

**Mindless Scrolls:
Social Kinds and Power in the Algorithmic Society**

By

László Kőszeghy

Submitted to

Central European University

Department of Philosophy

In partial fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisor: Professor Maria Kronfeldner

Vienna, Austria

2024

COPYRIGHT NOTICE

Copyright © Laszlo Koszeghy, 2024. Mindless Scrolls: Social Kinds and Power in the Algorithmic Society - this work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International license.

For bibliographic and reference purposes, this dissertation should be referred to as: Koszeghy, L. 2024. Mindless Scrolls: Social Kinds and Power in the Algorithmic Society. Doctoral dissertation, Philosophy Department, Central European University, Vienna.

I hereby declare that the dissertation contains no materials accepted for any other degrees in any other institutions and no materials previously written and/or published by another person unless appropriate acknowledgement is made in the form of a bibliographical reference.

László Kószeghy
The Hague, 28th September 2024

Abstract

This dissertation explains a contemporary shift in the conditions of existence of social reality. Using the conceptual framework of social ontology, I argue that our contemporary social reality depends on our minds – both collective and individual intentionality – in a weaker sense than the traditional picture of social reality holds. Specifically, I establish the following: (i) social kinds constructed through data-driven algorithmic systems and forms of algorithmic power are not dependent on collective intentionality; (ii) algorithmic nudging power is neither dependent nor is enhanced by collective intentionality; and (iii) some forms of algorithmic nudging do not require intentional actions from those over whom it is exercised. As such, this work explores the social ontological implications of the shift towards “algorithmic societies” – how our social reality is becoming “mindless”.

The dissertation consists of two parts. In Part I (Chapters 1 & 2), I focus on the question of how social kinds, including kinds constructed through data-driven algorithmic systems, are dependent on minds. In Part II (Chapters 3 & 4), I focus on how forms of power enabled by data-driven algorithmic systems relate to our traditional understanding of social power.

In Chapter 1, I introduce in detail the traditional picture of social reality, which I base on Searle’s theory of social construction and call Intentionalism. Intentionalism posits that the existence of social kinds depends on their being collectively recognized as existing. The question driving this chapter is whether Intentionalism has general coverage over social reality. After unpacking the most charitable form of Intentionalism by considering criticism the view has received, I conclude that Intentionalism must weaken the relevant senses of both intentional states and dependence and restrict its scope to *human* social reality.

In Chapter 2, I argue for two claims. First, socially significant real-time classifications constructed through data-driven algorithmic systems – which I will call Modulatory Algorithmic Kinds (MAKs) – qualify as “social kinds”. This is because, given their multiple shared properties, they can be effectively used for predictions. MAKs, then, qualify as social as they are mind-dependent, representation-dependent, and have large-scale, group-level consequences for social reality. Second, MAKs as social kinds challenge Intentionalism. This is because MAKs are not dependent on collective recognition: the relevant “social glue”, instead, is representation through digital data and algorithmic code.

In Chapter 3, I explain the forms of algorithmic power enabled by data-driven systems. I first define a minimal notion of social power, clarifying the relationship between social status, intentional action, and social power. I then examine two versions of what I call the Intentionality Constraint: power depending on intentional action and power depending on collective recognition. While I accept the first with caveats, I show that power exercised through data-driven systems challenges the second. Finally, I argue that exercising “nudging”

power through data-driven algorithmic systems is a form of social power which neither depends on, nor is enhanced by, collective recognition.

In Chapter 4, given how Foucault's notion of productive power is often contrasted with algorithmic power citing the lack of reflection in the latter, I ask in what sense algorithmic power is not productive. After unpacking productive power and its contrast with intimidation, I specify the sense of "reflection" that those over whom – non-productive – algorithmic power is exercised are lacking. I argue that reflection as the intentional states necessary for intentional action is required for the exercise of both productive power and intimidation, and identify behavioural nudging as the form of algorithmic power which can be exercised without this sense of reflection – thus being non-productive and distinct from intimidation.

Acknowledgement

At the outset, I would like to thank my supervisor, Maria Kronfeldner. Maria is a consistently efficient, uniquely creative, and deeply analytical philosopher as well as a caring and responsible mentor who has guided me throughout the slow and incremental process of approximating the ideal of a relentless and intellectually honest thinker. Among many things, one important glue which, I believe, formed the two of us into a good team is our shared meta-philosophical views. I agree with Maria's *ars poetica* which holds that philosophy's beauty lies in being the simplest – yet often most sincere – of arts. One is engineering, devising, creating, dissecting concepts – an activity for which, as she often puts it, pen and paper amply suffices. However, we also agreed that the stakes are high: philosophy, even its most radical forms, is a quest for truth. Until the very moment of my thesis submission, I kept learning thanks to her meticulous – if often crushing – feedback; I am certain that the meta-skills I have developed under Maria's supervision have enriched me in the noblest sense of the word.

I am, furthermore, thankful for the opportunity of having been her research assistant for three years as well as for her practical and emotional support when CEU surprised PhD students with an unexpected self-employed status – not always the most desirable social kind to belong to as a student. I remember taking a stroll together in the botanical garden next to the Belvedere at the beginning of the third year of my PhD program; I mentioned to her that I am having doubts as to whether I want to continue with academic philosophy right after my PhD – the encouragement and understanding I received is something I will always appreciate.

I would also like to thank my first-year advisor, David Weberman. Already as my MA thesis supervisor, he instilled in me the importance of making philosophy as simple and understandable as possible. This credo has guided my nearly six-year long teaching career in philosophy. As an advisor, David was uniquely insightful in pinpointing the crux of the issue in the content of my writing. I also enjoyed our discussions of Jewish culture in Detroit, Coen brothers as well as Charlie Kaufman movies, and Gadamer's ideas about our "flickering in the closed circuits of historical life". Let alone the conversation we had over beer after listening to Jacques Rancière rambling on about the politics of landscapes in French at a theater in Alsergrund. The last time I bumped into David, he told me: "I look forward to learning about where you end up after the PhD. You have so much to offer, Laci." These sentences have stayed with me. I hope you get better soon, dear David.

I can confidently say that, throughout the years I spent at CEU's Philosophy Department – first as an MA student and later as a PhD candidate – I have learned and gained insights from every faculty member as well as our past and present coordinators. I would, nevertheless, highlight one more person. I am grateful for the immensely useful discussions I had with Asya Passinsky about my thesis idea and drafts of certain chapters, as well as for her memorable, exciting

and inspiring courses in social ontology. Asya’s clarity of thought and knowledge of the extant literature yielded invaluable and helpful feedback on my thought-pieces.

I would also like to acknowledge the insights I have gained from the conferences and workshops I have attended throughout my PhD. Most notably, this includes the participants and organizers of the “Social Metaphysics Workshop” in 2022 at Duke University, the annual International Social Ontology Society conferences in 2022 at the University of Vienna and in 2023 at Stockholm University, and the “Social Ontology Faces the Future” conference at Boğaziçi University in Istanbul in 2024. I am also grateful for the insights I gained at the “Social Ontology of Power – Contemporary Perspectives” workshop, held at the Institute for Futures Studies in Stockholm in 2024, which I co-organized with Åsa Burman, Andrés Hurtado, and Cyrill Miksch.

And, of course, my dearest friends. Both my fellow PhD students, Emily, Fumi, Jamie, Kerem, Maria, Pelin, Vale, Vladimir, and many others, as well as my friends from outside philosophy, Andor, Atti, Ádám, Dani, Hamit, Soma, Szüzson, Ye In and others, kept me sane during times when I wanted to do anything but write a social ontological analysis of algorithmic societies.

Finally, my heartfelt gratitude also goes to my partner, Anna. She has stood by me from the beginning to the end of the most challenging – difficult, complex, and, at times, utterly exhausting – project I have ever undertaken. It is clear and simple that without Anna’s support and love, I could not have done it. Köszönöm szépen, Anna.

Table of Contents

INTRODUCTION	1
From collective intentionality to mindless scrolling	1
Ideal and nonideal social ontology	9
The mark of the social at a crossroads	13
Methodology, interdisciplinarity, and historical shifts.....	16
Brief chapter by chapter overview	20
PART I: SOCIAL KINDS AND THEIR MIND-DEPENDENCE	23
CHAPTER 1: RECOGNITION AND EXISTENCE: THE TRADITIONAL PICTURE OF SOCIAL REALITY	24
Introduction: The social and its puzzles	24
1.1. From mind-dependency to Intentionalism.....	27
1.2. Challenges to Intentionalism	42
Concluding remarks: Intentionalism revised	56
CHAPTER 2: BEYOND RECOGNITION: MODULATORY ALGORITHMIC KINDS AS SOCIAL KINDS	58
Introduction: Social ontology and blind spots.....	58
2.1. Marriages and their mind-dependency	61
2.2. From ancient bureaucracies to modulatory algorithms	65
2.3. Beyond recognition: The mind-dependency of MAKs	74
Concluding Remarks: Digital data and algorithmic code as a new glue of social reality	91
PART II: SOCIAL POWER AND ITS SUBJECTS	93
CHAPTER 3: BROADENING THE CONCEPT OF SOCIAL POWER: ALGORITHMIC FORMS OF SOCIAL POWER	94
Introduction: Social power, intentional status and social status	94
3.1. A minimal concept of social power.....	96
3.2. Intentionality constraints	103
3.3. Soft but powerful: Modulatory predictive nudging	118
Concluding remarks: Towards a subjectless power.....	122
CHAPTER 4: ALGORITHMIC POWER WITHOUT SUBJECT-PRODUCTION OR INTIMIDATION: THREE NOTIONS OF REFLECTION	125
Introduction: Foucault, subject production and algorithmic power	125
4.1. In what sense does power produce subjects?	127
4.2. In what sense does algorithmic power not produce subjects?.....	135
Concluding remarks: Lack of reflection on all fronts.....	154
CONCLUSION	157
BIBLIOGRAPHY	163

INTRODUCTION

*The rational sociality of the contract, dialectical sociality (that of the State and of civil society, of public and private, of the social and the individual) gives way to the sociality of contact, of the circuit and transistorised network of millions of molecules and particles maintained in a random gravitational field, magnetised by the constant circulation and the thousands of tactical combinations which electrify them.
But is it still a question of the socius?"*

Jean Baudrillard: *In the Shadow of the Silent Majorities... or the End of the Social*
(1983, 83; emphasis in the original)

From collective intentionality to mindless scrolling

You “mindlessly” scroll your social media feed, find yourself watching recommended videos on for hours, end up buying items that were suggested to you, wind up listening to songs that you did not choose. Your job or university application is automatically rejected in the first round, your insurance premium is adjusted based on your online behavior, your loan application is approved within minutes. Depending on where you live, even your “trustworthiness” may be assessed by the state.

In the following, I will be interested in the philosophical implications of such examples where social reality is quantified and classified through what I will call data-driven algorithmic systems. These systems – though they differ in important ways – include social media algorithms, recommendation systems, targeted advertising systems, music streaming algorithms, automated hiring systems, insurance risk assessment algorithms, loan approval algorithms, and social credit systems. The introduction of data-driven algorithmic systems into social reality, as I will show, brings about a historical shift towards “algorithmic societies”. My goal, then, is to better understand the ontology of our social reality as it shifts towards an algorithmic society.

Specifically, I will ask the following questions. What are the conditions of existence of the socially consequential classifications constructed through data-driven algorithmic systems? Do these classifications qualify as what philosophers call *social kinds*? Furthermore, are the ways in which various data-driven algorithmic systems affect us qualify as exercises of forms of *social power*? If so, what forms of social powers are being exercised upon us by these systems?

But let us start at the beginning.

*

Social ontology is a branch of philosophy which explores the conditions of existence of social reality in its most general forms. It raises questions about how elements of social reality such as money, marriage and gender exist. Within analytic philosophy, arguably the most famous formulation of one of the central questions of social ontology is provided by John Searle, being “puzzled” by how...

...there are portions of the real world, objective facts in the world, that are only facts by human agreement. In a sense there are things that exist only because we believe them to exist. I am thinking of things like money, property, governments, and marriages. [...] How is a socially constructed reality possible? (Searle 1995, 1–2)

Searle’s question, which echoes Kant’s famous how-possible questions, is rooted in a long-standing tradition of philosophical thought. This line of questioning can be traced back, for example, through the work of Margaret Gilbert on “social facts” (1992), and even further to Georg Simmel’s essay, “How is Society Possible?” (1910). Such “how-possible” questions assume that there are mind-independent and mind-dependent parts of reality and aim to provide an ontological explanation of how social reality can belong to the latter. In other words, their aim is to make sense of the idea that social reality, while being *real*, is, in some sense, *mind-dependent*: money, marriage and gender would not exist in a world without minds.

As such, it seems to me that much of social ontology is an exciting contribution to the age-old philosophical debate about the relationship between reality and our minds. Let’s unpack this a little.

In the most influential tradition in Western philosophy, reality is often viewed through the fundamental distinction between subjects and objects. While subjects have minds and a sense of self, objects have no minds, and, therefore, no sense of self. Unsurprisingly, humans are often cited as paradigmatic examples of being a subject, while examples for objects in the Western philosophical tradition are more diverse, with the most popular ones probably being billiard balls, inkwells, tables, and pieces of rock.

Surely, once probed more in detail, classifying things as subjects or objects and even the distinction between subject and object may become more and more difficult to clearly delineate. Take the case of animals or plants – or for that matter, what I will call data-driven algorithmic systems. Luckily, I do not need to settle these complex issues: here, I will simply

take humans to be subjects in the above-mentioned sense and remain neutral on what other beings may also qualify as subjects.

Now, the feature of the subject's mind which has long been in the center of philosophers' attention is its *intentionality*: its capacity to *direct* its states – such as its beliefs, emotions and desires – at the external world and thereby *represent* it. Representing reality, then, means being a stand-in for the relevant part of the external world. The states of the mind which represent reality are, in turn, known as *intentional states*. Take one of Searle's examples (1995, 9, 33): one's belief that "this is a mountain" is an intentional state which serves as a stand-in for the actual mountain in the external world, in other words, representing it. One important line of philosophical inquiry, then, has been driven by epistemological interest: exploring how intentional states relate to reality, more specifically, whether representations through our intentional states can ever correspond to reality.

Many, however, have been critical of the focus of philosophers on the epistemological question of correspondence with reality. Richard Rorty, for instance, famously criticized viewing the mind "as a great mirror", arguing that "the picture which holds traditional philosophy captive is that of the mind as a great mirror, containing various representations – some accurate, some not" (1979, 12). Those critical of the focus on correspondence between representations through our intentional states and reality aim to emphasize that the intentional states of human minds are characterized not only by the capacity to represent (more or less accurately) a pre-existing mind-independent reality. Intentional states, these critics argue, also, or perhaps rather, have the capacity to actively *construct* reality.

Accordingly, I am assuming that it makes sense to draw a distinction between intentional states, such as my belief that "this is a mountain", which merely represents a pre-existing reality, and intentional states which also construct a reality through representing it. For now, let's refer to this latter type of representing which is also a form of construction as representing *in a special way*. Thus, for the sake of simplicity, I will frame issues relying on the contrast between mind-dependent and mind-independent parts of reality, assuming there are indeed examples, such as mountains, which illustrate the latter. This means that I am setting aside more radical perspectives for now, such as a Rorty's which argue that all reality is mind-dependent – and therefore constructed. While I will not consider such radical perspectives, the points I will make throughout this dissertation will be of relevance even to them, as I will draw distinctions *between* various ways in which social reality can depend on minds.

This is how we arrive at the idea of a mind-dependent reality constructed through intentional states – and this is where social ontology enters the picture. In social ontology, it seems to be rather uncontroversial to hold that the intentionality of the human mind, far from being only a passive mirror reflecting the pre-existing mind-independent reality, indeed has the capacity

to construct a mind-dependent *reality*. Mountains would exist even if we did not form representations about them through intentional states such as beliefs; however, as opposed to the intentional state of believing that “this is a mountain”, the belief that “this is money”, “this is marriage” or, say, “this is a man”, in the right context, represents money, marriage and being a man in a special way which at the same time constructs the represented entities as such.

Social reality, according to the traditional picture of social ontology, turns out to be a “great mirror” itself – just not of the world “out there”, but the world “in here”. Already in the 18th century, the Neapolitan historian-philosopher Giambattista Vico argued that the fact that “the civil world itself has certainly been made by men” implies that “its principles therefore can, because they must, be rediscovered within the modifications of our own human mind. (Vico 1974 [1744], 331, quoted by Guala 2010, 245). The claim is not merely that we “must”, in some sense, *know* the “principles” of social reality but also that the *existence* of the “principles” of social reality *depend on* “the modifications of our own human mind”, or as we have called them, our intentional states.

This means that when social ontologists ask how elements of social reality such as money, marriage, and gender exist, they are asking: how exactly do our intentional states construct these entities? Money is not simply a piece of paper or metal, married people are not merely people living together or loving each other (perhaps especially not), and gender is not simply the presence of a set of biological characteristics. Money, marriage, and gender are kinds of entities, which, as I will explain in detail, philosophers often refer to as *social kinds*, whose members have a *social status*. For instance, pieces of paper or pieces of metal have the status of being a means of exchange and store of value, married people have the status of, say, having the right to joint tax returns or the responsibility of having shared debt, and gendered individuals have a status which defines how they are expected to behave.

These social kinds, and their respective statuses, it is argued, would not exist without subjects who have the relevant intentional states representing them. These intentional states, then, are not merely those of an individual subject but those of *many* subjects which are, in some form, related, constituting what is usually referred to as *collective intentionality*. It is, as I will show, this collective intentionality which is considered necessary for the existence of social kinds like money, marriage, or gender, along with their respective statuses. In other words, for money, marriage and gender to exist, we need to collectively have intentional states representing the relevant kinds and their associated statuses as existing.

We now have the terminology to formulate the unique insight of the traditional picture of social ontology. It is to have shown that social reality is a mind-dependent part of reality which is constructed through the representational capacities of the intentionality of the minds of many subjects, that is, through what is referred to as collective intentionality.

The existence of social kinds such as money, marriage, or gender, in turn, makes possible a new set of intentional actions – actions which require the intentional states representing the actions as existing. In our case, the relevant intentional actions are those which require intentional states representing the relevant social kinds (with their associated status) as existing. For instance, once there is some form of collective intentionality representing money, marriage, a given gender as existing, one can then buy or sell, wed or divorce, or live up to or challenge their gender status.

Understanding how social kinds and intentional actions relate to one another has been held key to model how *power relations* exist in social reality. What new ways of getting other subjects to do what one wants them to do are enabled once certain social kinds are constructed? How is social power related to various forms of intentionality – both individual and collective? Social ontology has aimed to shed light on these questions, emphasizing the crucial role of the subject’s intentionality in the existence and/or exercise of social power; both the role of collective intentionality in the construction of the social kinds, and their respective statuses, on which the exercise of power is often said to depend, as well as the intentionality of the individual subject in the exercise of power itself (cf. Searle 1995, 2010; Burman 2007, 2023).

We have seen that it is often held that the intentionality of subjects, more precisely, collective intentionality is necessary for social reality. Let’s call, then, collective intentionality the *building block* of social reality. Collective intentionality, as a building block, is held to be required to construct, in a way to be unpacked soon, what I will call *elements* of social reality, that is, various *social kinds* such as money, marriage, or gender – and their associated statuses – as well as *forms of social power*. I will, then, refer to collective intentionality as the *building block* of social reality, while social kinds and forms of social power as *elements* thereof.

My two basic units of analysis, then, will be “kinds” and “forms of power”. As I will show throughout the dissertation, neither of these is more fundamental than the other: there are forms of powers which depend on the existence of kinds as they depend on the existence of a given status, and vice versa, some forms of power, which lie in constructing social kinds and their associated statuses, ontologically precede the existence of kinds and their statuses. As I will show, the key to understanding how kinds and forms of power relate to one another lies in clarifying their relation to social statuses. Overall, I believe that, together, these two concepts give us a better understanding of social reality: *both* its time-slice picture which highlights various ontological dependencies between its elements and building blocks as well as social reality set in motion through various forms of action and behaviour.

*

Now, to show that the building block of social reality is some form of collective intentionality, authors often use, as an illuminating contrast, an imagined picture of social reality without the relevant forms of collective intentionality.

In a passage from his seminal work, *The Construction of Social Reality*, Searle notes that “From a God’s-eye view, from outside the world, all the features of the world would be intrinsic”, even the feature of how “people in our culture regard such and such objects” (Searle 1995, 12). God “could not see” money as money or married people as married and certain individuals as gendered in a certain way. Instead, he “would see us *treating*” certain objects and people as such (*Ibid.*, emphasis in the original).

Now, going further, take a memorable thought experiment by philosopher Rom Harré. By asking his reader to consider a university campus “with the sound off”, that is, without collective intentionality, he aims to show that universities, their faculty members, students and staff could not exist *as such* without the collective intentionality constituted by the intentionality of the relevant subjects.

With the sound off, we see hominids in a hive, moving about in mysterious but orderly ways. With the sound on, we see Georgetown University, its faculty and students engaged in the social production of knowledge, aided by the campus cops, the cleaners, the librarians, and more or less facilitated by the ‘administration’. (Harré 1997, 208)

I take it that the point in this thought experiment is not merely that there are different levels of description – as any quantitatively minded statistician or behaviorist anthropologist observing different cultures would attest. Taking Searle’s God’s-eye view, for example, one would see faculty and students *treating* each other as faculty and students. However, and this is the crucial point, this is only the case if there exists a “sound on” picture to begin with, that is, only if there exists some form of collective intentionality constituted by the intentionality of the relevant subjects. The point of the Georgetown University thought experiment, in turn, is *ontological*: without collective intentionality – whether viewed from inside, the “sounds on” picture, or from outside, the ‘God’s eye view’ – there is no social reality.

As we have seen, there is an ontological difference between there being a piece of paper and there being a dollar bill. Between some two of individuals moving their mouth and having just gotten married. Between an individual subject with certain biological traits and that individual subject being treated as a man or woman. This difference, then, needs an ontological explanation: as we have seen, for Searle, Harré, as well as much of the tradition of analytic social ontology, this explanation is provided through holding that a building block of social reality is subjects’ capacity of intentionality constituting *collective intentionality*.

The dissertation seeks to nuance the traditional picture of social reality which holds that social reality is mind-dependent in the sense that the existence of its elements depends on subject's capacity of intentionality, more precisely, that of *collective intentionality*. I will do so by introducing contemporary examples of social kinds and forms of social power to the extant debate in analytic social ontology. With this, I will show that in investigating what a "sound-off" human social world would look like, one no longer needs to resort to thought experiments.

Consider the following pictures:

In the early 2010s, Uber executives treated attendees at their parties to their "God view" of all Uber's cars tracked across the surrounding city. As part of the spectacle, they showed locations of currently active passengers; sometimes, for fun, they removed passenger anonymity. (Couldry–Mejias, 2019, 115)

In the Longgang district in Shenzhen, in the heart of China's booming economic region north of Hong Kong, we visit the smart city control centre. A giant monitor displays the data of the entire district in real-time: numbers of new residents by neighbourhood, to plan schools, water supply levels, power outages. All this information is collected, compiled, and evaluated using artificial intelligences. (DW Documentary, 2019)

I take it that Uber's spectacle as well as Shenzhen's smart city screen are examples of systems which aim to classify subjects and objects based on large scale, granular and real-time data through algorithmic analysis carried out by computers for various socially consequential purposes, which I will refer to as data-driven algorithmic systems. Though these systems differ in important ways, as I will explain more in detail, they include social media algorithms, recommendation systems, targeted advertising systems, music streaming algorithms, automated hiring systems, insurance risk assessment algorithms, loan approval algorithms, and social credit systems.

As I will show, data-driven algorithmic systems have the capacity to construct new kinds of social kinds – and thereby enable new forms of social power. This is *philosophically significant* even if this novelty – from a more nuanced historical perspective – might boil down to a matter of degrees in relation to earlier forms of quantifying and classifying social reality. I will, then, refer to a society whose elements include social kinds constructed by data-driven algorithmic systems and forms of social power enabled by data-driven algorithmic systems as an *algorithmic society*.

Specifically, the first claim I will argue for is that data-driven algorithmic systems have the capacity to construct social kinds which do not depend on collective intentionality. In other words, the "sound off" picture, the lack of collective intentionality, does not always entail a lack of social reality. This is because these algorithmic systems *represent* social reality through

the digital data and algorithmic code as stand-ins for the relevant part of the world. And just like in the case of the intentionality of subjects, the relevant work of representation through the stand-in of digital data and algorithmic code has ontological significance.

In other words, these short excerpts show not only our increased capabilities of describing or modelling social reality in a nuanced manner: recent technological developments in data gathering at an immense scale and granularity, as well as in a real-time manner, and its algorithmic analysis into dynamic categories, carry *ontological* implications as they are also new ways of constructing the elements of social reality. Now, as opposed to how Uber termed their – let’s call it – surveillance spectacle, these algorithmic systems often *also do not* show reality from what Searle called a God’s eye view. These algorithmic systems often do not “see” the intentionality-imbued, “sound on” picture from the outside, but represent social reality through quantifications and classifications – representations – of their own device. These considerations, then, will delineate for us two ways of representing reality: through the stand-in of the intentionality of subjects and through the stand-in of digital data and algorithmic code.

As I will argue, next to constructing new kinds of social kinds, data-driven algorithmic systems, through quantifying and classifying subjects and objects based on large scale, granular and real-time data through algorithmic analysis carried out by computers, also enable new forms of social powers. Specifically, data-driven algorithmic systems enable new ways to get other subjects to do what one wants them to do. What makes these new forms of social power, which I will call forms of algorithmic social power, unique is their lack of dependence on collective intentionality and/or intentional action. Accordingly, I will argue for two additional claims: that there are forms of algorithmic social power which do not depend on collective intentionality as well as the claim that, some forms of algorithmic power can be exercised even if the relevant subjects upon whom power is exercised do what others want them to do without engaging in an intentional action, and, therefore, without the intentional states necessary for the given action.

A good expression for such kinds of behaviour without intentional action is “mindless scrolling”. Now, I like this expression as it hints at the fact that in the algorithmic society, there are *two kinds* of building blocks of social reality which are shifting toward a degree of independence. Let me explain. First, one kind of building block, just like in the case of traditional societies, is *minds* and their capacity of intentionality and collective intentionality; in other words, subjects with “the sound on”. The second kind of building blocks, *scrolls*, that is, *representations*; historically, rolls of parchment with forms of writing whose existence as representations seemed to be dependent on minds – today, representation through digital data and algorithmic code with a degree of ontological independence from minds. Algorithmic societies, then, bring about *a* historical shift of *ontological significance* from social reality depending on subjects’ minds and their collective intentionality towards scrolls of digital data

and algorithmic code existing and “modulating” – in a sense to be specified – independently of subjects’ intentionality.

Unpacking a detailed ontological picture of how social reality is becoming indeed *mindless* is the subject matter of this dissertation. I will analyze the ontological significance of this historical shift in two steps. In Part I, comprising of the first two chapters, I will consider the conditions of existence of social kinds, focusing on their mind-dependence. Here, I will argue that data-driven algorithmic systems bring about social kinds which do not require collective intentionality for their existence. In Part II, comprising of the last two chapters, I will focus on the concept of social power. Through considering how power relates to social status, subjects and their intentionality, I will focus on how we should incorporate forms of power enabled by data-driven algorithmic systems to our understanding of power. I will argue that there are forms of algorithmic social power which do not require collective intentionality for their existence. Finally, I will show that there are forms of algorithmic power which also do not require that the relevant subjects upon whom power is exercised engage in intentional action, differentiating these power relations from cases of intimidation as well as what I will call, inspired by Foucault, cases of “productive power”.

While examples of social kinds constructed through, and forms of algorithmic social power enabled by, data-driven algorithmic systems are largely absent from the discourse of analytic social ontology, I will demonstrate that this historical shift can be fruitfully illuminated by the conceptual toolkit developed in this tradition of thought. Ian Hacking, in his classic work, *The Social Construction of What?*, writes, “(p)erhaps we will become aware of the ways in which machines are classifying us”, but does not follow up on the issue as it is, for him, “not pressing” (1999, 108). I will aim to convince my reader otherwise.

To see how my analysis of social kinds constructed through data-driven algorithmic systems and forms of algorithmic social power fits the theoretical landscape of social ontology, let us briefly review the main trends of the field.

Ideal and nonideal social ontology

The last decades within analytic philosophy have seen a booming interest in the questions of social ontology from a variety of angles (for overviews, cf. Schmitt 2003; Ásta 2015; Mason 2016; Mason–Ritchie 2020; Passinsky 2020a). Kevin Richardson, for instance, goes as far as to argue that “today, a new revolution is brewing. Analytic philosophy is in the midst of a *social* turn”. (Richardson 2023a, emphasis is in the original). In this vibrant field, there are several meta-level analyses of the various types of trends in social ontology (cf. Haslanger 2012; Richardson 2023b; Guala–Hindriks 2023; Koslicki–Massin 2023). Out of these, I find Åsa

Burman's (2023) classification particularly useful to highlight two important trends in analytic social ontology.

Burman (2023) argues that we can make a useful distinction between two trends of analytic social ontology: ideal social ontology and nonideal social ontology. Relying on her analysis, I will unpack these trends below, shedding light on important differences between their aims, assumptions and methodologies. As I will emphasize, Burman is highlighting contingent historical trends – not necessarily distinctions within the logical landscape of possible ways of doing social ontology. After unpacking ideal and nonideal social ontology, I will shed light on a shared assumption between them, which will help me situate my research in relation to these trends.

Let us begin with ideal social ontology. Ideal social ontology, as Burman shows (2023, 1–2), tends to assume that social reality is characterized by cooperation and egalitarianism within usually small communities. When ideal social ontologists theorize about power relations, they tend to highlight how people organize themselves into what I will call codified and/or transparent statuses such as “president” or “leader”. These statuses, as I will explain in detail, are codified as they are rule-based and/or they are transparent as they require conscious intentionality for their existence.

It is telling to review the paradigmatic examples of social phenomena and metaphors chosen by ideal social ontologists at the outset of their work. Margaret Gilbert's examples involve “two people going for a walk together” or a “poetry discussion group” (1992, 19). She adds that “I am happy to be committed to the existence of small, ephemeral collectivities. I believe that these are the bedrock of human social life” (1992, 17). In a similar vein, for Searle, “obvious examples” of social phenomena are cases of cooperation, using the example of a violinist in an orchestra contributing to a symphony (1995, 23).

Ideal social ontologists, then, aim to offer a picture of social reality by starting from what they deem “simple” or “obvious” cases and zooming out from them. This method of starting from *uncontroversially social* phenomena seems to be shared, for instance, by Simmel who focuses on “the strictly social” (1910, 381), Gilbert who discusses “social phenomenon proper” and is looking for that which we “intuitively” or “very naturally” think of as a social (1992, 41, 51, 54) and Searle who takes “intuitive features of social reality” as his starting point (1995, 31). However, as Burman rightly points out, central assumptions around the cooperative, egalitarian, codified and/or and transparent nature of social reality, in fact, already drive these theorists to pick certain examples, and not others, as “paradigmatically” social phenomena.

How one's example-choices are determined by prior assumptions about the nature of social reality becomes evident when examining the contrast between the aims, assumptions and

methods between ideal and nonideal social ontology (Burman 2023, 2–3). Nonideal social ontology takes as its starting point conflict, inequality and tends to investigate “messy”, larger – especially modern-day – communities. When theorizing about power relations, nonideal social ontologists highlight what I will call non-codified and/or non-transparent statuses, such as gender, race or disability (cf. Haslanger 2012; Ásta 2018). These statuses, as I will explain, are non-codified as they are not rule-based and non-transparent as they do not require conscious, merely unconscious, collective intentional states for their existence.

As such, whether acknowledged or not, the nonideal trend of analytic social ontology, given its focus on non-transparent social statuses, is in line with the classics of the Continental social theory. For instance, it resonates with Bourdieu’s idea that our “acquired dispositions”, which lead us to classify each other and be classified by others, “owe their specific efficacy to the fact that they function below the level of consciousness and language, beyond the reach of introspective scrutiny or control by the will” (1984, 466) as well as with Foucault’s ideas about a form of social power which is “exercised through its invisibility” (1995, 187 cf. 1980, 2000).

