, but the focus here is rather on the choice that underlies accepting the uncertainty in modeling complex situations and accepting the fact that automated processes, with codified values and judgements, co-shape reality.

Codification and automation become especially problematic when the to-be-executed task is itself specifically an ethical one. That is,

when there is at least an intrinsic ethical component to the functioning, such as in the case of the behavior of autonomous vehicles, an example that will be discussed in the next section.

COMPUTATIONAL POSITIVISM The use of algorithms, because of the previous two points — requiring well-defined input, as well as being codified, automated, hence involving a delegation of agency and responsibility — means a choice to subscribe to a specific, *positivist* epistemology. Closely related to especially the previous point, the positivist approach to large-scale phenomena is in fact a requirement or justification for the use of algorithms. This axiom, however, is not neutral or value-free, and has been contested widely in the field of sociology [90] on the basis of framing ""social facts" as things independently of their own construction.... "[A]ny allusion to observational 'social facts' begs the question [sic]: facts under which interpretative scheme?"" The danger of reification is therefore quite real, as algorithms necessarily produce these results due to their descriptive nature, or more specifically by looking for patterns, 'correlation', instead of causation [75].

This acceptance of a positivist outlook becomes especially problematic in cases of ambiguity, where clear resolutions of the tension are either absent *currently*, or in principle already a matter of personal or ethical 'preference'. Kraemer et al [81] have shown that, in a medical setting when developing algorithms for the interpretation of scans, researchers can have *differing rational reasons* for developing different algorithms. Hence, 'rationality' is not the crucial criterion that describes choices in the design-process of algorithms, so that instead of unearthing patterns, and letting the 'data' speak for itself, we return to the notion that 'raw' or unstructured data is an oxymoron [51] [52], and the conclusion that these patterns are potentially superimposed. Whereas this becomes mostly a problem when looking at social phenomena, it is however an inescapable consequence of the algorithmic approach, because epistemological positivism underpins algorithms in that it requires well-defined, 'objective' input [49], resulting in computational positivism.

3.3 CASES OF DEONTOLOGICAL ETHICAL CONCERN

The previous section has outlined in a general sense how the basic characteristics of algorithms make the choice for using algorithms morally contestable, because of the transformation of phenomena, the delegation of tasks and responsibilities, and the assumption of positivism on social phenomena. This section will analyze a number of cases in which this is of practical concern. Unlike the previous chapter, which dealt with adverse consequences of the use of algorithms, in this section the aim is to highlight fields of tension between algo-

rithmic approaches and the values they encroach. Rather than offering suggestions for resolving these tensions, the goal is to spark and inform discussions, because the nature of the deontological analysis adopted here is not so much about arguing for the preference of specific values, but rather about showing that — and in which manner — values are at stake. Two sets of examples will be illustrating this argument: (a) case 'revisited', a short look at the example of Google that also featured in the previous chapter, which highlights dynamics that were left untouched, as well as clarifying how a different ethical approach yields different conclusions, (b) *Autonomous Vehicles* (AVs), where the behavior in extreme situations entails codifying value-laden ethical behavior, and military autonomous drones, which by their very nature of exercising control as well as (potentially) causing physical harm are an ethically contestable presence. Together, they will bolster the case that prior to selecting an algorithmic approach, careful consideration on precisely what that choice means in terms of value-judgements ought to take place. The cases are explicitly not intended to be resolved through an analysis in this thesis, but urge a philosophical and ethical debate in the fields that the cases cover, as well as indicating that such questions and discussions — beyond the examples discussed here — have a place whenever considering algorithms impacting aspects of life we deem valuable.

CASE REVISITED Before looking at new examples that underscore the concerns outlined in the previous section, it is informative to revisit an example that was brought to the fore in the previous, consequentialist chapter. This is especially so because it highlights both the differences of the specific case, but also emphasizes the different *type* of answers that each methodological stance produces.

Google's practice of algorithmically sorting search results, while creating or allowing an environment in which people are unaware of the fact that the results are tailored to the specific user are consequences that are an exponent of specific choices made by the company. While it has been shown that these consequences can be negative, there is another dimension to the reliance on algorithms to perform such tasks. In extreme cases, this has resulted in biased, or even outright *racist* results, such as identifying pictures of black people as "gorillas" [14], or strongly associating positive traits with white-skinned people [108]. Rectifying the specific instances, the company further responded that while it deeply regretted the hurtful results, such unfortunate correlations were inevitable, due to the fact that Google's algorithms produce *descriptive* results, meaning that any existing biases in the data that is being sorted on relevance will be reflected in the 'search result'. This is, by Google's own admission, an inevitable corollary of delegating the sorting on relevance to algorithms. What makes this point distinct from a specific consequence

of the algorithms Google is using, is that there is no 'fix' on a purely algorithmic level, i.e. this is a problem that cannot be purely algorithmically solved. It should be noted that curation or relevance-sorting by humans will, given a large enough number of instances and diversity of queries and people receiving them, also produce controversial results, but that in such cases there is no 'hard' codification, as well as that there is no delegation of responsibility. In this case, what sets a deontological look at the fundamental choice to approach a situation algorithmically apart from a consequentialist approach, is that the former is capable of analyzing the inescapable 'conditions', enabling a meta-view of the ethical implications. Whereas the problems outlined in the previous chapter could be remedied by improving the way the algorithm functions, or by developing knowledge and the socio-technical infrastructure, this issue with the algorithm cannot be solved without changing the approach to something other than (just) an algorithm.