Furthermore, nonideal social ontologists, Burman explains, do not start with social phenomena which are uncontroversially social. Instead, their point is to “debunk”, say, a given kind, such as gender, race or disability, *as* being social – as opposed to being natural. This debunking, then, shows that the given kind, in fact, depends on collective intentionality, and as such is mind-dependent. Mind-dependence, then, nonideal social ontologists argue, implies that social kinds could rather easily be otherwise and, ultimately, that they are arbitrary.

As Ian Hacking helpfully explains, such claims within nonideal social ontology aim to show that the given kind “need not have existed, or need not be at all as it is (...it) is not determined by the nature of things; it is not inevitable” (1999 6). In fact, he goes as far as to argue that claims about being “socially constructed” *only* make sense if they are debunking in this sense:

You do not find books on the social construction of banks, the fiscal system, cheques, money, dollar bills, bills of lading, contracts, tort, the Federal Reserve, or the British monarchy. These are all contractual or institutional objects, and no one doubts that contracts and institutions are the result of historical events and social processes. Hence no one urges that they are socially constructed. (Hacking 1999 12, cf. 67)

Given our framework of various trends in social ontology, Hacking’s view is, strictly speaking, false: we have seen that *ideal* social ontologists indeed aim to explain how kinds which are uncontroversially social are constructed. However, his point helps us draw a contrast between ideal social ontologists and nonideal social ontologists such as Ásta’s project of providing “tools to think about oppressive institutions and practices” (2018, 127), or Haslanger’s “social

critique” (2012) which is carried out with the normative goal of highlighting arbitrariness and bringing about social change.

In sum, ideal and nonideal social ontology differ in terms of their aim, assumptions, and methods. While the former focuses on cooperation within small communities and codified and/or transparent power relations, the latter’s emphasis is on conflict and large – often modern – communities as well as non-codified and non-transparent power relations. Their methodological starting points also differ as ideal social ontologists aim to explain uncontroversial cases of social construction, while nonideal social ontologists theorize with the goal of achieving social change by debunking controversial phenomena as socially constructed.

Finally, ideal and nonideal conceptions of collective intentionality may differ in how codified the relevant content of the intentional states and/or how conscious the intentional states involved are stipulated to be: while ideal social ontology thinks of subjects as rule-constructing and consciously so, nonideal social ontology posits subjects driven by often non-codified ideology and having intentional states which are often deeply seated and unconscious.

An assumption, however, which is crucially shared between the trends of ideal and nonideal social ontology is that some form of collective intentionality is necessary for social reality.

*

Let us circle back to the historical shift towards an algorithmic society. Within analytic philosophy, to my knowledge, there has been no comprehensive analysis of social kinds constructed through, or forms of social power enabled by, data-driven algorithmic systems in either trend of social ontology. Thus, analytic social ontology, it seems to me, whether it follows an ideal or nonideal trend, unduly neglects these elements of social reality. Why is that?

It seems to me that there are three reasons. First, of course, the emergence of data-driven algorithmic systems is a recent event, and social ontology needs time to reflect on it and catch up. Second, remember that ideal and nonideal social ontology refer to existing historical trends in theorizing about social reality but not necessarily the logical landscape of possible ways of doing social ontology. As I will explain in detail, the ontological explanation of social kinds constructed through, and forms of social powers enabled by, data-driven algorithmic systems requires assumptions from both trends. As such, it seems that these new kinds of social kinds and forms of power end up being “in-between” the paradigmatic examples of the two trends of social ontology. Third, and most importantly, social kinds constructed through,

and forms of social powers enabled by, data-driven algorithmic systems pose a challenge to our very *notion of “the social”*.

Let me provide a preliminary idea.

The mark of the social at a crossroads

The quest after the necessary and sufficient conditions of something being “social” – the “mark of the social” (cf. Greenwood 1997) or the “real glue” (Gilbert 1992, 13) of the social – have long preoccupied philosophers. Already at the outset, one is faced with the diversity of phenomena and entities labeled as “social” manifest. Money, marriage, presidents as well as gender, race and disability are, to use Margaret Gilbert’s phrase, a “motley crew” (1989, 441).

This diversity has led many philosophers in the recent literature in analytic social ontology to simply provide a close-to circular definition – by simply redirecting their reader from one term with the qualifier “social” to another – or no definition at all. For example, “a *social* kind is a collection of phenomena defined by a property or feature that is a *social* property or feature” (Ásta 2017, 290, my emphases). Or: “we can take a *social* fact to be a fact that corresponds to a proposition that has any *social* entity as a constituent. It might have *social* objects as constituents, or it might have *social* properties as constituents, or both.” (Epstein 2018, 67, my emphases). Sally Haslanger explicitly acknowledges this difficulty:

I have no intention in this chapter to define the term ‘social.’ This may seem to be a cop-out; however, I believe that a rough and ready understanding of the term is sufficient for my purposes. A non-circular definition is not likely to be available, but this does not preclude an account that situates the term within a conceptual web. (Haslanger 2012, 197).

However, as we have seen, social ontologists indeed have assumptions – even if these are, at times, implicit – about the mark of the social. Let us, for now, focus on *necessary* conditions – I will highlight two key assumptions in this regard.

First, we have seen that both ideal and nonideal social ontologists – whether or not their examples are controversial, as in the case of the latter – contrast social entities with entities found “in nature”. Let’s be a bit more precise here. Social ontologists tend to accept that social reality is part of “nature” as many take a naturalist view on reality where nature is the only reality. Yet – remember – they tend to emphasize that reality has a mind-dependent and mind-independent part and argue that social reality is part of the former while paradigmatic examples of “natural” entities, such as mountains, are usually taken from the latter.

Thus, as we have seen, one way in which the contrast is drawn between the social and the non-social is through the contrast between the social and (paradigmatic examples of) the

natural. This, then, is done through explaining the relevant notion of mind-dependence which characterizes only the social. Mountains, then, can be contrasted to both money and marriage as well as gender and race based on the relevant sense in which their existence depends on minds – specifically, as we have seen, their capacity for intentionality.

The second assumption is, in fact, in most cases, explicit. It is the idea that in order for something to be labelled “social”, it must have something to do, as we have already noted, not just with an *individual* subject, but with many subjects – with a *group* of subjects. Hence the idea that it is not merely individual intentionality which is required for social reality but *collective* intentionality. As we have seen, both ideal and nonideal social ontology holds that some form of collective intentionality is necessary for social reality, even if their conceptions thereof may differ based on how codified the relevant content of the intentional states and/or how conscious the intentional states involved are stipulated to be.

Furthermore, conceptions of collective intentionality may also differ based on the ontological status of the relevant intentional states of individual subjects: whether collective intentionality is more than the sum of intentionality of individual subjects and whether there is a new form of subjectivity born through or with collective intentionality. It is also debated whose intentional states, in fact, count; that is to say, who is taken to be part of the relevant “collective” in the term collective intentionality. I will unpack these questions more in detail in the upcoming chapters. Regardless, however, of how the specifics of the relevant concept of collective intentionality is spelt out, it is argued that without collective intentionality of some sort – without the “sound-on” picture of Harré’s example about Georgetown University – there is no social reality.

Collective intentionality, then, captures the combination of the two key assumptions about what is required for social reality. First, mind-dependence (dependence on the relevant intentional states), which delineates social reality from non-mind-dependent parts of reality. Second, going beyond the level of individual subjects towards the level of groups. This, then, is how collective intentionality is held to be the *building block* which is required for social reality.

Now, as I mentioned above, I will argue that there are cases where collective intentionality of the relevant form is not required for the existence of social kinds and certain forms of social power. To explain the ontology of these elements of the algorithmic society, one needs to rely on assumptions about social reality from both ideal and nonideal social ontology, and so my analysis further shows that these are historically contingent trends rather than logically possible ways of providing an ontological explanation of social reality. In exploring social reality beyond collective intentionality, I thus show that we need to rethink even those assumptions about social reality which are shared by both ideal and nonideal social ontologists.

*

Assuming I am correct in holding that there are indeed social kinds and forms of social power which can exist without the relevant forms of collective intentionality, there seems to be two possible avenues to take from here. One could simply insist that the traditional picture of social reality according to which collective intentionality is a necessary condition of anything being social is right. If so, then I am wrong in labelling kinds constructed through, and forms of power enabled by, data-driven algorithmic systems as “social”. In other words, one would argue that the – alleged – social kinds and forms of social power I identify in what I call our algorithmic society are not, strictly speaking, “social”. The second possible avenue to take is to update what our concept of the “mark of the social” holds regarding the necessary and sufficient conditions of something being social and incorporate algorithmically constructed kinds and forms of algorithmic power into our picture of social reality.

Holding onto the traditional idea that collective intentionality is necessary for social reality, many social theorists indeed argue that the elements of algorithmic society are, in fact, not social. However, acknowledging, at the same time, the growing relevance and use of quantification and classification through data-driven algorithmic systems in our society often leads them to hold that it has become timely to speak of the “end of the social” or the “hollowing out of the social”, the “post-social” or “non-social” (e.g., Baudrillard 1983; Couldry-Mejias 2019; Lee 1997, 255, 258). These intriguing remarks on the nature of the historical shift we are witnessing all seem to assume that collective intentionality – of some sort – is a necessary condition of something being social.

By contrast, the second way to respond to the tension between, on the one hand, acknowledging the social nature of algorithmically constructed social kinds and forms of algorithmic social power and, on the other, our traditional take on the mark of the social, is to update the latter. So, how do we include in our social ontology both traditional social kinds and forms of social power which are dependent on collective intentionality as well as algorithmically constructed social kinds and forms of algorithmic social power?

I believe that the historical task of preparing the theoretical ground for the “post-social” is best served by zooming out from overly specific understanding of the mark of the social and updating it through *pointing out common elements between* traditional social kinds and forms of social power *and* social kinds constructed by, and forms of algorithmic social power enabled by, data-driven algorithmic systems. More concretely, I believe that we should not place sole emphasis on the *overly specific* form representation through the intentionality – intentional states – of minds as the required building block of social reality. Instead, we should consider representation, as a required building blocks of social reality, in a broader sense,

both representation through intentionality and representation through digital data and algorithmic code.

Thus, we can say the following. On the traditional picture of social ontology, representation through intentionality is the building block of social reality. Now, the algorithmic society seems to me to have two kinds of building blocks. First, the existence of the elements of the algorithmic society are still in some sense – to be unpacked – dependent on minds and their intentionality. This is so even if, as I will argue, they are not dependent on collective intentionality of the relevant form. Second, the existence of the elements of the algorithmic society depends on representations through digital data and algorithmic code. These representations, then, have a degree of ontological independence from the relevant intentionality on which the existence of the elements of the algorithmic society depends. Once again, algorithmic societies have, in this sense, two kinds of required building blocks: representation through intentionality and representation through digital data and algorithmic code.

Zooming out from an overly specific understanding of the building blocks of social reality, as I will show, helps us update the mark of the social – the necessary and sufficient conditions for something to count as social. The elements of algorithmic social reality, then, will be considered “social” as they meet the two key assumptions delineated above about social reality. This is to say, elements of algorithmic social reality meet the conditions of being, on the one hand, sufficiently different from *paradigmatic examples* of natural parts of reality, such as mountains, as they are mind-dependent and representation-dependent, and, on the other, being having large-scale, *group level* consequences. These conditions, as I will show, are crucial shared features of elements of both traditional and algorithmic societies.

Methodology, interdisciplinarity, and historical shifts

Let me now clarify my methodology to further situate my research among various kinds of social ontology projects. I am embarking on a project of conceptual analysis of social kinds and forms of social power and will aim to balance my reliance on empirical evidence with analyzing their “folk” or “intuitive” concepts. This is all the more important as we do not seem to have folk or intuitive concepts of the social kinds and forms of social power constructed through, or enabled by, data-driven algorithmic systems. Thus, while I will aim to use “messy” real life cases and retain their detailedness and complexity, these will be analyzed in the rigorous manner characteristic of analytic philosophy.

My general goal in conceptual analysis and concept creation is to serve social scientists. In this regard, I am in agreement with ideal social ontologists such as Gilbert who shows that

analytic social philosophy “may be expected to be a useful adjunct to social science” (1992, 4) or Searle who argues that social ontology should be “philosophy *for* social science”:

I believe it will deepen our understanding of social phenomena generally and help our research in the social sciences if we get a clearer understanding of the nature and the mode of existence of social reality. We need not so much a philosophy *of* the social sciences of the present and the past as we need a philosophy *for* the social sciences of the future and, indeed, for anyone who wants a deeper understanding of social phenomena. (Searle 2010, 5; emphases in the original)

Now, the commitment to be an underlabourer for social science leads me to endorse interdisciplinarity. To me, this means acknowledging that, as opposed to narrow field of analytic social ontology, in social theory in general – especially in sociology and media studies, as well as in Continentally-oriented philosophy – there has already been much work on how datafication and algorithmic analysis constructs new forms of social knowledge and various kinds of social entities (e.g., Baudrillard 1988; Deleuze 1992; Cheney-Lippold 2017; Couldry–Mejias 2019; Burrell-Fourcade 2021).

This means that, if one does not take into account traditions or branches of thought other than the – heavily Anglo-Saxon leaning – corpus of analytic social ontology, one may risk the possibility of simply renaming concepts that have been carved out by social theorists working in different traditions. I agree with Kevin Richardson who argues that analytic social ontology often runs the risk of “philosophical Columbusing—the act of pretending to discover (or mistakenly thinking that one has discovered) a philosophical insight that has already existed” (Richardson 2023a). To prevent such Columbusing, I will take inspiration from various branches of social theory and aim to produce a work of interdisciplinary scholarship.

I believe, then, that joining the insights of existing empirical research on “societies of algorithms” (Burrell-Fourcade 2021) as well as those of Continental philosophy about “the numerical language of control” (Deleuze 1992, 5) with the detailed analytic vocabulary that social ontologists have developed over the past decades might give us better conceptual tools and descriptions to start unravelling what the historical shift towards algorithmic societies amounts to.

While I will rely on other disciplines – and will not worry too much about policing the boundaries of “real” philosophy – I do believe that the slow, rigorous and meticulous *analytic* method of conceptual analysis has much to contribute to this discourse. As Gilbert put it, rather bluntly, “students of sociology or the philosophy of sociology may well need some form of analysis of these concepts in order to understand fully what they are talking about” (Gilbert, 1992, 55). Thus, I will be relying both on insights and examples brought in social scientific and Continental philosophy literature *and* the rigorous method of conceptual analysis of the analytic philosopher.

Furthermore, as mentioned above, there is by now several meta-level analyses of what social ontology is, can or should be (e.g., Haslanger 2012; Diaz-Leon 2012; Richardson 2023b). Without an elaborate reconstruction of the available taxonomies, I would like to briefly situate my project with regards to one particular trend in social ontology. We have seen earlier that nonideal social ontology is often characterized by normative theorizing aiming for social change. In her famous analysis of various aims a social ontology project can have, Sally Haslanger (2012) calls these nonideal projects “ameliorative” projects. Ameliorative projects, according to Haslanger, ask what the point or purpose of using the given concept is – or should be. In other words, for ameliorative projects, “normative input is needed” (Haslanger 2012, 376); one needs to formulate a legitimate goal. For instance, what Richardson calls “critical social ontology”, that is, the “study of social ontology that is done in order to critique ideology or end social injustice” seems to me to be one way to unpack such a legitimate goal (Richardson 2023b). The task, finally, is to assess what concept would suit the given goal best.

I agree with “ameliorative” or “critical” projects insofar as they take folk or intuitive concepts or intuitions as secondary. For instance, take Richardson’s remark that critical projects might benefit from “the ability to analyze social kinds, objects, and states of affairs that do not correspond to ordinary terms or concepts” or “analyze social structures that are not perspicuously described in ordinary language” (Richardson 2023b, 204). These are indeed in line with the goals of my project. I also agree that it might well be inevitable not to bring numerous lurking normative assumptions to one’s philosophy. Furthermore, I admit that it is beneficial to be transparent and open about one’s normative assumptions. And, needless to say, I agree with the particular goal of critical social ontology – ending social injustice.

I would, however, demur from calling my project “ameliorative” à la Haslanger or “critical” à la Richardson, as I do not believe we *always* serve even a value-laden goal best by value-laden theorizing. I value philosophy most when it aims to keep such commitments to the minimum. For one, I am not convinced whether one can prescribe how their concepts are to be, or will be, thought and used.

For example, I believe that one reason the discourse around the “end of the social” looks theoretically enticing to many is due to value-laden assumptions about the role of intentionality and collective intentionality in constructing social reality. Perhaps we are indeed losing something irreplaceable by gradually shifting from representations through collective intentionality as the only necessary building block of social reality towards also including the building block of representations through digital data and algorithmic code which construct elements of social reality which do not require collective intentionality for their existence. Nevertheless, I believe a “non-ameliorative” and “non-critical” analysis of this historical shift towards algorithmic societies is equally legitimate. Perhaps it is such “non-

ameliorative” and “non-critical” analysis which may provide us with the concepts necessary for the desired social change towards a more just society.

More on my lacking commitments.

I do not believe that each social ontology project needs to offer a general or exhaustive theory of social reality to be useful for social science. While I do believe that having a general theory of social reality is a worthy goal, my aim is more modest. I wish to help to “fill the gap” of a general theory of social reality by bringing examples of social kinds and social power constructed through, and enabled by, data-driven algorithmic systems and examining how they relate to the traditional picture of social ontology. This way, my work is a step towards general coverage but does not aim – or claim – to offer a general or exhaustive theory of social reality in itself.

I am also not doing “foundational” or “substantive” ontology (Sider 2011, cf. Barnes 2014). It is clear enough that collective intentionality, digital data and algorithmic code, and data-driven algorithmic systems are not the foundational building blocks of *reality*. However, I believe that shedding light on *some* dependency relations is a sufficiently ambitious goal for ontology. For instance, one of my main concerns will be to illustrate how exactly “the social” is dependent on “the mental”. I do not wish to further pursue whether the mental is fundamental or not. And I do not think one needs to prescribe as a condition for adequate philosophy that it delve into the fundamental level of reality.

Finally, I am not aiming to provide an ahistorical analysis of social reality. It is indeed the *historical* shift from social reality being dependent on the building block of collective intentionality of subjects to its being dependent on the building block of representations through digital data and algorithmic code which exist and change, in some sense, independently of human intentionality, which I would like to elucidate using the conceptual toolkit of analytic social ontology.

I intend this dissertation, then, to be part of those philosophical works which aim to provide a general description of a historical shift; in this, I am inspired by the Foucauldian, historically informed attitude to philosophy. Discussing Kant’s famous essay, “What is Enlightenment”, he writes:

It was the first time a philosopher proposed as a philosophical task to investigate not only the metaphysical system or the foundations of scientific knowledge but a historical event— a recent, even a contemporary event. When in 1784 Kant asked, “What is Enlightenment?” he meant, “What’s going on just now? What’s happening to us? What is this world, this period, this precise moment in which we are living? (Foucault 2000, 335)

“What’s going on just now? What’s happening to us?” are questions which prompt one to ask what conceptual challenges are presented by one’s “historical awareness of our present circumstances” (*Ibid.*)? What are the ontological assumptions about social reality which we need to rethink given new ways of quantifying and classifying – and thereby constructing – social reality? I believe that even if the philosopher sees their job as primarily offering an analysis of concepts in their most general forms, they must recognize that historical conditions have the potential to challenge even the most all-encompassing analyses. My hope is that analyzing the contrast between traditional and algorithmic societies will mutually shed light on both.

Brief chapter by chapter overview

In Chapter 1, which is the opening chapter of Part I, I will introduce in detail what I have so far been referring to as “the traditional picture of social reality”, which I will base on Searle’s theory of social construction and will call Intentionalism. I will ask whether Intentionalism has general coverage over social reality. Preliminarily, Intentionalism posits that the existence of social kinds depends on intentional states representing them as existing. I will aim to unpack the most charitable form of Intentionalism by taking into account Searle’s revisions as well as the criticism the view has received. I will revise the view by (i) restricting the scope of social kinds concerned to human social reality, (ii) weakening the relevant sense of intentional states, and (iii) weakening the relevant sense of dependence. I conclude that if Intentionalism is weakened on these key matters, it seems to retain its general coverage over social reality.

In Chapter 2, which is the closing chapter of Part I, I will argue for two claims. The first claim is that socially significant algorithmically generated real-time classifications – which I will call Modulatory Algorithmic Kinds (MAKs) – qualify as “social kinds”. MAKs qualify as kinds as they can be effectively used for the epistemic purposes such as prediction. MAKs, then, qualify as social as they are mind-dependent, representation-dependent, and have large-scale, group-level consequences for social reality. The second claim is that MAKs as social kinds challenge Intentionalism. This is because MAKs present a case of social kinds not being dependent on collective intentionality: the relevant “social glue”, is no longer tied to collective intentionality but to the representational capacities of digital data and algorithmic code. I then go one step further and argue that, in the case of MAKs, the relevant sense of mind-dependence is dependent on intentionality is not constitutive. MAKs, then shows that an ever-increasing part of our social reality is beyond the coverage of Intentionalism.

In Chapter 3, which is the opening chapter of Part II, my aim is to better understand the forms of algorithmic powers made possible by data-driven algorithmic systems. To do so, I first formulate a minimal conception of social power and provide a taxonomy of social powers. To formulate a minimal concept of social power, I will investigate one kind of constraint, the

Intentionality Constraint, which holds that the existence and/or exercise of social power is, in some way, dependent on intentionality. I will show that the Intentionality Constraint has two versions: the exercise of power being dependent on some form of intentional action and the existence of power being dependent on collective intentionality. While I will accept the first constraint with a few caveats, as to the second constraint, I will argue that exercising power via data-driven algorithmic systems brings about forms of social power where neither the power of bestowing statuses nor the statuses itself depend on collective intentionality – although they may be enhanced by it. Finally, I will argue that with the introduction of “nudging” into data-driven algorithmic systems comes a form of social power which is not only not dependent on but also not enhanced by collective intentionality.

In Chapter 4, I examine forms of algorithmic power through Foucault’s concept of “productive power.” Drawing on how productive power is contrasted with algorithmic power in the extant literature, often citing the lack of reflection in the latter, I ask in what sense algorithmic power produces – or does not produce – subjects. I first unpack Foucault’s notion of productive power and how he contrasts it with intimidation. After distinguishing three senses of “reflection” allegedly lacking in the exercise of algorithmic power, I argue that reflection, understood as the intentional states necessary for intentional action, is required for the successful exercise of both productive power and intimidation. I then identify the form of algorithmic power which can be successfully exercised without the relevant sense of reflection – thus being non-productive and distinct from intimidation – which I call behavioural nudging power.

Overall, the introduction of data-driven algorithmic systems into social reality entails that we should rethink the traditional picture of the ontology of social reality as social reality shifts towards an algorithmic society. In the algorithmic society, some elements of our social reality do not depend on the building block of representation through collective intentionality. Instead, they depend on two building blocks which, as I will unpack, are characterized by a degree of ontological independence: representation through intentionality (but not collective intentionality) *and* representation through digital data and algorithmic code.

Specifically, then, our social reality is becoming *mindless* in the following sense: data-driven algorithmic systems construct social kinds and enable forms of algorithmic social power which do not require collective intentionality for their existence. Moreover, they bring about forms of algorithmic power which also do not require intentional action on the part of those over whom power is exercised, thereby being different from cases of intimidation or Foucauldian productive power.

I have noted earlier that my project is distanced from “ameliorative” or “critical” social ontology and that my primary aim is to be an underlabourer for social science. Yet I also believe the concepts I am about to carve out will be useful not only for social scientists but

anyone aiming to conceptually navigate this ongoing historical shift towards algorithmic societies. I conclude, then, with a quote from Deleuze's "Postscript on the Societies of Control" (1992), a brilliant and foretelling essay to which this dissertation is much indebted:

There is no need to ask which is the toughest regime, for it's within each of them that liberating and enslaving forces confront one another. (...) There is no need to fear or hope, but only to look for new weapons. (1992, 4)

Part I: Social Kinds and their Mind-Dependence

CHAPTER 1: RECOGNITION AND EXISTENCE: THE TRADITIONAL PICTURE OF SOCIAL REALITY

In quite a different sense from that in which it is true of the external world, is society 'my representation', i.e., posited upon the activity of consciousness.

Georg Simmel: *How is Society Possible?* (1910, 375)

Introduction: The social and its puzzles

It sounds almost cliché to point out that the beauty of philosophy lies in its ability to make us wonder about things we tend to take for granted. Much more fun, however, it is to explore the particular moments when philosophers are utterly baffled by seemingly ordinary and understandable entities and phenomena. Take Hume, for instance, trying to wrap his head around the everyday practice of making promises:

'tis one of the most mysterious and incomprehensible operations that can possibly be imagined, and may even be compared to transubstantiation, or holy orders, where a certain form of words, along with a certain intention, changes entirely the nature of an external object, and even of a human creature. (1978 [1739], 524)

With similar bewilderment, John Searle ponders not only the existence of promises but also that of similarly mundane entities such as money, private property, and marriage:

In our toughest metaphysical moods, we want to ask [...] are these bits of paper really money? Is this piece of land really somebody's private property? Is making certain noises in a ceremony really getting married? Even, is making noises through the mouth really making a statement or a promise? (1995, 45)

Hume and Searle are both puzzled by the fact that physical things can be “transubstantiated” into something entirely different. More precisely, it seems that one aspect of reality that leaves these philosophers baffled is the *existence of social* entities and phenomena – such as “money”, “marriage”, or, for that matter, “promises” and their respective practices.

Let us now examine Searle's case more closely to get off the ground. He wondered how money, marriage or promises – which, as mentioned in the *Introduction*, I will call *elements* of social reality – can exist and exist as part of the same reality as described by physics and other natural sciences (1995, 2010). How is it possible that, for instance, money is part of the same reality which is governed by the principles of physics and chemistry? If we agree that

the natural sciences identify the fundamental building blocks of our reality – particles, fields of force, molecules, etc. – then it appears that money, marriage, or promises and other elements of social reality cannot be regarded as fundamental in the same manner. This raises the question: what exactly does the existence of money, marriage and other elements of social reality depend on? As Searle put it: “How do we get from electrons to elections and from protons to presidents?” (2010, 3).

This is a version of what philosophers call “placement problems”: aiming to make sense of the existence of certain entity or phenomena within a given theoretical framework; in this case, finding the place for the elements of our social reality within the, let’s call it, naturalistic framework provided by the natural sciences. Most philosophers following up on Searle’s placement problem, then, are not asking for a *causal* explanation of the relevant elements of social reality. For instance, they are not asking for an account of the evolution of money, from bartering through gold coins to cryptocurrencies, or a detailed history of the institution of marriage across different cultures, or how various forms of promise-making has evolved across various civilizations. Instead, they are asking for an *ontological* explanation: how is it possible that elements of social reality such as “money”, “marriage”, or “promises” *exist at all*?

Trying to find a place for these elements of social reality in our ontology requires us to settle what qualifies as “social”. As we have seen in the *Introduction*, besides the point that “social” elements of our reality concern not only individual subjects but collectives of subjects, in the case of human social reality, the “bridge” between, on the one hand, (non-social) natural entities and phenomena and, on the other, social entities and phenomena – elements of social reality – is to be identified in the human mind (e.g., Searle 1995, 2010; Khalidi 2015; Ásta 2018; Mason 2021). The rather uncontroversial consensus, then, is that the existence of elements of social reality is, in some sense, *mind-dependent* (Searle 1995, 2010; Thomasson 2003a; Guala 2014; Khalidi 2015; Ásta 2018; Passinsky 2020b; Mason 2021). Theories diverge, however, when it comes to the specifics of the mind-dependency of the elements of social reality; philosophers have proposed different versions of what I will call the “mind-dependency thesis”, varying both in *scope* and the *strength of the mind-dependency* they posit.

Now, at first glance, it seems that it is *human* social reality which prompted our puzzlement leading up to the placement problem: examining “social” behaviour in ants, wolves or chimpanzees might not elicit the same “tough metaphysical moods”. Nonetheless, the question remains: what common features make us consider social reality in humans and other species as “social”? I will, then, differentiate between two problems. First, there is the narrower issue of the Placement Problem of Human Social Reality, or simply, Placement Problem. The Placement Problem aims to find the place for (at least some) elements of human social reality in a naturalistic ontological framework (whatever specifics of that may be).

Second, there is the broader issue of the General Problem of Social Reality, or simply, The General Problem, which aims to uncover the general conditions of possibility for the existence of *any* form of social reality, independent of which ontological framework we take as foundational.

In this chapter, I will focus on the mind-dependency thesis which I will call Intentionalism. Preliminarily, Intentionalism posits that social phenomena are mind-dependent in the sense that the existence of the elements of social reality depends on intentional states representing them as existing. As such, Intentionalism seems to provide an answer to the Placement Problem. Accordingly, on this view, money, marriages, and promises exist partly because, through our intentional states, we collectively represent them as existing. The question that drives this chapter, then, is whether Intentionalism can also provide an answer to the General Problem. Does Intentionalism have general coverage over social reality? This point remains a topic of ongoing debate in the literature (Thomasson 2003a; Burman 2007, 2015; Searle 2010; Guala 2014; Khalidi 2015; Ásta 2018; Franda 2021). I will argue that while non-human elements of social reality shows that Intentionalism cannot provide an answer to the General Problem, if the Intentionalist revises their position on a few key matters, it can retain its *general coverage over human social reality*.

The structure of the chapter is as follows. The first section presents “kinds”, and their “properties” and their “instances” as my choice of units of analysis, followed by an explanation of how versions of the thesis that “social kinds” are mind-dependent vary according to scope and strength. Then, relying on Searle’s theory of social construction, I unpack Intentionalism in detail, explaining the role of collective intentionality in the construction of social kinds and their respective statuses. In the second section, I aim to charitably revise Intentionalism based on the challenges it has received. In three subsections, I argue for revising Intentionalism by restricting its scope to human social reality, revising the relevant sense of intentional states, and revising the relevant sense of dependence, respectively.

Before I proceed, a quick note with regards to the overall argument of Part I of the dissertation. In this chapter, I am not yet introducing examples of elements of an algorithmic society; I will analyze the ontological implications of social kinds constructed through data-driven algorithmic systems in the following chapter. Here, instead, I am working with traditional examples of the elements of our social reality. This will allow me to explore, in the next chapter, the challenges that the elements of the algorithmic society present to even a revised version of Intentionalism.

Now, we may begin.

1.1. From mind-dependency to Intentionalism

As we have seen, it is widely held that the way to solve the Placement Problem is to hold that some elements of social reality such as money, marriage depend – in some sense – on minds. Let's call this the mind-dependency thesis.

Mind-dependency thesis: the existence of some of the elements of social reality depends on minds.

Now, Intentionalism, as I will show, is a more specific version of the mind-dependency thesis which holds that the existence of all elements of social reality depend on collective intentionality representing them as existing.

Intentionalism: the existence of all elements of social reality depends on their being represented as existing through collective intentionality.

How do we get from the mind-dependency thesis to Intentionalism? In this section, I explain this gap by introducing Intentionalism in detail. I will primarily draw from Searle's theory of social construction (1995, 2006, 2010), which I take to be the most famous version of this position.

First, however, to prepare the ground for a detailed unpacking of Intentionalism, I will settle my unit of analysis and provide an abstract overview of how Intentionalism fits into the logical landscape of possible ways to unpack the mind-dependency thesis.

1.1.1 *The language of kinds*

I have so far been referring to money, marriage and promises using the rather neutral terms of social entities and social phenomena for which I introduced the term "elements of social reality". What exactly are these "elements"? In this dissertation, I will employ two units of analysis to unpack what the elements of social reality are. In Part I, I will focus on "social kinds", while in Part II, I will focus on forms of "social power". Here are, then, a few preliminary ideas on the language of *kinds*.