A recent, additional component to a deontological view of how Google raises concerns, is voiced by Kurasawa [83]. Kurasawa investigates is how algorithmic logic is altering our perception of the 'social', which he states now "[F]rames society as the mere aggregation of measurable individual tendencies and choices—thereby leaving behind any notion of collective or structural forces. Moreover, this same logic understands agency as a matter of calculated probabilities amongst a range of possible courses of action and decisions by the actor." In this sense, it furthers the computational positivist tendency outlined in the previous section. By necessarily doing this, alternative interpretations are lost by virtue of the algorithmic approach.

DELEGATION: AUTONOMOUS VEHICLES AND MILITARY DRONES

Recent developments in algorithms and corresponding hardware have propelled the development of increasing automation of vehicles. In 2007, six teams of researchers had been able to complete "Urban Challenge", set by the Defense Advanced Research Projects Agency (DARPA), the "first benchmark test for autonomous driving in realistic urban environments" [19]. There are two general classification systems for levels of automation, with the 'highest' level referring to a degree of automation that is called *autonomy*. For clarity's sake, this section will only deal with this level, although various forms of automation, such as automatic braking or cruise control, have already been widely implemented. The first description of autonomy in vehicles, by the *National Highway Traffic Safety Administration*, phrases it as: "The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied

vehicles." [96] The second prevailing standard is that by the *AdaptIVe Consortium* and describes the situation as "systems can accomplish the complete journey from origin to destination in a high automation modus, and can do so anywhere on-road that a human can legally drive a vehicle. Except activation, deactivation and determining waypoints and destinations, no human driver is required any longer." [12] Such vehicles are currently in development, with early results on certain models — who can 'technically' already perform this, but require the passenger to remain alert and capable of taking over control of the vehicle — being very positive in terms of safety and reliability [31].

Beyond the practical concerns regarding safety, such as the vulnerability of software in terms of bugs and security, researchers worry about situations in which harm to either structures or people is unavoidable [19] [18]. These situations will occur, given enough vehicles and time, and require decision models in order to achieve the result that is deemed most desirable. A growing body of literature on this topic of 'algorithmic ethics', i.e. coding for behavior in intrinsically ethical situations, often frames the problem either in consequentialist or deontological terms, but mostly offers a hybrid approach [19] [18]. Preliminary studies have been conducted on attitudes towards types of decisions being made by an autonomous vehicle in various theoretical life-threatening situations [19] [18]. These studies found that people largely follow consequentialist modes of ethical evaluation, e.g. saving a larger number of people is favorable, etc. However, there are differences among the strength of such convictions, especially when the life of the passenger is taken into consideration in the endangerment. Without going too much into the specifics[2], the existence of difference in preferred behavior of the autonomous vehicle is theorized to be translated into different algorithmic decision models, resulting in varying, possibly *competing* autonomous vehicles [18]. While these elements all point to a consequentialist analysis of the dynamics[3], the more fundamental question of whether it is morally permissible to delegate such ethical decisions to an automatic process *to begin with* should enter into the debate. This question is rarely asked, and much less addressed, and despite the relatively low attention this issue has attracted, being aware of it is key. While the occurrence of 'life-and-death' situations is theorized to go down [54] [19], the requirement of algorithms to be codified *prior* to the occurrence of such situations, is problematic because it does not allow for improvisation, and assumes that the pre-defined set of encoded variables is sufficient to judge any possible situation satisfactorily. A second point is the requirement for *well-defined input*, which clashes with something so existential and

2 For an overview of the results, see [19] and [18].

3 For an excellent overview of the need for a consequentialist ethical assessment of automated vehicles, see [54], which argues against common claims that such an ethical assessment is not required.

intangible as human lives. The deontological question therefore focuses on whether it is morally permissible to delegate such decisions *at all*, and whether we are willing to transform human lives, health, and safety into well-defined, formal input.

Similar dynamics are involved with the development of military drones[4]. The development of algorithms geared specifically for military purposes has resulted in autonomous systems that are reliably outperforming human operators [40]. Whereas with autonomous vehicles the occurrence of harm is an unfortunate, inevitable side-effect, for which efforts are undertaken to minimize the chances, for military purposes physical and mental harm can be explicit objectives. While the issues with codified behavior and well-defined input are the same, then, there is an additional level of salience to the delegation of meting out violence. So far, in actual conflict-situations, there has been a requirement for humans to be 'in the loop', so that the delegation is not complete, and that there is no "risk-free" war [102]. Yet, while fully autonomous drones have not seen any use outside of simulations, the notion that the required technology will be acquired by multiple opposing military groups points to the functional benefits of deploying fully autonomous weapons, rendering the ethical concerns that underpin the current practice of keeping humans involved impotent. This, then, requires an approach akin to that of nuclear weapons [95] [102], where the use of them is deemed unjustifiable, no matter the conflict, and hence strict regulation needs to be in place.

It is not an aim of this thesis to resolve the deontological questions introduced in this section, but rather to highlight that these questions are relevant in a theoretical *and* practical sense. Whatever the answers to these questions are, they will have to be made explicit, as situations in which existential values are at stake should be evaluated as thoroughly as possible.

3.4 CONCLUSION

The main aim of this chapter was to explore the ethical dimension of algorithms in a way that does not involve the consequences of specific instances, i.e. what the fundamental characteristics of algorithms entail. By looking at the requirements or conditions under which algorithms can function, deontological arguments *pro* and *contra* the algorithmic approach can be formulated. In this sense, this chapter's goal has been to fuel debate, and contribute to a heightened sensibility to the restrictions attached to the algorithmic lens. While the benefits and possibilities are clear, the algorithmic approach is far from a panacea, and should be fundamentally questioned. Unlike

4 As with autonomous vehicles, the focus will here be on the choice *preceding* the implementation of algorithms. For a good overview of possible consequences, see [102].

the analysis undertaken in the previous chapter, there have been no countervailing measures proposed. This is an explicit choice, because the positions in the debate are too numerous to list, as well as that the weighing of arguments, or the adherence to certain principles, is context-dependent, an attribute that does not belong in a strict deontological approach. The second, supporting, sub-question, "What are the effects of electing an algorithmic approach, regardless of practical implementation?" has been addressed, albeit in a manner that invites further exploration. This further exploration is called for primarily due to the fact that while this chapter has outlined a number of fundamental concerns, the field is developing in a way that calls for constant, updated reflection.