In the extant literature in social ontology, authors use various metaphysical entities as the basic unit of their analysis. Some talk about social "facts" (Searle 1995, 2010; Epstein 2018, 160), others of "objects" (Passinsky 2020b), and there are those who employ the language of social "properties" (Ásta 2018). Take, again, the example of money. Asya Passinsky (2020b, 431), talks about "money" as a social object, Ásta talks about the "property" of "being money" (2018, 17), while for Searle it is a social "fact" that "the piece of paper in my pocket is a five-dollar bill" (1995, 1). I will, in turn, adopt the language of social "kinds", and, where relevant,

their respective “properties” and their “instances” (e.g., Thomasson 2003a; Khalidi 2015, 2016; Ásta 2017). Accordingly, money is considered a social kind, while a five-dollar bill is an instance of it which has the property of being money, that is, roughly, being a unit of exchange and a store of value.

All these metaphysical entities evoke important debates around the relation between reality and our representations thereof. The language of kinds, specifically, also evokes debates around *classificatory practices*, with its historical roots going back at least to Aristotle’s essentialism, Locke’s nominalism (as well as the ensuing criticism from Leibniz) and Mill’s theory of taxonomies. Thus, my focus on exploring kinds, their respective properties and instances stems from my interest in thinking through the ontological implications of data-driven algorithmic systems bringing about new ways of quantifying and classifying social reality. (Given the following chapter’s emphasis on these new ways of quantification and classification, I will revisit the question of “kindhood” when discussing social kinds constructed by data-driven algorithmic systems.) In the case of kinds, then, the relevant debate around reality and our representations thereof – without delving into the history of the philosophical reflections on classifications – is whether and in what sense the world comes pre-structured into *kinds* of things.

Traditionally, the main positions in this debate are usually labelled as nominalism and realism. While nominalists admit that similarities and differences exist, they deny that any division of the world as represented by our categories would be inherent to it. As such, for the nominalist, all categories reflect only our contingent interests – and not an inherent structure of the world. Using the formulation of the famous nominalist, Locke, what we call kinds are “nothing else but abstract Ideas [...] with names annexed to them” or “abstract ideas we have in our minds”, which reflect merely the classifier’s own “various Care, Industry, or Fancy” (1975 [1689], 3.3.9/13/29).

Realists, on the other hand, argue that kinds are not just representations grouping things under a category: kinds of things exist *independently* of how we happen to classify them. To quote Leibniz, Locke’s famous critic on this topic: “I don’t know why you and your associates always want to make (...) species [for our purposes, “kinds” – L. K.] depend upon our opinion or knowledge. They are present in nature, whether or not we know it or like it.” (1996 [1765], 327). This means that, from a realist perspective, one can distinguish between the *representation* of a given grouping of things or people and the respective group of things or people itself existing “out there” in the world. The realist, then, suggests reserving the term “kind” for the latter. This also implies, for the realist, a distinction between categories which *merely* reflect the interests of the classifier and those which also correspond to kinds (as these kinds might, nonetheless, in being picked out, be in the interest of those representing them as such).

It is, of course, a vexed question what exactly safeguards the “realness” of kinds. The traditional realist view that kinds exist “independently” of us led philosophers to identify paradigmatic examples of kinds “in nature” such as gold, tigers, and water. However, as I will show below, social kinds such as money and marriages are mind-dependent – and therefore do not exist independently of us. Yet, as Searle and others have argued, social kinds such as marriages and money are elements of social *reality* – and social reality is part of reality (Searle 1995; Thomasson 2003a; Wilson 2007; Ásta 2018). As such, they also cannot be conceived of in a nominalist framework. A tension, then, emerges. Given that social kinds do not exist independently of us, the realist would have to hold that social kinds are, in fact, not real; but given that we have stipulated that social kinds are elements of reality, the nominalist also cannot accommodate them. This tension, of course, hinges on defending a form of realism which is incompatible with the relevant forms of mind-dependency which are to be unpacked below.

How much of the theoretical baggage of kinds am I intending to take on board?

I do believe the ontological distinction between kinds and any incidental category is a useful one. Yet I will not provide a detailed theory of realism here. I simply subscribe to the desideratum assumed by the Placement Problem according to which one needs to find the place for social reality in the broader ontology of reality set out by the naturalistic framework. This seems to me to be compatible with various views on what it is to be real; those interested in a more detailed theory of realism are welcome to bring their own.

Furthermore, I will remain uncommitted as to the precise ontological character of kinds *in general* (e.g., how many common properties members of a kind have, what forms of relation these properties have). For now, I will only stipulate that kinds, as opposed to any incidental category, should be able to serve epistemic aims such as predictions or explanations. This, as I will show in the following chapter, may be due to sharing multiple properties or for other reasons. For instance, if money is a kind, then we may be able to use kind membership in explanations and predictions where other pieces of paper or metal would not do *the relevant* epistemic work. This is because from membership in the kind money, we can infer much more than its material constitution: knowing that money can be exchanged for goods or services, or that it is accepted as a means of settling debts, allows us to make predictions about how the given instance of money will function in various contexts—unlike in the case of other pieces of paper or metal. Thus, although I subscribe to the ontological distinction between kinds and any incidental category, I will, for now, leave open the precise ontological character of kinds in general – and instead focus on the epistemic aims they may serve.

Let us now circle back to linking the mind-dependency thesis to Intentionalism.

1.1.2 Scope and strength of mind-dependency

Before we can bridge the gap between the mind-dependency thesis and Intentionalism, there is still one preliminary task left: *mapping the logical landscape of possible ways to unpack the mind-dependency thesis*. This subsection will address this task.

With our unit of analysis for Part I established, we can now update the mind-dependency thesis, which aims to provide an answer to the Placement Problem, accordingly.

Mind-dependency thesis: some social kinds, their properties and instances depend on minds

I will differentiate between mind-dependency theses based on scope and strength of the relevant sense of mind-dependence. Here is a preliminary overview. Regarding scope, there are monist and pluralist positions. Monists believe all social kinds are mind-dependent in the same way. Pluralists, in turn, accept that different social kinds are mind-dependent in different ways or that only some social kinds are mind-dependent. Regarding strength, I will focus only on differences within monist positions; more specifically, on the distinction between Intentionalism and what I will call Anti-Intentionalism.

These two distinctions—regarding scope and strength—will allow us to analyze the logical landscape of possible ways to unpack the mind-dependency thesis with the nuance such a project requires. In sketching the logical landscape of possible ways to unpack the mind-dependency thesis, I may sound a bit abstract; I will be more concrete once I start unpacking Intentionalism and placing examples and (putative) counterexamples in accordance with the logical landscape to be mapped here.

Let's start with the difference between monist and pluralist theories. Monists acknowledge that money, marriages, and other often-mentioned examples of social kinds such as leader, race, inflation seem to be, to say the least, a diverse lot. However, they argue that the diversity which social kinds seem to manifest is only surface-level: they hold that their mind-dependency thesis, in fact, has general coverage. On their view, mind-dependency is a necessary condition for *all* social kinds: money, marriages, leaders, race, and inflation are all mind-dependent. If any version of the mind-dependency thesis has general coverage, it is, then, not only a candidate for explaining the Placement Problem but also the General Problem as it purports to have uncovered the general conditions of existence for any social kind.

While subscribing to the condition of general coverage may seem controversial, a quick glimpse at the extant literature informs us otherwise (Thomasson 2003b, 277–278; Guala 2014, 61; Khalidi 2015, 111; 2016, 223; Mason 2021). For instance, Muhammad Ali Khalidi claims that “all social kinds can be said to be mind-dependent” (Khalidi, 2015, 111), or that

they are “unavoidably mind-dependent” (2016, 223). One reason general coverage of the mind-dependency thesis is uncontroversial to accept is that this is only one of the tenets of monist theories. In itself, general coverage is compatible also with a pluralist position, with differing versions of mind-dependency as disjuncts (Thomasson 2003b; Khalidi 2015; Mason 2021). It may very well be that, for instance, “money” and “inflation” are both mind-dependent *in some sense*, but they are in fact mind-dependent in different ways.

To get to monism, then, and exclude pluralism, general coverage needs to be coupled with the criterion of *unity*: that all social kinds are mind-dependent *in the same way*.¹ I will thus call the claim that all social kinds are mind-dependent in the same way a monist version of mind-dependency thesis. Pluralism, then, is any theory where unity is not met, whether general coverage is met or not; that is to say, pluralists may hold that all social kinds are mind-dependent but not in the same way or that some social kinds are mind-dependent and some are not.

Another reason general coverage of the mind-dependency thesis may be uncontroversial to accept is that the relevant sense of mind-dependency may be formulated in a sufficiently weak sense. This consideration leads us to the distinction between two versions of monist theories – Intentionalism and Anti-Intentionalism – which, nonetheless, both accept the general coverage of the mind-dependency thesis.

Intentionalism, as I will unpack in detail in the next subsection, holds that the existence of all social kinds depends on their being represented as existing through collective intentionality (Searle 1995, 2006; 2010; Ásta 2018). Here are, then, two examples of monist aspirations of Intentionalism from the literature: Searle holds that “[a]ll human social institutions are brought into existence and continue in their existence by a single logico-linguistic operation that can be applied over and over again” (2010, 62, my emphasis); Ásta, in turn claims that her “main objective is to offer a general framework to account for all social categories” (Ásta 2018, 86, cf. 3, 31, my emphasis).

Anti-Intentionalism, in turn, holds that the existence of no social kinds depends on their being represented as existing through collective intentionality (DeLanda 2006; Guala 2014). In this case, the fact that both monist theories accept general coverage stems from the fact that the relevant concept of mind-dependency in Intentionalism and Anti-Intentionalism varies in

¹ Here, some may object that there looms an important – but also widespread – philosophical problem about what exactly makes a notion unified: is it possible, for instance, to preserve unity through the *right or innocent enough* disjunctions? I am in agreement with Brian Epstein who, reflecting on the same issue, writes that while it is “possible to trivialize this question” and “play games” or pull “tricks”, unifying or distinguishing between notions at will, in fact, “this is a substantive question at the heart of social ontology: is the diverse furniture of the social world all explained in one way, by one generalized operation of our individual minds, or our collective minds, or our practices?” (2014, 48). I will, then, proceed by noting simply that the less disjunctive a mind-dependency account is, the more unified it is.

strength. Take Francesco Guala – whom I take to be the most famous and ardent contemporary defender of Anti-Intentionalism – arguing that:

Social science deals with phenomena that depend on mental states, just as biology deals with phenomena that depend on chemical substances. This is rather trivial and philosophically insignificant in itself (Guala 2014, 61).

He goes on to say that the position which I have labelled Intentionalism, “makes a much more interesting and contentious claim” as it argues that social kinds are “constituted by beliefs *about the kinds themselves*” (*Ibid.*, my emphasis).² This means that Guala, as an Anti-Intentionalist, puts forth a monist theory of the social, accepting that all social kinds are mind-dependent in the same way. However, he defines mind-dependency in a sufficiently weak manner – as some form of dependence on minds, without specifying that social kinds need to be represented as existing through collective intentionality – that it becomes uncontroversial to accept it, even within a monist framework. Anti-Intentionalism, then, is a less demanding way to be a monist.

We are now equipped with the distinction between monist and pluralist theories, along with the initial insights into Intentionalist and Anti-Intentionalist versions of monist theories, of the mind-dependency of social reality.

Having clarified my choice of unit of analysis (for Part I) and mapped the logical landscape of possible ways to unpack the mind-dependency thesis, it is now time to unpack Intentionalism more in detail in the following subsection. My aim, ultimately, is to see if Intentionalism can indeed be construed as a monist theory having general coverage over social reality, and as such, provide not only an answer to the Placement Problem but also the General Problem. This will be the task of the next section.

1.1.3 *Intentionalism unpacked*

Let’s start at the beginning, recapping some steps covered in the *Introduction*. Subjects have minds and – as will be emphasized in Part II – a sense of self. According to Searle, the building block of any social reality is the capacity of the subject’s mind called *intentionality* (1995, 6–7, 2010, 25 ff.). Intentionality, then, is taken to be compatible with a naturalistic ontology and

² Another interesting case of Anti-representationalism, it seems to me, is Manuel DeLanda. He argues that “...most social entities, from small communities to large nation-states, would disappear altogether if human minds ceased to exist. In this sense social entities are clearly not mind-independent (...) a realist approach to social ontology must assert the autonomy of social entities from the conceptions we have of them” (2006, 1; my emphasis). In other words, while he accepts that “human minds” are, in some sense, necessary for the existence of “social entities”, he rejects the claim that their existence is dependent on “the conceptions we have of them”. Although it is not entirely clear whether “autonomy of social entities” refers to the existence or the nature thereof, I take DeLanda to be making an existence claim, parallel to the first version of mind-dependence he explains (which is also an existence claim).

thus viable candidate serving as a bridge between the ontology of the natural sciences and social ontology; in other words, a viable candidate for answering the Placement Problem.

Of course, the concept of intentionality has a long history in philosophy, going back (at least) to authors such as Sartre, Husserl or Brentano. Intentionality is also often referred to as the “aboutness” of the mind or as the idea that mental states have an “object” or a “content” (Crane 1998, 2009). I will, however, not delve into historical investigations here. Instead, I will stick with my Searle-inspired definition from the *Introduction*, according to which intentionality is the mind’s capacity to direct its states – such as its beliefs, emotions and desires – at the external world and thereby *represent* it. These states, representing the world, as we have established, will be referred to as “intentional states”. Representing reality, then, means being in an intentional state which contains a stand-in for – i.e., is directed at – the relevant part of the external world. So, how exactly are we to understand the role of intentionality and intentional states in constructing social reality?

Take some examples. A belief is an intentional state: my belief that there is an approximately 40-centimeter-tall dog at my feet is directed at, or about, a dog and its property of being approximately 40 centimeters tall – something distinct from my mind itself. However, beliefs are not the only kind of intentional states: emotions such as hopes and fears, desires, or intentions (to do something – which, in a notoriously confusing manner, is not the same as the broader sense of *intentionality* just discussed) themselves are also intentional states which are directed at, or about, and so represent, a given state of affairs in the external world. For example, my desiring the company of my dog, hoping for good weather during our walk, and fearing an encounter with aggressive other dogs, and intending to take my dog for a walk are all about and so represent various (including future) states of affairs in the world.

Searle rightly notes that it is important to clarify the relationship between intentionality, consciousness and mental states in general (1995, 6–7; 2010, 26–27). For instance, there is an interesting debate as to whether all mental states are intentional or not. While most philosophers, including Searle (1995, 121), hold that some mental states, such as a headache, are not intentional in the relevant sense, others argue that in fact all mental states are intentional (cf. Crane 2009). I will leave this question open. There are, however, two other issues which are more interesting for my purposes.

First, there is the issue that not all intentional mental states seem to be conscious (Crane 2017). I will assume that a subject may have intentional states, such beliefs, emotions, and desires etc., which do not enter their conscious mental experience. Indeed, at times, to explain certain forms of behaviour, we ascribe unconscious intentional states such as “unconscious beliefs” – even if they are in contradiction with what the given person consciously believes. I will have more to say on non-conscious forms of intentionality in the

following sections of this Chapter, when discussing whether Intentionalism has general coverage over social reality.

Second, there is the question of whether all intentionality is mental. I made a tricky move here. On my view, intentionality, by definition, refers to the representational capacities of the mind. In this sense, the question of whether all intentionality is mental is rather quickly settled. However, if we focus on the property of something (e.g. the content of an intentional state) being a representation – being a stand-in for, i.e., being directed at or about the relevant part of the external world – then the question of whether there are other ways to represent the relevant parts of the external world than the intentionality of the mind becomes an intriguing one. For instance, it seems like road signs or paintings represent a given state of affairs in the world while they are not themselves mental. However, road signs and paintings as representations, one may argue, ultimately depend on the intentional states involved in their creation as well as their perception.

Without settling this far-reaching issue for now, I am noting that my aim in Chapter 2 will be to differentiate between mental forms of representation (for which I will reserve the term “intentionality”) and forms of representation which do not depend on intentional states in the relevant sense. This distinction will pave the path towards explaining the ontology of algorithmically constructed social kinds as well as how the building blocks of an algorithmic society differ from those of traditional ones.

For now, let us turn back to intentionality as the building block of social reality. How is, then, the intentionality of the mind required for the existence of social kinds? As Searle explains, social kinds are kinds of people behaving in a certain way which requires the existence of some form of a collective (Searle 1995, 2010; cf. Simmel 1910; Gilbert 1992). Often-mentioned examples of such social kinds include people going for a walk *together*, playing music *as part of a group*, as well as engaging in monetary *exchange* (Gilbert 1992, 17, Searle 1995, 23, 26, 2010, 44). These social kinds (groups of individual subjects behaving a certain way), it is further argued, would not exist without the relevant people having certain intentional states. What are the intentional states required for these social kinds? In a sense, it is held, people constituting the relevant collectives – going for a walk, playing music together, engaging in monetary exchange – must represent themselves through their intentional states as belonging to the given collective.

This conception of social kinds as intentionality-bound collectives has a long tradition. Take Georg Simmel, for instance, who argues that “the consciousness of constituting with the others a unity is the whole unity in question in the societary” (1910, 374). Margaret Gilbert, building on Simmel’s view, also agrees that “participants must see themselves as bound together in a highly specific way” (1992, 13). So, how are we to unpack the exact sense in

which the existence of social kinds requires people thinking of themselves as belonging to a given collective?

One way to do so is to hold that each person must have the relevant intentional states about the relevant intentional states of others (cf. Searle 1995, 24). More specifically, one needs to have intentional states which are directed at their walking together, playing music together, and at the relevant pieces of paper and metal being not simply material objects but also, say, means of exchange and store of value. Furthermore, and this is the crucial point, one must have intentional states which are directed at others sharing their intentional states about their walking together, playing music together, and that the relevant pieces of paper and metal are not simply material objects but also means of exchange and store of value. This is one way to unpack how the intentionality of individuals adds up to *collective* intentionality. Some, however, worry that this gives us too weak a conception of a social kind – and the relevant sense of a collective might also be too weak.

Searle, in fact, argues that collective intentionality is not merely a sum of individual intentional states about the intentional states of others. He argues that there is a special kind of intentionality, a “we consciousness” (1995, 24), or “we-intentionality” (2010, 46–47). Thus, the relevant intentional states, in the case of everyone involved, are “we are walking” or “we are playing music” or “we are engaging in monetary transactions”, and that ‘we’ is more than the two above-mentioned reciprocal intentional states (e.g., “I believe that you believe” plus “You believe that I believe”). Moreover, some take a step further and hold that we need to postulate even a special form of subjectivity such as “plural subject” which has intentional states of their own (Gilbert 1992, 15; cf. Searle 1995, 23 ff.; Burman 2007, 23–52; Franda 2022, 64). I will, however, not attempt to unpack here further what such a “we-intentionality” or “plural subjectivity” would be. This is because I will remain uncommitted about what exactly needs to be added to a set of intentional states of individual subjects to reach collective intentionality: whether a mere aggregate of individual intentional states about others’ intentional states suffices, or if we need to postulate a special form of “we-intentionality” or “plural subjectivity”.

The point so far, then, is that collective intentionality – of some sort – is held to be the building block of social reality as it is required for the construction of social kinds as elements of social reality. More precisely, Intentionalism, as a monist theory of the social holds that collective intentionality is a building block of *all* social kinds. Thus, we have reached our initial definition of Intentionalism.

Intentionalism: the existence of all social kinds depends on their being represented as existing through collective intentionality.

Let us see below the details of Searle’s Intentionalist view.

*

Searle holds that there is a crucial difference between social kinds such as people taking a walk together and people engaging in monetary transactions. He argues that, while admitting that the distinction is rather gradual than sharp, the latter belongs to a subset of more complex and more “codified” social kinds which he calls “institutional” kinds (1995, 26–29, 88; 2010, 90 ff.). Institutional kinds, on his view, include kinds such as money, marriages and presidents. The existence of institutional kinds, then, needs further explanation as it requires not only collective intentionality, but *collective intentionality of a special form*.

Now, as a preliminary note, in the next section, I will challenge Searle’s distinction between the special subset of social kinds which he calls institutional kinds and other social kinds. However, to do so, I will first aim to clearly and charitably reconstruct his views, sticking with his terminology for the time being. So, what exactly do we need, according to Searle, next to collective intentionality to explain the existence of what he calls institutional kinds?

On Searle’s view, institutional kinds are a subset of social kinds demarcating those social kinds which are associated with a *social status* (what Searle calls “status function”). I take social statuses to be a – borrowing the expression of another Intentionalist, Ásta (2018) – “constraints and enablements” which could not exist given the “sheer physics” of their wielder as they require being represented as existing to exist. Searle’s well-known example of an institutional kind and the respective social status is that of a boundary between two tribes (1995, 39; 2010, 94 ff.). He asks us to imagine a physical boundary, say, a great wall, separating the territories of two peoples. After some time, the wall begins to crumble and slowly deteriorates until only a “line of stones” of its original physical structure remains. At this point, members of the two tribes could easily trespass, yet they refrain from doing so. This is because they *recognize* the line of stones to *count as a boundary*.

Searle, then, argues that such a boundary is an instance of an institutional kind, and all institutional kinds adhere to the same logical pattern: that is to say, what goes for boundaries, also goes for money, marriages and presidents. The logical pattern goes as follows: through our capacity of collective intentionality, we collectively *recognize* “constitutive rules” in the form of “X counts as Y in C”³ where X is some pre-institutional entity, either a person or an object, Y is the (institutional) status and C is a given context. In this case, X is a line of stones, Y is a boundary, and C is the specific context where the territories of the two tribes are delineated. Thus, Searle holds that only institutional kinds involve imposing a social status on a person or an object. Furthermore, he points out that the step from X to Y requires *a special*

³ Although Seale refined his analysis somewhat in his later work (2010, 90–122), his basic formula is still a helpful heuristic to understand his theory of the construction of what he calls “institutional facts”.

form of collective intentionality which *recognizes* and thereby, *represents* the Y term, that status itself, as existing.

Notice that we have switched from the general term “collective intentionality” to a specific intentional state in question: “recognition”. This is in line with Searle’s choice of word for the relevant notion of intentional state. In his earlier work, he uses various terms to denote the relevant intentional states required for the existence of the elements of social reality and talks not only about recognition but also “belief” and “acceptance” (1995, 32), while in his later work, he mostly sticks with “recognition” (2010, 8, 57, 103–104). Following the terminology of his later work, I will also stick with the term “recognition”.

Searle argues that he uses the term “recognition” as he does not intend to imply that the relevant sense of collective intentionality entails that people would be in any sense approving about the social or institutional kinds they construct and maintain. It is sufficient, he explains, to “go along” with the given institutional structure (2010, 8, 57, cf. Passinsky 2020b, 437). This is because even to go along, it is necessary to possess the relevant intentional states representing the associated social statuses. Recognition, then, will refer to the intentional states representing the given status as existing, which then enables the intentional action of using or treating instances of institutional kind as such. For example, in the case of money, recognition involves subjects treating pieces of paper as currency, thereby attributing to them the status of being a means of exchange even if they aren’t approving about the concept of money or other institutional kinds which govern its use.

One may wonder whether it is possible to *pretend* to go along. Perhaps in Theatre of Pompey, the group of Roman senators who assassinated Caesar only pretended to go along with the institution of his dictatorship. However, this still allows the Intentionalist to show that even in such cases, one’s pretending to maintain a given institution requires some form of the recognition of the relevant institutional kinds their powers that one pretends to adhere to. It is *precisely* because recognition allows for such wide continuum of feelings harboured towards the institutions which are being *recognized* that politicians and pollsters aim to monitor approval ratings. In sum, I will also use the neutral term recognize or recognition to refer to the intentional states representing the given status as existing – whether approvingly or not. (I will have more to say on various forms of “going along” with social reality in Part II.)

Now, it important that Searle argues that this special form of collective intentionality, collective recognition, is required for every institutional kind and thus every social status, as they require *language*. As he writes, “all human institutions are essentially linguistic” (2010, 63, 12–13, 61–89). To explain his reason for holding that language is necessary for institutional kinds, he holds that one can make a rough and ready distinction between two kinds of intentional states: language-dependent and non-language-dependent ones (1995, 61 ff.). Of course, language is required to represent both kinds of intentional states. However, in

the case of language-dependent intentional states, Searle argues, language is required for the very existence of these intentional states as well. I will reserve, for now, the term recognition for language-dependent intentional states which are, as Searle argues, needed for the existence of institutional kinds.

Let's look at a pair of contrast cases. The intentional state that "she is president" is a language-dependent intentional state, i.e., recognition; one could not have such an intentional state without language since the relevant status requires representation through language-dependent intentional states to exist. Now, if the president has a dog, it may believe and feel various things about her such as believing that it hears the footsteps of its owner – but will not *recognize* her as president. Thus, while recognizing that someone is a president is an intentional state which is necessarily language-dependent, believing that one hears the footsteps of a person they know is not.⁴

Furthermore, Searle sets a rather high bar for a set of signs to count as language – one which, many might argue, only human language can meet. Without aiming to reconstruct the details of Searle's view on language, or provide my own account, the important thing to note is that he argues that languages must have the property of generativity (2010, 64–65). This means that languages should be able to generate, in theory, an infinite variation of sentences and thus intentional states. This generativity is needed to formulate intentional states about constitutive rules which cannot only represent the here and now and the near future as well as likely possible states of affairs, but also, for instance, the distant past or future, or even unlikely but possible states of affairs. This is why language is needed for those institutional kinds, such as that of a president, which come with a status which consists of complex rights and responsibilities, often referring to unlikely but possible states of affairs.

So far, we have seen that Searle holds the following:

- (i) institutional kinds are a subset of social kinds,
- (ii) while all social kinds require collective intentionality for their existence, only institutional kinds are associated with a *social status*,
- (iii) institutional kinds require a special form of collective intentionality, that is, collective recognition, which represents the given statuses as existing through constitutive rules,

⁴ Of course, the very word "president" is not necessary for the existence of secretaries of state. As Searle notes (1995, 52. cf. 37, 47, 52 ff.; 2010, 14), it suffices to possess the concept which denote language-dependent the status and powers which being a secretary of state endows one with. Given that the role varies across nations, the exact list of relevant concepts depends on the specific context we have in mind; but concepts such as nation, state, government etc. are good candidates for being required to be available for there to be presidents.

- (iv) the *recognition* which constitutive rules require are language-dependent intentional states since the rules are generative and only language can represent that generativity.

Now, to make the Intentionalist view as clear as possible, a few more points regarding the relevant sense of “dependence” are in order.

*

It seems to me that the Intentionalist claim that the existence of all social kinds, which includes institutional kinds, depend on being represented as existing through collective intentionality (collective *recognition* in the case of institutional kinds) is not denying the importance of actions and their causality. For example, take the institutional kind of money: money would presumably not exist if people collectively represented money as existing, but nobody ever engaged in the intentional action of selling or buying items. Remember that the jump from X (a piece of paper) to Y (money) *requires* the representation of the status; this means that the relevant sense of dependence is best construed in terms of some form of a necessary, but not always sufficient, condition for existence (cf. Searle 2006, 14; 2010, 10, 17).⁵ More precisely, then, Intentionalism holds that social kinds are mind-dependent in the sense that their existence *partly* depends on being represented as existing through collective intentionality.

Going further, the relevant sense of collective intentionality being a necessary condition for the existence of social kinds is a form of *ontological dependence*. Indeed, many in the extant literature hold that social ontology should deal with how social reality is constructed and maintained *ontologically*, not merely *causally* as the latter seems to be a messy empirical matter, to be elucidated by the social sciences (e.g. Searle 1995, 45, 156; Haslanger 2012, 131; Epstein 2016, 151–152; Schaffer 2017, 2451). In contrasting “causal” to “ontological”, I am following the terminology of the extant debate in analytic philosophy – this contrast may be somewhat misleading as, surely, causation is as ontological as anything can get. The difference will get clearer once I unpack the relevant notion of ontological dependence at hand.

Now, Searle calls the relevant sense of ontological dependence “constitution”. This means that the collective recognition which is required for the existence of the given social kind does not simply cause but also partly constitute its existence. Constitutive dependence, then, in contrast to causal dependence, is a form of ontological dependence which holds that, when X is constitutively dependent on Y, then, Y is part of X’s being an X. Thus, constitutive

⁵ Some have suggested that Intentionalists argue that collective intentionality representing the given social kind or its instances constitute a sufficient condition (cf. Guala 2014; also on this point, cf. Franda 2021, 41, Khalidi 2015, 98; 2022). That would mean that once the existence of a given kind is collectively recognized, it exists. This seems to me to be too mentalistic.

dependence is taken to mean that Y cannot be just any “accidental” part of X, it is necessarily part of being X. For example, the bones in a dog’s body are not causing the respective body; they constitute the body; they are a necessary part of its being a body (not considering more radical, body-without-organs-like scenarios for now). Put simply, a cause is usually held to be spatiotemporally distinct from its effect: it happens earlier and does not occupy the same space; however, a constitutive part is usually held to be at the same time at the same place as the constituted entity as they are in a part-whole relationship.

Many in the extant social ontology literature draw further and more refined distinctions between causal and ontological dependency as well as between various kinds of ontological dependency to clarify the relation between minds and social kinds. For instance, Rebecca Mason (2021) provides an analysis of the mind-dependence of social kinds – inspired largely by Kit Fine’s (1995) work – where she differentiates between “essential” and “modal/existential” forms of ontological dependence. Furthermore, there is a rich discussion in social ontology on applying the latest toolkit of contemporary analytic metaphysics such as the concept of “grounding” to clarify the relevant notion of dependence in the mind-dependency thesis as well as the notion of “social construction” in general (Mikkola 2015, 2016; Schaffer 2017, 2019; Griffith 2017, 2018; Epstein 2018, 2019; Passinsky 2020a, 2020b, 2021). For our purposes, however, a rather crude distinction between constitutive and causal dependence will suffice.

Now, to make this contrast between causal and constitutive dependency clearer in the case of social kinds, consider, again, the institutional kind money. Take money in the form of a piece of metal coin. Again, taking causation as a physical production process, we may say that the physical properties of the metal coin such as its unique shape, mint mark, inscriptions, engravings etc. are causally mind-dependent. It is rather unlikely, if not physically impossible, that “nature” would produce such a design without human intentionality (cf. Diaz-Leon 2015, 1143). Now, a metal coin would remain existing as the physical thing it is even if everyone’s mind were wiped blank. However, its status of being money would indeed be lost, given that the status is constitutively mind-dependent. It is, in other words, a constitutive part of what it is for a metal coin to be money that it is collectively recognized as money.

In Searle’s own wording, in the case of social kinds, “the attitude that we take toward the phenomenon is partly constitutive of the phenomenon”, illustrated with the example of the institutional kind of money: “the belief that the type is a type of money is constitutive of its being money”, or with cocktail parties and wars: “part of being a cocktail party is being thought to be a cocktail party; part of being a war is being thought to be a war” (1995, 33–34). It is, for Searle, this “noncausal character” (1995, 45) of the creation of social, including institutional, kinds which explains why we may perceive what Searle calls “an element of magic” in the construction of social, including institutional, kinds – and which motivated our Placement Problem in the first place. Intentionalism, then, seems to provide a compelling

answer as to how to place (at least some) social kinds into a naturalistic ontological framework through unpacking the relevant sense of mind-dependence. For instance, in the case of money, it holds that the relevant collective recognition is constitutive of the existence of money.

As mentioned above, on a standard understanding of a constitutive relation, it is assumed that the dependence relation at hand is synchronous. To illustrate this in the case of the relevant sense of mind-dependency of social kinds, Searle mentions that “*continued* human cooperation” is needed for the existence of social kinds and their associated social statuses (1995, 40, cf. 21), or that they “require collective intentionality, both for their initial creation *and* for their *continued* existence” (2010, 59, my emphases). Or, as Muhammad Ali Khalidi puts it, social kinds “require human minds to sustain them at all times” (2016, 239–241).

Ásta, in turn, reflects on synchronicity of the constitutive relation through highlighting the fact that one’s membership in social kinds changes depending on the context. Specifically, her examples show that as the relevant collective recognition changes when one is travelling, so does one’s membership in social kinds. As she puts it, “Gómez- Peña is mistaken for a tourist from Argentina in Chinatown, at the Bollywood Café he is ‘the wrong kind of brown’, in the Castro he is an older gay man, and in the financial district he is nobody” (2018, 1). This change in social kind membership, then, is explained by the fact that social kinds are constitutively dependent on the collective recognition of the relevant others in a specific context and the mind-dependence at hand is synchronic.⁶

I will have more to say on the relation between synchronicity and constitution in the next Chapter; for now, given the above, I will take the relevant sense of mind-dependence in Intentionalism as a (synchronic) constitutive form of ontological dependence.

Now that we have explained Searle’s Intentionalism through unpacking the relevant sense of intentionality in the construction of social kinds, identifying institutional kinds as a subset of social kinds as well as recognition as a subset of intentional states referring (so far) to language-dependent intentional states, and discussing the relevant sense of dependence, we may finally update our definition of Intentionalism as follows:

Intentionalism Searlean: all social kinds are mind-dependent in the sense that their existence constitutively (and synchronously) depends on being represented through some form of collective intentionality as existing.

⁶ The idea of synchronic ontological dependence seems to be endorsed even by non-Intentionalist mind-dependence theories of the social world, such as Manuel DeLanda’s: “most social entities, from small communities to large nation-states, would disappear altogether if human minds ceased to exist” (DeLanda 2006, 1).

A unique subset of social kinds, i.e., institutional kinds, are mind-dependent in the sense that their existence constitutively (and synchronously) depends on being collectively recognized as existing.

Let us now move onto the challenges that Searlean Intentionalism faces.

1.2 Challenges to Intentionalism

In this section, I will aim to reconstruct the most charitable version of Intentionalism to see if it can live up to being a monist mind-dependency theory of social reality. To do so, I will consider challenges that Intentionalism has received and revise the view accordingly. For the sake of clarity, I will break down the challenges to Intentionalism into three main areas. First, its scope; second, the relevant notion of “recognition”; and, third, the relevant notion of “dependence”.

I take it that the challenges might be carried out from an Anti-Intentionalist as well as a Pluralist perspective: one could argue that *no* social kind is mind-dependent in the sense that their existence constitutionally (and synchronically) depends on being represented through some form of collective intentionality as existing or that some social kinds are not mind-dependent in that sense, respectively. Against the Anti-Intentionalist, then, it is sufficient for the Intentionalist to show that there are some cases of social kinds where the Intentionalist explanation of their ontology holds. Against the Pluralist Challenge, however, the Intentionalist needs to show that the Intentionalist explanation holds for all social kinds. Thus, given that the Pluralist Challenge constitutes a weaker claim and so is more difficult to defend against, I will, when relevant, formulate the challenges from a Pluralist perspective.

1.2.1 *Revising scope: The social reality of non-human animals*

The first challenge is about scope. As a monist theory, Intentionalism claims to have general coverage over social reality. In other words, the question arises: how does intentionalism fare with addressing not only the Placement Problem, which aims to find the place for some (human) social kinds in the ontology of the natural sciences but also with the General Problem, which aims to uncover the general conditions of possibility for the existence of *any* social kind?

In the *Introduction*, we have mentioned that “the social” is often contrasted not only to the level of the individual but also to paradigmatic examples of “nature”. Now, we have also stipulated a naturalistic ontological framework (however its details may be spelt out) according to which nature is the only reality; consequently, social reality is a subset of (natural) reality. The contrast to nature, then, is more precisely termed as the contrast

between (natural) social reality and (natural) non-social reality. Can Intentionalism, then, have general coverage over all of social reality?

Let's begin with the issue of whether Intentionalism can accommodate non-human social kinds. Searle himself mentions examples of non-human social kinds such as hunting hyena clans (1995, 38, 122, 228), however, he holds that animals do not have institutional kinds – and even reflects on his exclusion of animals from the realm of institutional kinds (he talks of “facts” instead of kinds):

In this book I sometimes contrast humans and other animals. The point is not to make a plea for the superiority of our species but to analyze the logical structure of some distinctive human phenomena. If it should turn out that some other species also have income tax, presidential elections, divorce courts, and other institutional facts, I welcome them to the club. (2010, 7)

Most of us will very likely agree that animals tend not to, say, hold two-round presidential elections. However, as many have argued (Wilson 2007, Velasco 2023), excluding animals from the realm of institutional kinds may be too hasty. Institutional kinds have been defined as a subset of social kinds which have an associated codified and, therefore, language-dependent social status which endows one with certain constraints and enablements which does not depend on “sheer physics”. Institutional kinds, then, on Searle's view, depend on collective recognition, that is, language-dependent intentional states.

Many argue, in turn, that animals (just like, as I will argue below, humans – but let us set this aside for a second) may attribute social statuses in the relevant sense without language, constructing constraints and enablements not stemming from “sheer physics”, and thus construct not only social but also institutional kinds in the relevant Searlean sense (Wilson 2007, Velasco 2023). If this argument is right, then “recognition”, which, so far, was stipulated to be a language-dependent intentional state, is not the right term for the relevant sense of intentional state in the case of non-human animal social reality. Thus, in exploring the argument that animals also construct social statuses, I will remain neutral on what term we should use in the case of animals and will stick with the neutral term “intentional states”.

Critics of Intentionalism argue that Searle simply overlooks certain empirical facts and bring examples of animals being able to attribute social statuses which do not depend on “sheer physics” but on their collective – non-language-dependent – intentional states representing the status as existing. For instance, while Searle argues that “the alpha male wolf is treated with fear and respect because of his physical strength” (2010, 95), critics point out that the alpha male status of wolves is, in fact, not merely a result of a physical stimulus-response structure, but depends on the wolves' intentional states representing other wolves' position in the family; not to mention that in some cases, “wolf social organisation changes: an unrelated wolf can be adopted into a pack (...) It is the pack that assigns the alpha status to

the outside wolf and that collectively recognises that status” (Velasco 2023, 35). Another kind with an associated status in wolf societies which is not physically determined on a stimulus-response fashion is the marking of territories – not unlike Searle’s example of a boundary between two tribes (Wilson 2007, 144, Velasco 2023, 35). These examples show that some animals, such as canids, may very well be able to construct social statuses and, therefore, on Searle’s view, institutional kinds, without language.

If there are indeed social statuses without language, then we cannot retain Searle’s original distinction between non-institutional social kinds and institutional social kinds. As we have seen, Searle holds that the former have no associated status and are not dependent on language-dependent intentional states, while the latter are a subset of social kinds that are associated with a status and dependent on language-dependent intentional states, which we have called recognition. How should we proceed?

First, there is clearly a spectrum of social kinds, from walking or hunting together to engaging in monetary transactions, which depend, respectively, on a spectrum of intentional states, stemming from non-language dependent ones to language-dependent ones. Second, we should acknowledge that not all social statuses require language-dependent intentional states. Third, instead of the term “institutional kind”, I will talk of more and less “codified” social kinds and their associated statuses where “codified” social kinds require language-dependent intentional states for their existence. This implies that not only codified social kinds have associated statuses. As for the relevant sense of “recognition”, let’s keep it intact for now; I will have more to say on the issue of the language-dependence of intentional states in the next subsection.

For now, let us turn back to Searle and his views on non-human animal’s social reality. As we have seen, while Searle realized that by tying what he called “institutional kinds” to language, he excludes non-human animals (without realizing that he is wrong in doing so), he did accept that there are non-human animal social kinds which are non-institutional. Let us, then, consider his putative examples thereof.

As mentioned above, Searle refers to the example of a hyena clan going for a hunt, arguing that the hunting of hyenas involves effective coordination, which would not be possible without hyenas having collective intentional states along the lines of “we are hunting” (taking Searle’s view of collective intentionality as “we-intentionality”) (1995, 38, 122, 228). This is perhaps uncontroversial: clearly, many animals engage in forms of coordinated behaviour – such as elephant herds protecting their calves, or dolphin pods hunting fish – which, many may argue, would not be possible without collective intentionality of the relevant form. So, one may think that identifying collective intentionality (of a non-language-dependent form) as the building block of social reality has the advantage of being able to incorporate non-human social in one’s social ontology.

However, the question arises, what about coordinated, arguably social, non-human animal behaviour *without* collective intentionality (of any sort)? Without providing a clear or detailed taxonomy of animals with and those without the capacity of intentionality – and that of collective intentionality – take the often-mentioned example of the behaviour of so-called social insects such as ants and bees as part of their colonies and hives. Ants show remarkable coordination when foraging or nest-building; bees work together in accomplishing complex tasks such as constructing their hive. So, what should the Intentionalist say about the behaviour of these animals? First, one may point out that whether ants or bees in fact do or do not have intentionality is an open question itself. I agree; attributing intentionality to various animals depends not only on a clearly defined concept thereof but also, certainly, on empirical facts about the animals concerned.

Suppose, however, that the Intentionalist is pressed to accept that these arthropods lack intentionality as defined above. In fact, some may insist that there are convincing arguments which purport to show that we should reserve the term “intentionality” for cases where it makes sense to attribute “minds” as specifying the intentional states is needed to explain the relevant behaviour. Kim Sterelny, for example, argues that employing the language of intentionality plays no explanatory role in “simpler” systems where the relevant interactions are “chemically mediated” or can be “characterised physically” (1995, 254, 258). He happens to illustrate his points with ant colonies, where a single physical stimulus prompts nest hygiene.⁷ Another memorable example he brings is that of naked mole-rats inducing infertility in subordinate females:

The suppression of ovulation in subordinate female naked mole-rats is chemically mediated. These chemicals merely function to suppress ovulation: I see no explanatory point in supposing they have the imperative content “Thou Shalt Not Ovulate!” (Sterelny 1995, 254)

Perhaps Sterelny and others who are more restrictive when it comes to the concepts such as intentionality rely on outdated dichotomies between animals that merely react to the world they inhabit and those also representing it through the intentionality of their minds. And, further, perhaps such dichotomies will not be useful explaining the ontology of not only a colony of ants, a swarm of bees, a clan of hyenas, a herd of elephants, a pod of dolphins but also a parliament of owls, a gaggle of geese, a pack of wolves, a troop of monkeys or a school of fish. The point is simply this: is it ever justified to tie sociality to some form of collective intentionality, wherever we may take it to begin?

⁷ Cf. “There is a single and simple physical stimulus for nest hygiene, and nest hygiene itself is simple removal. The stimulus can be characterised physically rather than informationally. The behaviour takes place whether appropriate or not, just so long as the simple initiating condition is present. Ants remove dead nest mates when they detect oleic acid the decaying ant releases. Thus a live though acidic ant will be expelled, struggle how she will.” (Sterelny 1995, 258)

It seems to me that there are two options for the Intentionalist to proceed.

The first is to bite the bullet and insist that social reality is tied to intentionality, i.e., to hold that sociality begins where collective intentionality begins. Supposing ants lack intentionality in the relevant sense, this entails accepting that the “social” insects are not, in fact, social. In fact, some Intentionalists insist that holding that collective intentionality as the “mark of the social” in the sense that collective intentionality is necessary for the existence of social reality helps us separate intuitive cases of sociality from non-sociality where the qualifier “social” is merely an “illusion” or a “metaphor”. As Rom Harré writes, “A whiff of pheromone, the prevailing scent of the hive, is the causal substrate of the *illusion* of sociality that we have in ant watching” (Harré 1997, 208, my emphasis). Similarly, John D. Greenwood (1997, 7–8) argues that it is a virtue of the intentionalist view that it cuts nature (or, in this case, the social) at its joints, as it manages to exclude collective behaviour devoid of collective intentionality from the realm of the social. Thus, for him, holding that collective intentionality is the mark of the social in the sense that collective intentionality is necessary for the existence of social reality...

...also has the advantage that it leaves open the question of whether some animal populations and behaviors are social in nature, since it allows that some forms of animal behavior may be governed by nonlinguistic arrangements, *while denying sociality to those behaviors that are mechanical products of physiological or chemical processes.* (*Ibid.*, my emphasis)

This is also the route that Searle seems to take; while he acknowledges that he merely “stipulates” (1995, 122) that collective intentionality would be the mark of the social, he seems happy to exclude ant colonies or beehives from social reality.⁸

The other option is more cautious. First, one may point out that there is a difference between the Placement Problem and the General Problem, then explicitly restrict the scope of the investigation to which Intentionalism provides an answer to the Placement Problem. In other words, the Intentionalist can point out that that their project was motivated by the struggle to place human social phenomena such as marriage and money in the rest of our naturalistic picture of the world. The first step towards answering this question, the Intentionalist answer goes, seemed to be that *some* social kinds are mind-dependent in the sense that their existence depends on their being collectively recognized as existing; the “glue”, then, (Gilbert 1992, 13; Searle 2010, 9, 89) holding together these *human* social kinds, then, is a unique form of collective intentionality. How exactly social reality is constructed in the case of non-

⁸ Cf. “It is common, for example, to read that certain ant colonies have slaves or that beehives have queens. I think such manners of speaking are harmless metaphors, especially where the so called “social insects” are concerned, but it is important to keep reminding ourselves that for a community literally to have slaves or literally to have a queen, the participants would have to have the apparatus necessary to represent something as a queen or as a slave. Just behaving in certain ways, where behavior is construed solely in terms of bodily movements, is not sufficient for a community to have a queen or to have slaves.” (Searle 1995, 37)

human animals, they may argue, is a different question – one addressed by The General Problem.

The Intentionalist, then, would argue that they do not have to be troubled by the fact that some animals seem to have the ability of collective intentionality, even to the extent of having non-language dependent social statuses, while some other animals, which also exhibit complex collective behaviour, do not seem to have collective intentionality. Simply put, they remain neutral about non-human forms of sociality (cf. Gilbert 1992, 442–444).⁹ In other words, the scope of Intentionalism is restricted to human social reality.

However, to retain its position as a *monist* theory, the Intentionalist needs to do two things. First, they need to subscribe to a stronger claim than merely providing an answer to the Placement Problem: they need to provide an ontological explanation of the mind-dependence of *all* human social kinds. Second, they also need to restrict the general scope of monism only to human social reality. This may seem like an easy cop-out. However, I take it that the Intentionalist could point out that having general coverage over human social reality is ambitious enough in itself and thus merits the label monism of human social reality. Now, while I will take the General Problem to remain indeed general (covering both non-human and human forms of social reality), monism will henceforth refer only to general coverage of human social reality. Let us, then, proceed on this more cautious note.

Accordingly, our definition yielded by our first-stage revision of Intentionality at this point looks as follows:

Intentionalism first-stage revision: *human* social kinds are mind-dependent in the sense that their existence constitutively (and synchronically) depends on being represented through some form of collective intentionality as existing.

In the following, then, I will use “social kinds” as referring to human social kinds, unless indicating otherwise.

1.2.2 *Revising recognition: Mistakes, language and the unconscious*

In this section, I consider possible objections to Intentionalism which challenge the relevant sense of collective intentionality which I labelled collective recognition. Specifically, I will

⁹ It seems to me that this more cautious strategy is the one that Gilbert chooses: “Concentration on the human case here has been productive. It led to the articulation of the concept of a plural subject. If, in order to cope with our talk about animals, the plan had been to articulate a general concept literally applicable without doubt to nonhumans or nonthinkers, it is possible the concept of a plural subject would simply have been overlooked. (...) Meanwhile, nothing in what I have said entails that animal populations may not be plural subjects. (...) This depends upon facts about the inner life of animals. No attempt to make a judgment on this matter need be made here.” (Gilbert 1992, 444, cf. 442–444).

consider three challenges to the notion of recognition. The first one is about being mistaken; the second one is about not-being dependent on language; the third is about not being conscious.

Let's start with the challenge about being wrong. Guala famously formulated this as an Anti-Intentionalist objection, challenging all cases: "we can all be wrong about the nature of *any institutional kind*" (2014, 61–62, 67, my emphasis; Guala uses Searle's vocabulary of "institutional kinds" – in our terminology, I will take his view to apply to all "codified" social kinds), without the kind ceasing to therefore exist. From this, the challenger may argue, it follows that collective recognition is not *constitutive* of (and so not necessary for) the existence of any (codified) social kind. We can, to construe the challenge more charitably and make our job more difficult, take Guala's claim in a weaker version holding that everyone can be wrong about the nature of *some* (codified) social kind, without the kind ceasing to exist. What does "being wrong" about the "nature" of the kind amount to in this context?

The key here is to differentiate between the parts of – especially codified – social kinds. In the Searlean formula of X counts as Y in C, where X is some pre-social entity, either a person or an object, Y is the status and C is a given context, X and C together delineate certain membership conditions, while Y specifies the status itself, that is, the constraints and enablements brought about by membership in the given kind. Thus, we may differentiate between being wrong about membership conditions and being wrong about the status of members of a given social kind.

Guala's point, in turn, is about being wrong about the membership conditions (2014, 67). He argues that we can be wrong about membership conditions – what matters is how we "treat" the members of the given kinds.¹⁰ But it seems to me that the Intentionalist can accommodate the fact that we are often wrong about the membership conditions of a given social kind. Searle, in fact, explicitly admits that we can be wrong about *why* we recognize membership in any social kind. However, and I agree with him on this, having such misconceptions does not mean we cannot recognize the status which is associated with the given social kind. His example is that one might mistakenly believe that it is in order for something to be money it must be backed by gold, yet still treat it as a medium of exchange because they recognize that pieces of paper have the status to be a medium of exchange (1995, 47–48). For the existence of social kinds, then, it seems to me that what matters is not whether people understand *why* something is a member of the given social kind but the relevant status of the social kind.

Can we, then, be wrong about the very status associated with the given social kind? Only to some degree. Searle's example of his dog's inability to recognize the status of money

¹⁰ "What matters is not what type of attitude people have toward a certain class of entities (the conditions they think the entities ought to satisfy in order to belong to that class), but what they do with them in the course of social interaction." (Guala 2014, 65)

illustrates this point: while the dog may be trained to fetch dollar bills for food, it doesn't "recognize" them as currency because it lacks the intentional states representing money as having the status of a means of exchange (1995, 70). However, there are degrees of recognition corresponding to degrees of status. In the case of money, some may only recognize its status allowing the buying of goods, while others' recognition may also encompass its status in various forms of investment. In the case of the status of being a president, in turn, people may not recognize all the rights and responsibilities this status entails, but only part of the status. Accordingly, each level requires the relevant set of recognitions.

Perhaps there are "essential" statuses of money which need to be recognized for its existence, say, its being recognized as a medium of exchange for a multitude of goods and services. Perhaps if everyone recognized money *solely* as a means to acquire food, we might hesitate to call what they called "money" *money*. In a similar way, perhaps the essential feature of being a president involves the right to veto legislation passed by Congress and the responsibility to safeguard national security, but not, say, the right to appoint ambassadors to foreign nations. Determining these "essential" statuses, however, is the task of another project; my point is merely that different levels of recognition constitute different statuses.

In sum, while we can be wrong about membership conditions, we cannot collectively be wrong about (at least the "essential" features of) the relevant statuses.

*

The second challenge to the notion of "recognition" concerns the role of language and, as a corollary, language-dependent intentional states in constituting social statuses. We have so far stipulated, à la Searle, that the relevant sense of recognition is a language-dependent intentional state. However, even if Intentionalism aims only to explain the ontology of a "distinctive human phenomena" (Searle 2010, 7), that is, to only answer the Placement Problem, there is an important point of criticism to be made.

Not even in the case of humans do all social statuses require codified social kinds and so language-dependent intentional states. Thus, there are statuses involved both in cases of codified social kinds as well as in cases which Searle originally identified as "non-institutional", that is non-codified social kinds. Take Searle's own thought experiments about (what we would call) a leader of an early human tribe which do not possess language which could represent the given kind of a "leader" and its associated status. The example of such a leader is, in fact, intended to show that there are examples of social statuses – constraints and enablements not fully determined by intrinsic physical properties – in human social reality which depend on non-language-dependent intentional states (2010, 65 ff.). And perhaps we do not even need to go so far as to speculate about pre-linguistic caveman societies: already

in high school, there are “bullies” and “cool” people (cf. Ásta 2018, 18–21) even if the students do not have language-dependent intentional states representing the bullies and the cool ones as such.¹¹

Here, the behaviour of the people involved merely presuppose intentional states which represent in a way as to make distinctions between different kinds of people in a given community but not the existence of language-dependent intentional states. For instance, it seems enough to represent the here and now or the near future as well as likely possible states of affairs to represent these distinctions. Notice that both the leader in the cave-society and the bully in the school require intentional states representing instances of the relevant kinds, and the relevant intentional states are examples of constitution (rather than mere causation). This is because the relevant statuses and the accompanying behaviour (e.g., ways of respecting or being afraid of the person in question) are partly constituted by, and so are only possible if, the relevant intentional states (recognizing someone as a leader or a bully) are in place. Thus, humans can also construct social statuses without language.

Even in the simple social behaviour of taking a walk together, I am not convinced that there is no status involved. The status may not be “codified” and thus not be dependent on language-dependent, but it does carry constraints and enablements which does not stem from the sheer physical properties of my walking companion. For instance, one might save someone from tripping over if they are indeed walking together, while they might not do the same with someone with whom they are not walking together.

Do these humans “recognize” each other as leader, as a bully, as cool or as a walking companion? Perhaps this is merely a question of words; in any case, I choose to *broaden the notion referred by the term recognition to encompass both language-dependent intentional states (on which social statuses which are more codified depend) as well as non-language dependent intentional states (on which statuses which are less codified depend).*

Having thus broadened the relevant sense of recognition, one may differentiate between versions of Intentionalism regarding the role of *rules* in the imposition of social status. We have seen that there is a spectrum of social statuses, from more codified ones requiring language-dependent intentional states to less codified ones which do not require language-dependent intentional states. Now, there are debates around whether social statuses are imposed through the recognition of a codified rule which specifies the membership conditions

¹¹ No wonder scientists are prone to make analogies between the hierarchies of a hyena clan or a rhesus monkey community, or, for that matter the pecking order of hens, and the social kinds humans construct in high-school: “(h)yenas in clans, like eighth graders in middle school, are sorted into complex social hierarchies” (Lyon 2019) or “(l)ife for female rhesus macaques is a little like being trapped in high school” (Caruso 2016). The basis of the analogies – whether or not the above-quoted authors are aware of it as such – is that these are all examples of social kinds which are not language-dependent.

of status as well the constraints and enablements entailed by it or through the direct recognition of each instance of the status itself.

Searle holds that for less codified social kinds, such as “cocktail parties”, there may be no “official” or “formal” rules about the membership condition and status of cocktail parties (1995, 32–33, 53). This implies that each instance of such social kinds must be collectively recognized to exist. On the other hand, more codified social kinds, such as money, depend on the collective recognition of rules specifying the membership conditions and the constraints and enablements entailed by the associated status.

This implies that the relevant collective recognition is about the kind, thus an instance of the kind can exist even without collective recognition of each individual instance. As Searle puts it: “[a] single dollar bill might fall from the printing presses into the cracks of the floor and never be used or thought of as money at all, but it would still be money” (1995, 32). Ásta (2018, 27–31), on the other hand, argues that it is *never* rules which impose social status but the direct recognition of the relevant “others” representing one as an instance of a given kind. Thus, she argues that for every social kind, be it a cocktail party or money, each instance must be directly and collectively recognized to exist.

I will adopt the Searlean, more inclusive view where in some cases, social statuses are imposed through the intermediary recognition of a rule specifying the membership conditions and the status itself, and in others, through the direct recognition of the status of an instance itself. In Chapter 2, I will revisit the role of rules vs. direct recognition in the construction of statuses.

*

Finally, the third challenge concerning the relevant sense of recognition is the following. In the extant literature, several counterexamples of social kinds were provided against Intentionalism, including race or gender which aim to show that it is not just that the relevant intentional states need not be language-dependent, but they also do not need to be conscious (Thomasson 2003a, 2003b; Khalidi 2015, 2021, 2022; Burman 2019, 2023). Purportedly, these are social kinds which often do not require *conscious* collective recognition, through either language-dependent or non-language dependent intentional states, for their existence.

Take the example of “race”. Race is often brought up as an example for a social kind whose existence purportedly does not depend on being consciously collectively recognized as existing (e.g., Thomasson 2003a, Khalidi 2015). The idea is that members of a society can be racist even if nobody has any conscious intentional states about races. Think of a society in which people do treat each other differently based on skin colour – for instance, excluding them from certain job opportunities – even if they do not consciously represent races through

their intentional states. What can, then, the defender of Intentionalism say about this kind of putative counterexample?

Searle argues that Intentionalism can in fact provide an ontological explanation for unconscious racism (existing without racist language and *conscious* intentional states about races), denying the claim that in the case of unconscious racism there is no collective recognition of the relevant social kind (2010, 116 ff.). Recall that not all intentional states are conscious and to make sense of specific behaviors, we attribute unconscious intentional states, such as, in this case, the unconscious belief that “people of a different skin colour are not equal”, even when they contradict the individual’s conscious beliefs. Thus, the way to hold onto Intentionalism in the face of challenges such as merely unconsciously recognized race or gender statuses is – rather simply – to hold that members of a society do not have to have conscious intentional states about the social kinds (neither the membership conditions, nor the relevant associated statuses) which depend on their intentional states for their existence. In other words, the relevant sense of collective recognition does not need to be conscious.

Thus, even if the members of the society in question do not have conscious collective recognition of “races” (neither of the membership conditions or the relevant statuses), but their behaviour is such which we could call racist – since, for instance, it involves treating individuals with different skin colour disproportionately to their abilities – Intentionalism can still explain their racism. It is, then, possible to argue that even though people do not consciously recognize themselves as “racist”, it is explanatorily useful to postulate that they, even if unconsciously, recognize and thus represent “races” when they treat people of different skin color as being endowed with a different social status (cf. Searle 2010, 118–119). Postulating the existence of unconscious forms of intentionality – and, specifically in our case, unconscious recognition – helps us explain the behaviour of unconsciously “treating” something or someone *as* belonging to a given social kind.

Explaining the *exact* nature of unconscious intentionality goes beyond the scope of this chapter. For instance, it could be an interesting task to settle issues such as that of whether unconscious intentionality differs in nature from conscious intentionality, as argued by Tim Crane’s concept of a “worldview” (2017), or whether, following Searle’s argument (1995, 127 ff.), all intentionality is at least potentially conscious and facilitated by a set of “Background” capacities, or if, as Bourdieu’s notion of “habitus” or “knowledge without concepts” (1984, 466–484) holds, unconscious intentionality is a form of “embodied taxonomy” that lies “beyond the reach of introspective scrutiny or control by the will”. A definitive settling these issues is a task for another project; here, I am proceeding by merely assuming that there are *possible* explanations of people being treated differently which refer to unconscious intentionality. So, these putative counterexamples to Intentionalism are helpful in pointing

out that the relevant intentionality, i.e., the relevant sense of recognition in Intentionalism does not always have to be conscious.

In sum, in this subsection, we have shown that the relevant sense of recognition does not have to be correct about membership conditions, language-dependent or conscious. I will call this a *weakened* sense of recognition. In the following, I will use “recognition” in its weakened sense, unless I indicate otherwise.

Accordingly, our second-stage revision definition of Intentionality at this point looks as follows:

Intentionalism second-stage revision: (*human*) social kinds are constitutively (and synchronously) mind-dependent in the sense that their existence depends on being collectively recognized (*in a weakened sense*) as existing.

1.2.3 *Revising dependence: Direct and indirect dependence*

In this subsection, I consider challenges to the relevant sense of dependence in Intentionalism. I will consider two versions of such challenge. First, I consider the argument that rather than being constitutive of the existence of social kinds, there is merely a causal dependence between collective recognition and social kinds (Guala 2014, 63). Second, I will explore the argument that not all social kinds depend on collective recognition directly, some are indirectly dependent on collective recognition.

Let's start with the challenge that, at least in some cases, the alleged constitutive dependence between collective recognition and the existence of social kinds, is in fact a causal one which facilitates the coordination of beliefs among people. It holds that the role of the intentional states of people collectively recognizing the existence of a given (codified) social kind, such as money, is just causal as it is merely about facilitation and coordination of various practices. Guala points out that the collective recognition of money as a medium of exchange includes the belief about its being used as a medium of exchange in the future and so does not “hang up in the air” (2014, 64). The incentives for these beliefs, he argues, are causally created, for instance, through taxation where only a given form of currency is accepted. He then continues to argue that one needs to turn to the social sciences to better understand the specific causal structure behind the incentives people have for, say, holding a specific currency, and how various social kinds are causally related to one another.

However, all this seems to me to be compatible with Intentionalism. Intentionalism holds that such collective recognition is needed for (in this case, codified) social kinds such as taxes, states and money to exist in the sense that collective recognition is a constitutive part of the existence of taxes, states, and money. This is compatible with the view that, for instance, the

existence of taxes sets in motion manifold causal consequences, such as shaping the incentives of people. Nonetheless, certain actions are only possible once the given social kind, such as money or taxes exists, and social kinds can exist only if they are collectively recognized as existing. The claim, then, that the existence of social kinds depends on collective recognition is compatible with the claim that to uncover the causal consequences of social kinds, one needs to turn to social science. As Searle often mentions, the ontological analysis of the Intentionalist is supposed to merely capture the rather abstract “logical structure” of human societies; it is not meant to capture all the causal interactions that give rise to and happen between various social kinds.

*

The second kind of challenge concerning the relevant sense of dependence between collective recognition and social kinds is posed by counterexamples such as “inflation” or “recession” (Thomasson 2003a, 2003b; Khalidi 2015; Burman 2007, 2015, 2019; Franda 2021). As we have seen, we can draw a distinction between being wrong about the membership conditions or status of the given kind. Now, these cases show that it is possible not to have any kind of recognition of the existence of the given kind, neither of its membership conditions nor the associated status, without the kind ceasing to exist. It is, then, argued that these are, let’s call them, *external* social kinds whose existence does not require (the *internal* point of view of) collective recognition through which they would be represented as existing. In the case of recession, for instance, it is possible that there are longer periods of slowing economic activity (e.g., reduced production and spending, decreased investments in goods and services) without anyone recognizing it, and thereby representing it as existing. If these external counterexamples stand, that implies that the scope of Intentionalism is narrower than even its (second-stage) revised version – which delineated its scope to human social reality and a weakened sense of recognition – would suggest.

The first line of defense of the Intentionalist is to point out that external social kinds do not have an associated status. This, indeed, is Seale’s strategy (2010, 21, 23–24, 116–117). While he admits that recession, as an external kind, does not require collective recognition for its existence, he argues that it is simply a “systematic consequence” of social kinds which does not have an associated status – the ones we have discussed so far and referred to as elements of social reality (2010, 21, 23–24, 116–117). Unless, he adds, a law is passed about the relevant status specifying what the central bank is obliged to do in the case of recession, thereby also specifying the membership conditions of recession (deciding upon issues such as “how should we should define ‘periods of slowing economic activity’?; “should such a period be taken last for two consecutive quarters, or more?”; “should emphasis be put on declining consumer spending or more on rising unemployment rates?”).

Thus, the (revised) Intentionalist could point out that (given our second-stage revision) all social kinds come with an associated status (whether it is more or less codified). Then they could argue that given that external social kinds do not have an associated status, not involving them in the scope of Intentionalism does not pose a challenge to Intentionalism. However, challengers could still push Intentionalists as leaving out these external counterexamples from the scope of Intentionalism implies that Intentionalism leaves out kinds which are, in fact, those elements of human social reality which much of the social sciences are in the business of discovering and explaining. Social scientists, the argument goes, often go beyond the categories which a given community recognizes and discover external categories which are crucial to understanding social reality. As such, the argument goes, these kinds seem to be legitimate candidates for being social kinds.

Defenders of Intentionalism, most notably Åsa Burman (2007, 105–125), have indeed taken the challenge of external social kinds seriously. Burman has tried to deflect the challenge brought about by putative counterexamples to Intentionalism such as inflation or recession by claiming that, although their existence does not depend on collective recognition, they are, in fact, “macro-level” social kinds. Macro-level social kinds, then, are further dependent on “micro-level” social kinds which are indeed dependent on collective recognition for their existence (Burman 2007, 105–125; Searle 2010, 22–23, 116–119). In other words, it is conceded the existence of inflation or recession does not require collective recognition. They both, however, require, for instance, money to exist, which is, in turn, a social kind whose existence depends on being represented as existing through collective recognition. Take away money, and one takes away inflation as well.