4

ALTERNATIVE AND FUTURE APPROACHES

In this thesis the aim has been to show the ethical sides to algorithms both practically and fundamentally. While this has been fruitful in uncovering in which ways algorithms contribute to adverse effects and consequences, and has demonstrated that selecting an algorithmic approach is not a neutral or value-free choice, there are angles that have not been covered. In the following short paragraphs, a number of possibilities will be outlined that either change the methodology of similar projects, i.e. geared at describing the ethical dynamics brought about by algorithms, or by shifting the focus away from a philosophico-ethical approach to other academic disciplines. In each of these paragraphs, the benefits associated with such different approaches will be outlined, as well as a short rationale for not selecting that approach in this thesis. The aim of this brief section is not to be exhaustive in listing possibilities, but rather to show that research in this direction can be enriched and expanded in a multitude of interesting ways. This chapter contains a brief discussion of a number of open questions, or topics that have explicitly been left undiscussed, either because it went beyond the scope of this thesis, or because it is outside my range of expertise. This list does not aim to be exhaustive, but rather to give the reader a glimpse of the myriad possibilities in this developing field.

TECHNOLOGICAL MEDIATION THEORY In this thesis, the question of the status of (moral) agency of artefacts has been left largely untouched. However, in the philosophy of science, several different theories on this topic have been developed. Polar opposite answers of this question have proposed, with on the one hand the view that artefacts and technologies do not have any agency and are neutral and external tools, and on the other hand the idea that artefacts and technologies have agency in a symmetrical way to humans, and co-shape reality, hence bestowing a moral character on them. The latter view is known as *technological mediation theory*, of which a notable proponent is Verbeek [123] [121]. Conceiving of technologies and artefacts in such a manner opens up a new analytic field, because the dynamic nature of the relation between artefacts and humans, by virtue of allowing agency to reside in artefacts, results in different questions.

The specific focus on mediating in an epistemological sense could thus enrich the analyses by offering a heightened sensibility to the mutual shaping dynamics. The reason this has not been incorporated in this thesis is that it entails an additional level of complexity on a topic that is already clouded in opacity. Thus, for reasons of clarity, the analysis has left these issues to the sideline, to be able to focus more accurately on answering the research question.

STIGMERGY This thesis has used methods and ideas from the philosophy of computing, and of ethics. However, similar observations about the structuring effects of some algorithms have been made in other fields. For example, the feedback mechanisms involved with recommender-algorithms could also be viewed through *stigmergy*, the study of indirect communication in multi-agent systems. Stigmergy is a concept originally developed in the field of entomology [37] to indicate how certain organisms were capable of influencing one another without coming into direct contact. The clearest example of this is the way ants mark a successful route towards a desired good by leaving a trace of hormones, specifically pheromones. This way, even though direct language communication between ants has not been established, they have a means of communicating and hence coordinating behavior. Recently, this concept of conveying and coordinating without direct interaction has been extended to interaction processes outside of the field of entomology, and it has been posited that human behavior can to a large degree be explained through stigmergic algorithms [37] [26]. This cross-pollination of ideas in different disciplines opens up new fields of research where established practices in one area can lead to new insights in others. Specifically, studying how algorithms — beyond the artificially created ones that have been the primary focus of this thesis — are capable at describing naturally occurring behavior in multiple species, including humans, invites further research into more complex patterns, as well as the development of technological structures that facilitate such naturally occurring dynamics [114]. In the field of computer science this has been adopted under the umbrella term of "swarm intelligence" [10]. Moreover, preliminary work on how "leadership emergence" takes place has been approached from a stigmergic perspective [86], indicating that this type of algorithmic pattern-forming has multiple applications that are currently understudied. The methodological tools required for looking at algorithmic patterns from the viewpoint of stigmergy differ greatly from those used in this thesis. It is for this reason, i.e. preserving clarity, that this has not featured here, its potential notwithstanding.

TRANSFORMATION OF THE JOB MARKET With the increasing implementation of software-driven, algorithmically based, technologies,

a transformation of the job market has already been observed, and is projected to continue [84] [116]. Already we see that algorithmically driven technological applications — such as robots — are replacing or displacing work previously carried out by humans. However, recent research carried out by Deloitte [33] [34] has shown that this is not necessarily the case. In a similar vein, Gownder et al [57] published a report dubbed "The Future Of Jobs, 2025: Working Side By Side With Robots" which holds that advances in automation will *transform* rather than *replace* human labor. That is, similar to how the Industrial Revolution has shifted the balance in manual labor but not replaced human activity altogether, so will the advance of algorithmic operations shift the balance in cognitive labor, but not replace human activity. However, others, such as the World Economic Forum (WEF) [116], are less optimistic or even neutral. Calling it a "fourth Industrial Revolution", it projects a global "net employment impact of 5.1 million jobs lost to disruptive labour market changes over the period 2015-2020." Other estimates, such as those in a work called "The future of employment: how susceptible are jobs to computerization" [46] project that "about 47% of US employment is at risk." An interesting and alarming detail of what type of jobs are at risk, according to the WEF, is that jobs currently occupied by women are much more susceptible to replacement Thinking along those lines, former CEO of McDonald's, Ed Rensi, has stated in an interview with Fox Business [45] that if the proposed rise of the minimum wage in the United States to $15 per hour would indeed take place, robots would be a cheaper option in the restaurant industry. With a drop of irony it could be said that dropping out of school might no longer result in the proverbial 'flipping burgers', but instead in 'flipping switches', where the latter presumably requires fever people to execute. The work on this is all preliminary, and largely based on conjecture or questionable extrapolation, but since it is a developing field, this could have important ramifications that could also heighten sensitivities for regulation of development and implementation. The reason for not focusing on this aspect of algorithmically driven economic dynamics is precisely the contradictory nature of the projections, as well as that an analysis which is sensitive to wider economic trends falls outside of the methodological scope of the thesis. Conversely, however, the analysis of this paper could inform future economic projections because of the dynamics — abstracted from the specific examples — described in the previous chapters.