Thus, Intentionalist could argue that the relevant sense of dependence in Intentionalism is disjunctive. The disjunction, then is between direct dependence on collective recognition in the case of micro-level social kinds or indirect dependence on collective recognition where the given macro-level social kinds further depends on micro-level social kinds which depend on collective recognition. Therefore, the thinking goes, given that macro-level social kinds further depend on micro-level social kinds whose existence depends on being recognized as existing, Intentionalism indeed lives up to its monist mandate of having general coverage over human social reality.

This move may come with difficulties. Simply incorporating macro-level social kinds disjunctively, could, in turn, jeopardize the unity criterion of monism according to which the mind-dependence of all social kinds should be explained in the same way. The defender of Intentionalism should, then, explicitly state that it is sufficient if their general coverage only applies to the fundamental micro-level. This is because, if macro-level social kinds are taken to further depend on micro-level social kinds whose existence depends on being recognized as existing, then it is sufficient to cover social kinds whose existence depends on being

recognized as existing as they are more “fundamental” in this ontological framework. If this is the case, Intentionalists need to update their view accordingly.

Now, to hold onto monism, the Intentionalist has to show that their view is sufficiently different from that of the pluralist. The pluralist would also hold a disjunctive view which would be a disjunction of different and *unrelated* forms of mind-dependence. Thus, the Intentionalist could point out that the fact that there is a dependence relation *between* the two kinds of social kinds (macro- and micro-level), means that their disjunctive formulation is sufficiently different from a pluralist position. With this move, the Intentionalist has a justification for why their theory only covers “micro-level” social kinds; I will, then, assume that this is a plausible defense which manages to set the Intentionalist perspective apart from that of the pluralist.

Thus, to defend Intentionalism from external social kinds whose existence is said not to depend on collective recognition, the Intentionalist can, first, point out that that putative cases of external social kinds only become social once they indeed have an associated status. However, if they are pushed to accept that external social kinds are indeed social, then Intentionalists may point out that external social kinds are, in fact, macro-level social kinds which further depend on micro-level social kinds which are, in turn, dependent on collective recognition for their existence.

Therefore, we have shown that the relevant sense of dependence can refer to being directly dependent on collective recognition or indirectly dependent on collective recognition. Finally, our (now fully revised) definition of Intentionality is as follows:

Intentionalism fully revised: (*human*) social kinds are mind-dependent in the sense that their existence depends (either directly or indirectly) on being collectively recognized (*in a weakened sense*) as existing.

In the following, I will use the scope of social kinds, recognition, as well as dependence in their respective revised sense reached in the subsections of this section, unless I indicate otherwise.

Concluding remarks: Intentionalism revised

In this Chapter, I began by asking why philosophers seem to be puzzled about social entities and phenomena such as money, marriage or promises and their associated practices. I differentiated between two related but distinct problems: the narrower issue of The Placement Problem which to find the place for (some) human social phenomena in a naturalistic ontological framework, and the broader issue of The General Problem which aims

to uncover the general conditions of possibility for the existence of *any* social phenomena. I then showed that answers to The Placement Problem begin by putting forth some version of the mind-dependency thesis regarding the elements of human social reality. Throughout the chapter, I focused on whether Intentionalism, the mind-dependency thesis of social reality which holds that social phenomena exists only if it is collectively recognized as existing, can be considered a monist theory of the social having general coverage.

Here is, then, a brief overview of our journey so far. In the first section, I explained that while authors use various metaphysical entities in their analyses like social “facts,” “objects,” or “properties,” I am using the terminology of social “kinds”, their “properties” (which I later unpacked into “membership conditions” and “associated statuses”) and “instances” as these terms evoke debates around classification which will be relevant in later chapters. I expanded on the logical landscape of possible ways to unpack the mind-dependency thesis; regarding scope, I introduced monist and pluralist positions, while the strength of the relevant sense of mind-dependency was explained through the distinction between Intentionalism and Anti-Intentionalism. In the last subsection of the first section, I introduced Intentionalism as a monist theory of the mind-dependence of social reality in detail, explaining the role of collective intentionality, social statuses, and language in the construction of social kinds.

In the second section, I considered challenges brought against Intentionalism targeting its scope, the relevant notion of recognition and the relevant notion of dependence as well as potential responses of the Intentionalist. Where did it all lead us? I concluded that the Intentionalist needs to adjust their position on three key matters to still retain general coverage over at least some parts of social reality. First, regarding scope, they should restrict the scope of Intentionalism – and, to hold onto monism, the scope of monism as well – to social kinds in *human* social reality. Second, regarding the relevant notion of recognition, they should broaden their view to incorporate what I called a *weakened sense of recognition* which includes the possibility of being mistaken about membership conditions as well as non-language dependent, non-conscious recognition. Third, admitting that the relevant sense of dependence can refer to being *directly dependent* on collective recognition or *indirectly dependent* on collective recognition. These revisions yield a revised version of Intentionalism as a monist theory of the social.

Now we are ready to explore the challenges that the elements of the algorithmic society present to the ontological framework of Intentionalism – even to its revised version.

CHAPTER 2: BEYOND RECOGNITION: MODULATORY ALGORITHMIC KINDS AS SOCIAL KINDS

*It is enough to create new names and estimations and probabilities
in order to create in the long run new 'things'.*

Friedrich Nietzsche: *The Gay Science* (1974, aphorism #58, 122)

Introduction: Social ontology and blind spots

Projects in social ontology stem from the realization that analytic metaphysics has often unduly overlooked *social* kinds when enumerating the elements of reality. In the previous chapter, I identified two specific projects: The Placement Problem and the General Problem. While the former aims to accommodate uniquely human social phenomena such as money or marriage in a naturalistic ontological framework, the latter is interested in the general conditions of possibility for the existence of any social phenomena, including those beyond human societies, such as ant queens or hyena clans.

I, then, introduced Intentionalism, the view which holds that all social kinds are mind-dependent in the sense that their existence depends on being collectively recognized as existing. Having argued for several caveats to make the view more plausible, I reached the following conclusion: while even a revised version of Intentionalism cannot address the General Problem as it excludes the elements of the social reality of animals without the representational capacities of intentionality, it does seem to provide an answer to the Placement Problem *as well as* have general coverage over human social reality. In this chapter, however, I will show that Intentionalism faces a crucial challenge even when its scope is restricted to the elements of human social reality. This is because analytic social ontology seems to have its own blind spots: it neglects elements of social reality constructed through what I will call data-driven algorithmic systems.

Today, both government institutions and private companies have the capacities to track people's behaviour at an unprecedented scale and granularity, aided by increasingly sophisticated methods of computational algorithmic analysis. This results in dynamically changing categories used for diverse predictive purposes in key areas of our social lives such as criminal justice, law enforcement, border control, social services, insurance and credit scoring, targeted advertising, personalized recommendations, or tailored search results.

These technological developments raise important empirical and conceptual questions; indeed, in the social sciences as well as in Continentally-oriented social philosophy, there has already been substantial work on the social implications of new technologies in large-scale

data collection and its computational analysis (e.g., Deleuze, 1992; Cheney-Lippold 2011, 2017; Gillespie 2014; Burell 2016; Couldry & Mejias 2019; Brusseau 2020; Burell & Fourcade 2021). Take the intriguing empirical studies on how algorithmic decision making employed in hiring, educational or judicial systems affect social reproduction, inequality, and mobility etc. (Eubanks 2018; Noble 2018; Burell-Fourcade 2021). Or consider research projects on rethinking the concept of “social knowledge” in light of the datafied nature of contemporary society (Couldry & Mejias, 2019). Several authors, in turn, investigate concerns regarding the opacity surrounding algorithmic operations or offer critiques of the forms of “governance” made possible by these technologies (Deleuze 1992; Gillespie 2014; Pasquale 2015).

While there is growing research on what I have called algorithmic societies, a systematic project in analytic social ontology investigating the mode of existence of social kinds constructed through data-driven algorithmic systems is yet to be carried out. The arguments presented in this chapter are intended as a first step towards such a project.

I will argue for two main claims. First, socially significant algorithmically generated real-time classifications – which I will call Modulatory Algorithmic Kinds (MAKs) – are “social kinds” in the relevant sense. Second, accommodating (at least some) MAKs into our social ontology challenges the Intentionalist view regarding the specific mind-dependence of social kinds, that is, it challenges the view that all (human) social kinds are mind-dependent in the sense that their existence (directly or indirectly) depends on being collectively recognized (in a weak sense) as existing.

Here is my argument in brief. MAKs are kinds in the sense that, given their multiple shared properties, they can be reliably used for epistemic purposes such as prediction. The trickier part is the claim that MAKs are indeed social. While for the Intentionalist, the necessary “social glue” that binds social reality together is collective recognition, the existence of MAKs, as I will show, does not depend on collective recognition. Yet, I will argue that (i) being mind-dependent in some sense (not necessarily in the specific sense outlined by even a revised version of Intentionalism), (ii) being dependent on a form of representation (not through collective recognition but through digital and algorithmic code), as well as being (iii) consequential on a group-level, together, are sufficient conditions for being social.

Remember that, as we agreed in the *Introduction*, Intentionalism initially seems to delineate social reality from the rest of reality by holding that collective intentionality is the mark of the social. It is assumed that collective intentionality sets social phenomena apart from (non-social) natural phenomena, and those which exist only at the individual level. In the case of MAKs, while (i) and (ii) set MAKs apart from phenomena which are part of the non-social natural reality, (iii) sets them apart from phenomena which only concerns individuals. Thus, incorporating MAKs into our ontology of the elements of social reality shows that an ever-

increasing part of our social reality is mind-dependent in a weaker sense than even the revised Intentionalism would hold.

The structure of this chapter is as follows.

First, I will revisit the question regarding the role of rules in Intentionalism using the example of “marriage”. Then, I present the category “Married” used by Google to deploy targeted advertisements to its users, raising the question of whether it qualifies as a social kind, and if so, whether *any* version of Intentionalism can adequately account for it.

In the second section, I take a slight detour to explore what kind of category Google’s “Married” label is. I argue that Google’s “Married” is an example of the result of a process of kind-making where the contrast with earlier examples of macro-level social kinds or social kinds constructed through the outside perspective of scientists (even if it is only a matter of degree not of kind) is so stark that social ontology needs to address the issue. Specifically, Google’s “Married” is a Modulatory Algorithmic Kind (MAK), born out of the possibility of an unprecedented scale and granularity of real-time data-collection and its computationalized analysis.

In the third section, I circle back to the question of mind-dependency of social kinds, showing that instead of the representational capacities of collective recognition, MAKs depend on the representational capacities of digital data and algorithmic code. Therefore, MAKs cannot be incorporated into any version of Intentionalism. I even go one step further and argue that, in the case of MAKs, the relevant sense of mind-dependence is neither synchronic, nor constitutive. Nevertheless, for the reasons listed above, I will argue that MAKs are indeed *social kinds*.

One last comment before we begin. As I have hinted at in the *Introduction*, and will explain in more detail below, it seems to me that the fact that an analysis of the increasing datafication and algorithmization of social reality has fallen outside the purview of analytic social ontology has to do with some of the deeply held assumptions within the field.

Recall Åsa Burman’s useful distinction between ideal and nonideal social ontology from the *Introduction*. There is the tendency in classical texts of “ideal social ontology” such as those of Searle (1995, 2010) or Gilbert (1992) to take small, egalitarian and cooperative communities and the transparent statuses they collectively recognize as existing as paradigmatic cases of “the social”. Clearly, such assumptions will not inevitably lead the theorist to look for social kinds in, say, vast and complex digital databases of Big Tech companies.

Proponents of nonideal social ontology rightly point out the importance of considering larger communities, inequality, and conflict, especially in modern-day societies. Some of these insights of nonideal social ontology, in fact, inspired the challenges to Intentionalism which led us to revise it in the previous chapter. However, that some form of collective intentionality would be the “real glue of the social world” (Gilbert, 1992, 13) is a crucial assumption which is shared even by proponents of nonideal social ontology.

The messy real-life case of social kinds constructed through data-driven algorithmic systems, in turn, questions even this basic premise of some form of collective intentionality as a necessary building block of social reality. I believe, nevertheless, that the conceptual toolkit of analytic social ontology is useful in elucidating (at least some of) the ontological implications of the shift towards an algorithmic society. The concepts of traditional social ontology are relevant not despite, but rather precisely because of the contrast between traditional examples of social kinds and those constructed through data-driven algorithmic systems; between, as we will see, “marriage” and Google’s “Married”. My hope is that drawing such a contrast mutually sheds light on both kinds of social kinds.

2.1 Marriages and their mind-dependency

Thus far, I have introduced Intentionalism, which, relying on Searle, I defined as the view that all social kinds are mind-dependent in the sense that their existence depends on being collectively recognized as existing. My aim was to revise Searle’s ideas in light of the challenges it has received. Briefly, I highlighted three key caveats: restricting the *scope*, weakening the relevant notion of *recognition* and that of *dependence*. I restricted the scope of Intentionalism to uniquely *human* social phenomena; argued that the relevant sense of recognition is not necessarily correct about membership conditions and is not necessarily language-dependent or conscious; and showed that the dependence is in some cases indirect as there are macro-level social kinds which are not themselves dependent on collective recognition but are dependent on social kinds which are.

As a result of these revisions, Intentionalism, and its relevant concepts, will henceforth be taken in the following sense (unless specifically noted otherwise): the view according to which (*human*) social kinds are mind-dependent in the sense that their existence depends (either directly or indirectly) on being collectively recognized (*in a weakened sense*) as existing. I concluded that, with these caveats, Intentionalism seems to provide an answer to the Placement Problem – by unpacking the concept of mind-dependence in this manner, we can incorporate seemingly “magical” human social kinds such as money and marriage into our naturalistic picture of reality – as well as have general coverage over *human* social reality.

Below, using the example of marriage, I revisit the debate between two versions of Intentionalism – Searle’s and Ásta’s – focusing on whether social statuses are imposed through the intermediary recognition of a rule which specifies the associated status or through the direct recognition of the status itself. I am interested in this debate as I will employ the resulting concepts in my subsequent analysis of social kinds constructed through data-driven algorithmic systems. For the sake of clarity, I will call Searle’s Intentionalism, “Rule”, and Ásta’s Intentionalism, “Ad Hoc”.

2.1.1 *Two Intentionalist takes on marriage*

Take again the often-analysed case of the social kind “marriage”. Remember that social ontologists are not interested in historical analyses of marriages or various causal relationships pertaining to marriage. For example, they are not interested in how marriages evolved in various cultures or how marriage causally affects the life of married people – e.g., whether it makes one happier or bored or financially stable etc. Instead, they are interested in how it is possible that marriage exists. On *any* Intentionalist view, marriage is not simply the pre-existing behaviour of, for instance, two people living together and maintaining sexual exclusivity with one another. Marriage, instead, is a codified social kind which delineates a unique social status, whose existence is dependent on being represented through collective recognition as existing.¹² Unpacking the specifics of the relevant notion of collective recognition is where Rule and Ad Hoc versions of Intentionalism diverge.

According to Rule, the relevant notion of collective recognition is the recognition of *rules* about marriage (or the given codified social kind). As we have seen, these rules have a general formula which specifies that, in a given context, those meeting certain membership conditions (individuals meeting, e.g., minimum age requirement or gender requirements) count as having the social status (being married, thereby gaining access to the accompanying rights and responsibilities). Given that marriage is a codified social kind, the recognition of these rules seems to require language-dependent intentional states. In sum, according to Rule, we collectively recognize what Searle called “constitutive rules” which impose the social status of being married onto the relevant individuals.

How exactly does Ásta’s Ad Hoc view on marriage differ? Consider the following scenario. A couple is pronounced as married by the relevant persons even though they do not, in fact, meet the membership conditions (Ásta 2018, 29 ff.). Everyone agrees that there is a

¹² Here’s Searle on one’s (or merely his?) potential motivation for establishing such statuses: “Marriage begins with people simply living with each other, and in the case of monogamous marriage, having a sexual monopoly on each other. Why are we not satisfied with these arrangements? (...) why is it not enough that we just live together? (...) For one thing, we can have a much more stable system of expectations if we add this deontic apparatus; for another, we don’t have to rely on brute physical force to sustain the arrangements; and for a third, we can maintain the arrangements even in the absence of the original physical setup. For example, people can remain married even though they have not lived with each other for years.” (Searle 1995, 81)

discrepancy between the membership conditions described in the relevant rule and the status of the given couple *as recognized by others*. However, while the proponent of Rule would say that this discrepancy implies that they are not *really* married, Ad Hoc holds that they indeed *are*. According to Ad Hoc, what determines whether a person or an object is an instance of a given social kind, even in the case of (seemingly) codified social kinds, is *not* whether they fulfil a collectively recognized set of membership conditions of a given kind, but whether others in fact recognize them as being an instance of the given kind (2018, 28, cf. 24–31). In our case, then, the couple is indeed recognized to be married which makes them married even if they do not meet the membership conditions. Ásta argues that,

The couple may in fact not meet the eligibility requirements, but if the marriage conferrer judges them to, they become married anyway and get the certificate of their new institutional status to prove it. To be married is thus not to meet the various requirements for being married, but to be judged by a person in authority to do so. (Ásta 2018, 29, my emphasis)

While for Rule, the relevant metaphysical work is done by the recognized *rules*, Ad Hoc emphasizes the direct recognition of “others”. These “others”, then, are the ones with “authority”; in other words, as Ásta explains, it is not simply *collective* recognition that counts but the “authority” of the relevant people doing the recognizing (2018, 20). Nevertheless, the source of such “authority”, again, seems to bottom out in collective recognition of “others” (ibid.). Thus, although Ad Hoc emphasizes that social statuses are imposed directly by those *already* wielding some form of social power (as opposed to being imposed by collectively recognized rules à la Rule), the relevant social kinds will ultimately depend, again, on some form of collective recognition.¹³

I take it that the two versions of Intentionalism belong to different trends in social ontology. Rule appears to belong to ideal social ontology, describing the ontology of an egalitarian and cooperative community having transparent statuses; an institutionalized and bureaucratic civil society. Ad Hoc, in turn, belongs to the trend of nonideal social ontology, by making social status imposition depend on the ad-hoc recognition of the powerful, brings into question egalitarian and cooperative assumptions; the social reality of Ad Hoc centers around what Weber termed “the master of older social structures, who was moved by personal sympathy and favour, by grace and gratitude” (Weber 1978, 975).

As I will show, while some insights of both versions of Intentionalism capture certain features of social kinds constructed through data-driven algorithmic systems, neither manages to incorporate them. But one step at a time.

¹³ Ásta argues that Ad Hoc is preferable to Rule as it “fares better on simplicity”, given that it does not postulate an intermediary step of recognizing constitutive rules between how we get from collective recognition to the relevant social status, elucidating not “what meets the requirements to matter” but “what does in fact matter” (2018, 20, 30).

2.1.2 Google's "marriage"

A few weeks ago, when preparing for a course I taught on the role of the media and advertising industry in shaping public opinion, I checked the targeted advertisement categories that Google, "the world's most important information gatekeeper" (Gillespie 2014, 181), put me into.

Google's database tells me that my "Marital Status" is "Married". Which is interesting. I have never told Google that I was married. This, in turn, has to do with the fact that I am not, actually, married. But we have just seen that this is not such a simple matter to decide. As far as I know, there is no collectively recognized rule regarding the membership conditions for, and the constraints and enablements of the relevant status associated with, becoming "Married" according to Google. Not to mention that nobody, let alone those with authority, recognize me as being married. So, *prima facie*, I do not seem to be married, either according to Searle's Rule or on Ásta's Ad Hoc account of marriage.

After a few clicks, I learn why I am "Married" according to Google. My "signed in activity on Google services (such as Search or YouTube) is similar to people who've told Google they're in this category"; in other words, an algorithm classified me as being married based on the fact that my online behaviour is similar to those who "told Google" that they are married. Did the algorithm make a mistake, and if so, in what sense? Incidentally, I am also categorized as a "Homeowner" (I wish), a "Bread maker" (not much of a baker), and "Dogs" (admittedly, I do like dogs). Most importantly, the following question arises: what implications do cases of various kinds of data-driven algorithmically constructed categories hold for social ontology?

Some may simply shrug their shoulders and be hesitant to attribute ontological significance to such algorithmically constructed categories. Such skepticism can be due to two reasons. First, one may believe that Google's "Married" is not, in fact, a social kind. They could argue that Google's "Married" may not qualify as social and/or qualify as a kind. It may not qualify as *social* given that we have already set out necessary conditions for a kind being a social kind in terms of a specific sense of mind-dependence and algorithmically constructed categories do not meet these. Or, algorithmically constructed categories may not be what philosophers call a *kind*, only an *incidental* category employed by an advertising platform to enhance their productivity. The second reason one may be a skeptic about attributing ontological significance to Google's "Married" and related categories is that *even if* these categories are indeed social kinds, they may be incorporated into a revised version of Intentionalism as we have so far unpacked it.

In the following, I will argue that the positions of both kinds of skepticism are hasty conclusions.

In the next section, I will show that Google’s “Married” shows that there is now a way of kind-making which is different from traditional forms of quantification- or outside perspective-based kind making (even if, from a historical point of view, this is only a matter of degree not of kind) to the extent that the problem becomes relevant for social ontology. This contrast, then, stems from new technologies of large-scale, granular and real-time data-collection and analysis, particularly those relying on machine learning which made data-driven algorithmic systems possible.

In the final section, I will return to the issue of specifying the mind-dependency of social kinds and provide answers to the skeptics. I will defend the view that kinds constructed through data-driven algorithmic systems are indeed social kinds *and* that the fact these kinds are social kinds depends even the revised Intentionalist picture of social kinds, as these kinds are mind-dependent in a weaker sense than previously considered social kinds.

2.2 From ancient bureaucracies to modulatory algorithms

To begin to uncover the ontological significance of categories such as Google’s “Married”, an admittedly abstract and simplified historical contrast between traditional demographics and data-driven algorithmically constructed categories may prove helpful.

Quantitative social modelling and various forms of classifications through the external perspective of the bureaucrat, the social scientist or the marketing agent, are far from being new ways of kind-making (e.g., Hacking 1995, 1999, 2007). For instance, when discussing the history of bureaucracy, Max Weber shows that, while bureaucracy “fully developed” in the modern state, ancient Egypt, the Roman Principate and Emperor Qin Shi Huang’s China are also examples of states having “clearly developed and quantitatively large bureaucracies” (1978, 956, 964). In the 17th-century, in turn, Leibniz famously suggested that, in order to have a better grasp of the capacities at its disposal, the Prussian state should have an office of statistics, putting forth a “56-category evaluation” including various social categories such as “sex, social status, the number of able-bodied men who might bear weapons, the number of marriageable women, the population density and age distribution, child mortality, life expectancy, distribution of diseases and causes of death” (Hacking 1990, 18–19). And then, of course, there are famous 19th-century examples of state officials calculating “normalcy” and accumulating relatively large quantities of data about “deviant” behaviour, crimes and their locations, suicide as well as poverty and unemployment rates. The 20th-century, in turn, saw the rise of the data-collecting capacities of private companies which, by the 1970s, matched that of states (Couldry-Mejias 2019, 121 ff.).

So, why lump together the numerous developments in data gathering about society from ancient China to post-World War II marketing agencies? Because from the point of view of

social ontology, despite the varied methods used to classify people, it seems to me that Intentionalism has an answer to these cases. Remember the challenge of putative *external* social kinds from the previous chapter. Intentionalists can, on the one hand, incorporate these examples, as social kinds into their framework, either as social kinds with an associated status or as macro-level social kinds without an associated status. On the other hand, Intentionalists can also argue that the categories at hand are not social kinds given that there is no social status associated with the relevant group *and* these categories also do not qualify as macro-level social kinds which would further depend on micro-level social kinds with an associated status. Accordingly, there are two avenues the Intentionalist may take.

First, Intentionalists can show that the given categories can be incorporated into the Intentionalist framework as they qualify as social kinds. This is because, the Intentionalist could point out, that some of the above-mentioned categories are examples of what we have previously called macro-level social kinds. Macro-level social kinds, as shown in the previous chapter, are social kinds which are not dependent on collective recognition for their existence, but which further depend on social kinds which are. For instance, while “the number of marriageable women” may not depend on collective recognition, it indeed further depends on marriage which does.

Second, Intentionalists can show that as long as there is no social status associated with the relevant group *and* as long as it is not a macro-level social kind which would further depend on micro-level social kinds with an associated status, the given category is not a social kind. Take “life expectancy”. While “social” causes may contribute to differences in life expectancy, the Intentionalist could point out that social ontology aims to map ontological, not causal relations. However, if “life-expectancy” would have an associated status – e.g., in the dystopic, dehumanizing and disturbing visions of eugenics, as explored in fiction in the movie *Gattaca* – which would endow individuals on a collective level with constraints and enablements based on their, say, genetic makeup, then it *would* qualify as a social kind. The same goes for, for example, “able-bodied men who might bear weapons”, which, as such, also seems like a non-social kind. However, it is likely that this category would have an associated status with the constraints and enablements involving, e.g., being subject to conscription duty.

In sum, Intentionalism does not seem to be challenged by the external kinds yielded through traditional quantificational demographic methods. In the remainder of this section, I will explain how data-driven algorithmic ways of quantifying and classifying social reality introduce ontologically intriguing new ways of, borrowing Ian Hacking’s phrase, *kind-making*. There are, undoubtedly, important historical continuities with earlier methods of quantitative social modelling and classification through what we have called an “external” perspective, such as that of the bureaucrat, the scientist or the marketing agent. I am, thus, happy to admit that the difference between traditional methods of quantitative social modelling and

construction of categories through an external perspective and data-driven algorithmic ways of classifying social reality may, in fact, be one of degrees and not in kind. Nonetheless, the conceptual toolkit of social ontology reveals how even a difference of degree can have ontological significance which cannot be overlooked. This is because, as I will show, the mind-dependency of kinds constructed through data-driven algorithmic systems pose stronger challenges to Intentionalism.

2.2.1 *Algorithms as new kind-makers*

What information is showing up on one's news feed on various social media platforms, what items are recommended to one on online marketplaces, who shows up as a potential match on a dating platform, what movies or musicians are recommended on streaming platforms, whether one passes the first rounds of a hiring or university admissions process, where one's savings are invested, whether one is given a loan and with what conditions are all decisions with far-reaching social consequences which today are potentially made by what I have so far referred to as data-driven algorithmic systems.

Data-driven algorithmic systems, then, are systems which classify subjects and objects based on large scale, often granular-level and real-time data through algorithmic analysis for various socially consequential purposes. Algorithmic analysis, simply – which, again, suffices for our purposes – refers to a set sequence of computational process designed to achieve a desired outcome: in our case, the classification of subjects and objects. Computer scientists often mention, as a toy example, making an omelette as a basic algorithm – as one follows a set sequence of actions to achieve the desired outcome of a breakfast (cf. Hosanagar 2019). However, the way I use the term “algorithm”, I imply that these set sequences are carried out by computers. There are multiple ways to draw distinctions between various types of data-driven algorithmic systems; below I will draw two broad distinctions to show the different senses in which data-driven algorithmic systems can be data-driven.

Many social scientists argue that the technological developments of the past decades which enabled the digital mediation of various social domains bringing about data-driven algorithmic systems entail an important difference (again, I remain neutral on whether it is novelty in degree or in kind) when it comes to socially consequential quantification and classification of people and objects (Tufekci 2014; Cheney-Lippold 2017; Brubaker 2020; Burrell-Fourcade 2021). In the following, I will identify three related features of new forms of kind-making through data-driven algorithmic systems. These are: (1) data collection of unprecedented scale and granularity (2) vastly increased computational capacities for algorithmic analysis (3) both data collection and processing happening in a dynamic, real-time

manner.¹⁴ This, it is worth noting, is the zoomed-out perspective of the philosopher; I am necessarily omitting otherwise important details which may be of interest to, for instance, the computer scientist or the historian of various methods of social quantification and classification.

First, on data collection at an unprecedented scale and granularity. Consider that, reflecting on the nature of the human practice of classification, John Locke had once argued that the number of our categories employed must be limited as having one for every “particular Thing” would be “beyond the Power of humane Capacity” and “could not find a place in the most capacious Understanding” (Locke 1975 [1689], 3.6.2). Indeed, compared to a 17th-century, Leibnizian “56-category evaluation”, the *scale* of information, the sheer volume or magnitude of data being collected about society as well as the number of categories employed has exponentially increased. Governments and companies can operate databases with a much more expansive range of categories and easily track billions of individuals’ data points, constructing, if needed, just as many categories – going far beyond “the Power of humane Capacity”.

Even more importantly, there is an increase of *granularity*. As many scholars noted, our “data doubles” consist not merely of the data we provided explicitly and voluntarily on various platforms. For instance, beyond gathering basic demographic and/or self-declared information like age and gender, it is possible to obtain much more *granular* behavioural data, at times vividly referred as “digital footprints”, “digital breadcrumbs” (Zuboff 2019, 90) or “behavioural traces” (Cheney-Lippold 2017, 184), such as browsing and search history, scrolling speed, frequency and duration of device usage, patterns in online purchases, social media interactions, language use or even physical movements tracked via GPS data, let alone the various kinds of metadata such as timestamps, locations, and device identifiers (cf. Tufekci 2014; Pasquale 2015, 4; Burrell 2016, 1; Cheney-Lippold, 2017, 4; Zuboff 2019; Brubaker 2020; Burrell-Fourcade, 2021, 214 ff.).

This granular data, then, can be used to infer membership in various demographics or construct new categories. An-often mentioned area where one may be categorized algorithmically based on granular data is finance and mortgage services where one may be a “bad credit risk”, or “likely to default” and denied a loan they applied for (Pasquale 2015, 4–5, 32; Burrell 2016, 1, 3). Or, based on one’s browsing history, one might be categorized as a “captive traveller” or a “frequent flyer” by an airline company’s website and so be shown

¹⁴ Here are some formulations of “what is new” about the latest developments in quantifying the social world: “What is new today is not the quantification of the social or its measurement but the depth at which datafied social knowledge conducts its operations” (Coudry-Mejias 2019, 118). Or: “What is new about the categories that constitute us online is that they are unknown, often proprietary, and ultimately— as we’ll later see— modulatory” (Cheney-Lippold, 9, 30). “What is new is the use of behavioral trace data, including the things we do on devices and also things captured about us by devices in our environment, and the growing reliance on computing power to process these data” (Burrell-Fourcade 2021, 221).

higher prices for plane tickets (Rouvroy–Berns, 2013, xiii). As famously revealed by the Snowden-leaks, one’s citizenship could also be inferred from digital “breadcrumbs” to determine whether one is a “US-citizen” or not, which, in turn, determines what kind of privacy rights one has on US networks (Burell-Fourcade 2021, 228). In a 2013 study, by now considered as a “classic”, researchers claimed to be able to accurately predict, based on Facebook likes, among other traits, sexual orientation (88%), ethnicity (95%), religion (82%), or use of addictive substances (65–73%) inferred from relatively few data-points (Kosinski et al. 2013). Furthermore, not only people can be targets of the quantifying and classificatory work of data-driven algorithmic systems: search engines or social media services calculate the “relevance” of information to structure one’s search results or news feed (Gillespie 2014, 178).

Second, vastly increased computational capacities, in turn, enable the algorithmic processing of enormous amounts of data, identifying patterns and correlations in granular-level data which would be impossible for humans to discern. Seemingly uninteresting digital footprints enable algorithmic analysis to draw out various correlations which, in turn, may be used for socially significant future predictions (Burell 2016; Ananny–Crawford 2016; Couldry–Mejias 2019; Burell-Fourcade 2021).

Finally, not only are both public institutions and private companies able to gather enormous amounts of data with unseen granularity and analyse them algorithmically, but it is also possible to do the collection and the analysis of the data in a dynamic, real-time manner. Internal presentations at multinational corporations no longer begin with the presenter apologizing that their data is almost two weeks old: they rather employ real-time embedded data visualization tools.

So far, I have used the term “data-driven algorithmic systems”. To show in what sense these systems are indeed “data-driven”, let us draw a broad, yet for our purposes sufficient, distinction between two types of data-driven algorithmic systems (cf. Hosanagar 2019, 19–37).