CONCLUDING REMARKS

This thesis has aimed to seriously analyze algorithms in both practical and fundamental terms. In order to do this properly, the first chapter has been devoted to establishing a firm understanding of *what algorithms are*. This has been done both in a natural language approach, which has served to outline algorithms in a functional sense, as well as in a more formal approach, which highlighted the fundamental requirements and conditions, as well as theoretical limitations to their functioning. With this, the sub-question "What are algorithms and how do they function?" has been answered.

The next chapter looked at the consequences of current applications of algorithms, and have uncovered a number of patterns through which these can be adverse, according to a *consequentialist* ethical view. This has explicitly not been formulated in a way to strictly argue *against* algorithmic application, by using the notion of dual-use technology, which enabled the idea that good and bad consequences are not mutually exclusive, and the balance of these varies per context. The patterns or dimensions along which algorithms create negative effects have been illustrated by looking at examples that satisfied the two conditions of being, first, unique to algorithmic operation, and second, of societal concern. The first condition justifies the choice for the unit of analysis, i.e. provides the reason for looking at algorithms as opposed to other parts of the socio-technological frame in which they operate. It is also a logical consequence of the theoretical framework of the dimensions of possible negative consequences, as these point to algorithms as The second condition was necessary to exclude examples that are unique to algorithmic processing, *and* produce negative effects, but are rare or inconsequential to the point where they would not be relevant for ethical consideration. Through these examples, I have shown that in the two realms of government, and media and information, algorithms produce large-scale negative effects corresponding to the theoretical framework. A final example of a practical, specific technology rounded off the set of examples, to zoom in on the ubiquitous and pervasive nature of algorithms in a more applied context. The final section of Chapter 2 has been devoted to possible countermeasures to the negative consequences described and observed in the previous sections. Many of these measures focus on various approaches aiming to increase awareness on the parts of all stakeholders, e.g. developers, users, policy-makers, journalist, to combat the obscurity side of the "black box" description of algorithms. Work on these approaches is ongoing, and with recent legal developments such as the outlined EU-regulations these approaches

will be put to the test. With this, the sub-question "What are current applications of algorithms and what are the ethical *consequences* of these?" has been answered sufficiently, as well as addressed in a manner that opens up avenues for amelioration.

The consequentialist analysis has offered insight into how and where algorithms are currently producing negative effects, and has provided a number of potential directions towards amelioration of them. However, informative as this analysis may be, it does not cover the entire spectrum of possible ethical considerations associated with algorithms. That is, as a method or epistemological technology, algorithms come with a number of intrinsic requirements to function that themselves can be ethically problematic. These intrinsic requirements have been listed at the start of Chapter 3, which proposes to use the framework of *deontology* to evaluate the moral permissibility of accepting these requirements in various contexts. Again, the analysis should not be read as a simple statement against the use of algorithms, but rather in the light that being able to point out real ethical problems at the very least warrants serious deliberation on the conditions on the use of algorithms. The acceptance of the requirements mentioned in the first paragraph of Chapter 3 have subsequently been directly linked with leading to violations of deontological principles, that is, that the choice of accepting the requirements for, and results of algorithmic operation[1] is itself ethically contestable. This has formed an answer to the third sub-question: "What are the effects of electing an algorithmic approach, regardless of practical implementation?"

Taken together, these three chapters have contributed to answering, in parts, the main research question: "What are the ethical implications in practical and fundamental terms of the use of algorithms?" While the conclusions have not always been clear or complete, the aim of sparking debate, and informing the arguments that will be used in such debates surrounding the use of algorithms, has been achieved. In this sense, it has contributed to the science and literature on algorithms by offering a building block for future research and policy-development.

1 The results, that is, that are not context-dependent or expressed in terms of consequences, but rather inevitably a "result" of algorithmic operations to begin with, e.g. quantitative output.

BIBLIOGRAPHY

[1] D. Abhyankar and M. Ingle. "A Novel Mergesort". In: *IJCES International Journal of Computer Engineering Science* Volume1.3 (2011).

[2] Mohamed Tahir Ahmed and Shamsudin H. M. Amin. "Comparison of Face Recognition Algorithms for Human-Robot Interactions". In: *Jurnal Teknologi* 72.2 (2015). ISSN: 2180-3722. DOI: 10.11113/jt.v72.3887.

[3] Larry Alexander and Michael Moore. *Deontological Ethics*. 2015. URL: http://plato.stanford.edu/entries/ethics-deontological/.

[4] Korzybski Alfred. "A Non-Aristotelian System and its Necessity for Rigour in Mathematics and Physics". In: *Science and Sanity* (1933).

[5] M. Ananny. "Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness". In: *Science, Technology & Human Values* 41.1 (2015), pp. 93–117. ISSN: 0162-2439. DOI: 10.1177/0162243915606523.

[6] A. B. Arndt. "Al-Khwarizmi". In: *The Mathematics Teacher* 76.9 (1983).

[7] John Aycock. "Applied Computer History". In: (2015), pp. 105–110. DOI: 10.1145/2729094.2742583.

[8] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. "Political science. Exposure to ideologically diverse news and opinion on Facebook". In: *Science (New York, N.Y.)* 348.6239 (2015), pp. 1130–1132. ISSN: 0036-8075. DOI: 10.1126/science.aaa1160.

[9] Douglas Baldwin and Greg W. Scragg. *Algorithms and data structures: The science of computing*. 1st ed. Charles River Media computer engineering series. Hingham, Mass.: Charles River Media, 2004.

[10] Nimrod Bar-Am. *In Search of a Simple Introduction to Communication*. Cham: Springer International Publishing, 2016. DOI: 10.1007/978-3-319-25625-2.

[11] Solon Barocas, Sophie Hood, and Malte Ziewitz. "Governing Algorithms: A Provocation Piece." In: Available at SSRN 2245322 (2013).

[12] Arne Bartels, Ulrich Eberle, and Andreas Knapp. *System Classification and Glossary*. 2014.

[13] Daniel Bathgate. *Mobile-friendly web pages using app banners*. 2015. URL: https://webmasters.googleblog.com/2015/09/mobile-friendly-web-pages-using-app.html.

[14] BBC. *Google apologises for Photos app's racist blunder*. 2015. URL: http://www.bbc.com/news/technology-33347866.

[15] Austin Blanton et al. "A Comparison of Human and Automated Face Verification Accuracy on Unconstrained Image Sets". In: (2016).

[16] Andreas Blass, Nachum Dershowitz, and Yuri Gurevich. *When Are Two Algorithms the Same?* 2009. URL: http://arxiv.org/PS_cache/arxiv/pdf/0811/0811.0811v1.pdf.

[17] Andreas Blass and Yuri Gurevich. "Algorithms: A Quest for Absolute Definitions". In: *Bulletin of European Association for Theoretical Computer Science* 81 (2003).

[18] J. Bonnefon, A. Shariff, and I. Rahwan. "The social dilemma of autonomous vehicles". In: *Science (New York, N.Y.)* 352 (2016). ISSN: 0036-8075.

[19] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. "Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?" In: ([In press]).

[20] Philip Brey. "Ethical aspects of facial recognition systems in public places". In: *Journal of Information, Communication and Ethics in Society* 2.2 (2004), pp. 97–109. ISSN: 1477-996X. DOI: 10.1108/14779960480000246.

[21] Philip Brey and Johnny Hartz Søraker. "Philosophy of Computing and Information Technology". In: *Philosophy of Technology and Engineering Sciences* (2009), pp. 1341–1407. DOI: 10.1016/B978-0-444-51667-1.50051-3.

[22] Seth Flaxman Bryce Goodman. "EU regulations on algorithmic decision-making and a "right to explanation"". In: *ICML Workshop on Human Interpretability in Machine* (2016). URL: http://arxiv.org/abs/1606.08813v1.

[23] J. Burrell. "How the machine 'thinks: Understanding opacity in machine learning algorithms". In: *Big Data & Society* 3.1 (2016). ISSN: 2053-9517. DOI: 10.1177/2053951715622512.

[24] J. L. (Ed.) Chabert et al. *A history of algorithms: from the pebble to the microchip*. Springer Science & Business Media., 1999.

[25] J. Cheney-Lippold. "Jus Algoritmi: How the National Security Agency Remade Citizenship". In: (2016).

[26] Lars Rune Christensen. "Stigmergy in human practice: Coordination in construction work". In: *Cognitive Systems Research* 21 (2013), pp. 40–51. ISSN: 13890417. DOI: 10.1016/j.cogsys.2012.06.004.

[27] Alonzo Church. "A Note on the Entscheidungsproblem". In: *Journal of Symbolic Logic* 1 (1936).

[28] Jack B. Copeland. "Computation". In: *The Blackwell Guide to the Philosophy of Computing and Information*. Ed. by Luciano Floridi. Blackwell Publishing Ltd, 2004, pp. 3–17.

[29] Jack B. Copeland. *The Church-Turing Thesis*. Ed. by Edward N. Zalta. 2015. URL: http://stanford.library.sydney.edu.au/archives/sum2014/entries/church-turing/.

[30] Kate Crawford. "Can An Algorithm Be Agonistic? Ten Scenes of Contest in Calculated Publics". In: *Science, Technology & Human Values* 41 (2016).

[31] Daniel V. McGehee et al. *Review of Automated Vehicle Technology: Policy and Implementation Implications*. 2016.

[32] W. H. Dean. "What Algorithms Could Not Be: Doctoral dissertation, Rutgers University-Graduate School-New Brunswick". In: (2007).

[33] Deloitte. *From Brawn to Brains - The Impact of Technology on Jobs in the UK*. 2015. URL: http://www2.deloitte.com/uk/en/pages/growth/articles/from-brawn-to-brains--the-impact-of-technology-on-jobs-in-the-u.html.

[34] Deloitte. *Technology and People - The Great Job-Creating Machine*. 2016. URL: http://www2.deloitte.com/uk/en/pages/finance/articles/technology-and-people.html.

[35] Nicholas Diakopoulos. "Accountability in Algorithmic Decision-Making". In: *Queue* 13 (2015).

[36] Nicholas Diakopoulos. "Algorithmic Accountability". In: *Digital Journalism* 3.3 (2014), pp. 398–415. ISSN: 2167-0811. DOI: 10.1080/21670811.2014.976411.