The first type, often referred to as “expert systems” are ones where humans (pre-)specify the relevant rules which do not change based on incoming data. While these systems may classify subjects and objects dynamically into different categories depending on their data input – and so the membership of these subjects and objects in various categories may be in constant change – the rules specifying membership conditions and the relevant statuses do not change dynamically based on incoming data. For instance, a tax system might employ expert systems which may contain the following conditional: if one is married, then one is eligible for certain tax deductions. Here, this rule – which specifies both the membership conditions and the associated status – is pre-determined by the engineers of the system.

By contrast, the second type of data-driven algorithmic systems, which are commonly referred to as “machine learning systems”, dynamically change their relevant *rules* based on what they “learned” from incoming data. Now, a further distinction is often drawn between two sub-types of machine learning algorithmic systems.

The first one learns with the use of some pre-labeled training data. Take Google’s “Married” where the relevant rule of being “Married” is constantly updated by what “Married” people do. Google’s algorithms search for correlations between the behaviour of the labelled data (“people who’ve told Google they’re in this category”), and the behaviour of those yet unlabelled to predict whether the latter are “Married” or not.

The second one, however, does not depend on a set of pre-labelled data and constructs categories simply by clustering together based on correlations found in the relevant data – not unlike how scientists discover what are often called “natural kinds” or, as we have seen, macro-level social kinds. As I will unpack in more detail below, social media platforms may use such clustering algorithms to group users based on their behaviour, often relying on granular-level data (e.g., interactions, scrolling patterns, hover patterns, GPS tracking etc.). This analysis dynamically constructs categories which endow individuals with a predictive status of, for instance, being likely receptive to a specific kind of content.

I will have more to say on the “kindhood” of categories constructed through data-driven algorithmic systems as well as the distinction between various types of algorithmic systems in the remainder of this chapter.

*

Now that we have seen that within data-driven algorithmic systems, there are expert systems as well as machine learning systems – some of which learn from pre-labeled data, while others can learn without it – let us further analyse the different ways in which data-drivenness can lead to change in the given system. Deleuze, in his 1992 vision of a world run by “numerical language of control”, writes of a society whose elements are described as being “a self-deforming cast that will continuously change from one moment to the other, or like a sieve whose mesh will transmute from point to point” (1992, 4–5). Now, as we have seen, his vision seems less and less fiction-like. Deleuze, while not providing an explicit definition, often refers to the phenomena of systems dynamically changing based on incoming data as “modulation” (*Ibid.*). I will here borrow his term and unpack it to refer to a special form of change which characterizes data-driven algorithmic systems.

As an example, take Cheney-Lippold’s remark that “as a user surfs from one site to another online, that user might change from being 92 percent ‘male’ to 88 percent ‘male’ within a few seconds” (Cheney-Lippold, 2017, 26, cf. 34, 170). In my case, while writing this chapter, I

oscillated twice, back and forth, between being “Married” and being “In a Relationship” according to Google. Now, this change may either be because my behaviour has changed, and I have been reclassified, or because the relevant rules driving the classification have changed. I will, accordingly, differentiate between two level of modulation: on lower levels, modulation refers to the membership being changed due to change in the data associated with a given subject or object. This is the case with what we have called expert (data-driven algorithmic) systems. On a higher level, modulation refers to the constant changing of the *rules* themselves which specify the relevant membership conditions and statuses. This kind of change, in turn, characterizes machine learning (data-driven algorithmic) systems.

As we have seen, Google’s advertisement categorization system is a data-driven algorithmic system which is characterized by modulation on a higher level: the relevant rules are constantly changing based on incoming data. The rules, then, may modulate in the following ways: either the membership conditions, the associated status, or both may change. For instance, in the case of a machine-learning-system-based spam filter, the rules defining what qualifies an email as “spam” might change, but the consequences of being classified as spam (e.g., moving to a spam folder) remain the same. Or the membership conditions in categories like “frequent flyer” may remain the same (one needs to fly X times) but the associated status may change based on data about the behaviour of flyers (do they want priority boarding or lounge access?).

The highest level of modulation happens when both the membership conditions *and* the associated status is changing in a data-driven manner. Take, again, Google’s “Married”. Here, not only the membership conditions but also the associated status is up to the *behaviour* of those classified as such – and so both are constantly changing. Perhaps membership conditions simply involve searching for wedding-related content or visiting websites about family life – and the status associated with this category involves displaying ads for family vacations or joint bank accounts. But Google may gather more granular data of those who are “Married” and so the rules – both the behaviour data used to infer membership in the category as well its associated status – may change: data may reveal that “Married” people search for cat-related contents and so the category may, in fact, resemble that of, say, “cat-lovers”.

Understanding various forms and levels of modulation, then, is the first step toward understanding the ontological difference between being married and being “Married” (according to Google). Thus, to capture categories constructed through data-driven algorithmic systems which rely on (1) data collection of unprecedented scale and granularity, (2) vastly increased computational capacities for algorithmic analysis, (3) and both data collection and its algorithmic processing happening in a dynamic, real-time manner, I will use the term Modulatory Algorithmic Kinds (MAKs).

Accordingly, I should now say a few words on why these categories qualify as what philosophers call *kinds*.

2.2.2 *In what sense are MAKs kinds?*

Remember how philosophers often distinguish between kinds and just any incidental category. Kinds carry ontological significance as they show us, in some sense, the structure of the world, while incidental categories can be any grouping of things we happen to come up with. How do we know whether a category is indeed a kind – an element of the structure of reality – or simply an incidental grouping of things?

As we have seen in the previous chapter, traditionally, there is a contrast between categories, that is, mere representations of groupings and, kinds which are those *groupings themselves* which exist independently of whether they are represented or not. Accordingly, some categories correspond to kinds and some do not; it was the latter which we have called incidental categories. For instance, while the chemical elements of the periodical table seem to be kinds as they inform us about the structure of the world, less revealing, perhaps, about the ultimate joints of reality, is the category of “items on my desk at this moment”. But why exactly? In what exact sense does the category of “items on my desk at this moment” not pre-exist before being represented as such?

As we have seen, drawing the contrast between kinds and incidental categories is not all that easy. The most notable difficulty, in our case at least, is that social kinds are a special subset of kinds whose existence is indeed dependent on being represented as existing (through, sticking with Intentionalism, collective recognition). Dependence on being represented as existing, thus, is not sufficient for calling a category incidental: it does not establish that the given grouping is not a kind. As I have mentioned in the previous chapter, while there are various proposals for what makes a kind non-incidental, I argue that one legitimate sufficient condition for kindhood is epistemic usefulness which may stem from various underlying ontological states of affair. Let’s go step by step.

One may hesitate to call MAKs *kinds* because of their changing and prolific nature. Skeptics would argue that the fact that the membership of members, as well as the membership conditions of, and the status associated with, MAKs, as well as the number of MAKs, is constantly changing raises doubts as to their kindhood. Such skepticism appears to be rooted in assumptions holding that kinds should have members whose membership is immutable and that these kinds possess stable, unchanging membership conditions (or even unchanging statuses in the case of social kinds) as well as the idea that the number of kinds remains fixed.

However, these assumptions have been widely shown to be empirically implausible (for overviews, cf. Boyd 1991; Dupré 1993; Khalidi 2013; Mattu-Sullivan 2021). Take biological

categories, such as lion or wolf, which are usually held to be a paradigmatic example of kinds. Given the fact of the evolution of species and the variation within species, we no longer tend to think that the membership of individual organisms within a species cannot change over time; we also do not think of species as having unchanging membership conditions; and, given and the fact that hybridization is possible (think of ligers or wolfdogs), we no longer think that there is a fixed a fixed number or species. Surely, *some* form of ontological stability of kinds may enhance their *epistemic* use; however, instead of focusing solely on defining strict and fixed ontological criteria for kinds, the growing consensus seems to be shifting towards also specifying the epistemic requirements that a kind must meet. The relevant question to ask, then is this: in what sense will the given category advance our knowledge about its (putative) members?

Now, MAKs, unlike “items on my desk right now”, share more similarities beyond just our interest in them: they are defined by multiple shared properties. This is what enables their usage for various *predictive* purposes in areas such as hiring, finances, marketing, crime prevention etc. So, the sole ontological commitment in this case is the criteria of having multiple shared properties, at least at a given point in time, which then enables far-reaching epistemic use. For my purposes, then, this epistemic usage for the purpose of wide ranges of prediction, stemming from having multiple shared properties, is one (even if not the only) legitimate condition which draws the relevant boundary between incidental categories and kinds.

Incidentally, this is in line with how John Stuart Mill (1974 [1843], 6.4) – to whom of the earliest use of the term “kind” in this philosophical context is usually attributed – defined kindhood. He emphasized the epistemic value of the projectability of kinds, pointing out that kinds are those categories which group things which have multiple properties in common and, therefore, they allow for inductive inferences and predictions regarding them.

By a kind, it will be remembered, we mean one of those classes which are distinguished from all others not by one or a few definite properties, but by an unknown multitude of them. The class horse is a kind, because the things which agree in possessing the characters by which we recognize a horse, agree in a great number of properties, as we know, and, it cannot be doubted, in a great many more than we know. (1974, 6.4, 703–704)

With focusing on multiple shared properties, Mill aimed to salvage the idea that some categories are less incidental than others while also distancing himself from *more demanding* ontological criteria of kindhood. I will, then, follow Mill in holding that the one legitimate criterion for kindhood lies in the epistemic use for reliable predictions which is enabled by the ontological fact of having multiple shared properties.

Furthermore, such predictions, as we have seen, do not merely hang “up in the air” but are fed back onto the relevant subjects and objects in the form of, e.g., job offers, credit scores, recommendations, and targeted policing. Thus, the relevant predictions of data-driven algorithmic systems go beyond being mere *epistemic* tools of discovery; through the kinds and the associated statuses they construct, data-driven algorithmic systems construct the elements of the social reality of the algorithmic society. As such, membership in MAKs are consequential not only on the level of individual subjects but on the level of groups.

However, even if one accepts that MAKs are *kinds*, the more controversial question seems to be whether MAKs are also *social*. According to Intentionalism, if a kind is a social kind, then it is mind-dependent in the sense that its existence is dependent on collective recognition. I will argue that this conditional is false by showing that MAKs are indeed social *and not* mind-dependent in the relevant sense.

2.3 Beyond recognition: The mind-dependency of MAKs

We are now one step closer to having a better understanding of how Google’s “Married” and other MAKs differ from earlier quantification-based classificatory methods and traditional demographics – even if the difference, from a historically more nuanced perspective, may boil down to a difference in degree. We have, then, shown that, given that they share multiple properties, MAKs are epistemically useful for predictive analytics, and, if so, they qualify as kinds. Furthermore, membership in MAKs come with significant consequences on the level of groups. In this section, I switch back to the vocabulary of social ontology and explore – in order to see whether MAKs are *social* kinds or non-social kinds – in what sense MAKs are mind-dependent.

2.3.1 *Representation without collective recognition*

In the first section, through the example of marriage, I introduced two versions of Intentionalism: Rule and Ad Hoc. Rule holds that the existence of (codified) social kinds depends on the collective recognition of constitutive rules which impose the given status onto subjects or objects. Accordingly, if a couple meets the relevant collectively recognized constitutive rule, then they are married. On the Ad Hoc view, it is not rules, in any social kind, which impose social status but the recognition of those who have authority in the relevant context; while their authority depends on collective recognition, on this picture, if those who have the relevant authority deem a couple to be married, then they are married, regardless of whether they meet the membership conditions in line with any collectively recognized rule or not.

I have, then, hinted at the issue that neither version of Intentionalism seems to have the resources to provide an ontological explanation for Google's "Married" category. Let us now delve deeper into this question. I will show that some insights of Rule and Ad Hoc may be worth salvaging for our analysis of MAKs.

What would the proponent of Ad Hoc say about Google's "Married"? Take someone who was, at least for certain periods of time, undoubtedly Google-Married, such as myself. Nobody needs to recognize me in any way *as* being Google-Married to be Google-Married. Yet, it also seems that I would not be Google-Married if I were not categorized, that is, *represented* as such. This is because the relevant representation endowing my behaviour (which meets a given set of membership conditions) with a status is not constructed through human intentionality, more specifically, collective recognition, but through digital data and algorithmic code.

This leads us to a crucial distinction between two referents of representation. In the previous chapter, representation was thought only in terms of human intentionality, the capacity of the mind to direct its states – such as its beliefs, emotions and desires – at the external world. Here, the relevant work of constructing a stand-in for the relevant part of the external world which is, then, directed at something beyond itself, is done via digital data and algorithmic code, not human intentionality. Thus, MAKs seem to be "representation-dependent" without being mind-dependent in either the Rule or the Ad Hoc Intentionalist sense.

However, the proponent of Ad Hoc could point out that their theory is better positioned to explain the existence of kinds such as Google-Married insofar as, just like in the cases of Ad Hoc, *every instance* of Google's "Married" needs to be represented as such to be Google-Married. To illustrate, suppose there is a temporary glitch in the algorithmic system, and it does not categorize someone as "Married" even though their behaviour matches the relevant pattern. Can this putative instance of the given kind and the relevant representation go apart in this case? It seems counterintuitive to me to say that one has the status of "Married" without being algorithmically constructed as member of the given kinds, that is, represented as such.

However, in the case of algorithmic construction of kinds, unlike in the case of Ad Hoc, the recognition of those with authority cannot ontologically trump the relevant rules. Instead, the algorithmic code itself, which is changing based on incoming data, as we have seen, *is* a set of sequences of rules. Thus, in the case of algorithmic systems, there is no distinction between rules and status attribution.

So far, we know that MAKs are representation-dependent and that the relevant status attribution is rule-based. Thus, it may seem as if Rule could be more viable candidate for explaining the ontology of MAKs. However, according to Rule, the existence of social kinds

depends on our *collectively* recognized constitutive rules which then associate the status in question with the relevant subjects or objects.

Now, for Searle, as we have seen, the collective recognition of rules – in the cases we have called codified social kinds – is enabled by *language*. Following the tradition which holds that there is no such thing as a “private language”, he argues that the relevant language is “publicly understandable”:

So when I say that language is partly constitutive of institutional facts, I do not mean that institutional facts require full-blown natural languages like French, German, or English. My claim that language is partly constitutive of institutional facts amounts to the claim that institutional facts essentially contain some symbolic elements in this sense of “symbolic”: there are words, symbols, or other conventional devices that mean something or express something or represent or symbolize something beyond themselves, in a way that is publicly understandable. (1995, 60–61, cf. 2006, 19)

Let’s, for the sake of argument, take digital data and algorithmic codes as language in the relevant sense. To what extent the “language” of digital data and algorithmic codes which constitute the representations on which MAKs depend is in fact “publicly understandable” is an intriguing empirical question. It seems reasonable to say that today, most people are simply unaware of the existence of categories such as Google’s “Married”. Furthermore, there is the issue of states and companies deliberately keeping their datasets and algorithmic codes secret as well as the lack of technical literacy among the public, both of which require extensive empirical research (cf. Burrell 2016).

However, what ultimately matters here is a conceptual point: *even if* there was a collectively recognized rule of being “Married” according to Google, the relevant algorithms specifying what “Married” would depend only on behavioural data and work independently of people’s recognition. Thus, what is important for us is that the relevant representations of digital data and algorithmic code seem to be able to exist and function without being *collectively* recognized.

Take, again, the challenges to the relevant notion of recognition in Intentionalism which we considered in the previous chapter. Kinds such as money, leader or race were brought as examples to show, respectively, that the relevant sense of recognition is not always correct about membership conditions, not always language-dependent, and not always conscious. In the case of MAKs, however, the issue is not that the relevant collective recognition is mistaken about membership conditions, non-language-dependent or non-conscious. It is rather that the relevant representations of digital data and algorithmic code seem to be able to exist and function without being *collectively* recognized. In other words, even a revised form of Intentionalism which has weakened the relevant sense of recognition cannot incorporate MAKs into its framework. Once the given algorithmic codes are programmed, MAKs only

require the representational capacities of data to function. The production and analysis of such data, however, is not dependent on collective recognition in the relevant sense.

As far as algorithmic systems are concerned, it seems, we might as well be philosophical zombies, i.e. thought-experiment human beings capable of carrying out the same behaviours as their non-zombie counterparts but without having a mind, and therefore, any forms of intentionality or internal experience. I will have more to say on the implications of this point in Part II.

One might point out that perhaps accepting Google's *Terms of Service* counts as indirect collective recognition. However, a simple thought experiment shows that that does not seem to create an ontological difference in what the existence of Google's "Married" is dependent upon: even if Google gathers data about which they do not inform us in their *Terms of Service*, the MAKs based on that data can exist and function. Incidentally, there is intriguing research on how people do not, in fact, read Terms of Service of popular online services (Cakebread 2017). As Alexis C. Madrigal (2012) succinctly puts it:

So, each and every Internet user, were they to read every privacy policy on every website they visit would spend *25 days out of the year just reading privacy policies!* If it was your job to read privacy policies for 8 hours per day, it would take you 76 work days to complete the task.

But, once again, the philosophical point is that even if everyone indeed recognized Google's "Married" and other MAKs, while it might be a politically important change, it would not make a difference in the ontology of MAKs.

It follows, then, that even though MAKs depend on representations which are rule-based, Rule Intentionalism also cannot incorporate MAKs into its framework.

The existence of MAKs may not be dependent on *collective* recognition, but it does seem to be, in some sense, dependent on the recognition of their engineers. This, one may argue, seems to align with the Ad Hoc view that we have a smaller group of *powerful* people whose recognition has the special ontological role of attributing social statuses. (I will unpack the relevant sense of "power" in Part II where I focus on the analysis of social power.) However, in Ad Hoc, this power stems from *authority* which is dependent on the collective recognition of others. For instance, the "marriage conferrer" (Ásta 2018, 29) may gain their authority from others recognizing their status.

In contrast, in the case of MAKs, the engineers' power which lies in their capacity to encode a data-driven algorithmic system which would construct various MAKs does not further depend on *collective* recognition. It is worth noting that it is not, in itself, the contrast between the *few* engineers compared to the *many* subjects whose behavioural data is quantified and

classified through the data-driven algorithmic systems which implies that the quantifier “collective” is unwarranted here. The point, instead, is the following: while it may seem that data-driven algorithmic systems can only be constructed in a social reality which comes with existing codified social kinds such as, for instance, “companies”, and the relevant associated statuses carrying certain forms of authority, in fact, a data-driven algorithmic system which is not part of a system of collectively recognized codified social kinds seems conceivable. Once again, as opposed to Ásta’s “marriage conferrer”, the engineers designing the algorithmic system do not have the capacity of endowing individuals with a given status due to a collectively recognized “authority”. I take it that this shows, once again, that the kinds constructed through the algorithmic systems they design are not dependent on collective recognition.

In sum, MAKs cannot be incorporated into either Rule or Ad Hoc versions of Intentionalism. Nevertheless, MAKs, then, just like traditional social kinds whose existence is explained by Intentionalism, are representation-dependent as they depend on the representational capacities of digital data and algorithmic code. Now, some features of Rule or Ad Hoc have proved to be useful to elucidate *certain aspects* of the ontology of MAKs. Just like cases of Ad Hoc, every instance of MAKs needs to be represented to exist as such. Furthermore, MAKs, like Ad Hoc cases, are dependent on the recognition of those who are – if not in “authority” per se – in a *powerful* position; in this case, the engineers of the given system. Moreover, just like in the cases of Rule, MAKs are rules-based. We have, then, outlined the unique ontology of MAKs which are representation-dependent but not through collective recognition.

Before I unpack how exactly MAKs are dependent on the recognition of their engineers, I take a brief *excuse* to elucidate below the lack of dependence on collective recognition in the specific cases of social kinds of people recognizing, or not recognizing, themselves as such.

*

Having agreed that there is representation through collective intentionality, more precisely, collective recognition (whether à la Rule or Ad Hoc), and representation through digital data and algorithmic code, I find it helpful to differentiate a few modes of existence of social kinds, specifically social kinds of people. When discussing the notion of “kindhood”, we have already delineated the kinds themselves, as well as representations (categories) which correspond to kinds and representations (categories) which are incidental and so do not correspond to kinds.

Here I will make a further distinction. Let’s separate the following claims:

- (1) whether people of the given kind exist;
- (2) whether a group-level representation corresponding to that kind of people exists;

(3) whether people of the given kind are collectively recognized as existing (as members of the given kind)

The Intentionalist picture holds – given that social kinds of people need to be represented as existing to exist – that (3) is partly constitutive of (1). In other words, the existence of people of a given kind depends (in the relevant sense) on their being collectively recognized as existing. Furthermore, Intentionalism often uses (2) and (3) interchangeably, as the group-level representations of various kinds of people taken to be ones which are collectively recognized – either consciously or unconsciously. In the case of MAKs, however, while (1) depends on (2), it does not depend on (3) – as (2) and (3) are now distinct.

Now, to further elucidate the difference between MAKs (of people) and traditional social kinds of people, I will draw a contrast using the insights of Ian Hacking (1995, 1997). Hacking was famously interested in a subset of (3); let's call it (3)*:

(3)*: whether people of the given kind are collectively recognized as existing (as members of the given kind) *by themselves*

Hacking's aim is to show that the existence of a group-level representation (2) enables (3)*, that is, that people collectively recognize themselves as belonging to a given category. This, then, leads to (1), that is, the existence of the given kind of people, as people may collectively begin to intentionally act in accordance with their representation. (Or, as he noted, it may also lead to people resisting their representation and thus the construction of a new kind of people.)

Using illustrative examples, he contrasted historical periods when a (2), that is, group-level representation referring to a given kind of people, did not exist, with periods when these representations did exist. In the former case, Hacking argues, (3)* was also not true, that is, people of the given kind were not collectively recognized as existing (as members of the given kind) *by themselves*. This is because (1) was not true either: people of the given kind did not exist. As he puts it, the given kind did not exist as an experienceable “way to be a person”. He quotes Arnold Davidson on the example of the category of “the pervert”:

Were there any perverts before the latter part of the nineteenth century? According to Arnold Davidson, “The answer is NO. ... Perversion was not a disease that lurked about in nature, waiting for a psychiatrist with especially acute powers of observation to discover it hiding everywhere. It was a disease created by a new (functional) understanding of disease.” Davidson is not denying that there have been odd people at all times. He is asserting that perversion, as a disease, and the pervert, as a diseased person, were created in the late nineteenth century. Davidson's claim, one of many now in circulation, illustrates what I call making up people. (1999, 161, cf. 2007, 299).

Hacking's point, more concisely, is that in the case of the category of “the pervert”, in an important sense, at some point in time, there was no (2) which would have led to (3)*, and therefore there was no (1) either.

The contrast between these claims is perhaps even starker in the following passages:

(A) There were no multiple personalities in 1955; there were many in 1985.

(B) In 1955 this was not a way to be a person, people did not experience themselves in this way, they did not interact with their friends, their families, their employers, their counsellors, in this way; but in 1985 this was a way to be a person, to experience oneself, to live in society. (Hacking 2006)

And he offers the same contrast for “autism” (2007, 303):

(A) There were no high-functioning autists in 1950; there were many in 2000.

(B) In 1950 this was not a way to be a person, people did not experience themselves in this way, they did not interact with their friends, their families, their employers, their counsellors, in this way; but in 2000 this was a way to be a person, to experience oneself, to live in society.

Here, Hacking seems to contrast (1) and (3*) – the (A) and (B) claims in the preceding passages, respectively. For Hacking, it is social scientists and medical practitioners who construct (2), that is, the group-level representation of a given kind. He then assumes that humans are unique as, after a while perhaps, they recognize the categories which have been constructed for them and react to these through intentional actions informed by their views on the given category. In other words, he assumes that after (2), the way towards (1) is through (3)*.

However, as we have seen, in the case of MAKs constructed through data-driven algorithmic systems, it is *often* the case that one does not recognize (consciously or not) the MAKs they belong to. Regardless of, and besides, the fact that the existence of MAKs do not depend on collective recognition, the point here is that people often belong to MAKs (with the associated statuses and their constraints and enablements) without in any sense recognizing that they do so. MAKs are not dependent on any form of collective recognition, even if we take recognition in its weakened form.

Furthermore, even if people *did* recognize their belonging to MAKs, it may still not be clear how the statuses associated with MAKs enable “a way to be a person, to experience oneself, to live in society” – an intriguing question to be settled elsewhere. So, using Hacking’s formulation, one could say:

(A) There were no Google-Married people in 1950; there are many today.

(B) In 1950 Google-Married was not a way to be a person, people did not experience themselves in this way, they did not interact with their friends, their families, their employers, their counsellors, in this way; but today, when Google-Married people exist as Google-Married exists as a representation, it is *still not* a way to be a person, to experience oneself, to live in society, and it is not clear how it *would* be.

The contrast between, on the one hand, MAKs, and, on the other, Hacking’s cases of social scientists or medical practitioners classifying people – who, in turn, begin to recognize the

categories they have been put into and react to these – is informative in the following sense. We have already agreed that the difference between the classificatory methods of data-driven algorithmic systems and earlier quantification-based methods or other ways of classifying people from the external perspective of the bureaucrat, the social scientist or the marketing agent may very well be one of degrees. Even so, it is now possible both to construct *kinds*, which have a significant impact on the level of groups more quickly and on a broader level as well as to make these kinds – in want of a better term – more difficult to be recognized by those who are being categorized through them.

Moreover, as we have seen, in the case of MAKs, whether or not (3) or (3)*, people of the given kind are collectively recognized as existing (as members of the given kind) *by others or themselves*, (1) the given kind of people do exist as their existence is partly constituted by (2) being represented as existing on a group-level (through digital data and algorithmic code). In the case of MAKs, then, being represented as existing on a group-level (2) and being collectively recognized as existing (3) or (3)*, come apart, while in Hackingian cases (2) eventually leads to (3), and, especially, (3)*.

Let us now circle back to the question of how exactly MAKs are dependent on the recognition of their engineers.

2.3.2 Engineers' recognition: causal vs. constitutive mind-dependence

To unpack how MAKs depend on the minds of their engineers, we will need some distinctions which we have introduced earlier. In the previous chapter, we have said that the relevant notion of dependence in “mind-dependence” in Intentionalism is a form of *ontological* dependence on collective recognition as opposed to mere *causal* dependence.

Then we have unpacked the relevant notion of ontological dependence as a relation of synchronous constitutive dependence. The dependence relation is synchronous as collective recognition is required not only for the initial construction but also for the maintenance of social kinds. The dependence relation, in turn, is constitutive as collective recognition regarding a given social kind is held to be part of what makes the given social kind what it is.

Furthermore, earlier in this chapter, we differentiated between various types of data-driven algorithmic systems, among them expert systems and machine learning systems. Both carry ontological significance regarding the relevant sense of mind-dependence: in simple terms, expert algorithmic systems diverge from synchronicity, while the machine-learning systems also diverge from constitutivity.

Now, the existence of MAKs, as we have seen, depends on *representations*. The relevant sense of representation, however, is not through collective recognition, but through digital

data and algorithmic code. Using the relevant forms of dependencies introduced earlier, we can more precisely state that the existence of MAKs is representation-dependent as it synchronically and constitutively depends on digital data and algorithmic code.

Let's examine, then, in what sense MAKs are *mind-dependent*. The fact that MAKs are social kinds which are not dependent on *collective* recognition is in itself sufficient to challenge the Intentionalist picture of social kinds; but let us not stop here. We have seen that MAKs seem to be dependent on the recognition of their engineers. How should we characterize the relevant notion of dependence, given our earlier distinctions?

Let us begin with synchronicity. It seems to me that the relevant mind-dependence is no longer characterized synchronicity in the case of not only machine-learning but also expert data-driven algorithmic systems. Take the above-mentioned example of an expert algorithmic system designed for classifying people for tax purposes. Such a system, if all goes well, still puts people in the correct tax bracket, even if no human minds existed anymore, even if we all turned into philosophical zombies. The same seems to apply to machine-learning-based data-driven algorithmic systems: machine-learning-based social media feed curating algorithms would continue classifying users into MAKs, again, even in a world without human minds. Neither the tax system, nor Google's "Married" needs any form synchronous recognition to be "sustained".

Here, the claim that no "mental sustenance" is needed is to be unpacked in the sense that mental states do not *constitute* the continued existence of these algorithmic systems. It is, then, *not* a causal claim: I do not want to conceal or deny that human "ghost" labour, to a greater extent than usually admitted by those who own these systems, plays a crucial causal role in maintaining, updating, fixing the algorithmic systems at hand (Burell-Fourcade 2021, 219 ff.).

However, the construction of both the tax system and the social media algorithm were, at some point in time, clearly dependent on human minds. Thus, even before unpacking the relevant sense of dependence – whether it is causal or constitutive – we can note that the mind-dependence relation at hand is *diachronic*. Does the fact that the relevant sense of dependence is diachronic immediately make it causal?

We have mentioned in Chapter 1 that the standard understanding of constitutive dependence thinks of the relation of constitution in synchronous terms. However, some argue that there are cases of *diachronic constitution*. For instance, Brian Epstein argues that a footprint is dependent upon a foot having left a mark in the ground *at some point* (2018, 153–154). While the foot's leaving a mark is a causal event, the relationship between the foot's leaving a mark and the footprint being a footprint is an ontological relation of constitutive *and* diachronic dependence (*Ibid.*).

We could find cases analogous to the footprint example even in the case of various social kinds. Consider, again, the difference between the existence of a kind and the existence of one its instances. Going back to our example of marriage, we have seen that the existence of the social kind of marriages depends synchronously on collective recognition. However, we can imagine a specific instance of marriage where the married couple as well as those around them no longer think of them as married. It seems that, given that marriage is a codified social kind – a legal institution – and its instances require an official divorce for the status of being married to be ended, this couple would still be married. This is because their status of being married is partly constituted by the past event of having been announced as married by the marriage conferrer or priest.

All in all, the question of whether there is such a thing as diachronic constitution hinges on, primarily, whether one accepts that the being of a subject or an object may be constituted by temporally distinct parts. This is a long and difficult metaphysical issue which I do not wish to delve into at this point. In the following, I will assume, without providing any further elaboration, that the ontological dependence relation of *constitution* can be both synchronic and diachronic; in any case, not much will hinge on this point.

Thus, going back to our examples of the algorithmic tax system and the social media feed curating algorithmic system, simply noting that the mind-dependence relation at hand is diachronic is not sufficient to imply that it is causal, rather than constitutive – in case one stipulates that there is diachronic constitutive dependence as well. An algorithmic tax system with the encoded rule that married people can apply for a tax deduction would be, in turn, an example of a diachronic *and* constitutive dependence. This is because the rules, at a point in time, were pre-set – and so constituted – by their engineers, which is an intentional action which requires the recognition of these engineers.

But we can go a few steps further. Data-driven algorithmic systems which rely on machine learning, as we have seen, rely on rules and categories, which were, in an important sense, not pre-set by human recognition. In the case of Google’s “Married”, the engineers’ recognition was required for pre-labelling a specific group, those who self-declared as being married as well as encoding the rule that however the given group behaves, that will specify its future membership conditions, and the relevant status associated with the kind. Google’s “Married”, then, tracks the behaviour of those who have “told Google” that they are married, and if the behaviour of others is similar, they are clustered together, and are shown similar content.

However, as we have seen, there are cases where algorithmic systems form clusters based on behavior similarities without pre-labeled data, relying solely on the behavioral correlations extracted from various granular-level data. Here, property clusters are drawn by the

algorithmic system, not preset by human intentionality. I believe this comes with further ontological significance as it shows that the relevant sense of mind-dependence in the case of these data-driven algorithmic systems and the MAKs they construct diverges from not only synchronicity but also from constitutivity.

In both cases of machine-learning-based data-driven algorithmic systems (ones operating with pre-labelled data and those without pre-labelled data), the relevant rules of the system are subject to constant change based on incoming data and therefore are not constituted by the recognition of their engineers. In these cases, I argue, the dependence on the recognition of the engineers is merely *causal*.

Some may argue that causal dependence may not capture the important distinction between things which can only be caused by some form of human intentionality and things which can be caused by other causes as well. Perhaps the world is such that algorithmic systems are necessarily caused by human intentionality: if so, then there could not possibly have been algorithmic systems and MAKs in a world where human minds never existed. I will, however, not go into details regarding the modalities of causation. Suffice it to say that my claim that kinds constructed through machine-learning-based data-driven algorithmic systems depend causally (as opposed to constitutively) on the recognition of their engineers is compatible with different ideas regarding the modality of the relevant causes. What matters for my purposes is to show that, in the case of the relevant MAKs, not only the synchronic but constitutive dependence is brought into question regarding the relevant sense of mind-dependence.