[37] Margery J. Doyle and Leslie Marsh. "Stigmergy 3.0: From ants to economies". In: *Cognitive Systems Research* 21 (2013), pp. 1–6. ISSN: 13890417. DOI: 10.1016/j.cogsys.2012.06.001.

[38] Michael C. Dunne and Renfrey B. Potts. "Algorithm for Traffic Control". In: (1964).

[39] Robert Epstein and Ronald E. Robertson. "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.33 (2015), E4512–21. ISSN: 1091-6490. DOI: 10.1073/pnas.1419828112.

[40] Nicholas Ernest and David Carroll. "Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions". In: *Journal of Defense Management* 06.01 (2016). ISSN: 21670374. DOI: 10.4172/2167-0374.1000144.

[41] Martín Escardó. *Foundations of Computer Science*. Birmingham, January 11th 2005.

[42] Motahhare Eslami et al. ""I always assumed that I wasn't really that close to [her]"". In: (2015), pp. 153–162. DOI: 10.1145/2702123.2702556.

[43] Executive Office of the President. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. 5-2016.

[44] John Forge. "A note on the definition of "dual use"". In: *Science and engineering ethics* 16.1 (2010), pp. 111–118. ISSN: 1471-5546. DOI: 10.1007/s11948-009-9159-9.

[45] Fox Business. *Fmr. McDonald's USA CEO: $35KRobots Cheaper Than Hiring at $15 Per Hour*. 2016. URL: http://www.foxbusiness.com/features/2016/05/24/fmr-mcdonalds-usa-ceo-35k-robots-cheaper-than-hiring-at-15-per-hour.html.

[46] Carl Benedict Frey and Michael A. Osborne. "The Future of Employment: How susceptible are jobs to computerisation." In: (2013).

[47] Judith Gal-Ezer and Ela Zur. "The efficiency of algorithms—misconceptions". In: *Computers & Education* 42.3 (2004), pp. 215–226. ISSN: 03601315. DOI: 10.1016/j.compedu.2003.07.004.

[48] GCCS. *The Ethics of Algorithms: from radical content to self-driving cars: Final Draft Background Paper*. URL: https://www.gccs2015.com/sites/default/files/documents/Ethics_Algorithms-final%20doc.pdf.

[49] Gonzalo Genova and M. Rosario Gonzalez. "Teaching Ethics to Engineers: A Socratic Experience". In: *Science and engineering ethics* 22.2 (2016), pp. 567–580. ISSN: 1353-3452. DOI: 10.1007/s11948-015-9661-1.

[50] Tarleton Gillespie. "The Relevance of Algorithms". In: *Media Technologies* (Forthcoming).

[51] Lisa Gitelman. *Raw Data is an Oxymoron*. MIT Press, 2013.

[52] A. Goldberg. "In defense of forensic social science". In: *Big Data & Society* 2.2 (2015). ISSN: 2053-9517. DOI: 10.1177/2053951715601145.

[53] Dina Goldin and Peter Wegner. "The Church-Turing Thesis: Breaking the Myth". In: *Conference on Computability in Europe* (2005).

[54] Noah J. Goodall. "Machine Ethics and Automated Vehicles". In: (), pp. 93–102. DOI: 10.1007/978-3-319-05990-7{\textunderscore}9.

[55] Don Gotterbarn and James Moor. "Virtual decisions". In: *ACM SIGCAS Computers and Society* 39.3 (2009), pp. 27–42. ISSN: 00952737. DOI: 10.1145/1713066.1713068.

[56] Government Accountability Office. *Face Recognition Technology: FBI Should Better Ensure Privacy and Accuracy: FBI Should Better Ensure Privacy and Accuracy: Report to the Ranking Member, Subcommittee on Privacy, Technology and the Law, Committee on the Judiciary, U.S. Senate*. 2016.

[57] J. P. Gownder et al. *The Future Of Jobs, 2025: Working Side By Side With Robots: Automation Won't Destroy All The Jobs, But It Will Transform The Workforce — Including Yours*. 2015.

[58] James Grimmelmann. "Some Skepticism About Search Neutrality". In: *The Next Digital Decade: Essays on the Future of the Internet* (2010).

[59] Yuri Gurevich. *What is an Algorithm?* 2014. URL: http://research.microsoft.com/en-us/um/people/gurevich/Opera/209a.pdf.

[60] Wolfgang Hatzack and Bernhard Nebel. "The Operational Traffic Control Problem: Computational Complexity and Solutions". In: ().

[61] Patrick Heavey. "Synthetic biology ethics: a deontological assessment". In: *Bioethics* 27.8 (2013), pp. 442–452. ISSN: 0269-9702. DOI: 10.1111/bioe.12052.

[62] D. Hilbert and Ackermann. W. *The Principles of Mathematical Logic*. Chelsea Publishing Company, 1950.

[63] Peter Buell Hirsch. "How to pitch an algorithm". In: *Journal of Business Strategy* 36.4 (2015), pp. 56–59. ISSN: 0275-6668. DOI: 10.1108/JBS-05-2015-0047.

[64] Bjorn Hofmann et al. "Integrating ethics in health technology assessment: many ways to Rome". In: *International journal of technology assessment in health care* 31.3 (2015), pp. 131–137. ISSN: 0266-4623. DOI: 10.1017/S0266462315000276.

[65] Harald Holone. "The filter bubble and its effect on online personal health information". In: *Croatian Medical Journal* 57.3 (2016), pp. 298–301. ISSN: 0353-9504. DOI: 10.3325/cmj.2016.57.298.

[66] Brad Hooker. "Ethics in Conflict". In: *Science and Technology Ethics*. Ed. by Raymond E. Spier. Routledge, 2002, pp. 89–106.

[67] Kartik Hosanagar et al. *Will the Global Village Fracture into Tribes? Recommender Systems and their Effects on Consumer Fragmentation*. 2013.