We have so far said, simplifying for the sake of clarity, that expert algorithmic systems depend constitutively (if diachronically) on their engineers' recognition, while machine-learning-based data-driven systems depend causally on their engineers' recognition. It is, perhaps, more precise to say, however, that there is a spectrum between expert and machine learning types of data-driven algorithmic systems which corresponds to a spectrum between constitutive dependence and causal dependence on the recognition of their engineers. The more "expert-like" the algorithmic system is, the more the relevant dependence can be characterized as constitutive dependence given that the membership conditions and the statuses of the relevant social kinds were constituted by the relevant intentional states of their engineers. On the other hand, the more machine-learning-based the algorithm is, the more the dependence can be characterized by a causal, rather than constitutive, dependence given that the membership conditions and the statuses of the relevant social kinds may modulate *beyond recognition*. Even beyond the recognition of their engineers.

Given the distinctions we have drawn so far, it seems like the existence of MAKs is (1) representation-dependent as it synchronically and constitutively depends on digital data and algorithmic code; (2) the existence of (at least an important set of) MAKs is also mind-dependent in the sense that it *causally* depends on the intentional states of their engineers.

These observations, in turn, will help us further explain how MAKs challenge even a revised Intentionalism.

*

Let us now revisit the challenges that the Intentionalist faced which drove them to revise their view. When considering the challenges regarding the relevant sense of recognition, we have agreed that the Intentionalist can accept that we may collectively be wrong about the membership conditions of a given social kind. The key is to hold that we cannot be collectively wrong about (at least what we have called “essential features” of) its status. When we considered macro-level social kinds, we agreed that they are unique as everyone could also be wrong about their very existence. Yet, as we have seen, the defender of Intentionalism can point out that some macro-level social kinds often do not have an associated status on their own and, more importantly, they further depend on social kinds which do, and which depend on being collectively recognized as existing.

However, in the case of MAKs, which are associated with clearly defined – if, in some cases, ever-evolving – statuses (such as the constraints and enablements regarding what content shows up on one’s social media feed), we can *collectively* be wrong not only about its membership conditions but also about the associated status, and even their existence.

Could even the engineers be wrong about the membership conditions, associated status and existence of a kind such as Google’s “Married”? As we established, in the case of machine-learning-based data-driven algorithmic systems and the MAKs they construct, the dependence between the engineers’ recognition or other intentional states and the membership conditions and associated status of the relevant kinds is only causal as the relevant membership conditions and associated status may modulate beyond their recognition. Thus, from the fact that the dependence relation in the case of MAKs is only causal and not necessarily constitutive, it follows that they have membership conditions and statuses which may modulate beyond what their engineers recognized or envisioned.

Engineers can also be wrong about the very existence of certain MAKs as new MAKs may be constructed by the algorithmic systems without constitutive or synchronous dependence on the intentional states of the engineers. Take the illustrative case of tech journalist Alexis C. Madrigal (2014) who set out to decipher the logic behind Netflix’s specific and complex movie-classification system. At first, he thinks about the relevant kinds in constitutive terms:

I could tell someone had gone down this road before. A single human brain had had to make the decisions that we had. How many adjectives? How long should they be? And even more basic: what should the adjectives be? Why cerebral and not brainy? Why differentiate

between gory and violent? (...) The words themselves were carefully chosen. By whom?
(Madrigal 2014)

Then Madrigal highlights a curious phenomenon he calls “The Perry Mason Mystery.” Despite what one would expect, Raymond Burr, from the 1950s series Perry Mason, and Barbara Hale, his co-star, outrank modern Hollywood icons like Meryl Streep and Samuel L. Jackson in Netflix’s otherwise sensible list of “mostly expected Hollywood stars”. This shows that, in data-driven algorithmic systems, categories – MAKs – might emerge as an unintended and unforeseen consequence, independently of the recognition of their engineers. The explanation for the existence of these categories provided by Todd Yellin, the leading engineer of the system, is telling: “[it is] inexplicable with human logic. It’s just something that happened” (Ibid.). This shows that causal dependence is compatible with being wrong about the existence and membership conditions and associated status of a given kind.

Another way in which MAKs differ from the examples which challenged Intentionalism and led to its revision in our previous chapter is this: when considering the challenge of macro-social kinds, such as inflation or recession, the revised version of Intentionalism holds that they are nevertheless indirectly dependent on social kinds which are dependent on collective recognition, such as money. Along these lines, one could argue that MAKs are ontologically dependent on their “offline” counterpart. So even if Google’s “Married” is only causally dependent upon the intentional states of its engineers, it is indirectly dependent on *marriages* which is a social kind dependent on collective recognition which is constitutive and synchronous.

I believe this is true about *some* MAKs: MAKs such as “Google-marriage” indeed indirectly depend on social kinds which are dependent on collective recognition. However, even this indirect dependence is weaker than in traditional examples of macro-kinds. While Google-marriage seems to be indirectly dependent on the existence of marriage *at some point in time*, the dependence in the case of MAKs is diachronic: no “sustained” or “continued” dependence is required. It indeed seems possible that there could be Google’s “Married” even if people stopped getting married and marriage became merely a relic of the past: Google’s “Married” tracks a pattern of behaviour, not *marriage*. Just like I am “Married” but not married, people could be still clustered together based on tracking patterns of similarity in their online behaviour without anyone being married. So, it seems to me that, if marriage stopped being a social kind, Google’s “Married” could still function according to its encoding and perhaps capture important similarities.

Furthermore, remember that we differentiated between two types of machine-learning-based data-driven algorithmic systems. The first one learns with the use of some pre-labeled training data (“people who’ve told Google they’re in this category”). The second one, however, does not depend on any set of pre-labelled data and construct kinds simply by clustering together based on correlations in behavioural data. Kinds constructed by the latter

type of data-driven algorithmic system, then, do not seem to depend, even diachronically, on an “offline” social kind counterpart.

For instance, in a project inspired by Judith Butler’s theory of gender performativity, researchers created categories based on similarities in language use in Twitter tweets using clustering algorithms without prelabelled data (Bamman et al. 2014, cited by Cheney-Lippold 2017, 79 ff.). Previous research would often frame the relationship between gender and language use as follows: how do men and women use language, presupposing collectively recognized binary gender categories. By contrast, these researchers ran a clustering algorithm on millions of Twitter tweets by more than ten thousand authors and created 20 clusters, named C1 to C20, with “similar lexical frequencies”, and while some of these “strongly skewed with respect to gender”, there was great variety between clusters, showing that “there are multiple linguistic styles which enact each gender” (Bamman 2014, 20–22). Implementing such real-time clustering classification methods in a data-driven algorithmic system such as the recommendation system on a social media platform constructs MAKs independent of an “offline” social kind counterpart which would depend on collective recognition (cf. Burrell 2016).

One may still point out that kinds born out of data-driven algorithmic systems without prelabelled data still depend on other social kinds which depend on recognition, such as some basic labels in initial datasets (e.g., in the gender case above, they used categories such as parts of speech) or bits of the code employed. However, the point is to show that a given social kind is further dependent on social kinds which are mind-dependent in the sense that they are dependent on collective recognition which is constitutive and synchronous. The basic labels or the code, however, does not seem to be mind-dependent in this way. Thus, the distinction between MAKs which depend on a pre-existing social kind (Google-marriage/marriage) and MAKs which do not (C20 does not depend on any collectively recognized gender) still holds.

In sum, given that MAKs are mind-dependent in a non-collective, non-synchronous, and non-constitutive way, they seem to pose new challenges to even a revised version of Intentionalism.

2.3.3 *The “end of the social”?*

While MAKs, as we have established, do not depend on collective recognition, and (in some cases at least) constitutive or synchronous recognition of their engineers, we have also established that the tradition of social ontology which yielded the view of Intentionalism, both ideal and nonideal trends thereof, holds that collective recognition is the *mark of the social*. More precisely, we have defined the *mark of the social* as the necessary and sufficient conditions of something being social. Collective recognition, then, turned out to be a

necessary building block of the existence of social kinds, the elements of social reality. In this last subsection, I *also* examine the sufficiency part of the mark of the social: not just the question of what is required for something to be social but also that of what it takes for something to be social.

We have seen in the *Introduction* that social ontologists have two main intuitions about the mark of the social. First, even if one accepts a naturalistic framework where only reality is nature, one needs to contrast the social part of nature (marriages) with the non-social part of nature (mountains). This is where the notions of mind-dependence, intentionality and recognition entered the picture. Second, the social is contrasted with the individual subject; it, instead, is related to groups of subjects. Given that Intentionalism holds that the relevant groups are, in fact, *collectives*, this is where the notion of a *collective* came to the focus.

Collective intentionality, and, specifically, collective recognition, then, emerged in the Intentionalistic picture of social reality as a concept which captures these two intuitions and serves as a necessary condition for all social kinds. Indeed, holding onto the traditional idea that collective recognition is a necessary building block of social reality, many authors draw the conclusion that MAKs are not social. One could accept the causal role of data-driven algorithmic systems in structuring social reality but insist that, given the lack of collective recognition involved, we are only faced with an “illusion of sociality” – just as in the case of, as we have seen in the previous chapter, insect “societies” whose elements do not depend on collective intentionality (Harré 1997, 208; cf. Greenwood 1997, 7–8). Consequently, now that quantification and classification through data-driven algorithmic systems is becoming widespread *in* our social reality, some indeed hold that it has become timely to speak of the “end of the social” (e.g., Baudrillard 1983; Lee 1997; Couldry-Mejias 2019).

Here are a few examples of the discourse around the end of the social. Already in the 1980s, Baudrillard, in analyzing mass media and information production argued that it brings about “the end of the social” and that “the social regresses to the same degree as its institutions develop” (1983, 66). This discourse has been even more amplified by the growing number of analyses of algorithmic societies. Raymond M. Lee, for instance, endorses the vocabulary of the “post-social” and “non-social” when arguing that our historical position “provides a vantage point to understand why some contemporary thinkers have jettisoned the social to prepare the ground for coming to terms with the powers of *the non-social*” (Lee 1997, 255, 258, my emphasis). John Cheney-Lippold, in turn, shows that “our social identities, when algorithmically understood, are really not social at all” (2017, 5, cf. 46). Nick Couldry and Ulises A. Mejias argue that the datafication and algorithmization of social reality is leading to the “hollowing out of the social” (Couldry-Mejias 2019, 118, 115–152). Others also warn that the end of the social is signaled by the fact that “we are losing the common shared knowledge that once bound us together as citizens” (Bartoletti 2020, 9). Finally, Antoinette Rouvroy

famously argues that algorithmic categories are not “a result of common exchange” (Rouvroy–Berns 2013, v). In an interview, she formulates this idea as follows:

...individuals are no longer profiled according to socially experienced categories (ethnic origin, gender, work experience, etc.) in which they could recognize themselves, through which they could assert collective interests, but on the basis of “profiles” automatically produced according to their numerical trajectories and interactions, which no longer correspond to any socially experienced category. (Rouvroy 2016, my translation)

Thus, while in the social reality of Intentionalism, “we see the other person not simply as an individual, but as colleague or comrade or fellow partisan; in a word, inhabitant of the same peculiar world” (Simmel 1910, 380), the ever-expanding digital infrastructure, as Eran Fisher argues “puts us at a risk of entering a *post-social cosmos* where individuals have a harder time identifying each other as sharing a similar category” (Fisher 2022, 46, my emphasis).

Baudrillard, however, at one point, adopts a tone more open to the possibility that we are, in fact, experiencing a new form of sociality:

The rational sociality of the contract, dialectical sociality (that of the State and of civil society, of public and private, of the social and the individual) gives way to the sociality of contact, of the circuit and transistorised network of millions of molecules and particles maintained in a random gravitational field, magnetised by the constant circulation and the thousands of tactical combinations which electrify them. But is it still a question of the *socius*? (1983, 83)

I would answer Baudrillard’s last question in the affirmative: I consider MAKs as *social kinds*.

As we have seen – although this is not a sufficient condition in itself – they have consequences which are significant on the level of groups, thereby living up to one of the main intuitions of social ontologists of what counts as social. Ásta, for instance, argues that social ontology should deal with those “categories that matter to our social life” (2018, 29). MAKs clearly meet this criterion of group-level or “social” significance (no wonder social scientists use terms like “socially consequential mechanisms of classification” [Burrell 2016, 1] or “public relevance algorithms” [Gillespie 2014, 168]). Yet having consequences significant on the level of groups alone may not be sufficient to call something social. Kinds such as freshwater or earthquake also “matter to our social life” in the sense that they have consequences significant on the level of groups, but we may hesitate to call them *social kinds*.

My argument is that MAKs are social kinds – even if they are not dependent on collective recognition or any intentional states in a synchronous and constitutive way – as they have relevant similarities with the Intentionalist understanding of social kinds. In other words, collective recognition is too specific a form of mind-dependence and representation-

dependence to specify it as the mark of the social – a necessary and sufficient condition for something being social. Let’s unpack this a little.

It seems to me that MAKs capture both intuitions of social ontologists regarding what counts as social. That is to say, not only the intuition that they concern group-level consequences, but that MAKs are also sufficiently different from non-social parts of nature. First, the existence of even the most high-level modulating MAKs is also mind-dependent in the sense that it is causally dependent on the intentional states of their engineers. Second, MAKs introduce statuses onto subjects and objects which are dependent on forms of *representations*.

In the algorithmic society, as we have seen, there are now two forms of representations as building blocks of social reality: beyond representation through collective recognition there is also representations through digital data and algorithmic code. This growing independence of the two forms of representation as building blocks of social reality is nicely illustrated by Raymond M. Lee when he discusses “codes of dominance” as the “driving force of social life”:

It is this agenda that deflates the primacy of the subject and elevates discursive structures (or one could say “codes of dominance”) as the driving force of social life. It is in this sense that society is said to be “decentered,” without the individual or self-consciousness central to the notion of civil order. (Lee 1997, 257)

Thus, the following shared features of elements of both traditional and algorithmic societies establish that MAKs are social kinds: being consequential on the level of groups, being mind-dependent (even if only causally), and, crucially, being (constitutively and synchronically) dependent on some form of representation – in this case, through digital data and algorithmic code.

Dependence on collective recognition may indeed mark an important difference – but not one between social and non-social kinds, but one between various kinds of social kinds, all of them being elements of social reality. Thus, I’d argue that we need not commit ourselves to the idea that kinds falling outside of the scope of traditional Intentionalist theories of the social cannot be social. The inclusion of MAKs into the realm of social kinds, then, is an example of how historical shifts, such as developments in technology, may challenge even our most-encompassing conceptual analyses.

One of the pioneers of analytic social ontology, Sally Haslanger, when elucidating her view on the role social ontology in debunking putative “natural” kinds such as gender or race as social kinds, writes that “an important first step is to make the category visible *as a social category*. This sometimes requires a rather radical change in our thinking” (Haslanger 2012, 132, my emphasis). Such a change in our perspective is what is required to view MAKs as social kinds.

And such a change is what I think we need in order to understand the algorithmic society which is in the making today.

Concluding Remarks: Digital data and algorithmic code as a new glue of social reality

I began with reiterating the findings of the previous chapter that a revised version of Intentionalism seems to have general coverage of the social world. After unpacking two ways in which the Intentionalist can interpret the relevant role of collective recognition in kind-making, Rule and Ad Hoc, I showed that I happen to be categorized as “Married” by Google, while, *prima facie*, not being *married* on either Intentionalist accounts. This prompted us to ask three questions. What kind of category is Google’s “Married”? In what sense is it mind-dependent, as a corollary, can it be incorporated into the Intentionalist framework? Is it a social kind?

Here are my responses, in brief.

First, I identified Google’s “Married” category as a Modulatory Algorithmic Kind, characterized by data collection of unprecedented scale and granularity, vastly increased computational power for algorithmic predictive analytics, and both data collection and analysis happening real-time. I then discussed various types of data-driven algorithmic systems, and, accordingly, various levels of modulation which may characterize the MAKs they construct. I also argued that regardless of the changing and prolific nature of MAKs, they qualify as *kinds* given that taking a Mill-type of view of kinds which emphasizes their epistemic role in prediction given their multiple shared properties.

As to the second and third questions, I showed that while MAKs are mind-dependent, they are not dependent on collective recognition for their existence. Furthermore, the relevant sense of dependence (at least in the case of some MAKs) is causal, that is to say, non-synchronous (as they do not require recognition for continued existence), and non-constitutive (as some MAKs can modulate beyond what human minds might have previously recognized). Then, I showed that given MAK’s consequences on the level of groups, their mind-dependence (as they depend on the intentional states of their engineers), their representation-dependence (as they depend on digital data and algorithmic code), they indeed qualify as *social* kinds.

MAKs are elements of a society which involves elements where the relevant “social glue” is representation through digital data and algorithmic code and not via collective recognition of subjects; a society which is increasingly not solely a matter of collective, synchronic and

constitutive role of human intentionality, as has been traditionally postulated by social ontology. As such, MAKs seem to pose new challenges to Intentionalism.

I believe that this is time for the Intentionalist to admit that it cannot hold onto its ambition of being a monist theory, even if the scope of monism is restricted to human social reality. Thus, it seems that a pluralist framework is in a better position as it would be able to provide an answer the General Problem (which aims to uncover the general conditions of possibility for the existence of any social kind) without thereby losing Intentionalism's valuable answer to the Placement Problem (which aims to accommodate uniquely human social kinds into a naturalistic ontology).

A tentative proposal for such a pluralist picture, then, holds that there are (1) some social kinds which are not representation-dependent, and, therefore, not dependent on collective recognition (e.g., ant queens), (2) social kinds which are representation-dependent in the sense that they are dependent (either directly or indirectly) on collective recognition (money, marriage), (3) as well as social kinds which are representation-dependent but are not through collective recognition (e.g., MAKs). By going pluralist, we could also legitimately welcome back into social reality the social animals we excluded in the last chapter.

The next task, then, is to better understand the new forms of social *power* which data-driven algorithmic systems, through the MAKs they construct, enable and which characterizes an increasing part of social reality. Media scholar John Cheney-Lippold rightly remarks that, at present, "we lack the vocabulary needed to enact a politics around our algorithmic identities" (2017, 30). The first step was to understand the underlying ontology of new ways of kind-making. Now, it is time to move onto the analysis of the forms of social power enabled by data-driven algorithmic systems.

Part II: Social Power and Its Subjects

CHAPTER 3: BROADENING THE CONCEPT OF SOCIAL POWER: ALGORITHMIC FORMS OF SOCIAL POWER

Control is not discipline. You do not confine people with a highway. But by making highways, you multiply the means of control. I am not saying this is the only aim of highways, but people can travel infinitely and “freely” without being confined while being perfectly controlled. That is our future.

Gilles Deleuze: *Two Regimes of Madness* (2006, 322)

Introduction: Social power, intentional status and social status

The upshot of the previous two chapters is that the scope of even a revised version of the Intentionalist picture of social reality – which holds that (human) social kinds are mind-dependent in the sense that their existence depends (directly or indirectly) on collective recognition (in a weak sense) – needs to be further narrowed. This is because, as we have seen, through data-driven algorithmic systems such as social media platforms, streaming platforms, credit card companies, or search engines, subjects and objects are classified into Modulatory Algorithmic Kinds (MAKs) which, as I argued, are social kinds but are not mind-dependent in the Intentionalist sense – the relevant mind-dependence being neither collective, nor synchronous, nor (in the case of high-level modulating MAKs) constitutive.

In this chapter, I will focus on the concept of social power to ultimately show what forms of algorithmic powers are enabled by data-driven algorithmic systems. To do so, I need to clarify the relationship between social power, intentional action, and social status. Relying on my preceding analysis of the mind-dependency of social kinds, I will argue that exercising power via data-driven algorithmic systems brings about new forms of social power where neither the power of bestowing statuses nor the powers stemming from the statuses themselves depend on collective recognition. Yet, as I will show, these are social powers nonetheless as they meet the criteria of what I will call a minimal concept of social power. Furthermore, I will show that in the case of what I will call algorithmic forms of *nudging* power, the power of bestowing statuses is not only not dependent on, but also not enhanced by, collective recognition.

Here is how I will proceed.

In the first section, I will begin formulating a minimal concept of social power which I will rely on in the subsequent analysis of various forms of social power. This minimal concept holds that social power is a capacity of getting other subjects to do what the powerholder wants

them to do. In the second section, I will investigate one kind of constraint on a minimal concept of social power which holds that power is, in some way, dependent on intentionality; I will call it, following Searle, the “Intentionality Constraint”, and I identify two versions thereof. The first holds that the exercise of power is dependent on some form of intentional action. The second holds that the existence of power is dependent on statuses which depend on collective intentionality. I will accept the first constraint (with a few caveats) and will challenge the second. Regarding the first constraint, as I will show, the exercise of social power requires intentional action on the part of the powerholder. As to the second, I will show that in an algorithmic society, where statuses and collective intentionality become ontologically independent, forms of algorithmic power are an exception to this constraint.

Throughout my analysis, I will develop two – non-exhaustive – taxonomies of social power: one addressing the ways in which power can be hidden, and another addressing the types of statuses which social power can construct and depend on. This will then be supplemented by a third taxonomy in the following chapter.

One last remark before we begin. In a similar vein to the previous chapter, I aim to take the project of “non-ideal sociology” one step further. Ideal social ontologists often assume the following about social power: that social power always depends on pre-existing statuses; that social power is transparent in the sense that individual subjects have conscious intentional states about each other’s statuses; that the statuses on which social powers depend are codified and rule-based. Usual examples of ideal social ontologists tend to be those powers stemming from rights and responsibilities associated with legal categories, such as “citizen” or “president”.

By contrast, non-ideal sociologists such as Åsa Burman (2007, 2023) have done important work broadening our perspective by focusing more on forms of power which are, in some sense, *hidden*. They focus on powers stemming from non-transparent statuses which depend on intentional states of which subjects are not conscious as well as less codified and rule-based statuses. Their examples tend to include gender, race and various forms of disability.

Even in non-ideal social ontology, however, there has yet to be comprehensive analysis of the kinds of social powers – and their associated forms of “hiddenness” – which are made possible by data-driven algorithmic systems. In this chapter, I will continue the line of thought which explores hidden forms of power by identifying what I will call algorithmic forms of social power.

3.1 A minimal concept of social power

Many authors writing on social power, both in analytic philosophy and in other disciplines, begin their treatise by remarking that the English word “power” refers not only to *social* power – whatever that may turn out to be – but to various other kinds of phenomena, for instance, in physics, engineering or computer science (Lukes 2005, 61–62; Burman 2007, 127–128; Searle 2010, 145–146). Discovering what is common between the electrical power of the lightbulb, the horsepower of the car, the processing power of a CPU *and* social power, however, is a task for another inquiry. Here, to get off the ground, I am assuming, along with many power theorists, that while the (English) word “power” is used in many contexts, there is a distinct phenomenon which we can call “social power”. As Searle puts it, there exists “a specific form of power that goes with human social (...) reality” (Searle 2010, 145). Thus, I will focus solely on *social* power; whenever I use the term “power” without a qualifier, it refers specifically to social power.

To begin carving out a *minimal* concept of social power, let us start with examining what seems to me the main challenge in conceptualizing social power: the tension between, on the one hand, incorporating hidden forms of social power and, on the other, setting legitimate constraints for the concept.

A crucial insight of 20th-century and contemporary analyses of social power is that we should broaden our conception of power by focusing on hidden forms thereof (Lukes 2005, 2021; Foucault, 1980, 1982, 1995, 2000; Burman 2007, 2023). I take it that various forms of social power can be categorized in many ways; one important taxonomy emerges when we consider the ways in which power can be hidden. As I will unpack below in detail, power can be hidden in the following ways: (i) social power can exist without observable behaviour, (ii) without conflict or (iii) without being consciously represented through the relevant intentional states of those over whom power is exercised or those who exercise power. It is, then, often argued that these hidden forms of power are crucial – if not the most crucial – forms of social power. Thus, if this is true, then if the concept of power aims to be comprehensive and capture its most crucial forms, it must also incorporate its hidden forms.

Now, it is also crucial that we should establish legitimate constraints to avoid over-broadening the concept. This is because, as I will show below, broadening the concept of social power comes with methodological difficulties and may run the risk of incorporating any kind of causal interaction between subjects.

In this section, I will focus on broadening the notion of social power to incorporate hidden forms thereof; the next section is dedicated to finding legitimate constraints.

3.1.1 Power-to and power-over

It is often noted that there are two terms corresponding to two “core concepts” of social power in the literature: “power to” and “power over” (Morriss 2002, 32–35; Lukes 2005, 34; Burman 2007, 141–143, 2023, 210).

Those who propose that the core concept of social power is “power to” argue that social power-to refers to a *capacity* of a social actor to engage in some action (e.g., Morriss 2002, 34, cf. Lukes, 2005, 64, 69; Searle 2010, 145–146). An often-cited example is the prime minister’s power to dissolve parliament (Moriss 2002, 32). The point is that power does not *only* refer to its exercise through an action – power as a capacity does not need to be exercised to exist. In our example, the prime minister has the power to dissolve parliament even if it is not being exercised – even if it is never exercised.

On the other hand, proponents of the “power-over” view focus on the fact that power is a relation between social actors and their actions. In the prime minister case, they would put emphasis not on the inherent capacities of the prime minister but on the fact that exercising these capacities means having power over the actions of members of the parliament. In other words, the key insight of power-over view is that that social power is inherently *relational*.

Now the debate is often framed as a dichotomy between the “power-to” and the “power-over” view (Morriss 2002, 32–35; Lukes 2005 34; Burman 2007, 141-143, 2023, 210). This is because it is argued, on the one hand, that the relational nature of social power is overlooked by the “power-to” view, and, on the other hand, the “power over” view overlooks power as a capacity as it focuses only on observable behaviour (cf. Oppenheim 1981, 31; cf. Burman 2007, 141 ff., Morris 2002, 40). I will not delve into what theorists of power, in fact, tend to overlook; the point I will make, which I take to be rather straightforward, is that these positions are logically compatible.

The reason we are often presented with a dichotomy between “power-to” vs. “power-over” is that important questions – which, in turn, will help us devise a taxonomy of the various ways in which power can be hidden – are not kept separate. Relying on Burman’s “central” questions about the concept of social power (2007, 134 ff.), I believe we need to distinguish (at least) three questions: whether power must always be exercised to exist, whether power necessarily involves some form of conflict between subjects and whether those over whom power is exercised or those who exercise power represent the power relation at hand through their conscious intentional states. This may sound rather abstract at this stage; let us, then, put flesh on these points. Doing so, as I will show, will reveal that “power-to” and “power-over” are not only compatible but both views capture crucial forms of social power; as such, they should be combined.

3.1.2 *Hidden forms of power*

Proponents of the “power-to” view are right, I believe, in thinking that power as a capacity exists not only when it is being exercised through some form of observable behaviour. Specifically, focusing only on observable behaviour may hinder our ability to recognize two hidden forms of power.

First, conceptualizing power as a capacity highlights the kind of power which stems from *inaction* (cf. Bachrach and Baratz 1962, cited by Lukes 2005). Think of a member of parliament refraining from voting. Here, by choosing not to act, the member of parliament may sway a decision in a certain direction. For a more specific example, take the case of Hungarian members of parliament delaying the ratification of Sweden’s NATO-membership. The point here is that power includes things that have not been done or said, things that did not happen; inaction, then, becomes visible as a form of social power, precisely because we recognize one’s capacity to act otherwise.

The second hidden form of power that becomes visible once we conceptualize power as a capacity are various forms of *intimidation* where *the relevant power* does not have to be exercised to elicit compliance (Searle 2010, 149, 158; Burman 2007, 144). Consider a mafia leader intimidating a member of parliament through verbal threats to ensure they vote in the direction the mafia leader wants them to. Or think of a NATO again; NATO does not have to exercise its military capacities to deter potential adversaries, as the perception of NATO’s military capacities and the fear it elicits is sufficient to secure compliance. If only the exercise of (the relevant) power counted as power, then we would not be able to account for the fact that people can be intimidated through their perception of power and fear of some form of punishment – without the exercise of the relevant power.

Thus, the insight of the “power-to” view to think of power as a capacity rather than merely as an observable exercise of a capacity sheds light on hidden forms of power: those which lie in inaction as well as those lying in intimidation. This seems to be one good reason to think of power as a capacity which does not necessarily have to be exercised to exist.

Having the capacity to do something, however, might not be sufficient to capture the *social* nature of power. For example, one can easily think of a multitude of capacities which one can have and exercise alone, say, on a deserted island, having no effect on any other subjects. These capacities, whether exercised or not, do not strike us “social” powers. This shows that having a capacity to do something may be necessary but not sufficient for capturing the social nature of power. We should, at this point, turn to the insight of proponents of the “power-over” view that social power is inherently relational. Accordingly, social power is not merely

a capacity to do something, but a capacity to have an effect *over* the actions of another subject or other subjects. Let's formulate this initial definition explicitly to get off the ground.

Social power initial: a capacity to have an effect *over* the actions of another subject or other subjects.

I will call the relation between those having power over others and those over whom others have power a *power relation*. While the simple formula of one subject having power over the actions of another subject or other subjects seems to imply a potentially oversimplified one-way direction of power relations, one does not need to subscribe to that. As many have pointed out, power is often a two-way street where, within a power relation, the various parties have power over each other. However, for our purposes at least, I will assume that having power over each other is analyzable as the sum of two one-way streets: one having power on another – and vice versa.

Accordingly, we will need also terms to refer to those having power over others and those over whom others have power. Let's call those subjects having power over others as *powerholders*. Now, finding a term for those over whom others exercise power leads to an interesting terminological ambiguity. As I will further unpack in the next chapter, the term "subject" often does not only refer to, as we have already established, a being with a mind, intentionality, and a sense of self, but also to those over whom others have power. To differentiate between these meanings, I will refer to *those* subjects over whom other subjects (powerholders) have power as *sub-jects* – a term drawing on its Latin etymology of "being thrown under". Thus, some subjects are powerholders, others are sub-jects – and, often, subjects are indeed both.

Now, onto the relevant notion of "over" in "having an effect *over* the actions of another subject or other subjects". It is often held that the relevant sense of "over" necessarily involves some form of *conflict*. Steven Lukes, in his early work on power, argues that viewing power as "power-over" rather than "power to" is important to highlight "the conflictual aspect of power – the fact that it is exercised *over* people" (Lukes 2005, 34, emphasis in the original). Others seem to hold that focusing on conflict is, in fact, a weakness of the power-over view as it makes the concept of power too narrow.¹⁵ What exactly is meant by the rather vague term of "conflict" here?

To clarify the relevant sense of conflict and the role it plays in social power, let's draw a rough distinction between conflict on the level of behaviour and on the level of intentional states. That is to say, there is a difference between engaging in conflict on the level of behaviour,

¹⁵ Morriss, for instance, criticizes the "power-over" view as it is only about "affect(ing) others in various nasty ways" (Morriss, 2002, 33), rather than social power in general. In other words, he seems to assume that some form of conflict of interest is necessary for the "power-over" view.

such as in some form of political protest, or simply harbouring negative intentional states – beliefs, desires or emotions – about the relevant others. Furthermore, one can ask whose behaviour or intentional states in the power relation – that of the powerholder or the subject – are characterized as conflictual. Finally, one could also ask whether, regardless of the behaviour and intentional states of those involved, the power relation can be characterized by a conflict between the interests of those involved (cf. Lukes 2005, 30, 64; Burman 2007, 145).

Now, it seems to me that while these are useful distinctions to draw, a minimal conception of social power, remains neutral with regards to the specifics of whose behaviour or intentional states is characterized by conflict and whether, regardless of the behaviour and intentional states of those involved, there is a conflict between the interests of those involved. Without explaining every possible kind of scenario where power involves, or lacks, conflict of a given kind, I will here briefly mention one pair of contrast cases: that of intimidation and productive power.