[68] Sherif E. Hussein and Mahmoud Abo El-Nasr. "Resources Allocation in Higher Education based on System Dynamics and Genetic Algorithms". In: *International Journal of Computer Applications* 77, No. 10 (2013).

[69] Y.S. Hwang et al. *Categorization of search results*. US Patent 8,498,984. 2013. URL: https://www.google.com/patents/US8498984.

[70] Lucas D. Introna. "Shaping the Web: Why the Politics of Search Engines Matters". In: *Ethics and Information Technology* 2.1 (2000), pp. 1–2. ISSN: 13881957. DOI: 10.1023/A:1010037326272.

[71] Iyengar, Siriam M. Svirbeli, J. R. "The Medical Algorithms Project". In: (2009).

[72] D. Jungnickel. *Graphs, networks and algorithms*. Springer Science & Business Media, 2006.

[73] Natascha Just and Michael Latzer. "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet". In: *Media, Culture & Society* (2016).

[74] Noel Kalicharan. *Learn to program with C*. [Berkeley, CA] and New York, NY: Apress and Distributed to the book trade worldwide by Springer, 2015.

[75] T. Karppi and K. Crawford. "Social Media, Financial Algorithms and the Hack Crash". In: *Theory, Culture & Society* 33.1 (2015), pp. 73–92. ISSN: 0263-2764. DOI: 10.1177/0263276415583139.

[76] Hyun-jung Kim and Kyung-shik Shin. "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets". In: *Applied Soft Computing* 7.2 (2007), pp. 569–576. ISSN: 15684946. DOI: 10.1016/j.asoc.2006.03.004.

[77] Takashi Kimoto and Kazuo Asakawa. "Stock market prediction system with modular neural networks". In: (1990).

[78] Rob Kitchin. "Thinking critically about and researching algorithms". In: *Information, Communication & Society* (2016), pp. 1–16. ISSN: 1369-118X. DOI: 10.1080/1369118X.2016.1154087.

[79] Migga Joseph Kizza. *Ethical and Social Issue in the Information Age*. 4th ed. Springer, 2010.

[80] Donald Knuth. *The Art of Programming: Volume 1: Fundamental Algorithms*. Addison-Wesley, 1973.

[81] Felicitas Kraemer, Kees van Overveld, and Martin Peterson. "Is there an ethics of algorithms?" In: *Ethics and Information Technology* 13.3 (2011), pp. 251–260. ISSN: 1388-1957. DOI: 10.1007/s10676-010-9233-7.

[82] Mario Krenn et al. "Automated Search for new Quantum Experiments". In: *Physical Review Letters* 116.9 (2016). ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.116.090405.

[83] Fuyuki Kurasawa. *The Algorithmic Shift: Theorizing the Social Implications of Big Data and Crowdsourcing*. 2014. URL: https://isaconf.confex.com/isaconf/wc2014/webprogram/Paper50155.html.

[84] Jaron Lanier. *Who Owns the Future?* Simon and Schuster, 2014.

[85] S. Lash. "Power after Hegemony: Cultural Studies in Mutation?" In: *Theory, Culture & Society* 24.3 (2007), pp. 55–78. ISSN: 0263-2764. DOI: 10.1177/0263276407075956.

[86] Ted G. Lewis. "Cognitive stigmergy: A study of emergence in small-group social networks". In: *Cognitive Systems Research* 21 (2013), pp. 7–21. ISSN: 13890417. DOI: 10.1016/j.cogsys.2012.06.002.

[87] Iacopo Masi et al. "Do We Really Need to Collect Millions of Faces for Effective Face Recognition?" In: (2016).

[88] Matt McGee. *EdgeRank Is Dead: Facebook's News Feed Algorithm Now Has Close To 100K Weight Factors*. 2013. URL: http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908.

[89] Seumas Miller, Michael J. Selgelid, and Koos van der Bruggen. *Report on Biosecurity and Dual Use Research*. 2011.

[90] Marcus Morgan. "Humanising Sociological Knowledge". In: *Social Epistemology* (2016), pp. 1–17. ISSN: 0269-1728. DOI: 10.1080/02691728.2015.1119911.

[91] Evgeny Morozov. *Don't Be Evil*. 2011. URL: https://newrepublic.com/article/91916/google-schmidt-obama-gates-technocrats.

[92] Yiannis N. Moschovakis. "What is an Algorithm?" In: *Mathematics unlimited–2001 and beyond* (2001).

[93] Chantal Mouffe. "Deliberative Democracy or Agonistic Pluralism?" In: *Social Research* (1999).

[94] C. E. Mutlu. "Of Algorithms, Data and Ethics: A Response to Andrew Bennett". In: *Millennium - Journal of International Studies* 43.3 (2015), pp. 998–1002. ISSN: 0305-8298. DOI: 10.1177/0305829815581536.

[95] Vincent C. Müller and Thomas W. Simpson. "Autonomous Killer Robots Are Probably Good News". In: (2014).

[96] National Highway Traffic Safety Administration. *U.S. Department of Transportation Releases Policy on Automated Vehicle Development*. 2013. URL: http://www.nhtsa.gov/About+NHTSA/Press+Releases/U.S.+Department+of+Transportation+Releases+Policy+on+Automated+Vehicle+Development.

[97] NETMARKETSHARE. 2016. URL: https://www.netmarketshare.com/.

[98] Folorunsho Olaiya and Adesesan Barnabas Adeyemo. "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies". In: *International Journal of Information Engineering and Electronic Business* 4.1 (2012), pp. 51–59. ISSN: 20749023. DOI: 10.5815/ijieeb.2012.01.07.

[99] Eli Pariser. *Did Facebook's Big New Study Kill My Filter Bubble Thesis? Not really. Let's dive into it and see why not.* 2015.

[100] Eli Pariser. *The Filter Bubble: What the Internet is hiding from you*. New York: Penguin, 2012. ISBN: 978-0241954522.

[101] Frank A. Pasquale. "The Algorithmic Self". In: *The Hedgehog Review* 7.1 (2015).

[102] Patrick Lin, George Bekey, and Keith Abney. "Autonomous Military Robotics: Risk, Ethics, and Design". In: (2008).

[103] Martin Peterson and Andreas Spahn. "Can technological artefacts be moral agents?" In: *Science and engineering ethics* 17.3 (2011), pp. 411–424. ISSN: 1353-3452. DOI: 10.1007/s11948-010-9241-3.

[104] Thomas Prevot et al. "Toward Automated Air Traffic Control—Investigating a Fundamental Paradigm Shift in Human/Systems Interaction". In: *International Journal of Human-Computer Interaction* 28.2 (2012), pp. 77–98. ISSN: 1044-7318. DOI: 10.1080/10447318.2012.634756.

[105] Shiwani Rana and Roopali Garg. "Evaluation of Student's Performance of an Institute Using Clustering Algorithms". In: *International Journal of Applied Engineering Research* 11, No. 5 (2016).

[106] David B. Resnik. "What is d̈ual useëresearch? A response to Miller and Selgelid". In: *Science and engineering ethics* 15.1 (2009), pp. 3–5. ISSN: 1471-5546. DOI: 10.1007/s11948-008-9104-3.

[107] Michael Rundle. *Facebook's political 'echo chamber' is your fault, not theirs*. 2015. URL: http://www.wired.co.uk/article/facebook-echo-chamber-study.

[108] Fiona Rutherford and Alan White. *This Is Why Some People Think Google's Results Are "Racist"*. 2016. URL: https://www.buzzfeed.com/fionarutherford/heres-why-some-people-think-googles-results-are-racist?utm_term=.kueMWP7yA#.jmdej0y8b.

[109] Christian Sandvig, Kevin Hamilton, and Karrie Karahalios. *Can an Algorithm be Unethical?* 2015.

[110] Christian Sandvig et al. "Re-Centering the Algorithm". In: *Governing Algorithms: A Conference on Computation, Automation and Control* (2013).

[111] Walter Sinnott-Armstrong. *Consequentialism*. Ed. by Edward N. Zalta. 2015. URL: http://plato.stanford.edu/entries/consequentialism/.

[112] T. Striphas. "Algorithmic culture". In: *European Journal of Cultural Studies* 18.4-5 (2015), pp. 395–412. ISSN: 1367-5494. DOI: 10.1177/1367549415577392.

[113] Cass Sunstein. "Is Deontology a Heuristic? On Psychology, Neuroscience, Ethics and Law". In: (2013). URL: https://dash.harvard.edu/bitstream/handle/1/13548959/sunstein_is_deontology_a_heuristic.pdf?sequence=1.

[114] Ioan Susnea. "Engineering Human Stigmergy". In: *International Journal of Computers Communications & Control* 10(3) (2015), pp. 420–427.

[115] Tsjalling Swierstra and Arie Rip. "Nano-ethics as NEST-ethics: Patterns of Moral Argumentation About New and Emerging Science and Technology". In: *NanoEthics* 1.1 (2007), pp. 3–20. ISSN: 1871-4757. DOI: 10.1007/s11569-007-0005-8.

[116] *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*. January 2016.

[117] Jean A. Thomas. "Deontology, Consequentialism and Moral Realism". In: *Minerva* 19 (2015).

[118] Alan Mathison. Turing and B. Jack Copeland. *The Essential Turing*. Oxford University Press, 2004.

[119] Raymond Turner. *The Philosophy of Computer Science*. 2016. URL: http://stanford.library.sydney.edu.au/archives/sum2014/entries/computer-science/#Compu.

[120] Allie VanNest. *Facebook Continues to Beat Google in Sending Traffic to Top Publishers*. 2015. URL: http://blog.parsely.com/post/2855/facebook-continues-to-beat-google-in-sending-traffic-to-top-publishers/.

[121] Peter-Paul Verbeek. "Expanding Mediation Theory". In: *Foundations of Science* 17.4 (2012), pp. 391–395. ISSN: 1233-1821. DOI: 10.1007/s10699-011-9253-8.

[122] Peter-Paul Verbeek. "Persuasive Technology and Moral Responsibility: Toward an ethical framework for persuasive technologies". In: *Persuasive* 6 (2006).

[123] Peter-Paul Verbeek. *What Things Do: Philosophical reflections on technology, agency, and design*. Penn State University Press, 2005.

[124] Amy van Wijnsberghe and Michael Nagenborg. "Civilizing Drones by Design". In: Routledge, 2016. Chap. 8.

[125] Michael J. Wilber, Vitaly Shmatikov, and Serge Belongie. "Can we still avoid automatic face detection?" In: (2016).

[126] Hai Yang and Sam Yagar. "Traffic Assignment and Signal Control in Saturated Road Networks". In: *Elsevier Science Ltd* 29A, No. 2 (1995).

[127] N. S. Yanofsky. "Towards a Definition of an Algorithm". In: *Journal of Logic and Computation* 21.2 (2011), pp. 253–286. ISSN: 0955-792X. DOI: 10.1093/logcom/exq016.

[128] M. Ziewitz. "Governing Algorithms: Myth, Mess, and Methods". In: *Science, Technology & Human Values* 41.1 (2016), pp. 3–16. ISSN: 0162-2439. DOI: 10.1177/0162243915608948.