For the sake of clarity, I will call the conflict on the side of the subject, whether on the level of behaviour or intentional states, “resistance”. In the case of intimidation, in some cases due to the fact that *the relevant power* is so intimidating that it does even have to be exercised through observable behaviour to elicit compliance, there is often no resistance, at least on the level of behaviour. However, there could still be resistance on the level of intentional states. Given our example above, while there may not be any direct attack on NATO members, representatives of other countries may harbour negative intentional states toward them.

However, not all cases where the relevant power is exercised involve resistance, not even on the level of or intentional states. In a special kind of case, which I will call, following Foucault, “productive power”, resistance can also be prevented by the exercise of power itself; think of the kind of power which lies in changing people’s – subjects’ – intentional states, such as beliefs, emotions and desires, in order to make them have the intentional states which align with the powerholder’s goals. Perhaps the most tangible example is how the media can be used as a propaganda machine, using Walter Lippmann’s (1998 [1922]) famous terms, can “manufacture consent”, that is, shape public opinion to secure not only compliance of any sort but *willing* compliance with the goals of the relevant powerholders.

As I will show in detail in Chapter 4, the form of power exercised to prevent resistance both on the level of behaviour and intentional states is at the center of much of the work of Foucault (e.g. 1980, 1982). He referred to this form of power as “productive power” as it “produces subjects” through changing their beliefs, emotions and desires so that they align with those which the powerholder want them to have. As Steven Lukes – who, inspired by Foucault, aimed to capture this productive form of power with the vocabulary of analytic philosophy – put in his rhetorical question: “is it not the supreme exercise of power to get

another or others to have the desires you want them to have – that is, to secure their compliance by controlling their thoughts and desires?” (Lukes 2005, 27–28).

So far, we have seen that asking whether power as a capacity is exercised is distinct from the question of whether the given form of power involves conflict and if so, by whom and on what level. We established that beyond those forms of hidden power stemming from the lack of exercise of (the relevant) power, there is another hidden form of power which lies in the lack of conflict in the sense of lack of resistance on the level of behaviour and/or intentional states.

Now, onto the third sense in which power can be hidden. The relevant question to ask in this case is whether either the powerholder or the sub-ject have conscious intentional states representing the power relation as existing. Let’s recap our taxonomy of intentional states from Part I, to be able to find the place for such conscious intentional states representing the power relation as existing. We have, regarding the existence of social kinds, talked about the intentional state of “recognition”, a form of intentional state representing the given social kind as existing. We have agreed that a revised Intentionalism accepts that recognition need not be conscious. Here, to pinpoint the specific kind of recognition we need, i.e., conscious intentional states representing the power relation as existing, I will use the term *awareness*. I will have more to say on specific kinds of intentional states and how they relate to various forms of power in the next chapter.

In Part I, in the context of Intentionalism, I have also noted that the relevant sense of recognition is compatible with “enthusiastic acceptance” as well as mere “going along”; in a similar vein, given our distinctions above, we can say that awareness is compatible with the presence and absence of conflict on the level of intentional states. In other words, while the presence of awareness, either by the powerholder or the sub-ject, is often related to the question of whether conflict on the level of intentional states is present as well, these two are separate issues.

First, while one may be intimidated by others and there may even be conflict on the level of intentional states, one may not have awareness of the power relationship at hand. For example, imagine a workplace of a pervasive and accepted culture of fear where employees comply with their management as a result of being intimidated, without being explicitly aware that their compliance stems from such a power relation. Second, take the above-mentioned example of media propaganda. One may be aware of the fact that one’s beliefs and desires are being strategically changed by the relevant powerholders. Yet, one may not harbour any negative – conflictual – intentional states towards the powerholders; one may consider such “manufacturing of consent” necessary and justified. These two examples show awareness of the power relation and conflict on the level of intentional states (either by the powerholder or the sub-ject) can exist independently of one another. Thus, the third sense in which power

can be hidden is the exercise of social power where either the powerholders or the sub-jects are not aware of it.

Let us see where we have gotten so far. We have noted that there are two basic formulations of social power in the extant literature: “power-to” and the “power-over”. I showed that a minimal concept of social power should adopt insights from both views. Then I argued that asking the right questions reveals a taxonomy of hidden forms of social power – our first taxonomy of power. First, we established that, as the “power-to” view holds, social power is a capacity which does not have to be exercised through observable behaviour to exist. This shed light on two forms of hidden social power: cases of inaction and intimidation. Then we established, as the “power-over” view emphasizes, that social power is relational. This led us to two additional senses in which power can be hidden. First, when there is a lack of conflict in the power relation, either on the level of behaviour or also on the level of intentional states. Second, when there is a lack of awareness, either by the powerholder and/or by the subject, of the power relation at hand.

Given that we stipulated that the concept of social power should encompass the hidden forms of power, we need to update our concept of power accordingly. Overall, we need to broaden our concept of social power as follows.

Social power ^{updated}: a capacity to have an effect over the actions of another subject or subjects – who thereby become sub-jects – regardless of

- (i) whether the relevant capacity is exercised or not, that is, whether it is a case of inaction or intimidation;
- (ii) whether conflict, either on the level of behaviour or on the level of intentional states – or even on the level of interests – is present;
- (iii) whether either the powerholder or the sub-ject is aware of the power relation at hand.

While I will not explicitly mention (i)-(iii) in my further analysis of the definition of social power, I will assume that these “broadenings” are in place.

There are, however, legitimate worries about such a minimal concept of social power. For instance, there may be methodological difficulties in the application of such a minimal concept in empirical research. In the case of power as inaction, it may be difficult to establish the relevant counterfactual scenarios are: when does one have the capacity to do otherwise? In cases of intimidation, it may also not be easy to tell whether one is acting in a certain way because they are afraid of potential punishment or because of some other reasons. In cases of “productive power”, it may be difficult to uncover the causal history of the intentional states of the subject, settling the issue of whether their intentional states have been produced by others or not.

For our purposes, however, it is not necessary to provide guidance regarding these issues. It is sufficient to establish that the cases which these concepts delineate exist. For instance, it is sufficient to establish that in some cases, there is power in inaction; there are cases where sometimes people are indeed intimidated; there are also cases where people's intentional states (such as their beliefs) – often contrary to their interests – have been shaped in a way such that they align with the goals of the powerholder, at times even without them being aware of this power relation.

While I am not considering methodological difficulties here, I will consider some necessary clarifications which hold that the minimal concept of social power requires constraints to avoid being counterintuitively broad. In the following, I will examine one kind of constraints which our minimal concept of social power may require and, eventually, see how it fares with the new forms of power which are brought about through data-driven algorithmic systems.

3.2 Intentionality constraints

Many may point out that, as it stands, our concept of social power is too broad: not any capacity to have an effect over the actions of another subject seems to count as social power.

One might, for example, point out that without clarifying the nature of the “effect” in our definition of power, we may unintuitively incorporate effects too weak to count as exercises of power. For instance, a slave may bring a smile to their owner's face via their hard work, thereby having an effect on another subject's actions, but this would likely change nothing important regarding their power relation – it would not make the owner into a sub-ject – since the smile doesn't affect the inequality which stems from their respective statuses.

Is it sufficient, then, some may ask, if the effects are minor or should the effects in fact be so strong as to change the power relations at hand? In a similar vein, one may ask if it is sufficient if the effects only pertain to one's immediate peers or they should in fact spread, say, over a whole continent, having a larger scale? These are indeed crucial questions to clarify. However, for our purposes, I believe we do not need to worry about specifying the strength and scale of the relevant effect. This is because, in the cases of algorithmic power I will be focusing on, as I will show, both the strength and scale of the relevant undoubtedly qualify as *powerful enough*.

Another clarificatory point, the one I will focus on, holds that social power is linked to some form of intentionality (Searle 1995, 2010, Burman 2007, 2023, Taylor 1984). For instance, Searle argues that any “respectable” discourse on social power should hold that there is a “logical” connection between intentionality and power, captured by the motto, “no

intentionality, no exercise of power” (2010, 151 ff.). I will call this view, following Searle, the Intentionality Constraint. In the following two subsections, I will distinguish between two senses in which intentionality is held to be necessary for the existence of social power. I will then evaluate how the forms of power that are made possible by data-driven algorithmic system fare with these constraints.

3.2.1 *The First Intentionality Constraint: Do as I want!*

In the following, I will unpack the first intentionality constraint on a minimal notion of social power which helps us clarify the relation between power and intentional action.

One could point out that I have been, in fact, assuming a more robust concept of social power than I have so far admitted. Already in the examples so far mentioned, the argument would follow, the term “effect” was used in a specific sense. To illustrate, take the following simple case: somebody trips over and accidentally pushes another person who almost also trips over but then manages to hold onto something firm. Or take Stanley Benn’s more elaborate example: consider a careless smoker who, although never achieving anything in life that they in fact want, once accidentally causes a wildfire by throwing away a not-yet fully extinguished cigarette butt which indeed has an effect on the actions of other as they desperately try to save their houses from burning down (Benn 1967, 426, cited by Morriss 2002, 20). These scenarios seem to satisfy the criteria of having an (exercised) capacity of having an effect on the actions of another subject. We would, however, demur to call these instances of social power. This is because of our intuition that the relevant notion of “effect” should not encompass any form of causal impact – power, it seems, is not a matter of accident or luck.

This leads us to the first way in which power is often linked to intentionality, that is, the First Intentionality Constraint. The First Intentionality Constraint unpacks the “capacity to effect” as follows: the capacity to *get another subject or other subjects to do what one wants them to do*. This way, we can rule out examples involving accidental behaviour like tripping over and pushing someone or causing fire with a cigarette butt as exercises of power. More precisely, the First Intentionality Constraint holds that the exercise of social power requires *intentional action on the part of the powerholder*: it requires the intentional state of wanting one’s sub-ject(s) to engage in a given course of action as well as action driven by these intentional states, to actually “get” others to do what one wants them to (cf. Morriss 2002, 27; Burman 2007, 134–135, 148; Searle 2010, 147).

First Intentionality Constraint: the exercise of social power requires intentional action on the part of the powerholder in the sense of getting another subject or other subjects to do what the powerholder wants them to do.

I will, for now, remain neutral on whether the actions which the powerholder wants the subject to engage in are also intentional actions; I will revisit this question in detail in the next chapter.

Assuming that the exercise of power is linked to intentional action on the part of the powerholder is, in turn, endorsed by many in the literature. Robert Dahl, for instance, formulates an “intuitive idea of power” which he unpacks as follows “A has power over B to the extent that he can get B to do something that B would not otherwise do” (1957, 202–203). Steven Lukes, referring to the specific case of productive power, writes about the “power to get another or others to have the desires you want them to have” (2021, 32). Searle also takes this route when writing about “the ability of an agent of power to get subjects to do what the agent wants them to do whether the subjects want to do so or not” (2010, 147). I am assuming, then, that “getting” others to do something, in all the quotes above, refers to an intentional action.

Still, there are some difficulties.

First, we have already agreed that power can be hidden in the sense that both the powerholder and/or the subject may not be aware of the power relation. How can one have wants or desires about a desired action carried out by another subject or other subjects and exercise it intentionally if they are not, in some sense, aware of them? In Chapter 1, we have postulated that there is *unconscious* intentionality; it is, therefore, possible to have wants and desires regarding actions taken by other subjects and intentionally exercise one’s power to get others to take those actions, even if one is unaware of doing so. Searle, for instance, also takes this route and talks of cases of unconscious exercises of power (2010, 151). An example of he provides is regarding the “social pressures” he would face if, for instance, he wore high-heels and female clothes; he argues that even if people would not consciously intend to exercise power on him, their behaviour would still somehow convey intentional states along the lines of “Conform!” (2010, 155 ff.).

But cannot every behaviour be described through the language of having intentional states? Take the example of ant colonies from Chapter 1. Ants can be said to have the intentional state of “Gotta keep the place tidy!” when they are engaged in nest hygiene – even though they are merely responding to chemical signals like oleic acid released by their decaying colleagues. I will not settle the details of this issue. It does seem, however, that in the case of much of human behaviour, unlike ant behaviour, unconscious intentional state attribution may not only be, in some sense, explanatorily useful but seems to be the best explanation to make sense of the relevant behaviour. As nonideal social ontologists have pointed out, this is how we can make sense of behaviour which unconsciously follows intentional states such as deeply held expectations about, for instance, gender or race. Putting aside important methodological difficulties pertaining to how we are to attribute intentional states in various

contexts, I am merely committed to here, along with Searle, the idea that the First Intentionally Constraint does not imply that the relevant forms of intentionality are conscious – just that some form of intentional action is required.

Second, and more importantly, regardless of whether the relevant parties are aware of the exercise of power, we often have an effect on other subjects in unintended ways. As we have seen, accidental behaviour like tripping over and pushing someone or causing a fire with a cigarette butt do not seem to count as exercises of power. The defender of the First Intentionality Constraint could argue that there is a difference between accidental behaviour and unintended consequences of intentional exercises of social power. I am happy to weaken the First Intentionality Constraint, in line with Searle, to accommodate unintended consequences of intentional exercises of power (2010, 150–151).¹⁶ I will, however, remain neutral on whether we can identify *every* case of unintentional exercise of power as an unintended consequence of an intentional (whether conscious or not) exercise of power.

Now, it is common to differentiate, broadly, between two types of unintended consequences (cf. Hosanager 2019, 21). The first type of unintended consequence is one where the intended goal is carried out, but the process involves, potentially unforeseen, side effects, either benefits or drawbacks. An interesting example of this kind of unintended consequence is the power of policymakers who regulate schooling. For instance, as historian Richard Reeves highlights, schools starting at an earlier age negatively affect boys’ performance whose intellectual capacities develop later (cf. Reeves 2023). I take that it is uncontroversial that the relevant education policymakers did not have unconscious intentions to make schools harder for boys. Furthermore, it also seems reasonable to assume that the original goal of policymakers with setting the starting age of school was not to disadvantage girls in their academic development or to create gender-based differences in education. The starting age of school negatively affecting boys’ performance is, then, simply an unforeseen side effect, more precisely, an unforeseen drawback.

The second type of unintended consequence is usually called “perverse results” (Hosanager 2019, 21). These occur when the result of the exercise of power, in fact, turns out to be the opposite of the original goals. For instance, the outcome of a policy might actually end up making the original problem worse, instead of solving it. Thus, in contrast to unintended side effects, perverse results undermine the original goals of the policy. A famous anecdote – used by several authors – illustrating perverse results comes from colonial India, where, as the

¹⁶ Searle also makes a distinction between “unintentional exercise of power” and “unintentional exercise of influence” and argues that the latter is not a case of power: “I am not exercising power over my students if they imitate (or try to avoid) my style of speaking because I am not intentionally trying to influence their behavior” (*Ibid.*). However, his argument relies on a narrow interpretation of conscious intentional states which excludes, first, the possibility of unconscious intentional states, and, second, unintended consequences. This contradicts his (and our) initial aim of broadening the relevant concept of intentionality to meet the challenge raised by cases of unconscious exercise of power and unintended consequences.

story goes, a policy was introduced to control the rat population (cf. Hosanagar 2019, 22 who also introduces a “cobra” version). The policy involved offering a bounty for rat tails, thus rewarding people for killing rats and turning in their tails. The idea, of course, was that this would reduce the rat population. However, instead of reducing the rat population, the policy, according to the anecdotes, backfired: people began breeding rats for the sole purpose of collecting their tails to claim the reward, which, in turn, increased the rat population rather than reducing it. The result of the exercise of power through these policies, then, turned out to be the opposite of the original goals, thereby making it a case of perverse results.

Can both kinds of unintended consequences of power (e.g., via policy decisions) be accommodated in the first intentionality constraint view? We have said that unintended consequences of intentional exercises of power, whether foreseen or unforeseen, still count as an exercise of power. However, perverse results may pose an extra difficulty as in these cases, the original goal is not carried out. Therefore, in cases of exercises of social power where the original goal is what the relevant subject as a powerholder wants their sub-ject or sub-jects to do, it is an open question whether we can still hold that the powerholder indeed has the power to get other subjects to do what they want them to do. One example of perverse results, where the powerholder is trying to exercise their power and does act accordingly, but does not achieve their intended goal, may, in fact, be due to resistance of the sub-jects. The question, then, is whether we have reasons to attribute the relevant capacity to the powerholder regardless of occasional perverse results. Which cases – and how many or how frequent – of perverse results may imply a lack of power and which do not, however, is an issue for another project.

In sum, we have accepted the First Intentionality Constraint which holds that the exercise of social power requires intentional action on the part of the powerholder in the sense of getting another subject or other subjects to do what the powerholder *wants* them to do. The relevant intentional states of the powerholder – required by the intentional action of getting another subject or other subjects to do what one wants them to do – may be unconscious and they may have unintended consequences, both side effects and perverse results. I will, then, in the following, assume that these additional stipulations about the relevant sense of intentional action of the powerholder are in place and will not always highlight them explicitly. Finally, I note here again that I have remained neutral on whether intentional action is also required on the part of the sub-ject; I will unpack this issue in the next chapter.

*

Now, can forms of power which are made possible by data-driven algorithmic systems be made consistent with the First Intentionality Constraint? Take, for instance, the engineers of social media platforms or streaming platforms who have power over their users as they have

the capacity to get them to spend much of their time – and, ideally, money – on these platforms.

One may point out that, while unintended consequences, both side effects and perverse results, can occur in both non-algorithmic and forms of algorithmic power, forms of algorithmic power are especially prone to it. This is because, as we have seen in the previous chapter, the power which lies in constructing kinds through data driven algorithmic systems is about kinds which do not depend constitutively or (in some cases) synchronically – only causally – on the intentional states of the minds of their engineers.

Indeed, a prominent problem in the literature around data-driven algorithmic systems, called the “Alignment Problem” (cf. Christian 2020) addresses the worry that given their self-learning and so self-modulating nature, the behaviour of these algorithmic systems can easily be misaligned with the goals of their human engineers. An example of potential side effects of social media algorithms that prioritize “engagement” involve the creation of echo chambers and the prioritization of attention-grabbing over quality content. An example of perverse results, then, is an algorithm designed to assist in the hiring process in identifying the best candidates which may be, given the historical hiring data used to train the algorithm, biased in many ways (e.g., gender-wise), thus misaligning with the original goal of finding the best candidates.

Assuming there indeed is an increased potential for unintended consequences in the case of power exercised through data-driven algorithmic systems, both side effects and perverse results, one may ask whether that leads to any important philosophical implications.

Some seem to argue that the potential for unintended consequences which stem from the self-learning, and so modulatory nature, of data-driven algorithmic systems should drive us to attribute power to the algorithmic systems themselves. Indeed, some may go as far as to argue that social power lies, in fact, not only with humans and their organizations, but also, or perhaps increasingly, with the “algorithms themselves”. In other words, as opposed to linking the effects of algorithms to human power, they attribute power to the algorithmic systems themselves. Indeed, in the social scientific literature, one encounters papers titled “The social power of algorithms” (Beer 2017) as well as claims about algorithms “ruling our world” (Steiner 2012) or the idea that we live in “society in which power is increasingly in the algorithm” (Lash 2007, 71).

Could we attribute power to algorithmic systems given our definition of power? Given our definition of power, it seems that the powerholder needs to be a human subject, or at least a being with the capacity of intentionality as they need to be able to want another subject or other subjects to act in a certain way. It seems straightforward, then, that individual humans can be wielders of social power. For instance, Foucault is often quoted to show that even his

view, with all its emphasis on hidden forms of power embedded in “institutions” and “structures”, puts people in focus as the wielders of power:

For let us not deceive ourselves: if we speak of the power of laws, institutions, and ideologies, if we speak of structures or mechanisms of power, it is only insofar as we suppose that certain persons exercise power over others. (Foucault 2000, 337)

Others, however, hold that it is quite natural to say that nations, groups, institutions, and organizations may also be powerholders (cf. Dahl, 1958, 203; Lukes 2005, 71). Yet others hold that if groups and institutions can be powerholders, then algorithmic systems, given that they are relevantly similar, are also powerholders. Christian List, for instance, emphasizes that the relevant basis for the comparison is that “both phenomena involve non-human goal-directed agents that can make a difference to the social world” (List 2021, 1213).

Given our First Intentionality Constraint, attributing power to groups and algorithmic systems hinges on whether we can attribute the relevant intentional states to not only individual humans but also to nations, groups, institutions – and algorithmic systems. I will, however, not settle this issue here; I take it that what I will say about forms of algorithmic power is compatible with any of these views.

In sum, the tendency of power exercised through data-driven algorithmic systems – algorithmic power – to lead to unintended consequences (both side effects and perverse results) is compatible with accommodating forms of algorithmic power within the minimal notion of power, as constrained by the First Intentionality Constraint. Furthermore, I remain neutral on whether algorithmic systems themselves can be powerholders or whether we should always take human subjects as powerholders.

In the following subsection, I will unpack the Second Intentionality Constraint on a minimal notion of social power which will, in turn, clarify the relation between power and social status.

3.2.2 The Second Intentionality Constraint: Power and status

In this section, I will unpack what I will term the Second Intentionality Constraint, according to which the existence of social power depends, either directly or indirectly, on collective recognition. This constraint has been defended most famously by Åsa Burman (2007, 2023), who seems to be inspired by the Intentionalist understanding of social reality which I unpacked in Chapter 1. Intentionalism, remember, holds that the existence of social kinds depends, either directly or indirectly, on collective recognition. We have also seen that social statuses are associated with membership in a given social kind (Burman 2007, 150; 2023, 188–189).

Combining Burman’s insights with Intentionalism, I will take The Second Intentionality Constraint to hold that the existence of social power depends on collective recognition as the existence of social power depends on the existence of social statuses and the existence of social statuses, in turn, depends on collective recognition.

In Chapter 2, however, I have shown that even a revised version of Intentionalism cannot accommodate MAKs (Modulatory Algorithmic Kinds) whose existence is not only not dependent on collective recognition as the existence of MAKs only depends on the intentionality of their engineers. Furthermore, the existence of (at least some) MAKs depends on the intentionality of their engineers only causally, as opposed to being mind-dependent constitutively and synchronically as, according to Intentionalism, traditional social kinds are.

Below, I will rely on my preceding analysis of the mind-dependency of social kinds and MAKs, to analyze forms of algorithmic power. I will argue that forms of algorithmic power – while their exercise involves bestowing statuses, just like in the case of traditional forms of social power – are not dependent on collective recognition and so do not meet the Second Intentionality Constraint. This, I contend, shows that forms of algorithmic power differ from an important set of traditional forms of social power.

*

Let us begin with explaining the forms of traditional social power whose existence depends on the construction and maintenance of statuses. Earlier, I have said that membership in social kinds is associated with social statuses. I, then, defined social status as those constraints and enablements which could not exist given the “sheer physics” of their wielder as they require being represented as existing to exist. Now, as I will explain, the forms of social power whose existence depends on the construction and maintenance of statuses are exercised through having social statuses associated with membership in various social kinds.

In the extant social ontology literature in analytic philosophy, the most famous taxonomy of various forms of social power is provided by Åsa Burman (2007, 2023). Here, rather than adopting or reconstructing her taxonomy of forms of social power in its entirety, I will focus on the forms of social powers which she holds to be fundamental as they are, as she argues, directly dependent on collective recognition. These are, as I will explain below, “deontic powers” and “telic powers”. But, first, a quick note on Burman’s approach and my own.

Burman aims to provide *one overarching* taxonomy of social power which “capture(s) central dimensions of social reality” – although she does note multiple times that her proposal is “tentative” and “may not be exhaustive because there might be other categories of social power” (2023, 4, 201–202). By contrast, in this and the following chapter, my aim is to provide *several* possible ways to carve out taxonomies of social power depending on the specific

question driving our categorization. Earlier in this chapter, I have already outlined one possible taxonomy of social powers addressing *the ways in which power can be hidden*. Here, I am about to outline a second taxonomy – inspired by, and relying on, Burman’s work – which addresses *the types of statuses which social power can construct and depend on*. Before I proceed, a quick word of caution: I consider neither of these taxonomies exhaustive.

I will begin with deontic powers. Deontic powers refer to the capacity to get other subjects to do what one wants them to do through imposing statuses where the relevant constraints and enablements are rights and obligations (e.g., authorizations, permissions, requirements, or penalties). Deontic powers are dependent on collective recognition to exist because rights and obligations depend on statuses associated with social kinds to exist (cf. Burman 2023, 209–213). Both the powerholders and subjects of the exercise of deontic power, then, are required to recognize the relevant statuses and comply independently of what they want or desire (cf. Searle 2010, 123 ff., Burman 2023, 54).

Take, for instance, the codified social kinds and the associated statuses of being married, being a CEO or being a citizen. A married person has the deontic power in the form of the enablement of the right to make decisions on behalf of their spouse in certain matters; CEO has the deontic power in the form of the enablement of the right to hire and fire employees; citizens (of most democratic countries) have the right to vote or express opinions. As Burman shows, deontic powers made possible by these codified social kinds are most often highlighted in ideal social ontologists such as Margaret Gilbert, John Searle or Raimo Tuomela (Gilbert, 1996, 8; Searle 1995, 2010; Tuomela 2002, 223; cited by Burman 2023, 177).

I will leave open the question of whether there are also deontic powers made possible by less codified social statuses, such as gender or race; that is, I will remain neutral on whether the constraints and enablements bestowed by membership in social kinds such as gender and race may qualify as rights and responsibilities even if they are not codified in some institutional or legal framework (Searle, for instance, seems to suggest that non-codified deontic powers indeed exist, cf. 2010, 118).

Burman then makes the case that what she calls “telic power” also needs to be incorporated into the theory of social power (Burman 2007, 14, 95 ff.; 2023, 6, 171, 176-198; her argument is adopted by Searle as well: 2010, 145). Burman’s notion of telic power refers to having and/or exercising the statuses which one is endowed with based on one’s “perceived” distance from an “ideal” which, in turn, “effects certain outcomes” (2023, 213–214). Translating her notion into our terminology, we can take telic power to be getting another subject or other subjects to do what one wants them to do through imposing graded social statuses which are defined in relation to how closely they are perceived to approximate social “norms” and “ideals” (*Ibid.*).

For instance, take the shared norms about what a “an ideal married couple”, a “good student” or an “ideal man” is like. The status is derived, then, from how well the given subject is perceived to approximate the given ideal. A couple perceived to be, say, stable and mutually supportive may be granted enablements in the form of, for instance, easier access to an apartment for rent in a competitive housing market. A student perceived to be closer to this ideal (e.g., through good grades, discipline, or active participation) may be granted enablements in the form of more opportunities or trust by their teachers. Similarly, men who are perceived to be closer to the “ideal man” – for the sake of argument, let’s say, by being strong, assertive, and responsible – may have the enablement of being more likely to be endowed with further, leadership-involving (perhaps deontic) statuses.

Telic powers are, then, dependent on collective recognition to exist because these graded social statuses, which are defined in relation to how well they approximate social “norms” and “ideals” depend on collective recognition to exist. The exercise of telic power, then, also requires its powerholders and sub-ject to recognize “norms” and “ideals”; they recognize that they should live up to these ideals and/or expect others to do so.

Burman’s proposed distinction between deontic and telic powers, however, needs some further clarification.

First, in case one holds that not all deontic powers are dependent on codified statuses, the distinction between deontic and telic powers, then, will not boil down to one between codified social statuses versus those social statuses which are defined in relation to a less codified “norm”. In fact, regardless of the position we take on whether there are more and less codified statuses which enable the exercise of deontic powers, there indeed seem to be more and less codified statuses which enable the exercise of telic powers. Being a good student is measured against, for instance, as we have seen, the codified norm of getting good grades (however, in most cases, a student gets no new deontic rights by getting an A instead of an A-). On the other hand, living up to being an “ideal man” depends on less codified norms of the given community about what men should be like.

Second, having a status based on having telic power often comes with deontic powers. Take Burman’s own example of a medical research committee granting postdoctoral scholarships where “a female applicant had to be 2.5 times more productive than the average male applicant to receive the same competence score” (Wennerås–Wold 1997, 342, quoted by Burman 2023, 202–203). Here, using our terminology, male and female applicants were subjects of telic power (being perceived as being closer and further away, respectively, from the ideal of a good researcher), and, based on that, were granted – or not – the deontic power of having the right for receiving a postdoctoral scholarship.

It seems to me, then, that the most useful contrast between deontic and telic forms of power has to do with the nature of the relevant statuses that their existence depends on. While deontic powers depend on statuses which are defined by right and responsibilities which tend to come as binaries (one is either married or not, either a CEO or not, either a citizen of a given state or not), telic powers depend on statuses which are defined by a perceived distance from norms and ideals and thus come in degrees (one can be a good student or an ideal man to various degrees) (cf. Burman 2023, 191). Thus, even if a clear separation of deontic and telic powers is not easy to make, the contrast between being dependent on statuses which come in binaries and degrees, respectively, suffices – for my purposes – to tell apart these two forms of power.

What matters ultimately for my purposes is that the existence of both deontic and telic powers necessarily depends on collective recognition. This is because the existence of statuses – endowed either through membership in a binary social kind (married, CEO, citizen) or in one which is graded in accordance with a norm or an ideal (ideal couple, good student, ideal man) depends on collective recognition. Furthermore, both deontic and telic powers work through the intentional actions of their powerholders and sub-jects, for instance, as they exercise their rights, fulfill their responsibilities and strive to live up to norms and ideals.

These intentional actions, according to Burman, are always driven by “normative reasons”, involving a social form of “ought” (2023, 212, 214). I would, however, given our preceding analysis of various forms of power in the first section of this chapter, draw a few more distinctions. Remember that we have differentiated, if only briefly, between intimidation and productive power. Accordingly, we can draw a distinction between whether the types of intentional action which the sub-jects of intimidation and productive power engage in: whether these sub-jects comply unwillingly, because they are afraid of potential punishments, or whether they comply willingly – out of some form of “ought” – endorsing the relevant statuses because their beliefs and desired have been changed to align with what the powerholder wants.

Of course, the boundaries of willing and unwilling compliance are not always easy to draw. For example, one may comply out of *prudence*, recognizing the benefits of compliance without being intimidated and without endorsing the set of collectively recognized statuses which enable the exercise of deontic and telic powers. In the next chapter, I will further explain what such endorsement amounts to; I will henceforth refer to the endorsement of a power relation as legitimate as having *internalized* intentional states about the relevant set of statuses. I will also have more to say on forms of power which require compliance of any sort (e.g., unwilling, prudential, or internalized), and those which specifically require internalized willing compliance in the next chapter. The point I’d like to make here, then, is merely that both deontic and telic powers work on sub-jects through collectively recognized

statuses which provide them with reasons for intentional action – regardless of where they fall on the spectrum of compliance, from willing to unwilling.

*

We have now seen how, on Burman’s view, the existence, and thus, exercise of social powers – both deontic and telic forms thereof – depend on the existence of statuses. Let us consider a few examples as contrast cases to further clarify this idea. Following Searle, Burman contrasts social power with “brute” power, for example, as she puts it, the “ability to beat someone up” which “works through the intrinsic features” of the person who has it (Burman 2007, 148–150; cf. Searle 1995, 24). In other words, even though brute power refers to a capacity which can be intentionally exercised over other subjects to get them to do what one wants, on this view, it still does not qualify as a social power as its existence or exercise does not depend on the existence of a status.¹⁷

In a similar vein, Burman also aims to differentiate social power from “brute mental power” (Burman 2007, 152–153). She argues that if one manipulates another subject into doing what one wants using “only” one’s “intellectual powers” but not one’s social status, then – even if this case is an example of an intentionally exercised capacity having an effect over another subject – it does not qualify as an exercise of social power as the existence of such power, again, does not depend on a pre-existing status.

This might strike some as unintuitive. Both “brute power” and “brute mental power”, it seems, are indeed capacities which, when exercised, are not only having an effect over another subject but are, in fact, exercised to establish power relations. Think, now, of a group of shipwrecked people. This is a scenario explored by numerous stories; let’s take Ruben Östlund’s “The Triangle of Sadness” (2022) as inspiration for our purposes. In this specific scenario, the survivors at hand were previously on a luxury yacht, some of them formerly rich and famous – others, servants, and maids. After the shipwreck, they start renegotiating their power relations with one another, and the former servants and maids, thanks to their “brute powers” of fishing and fighting and “brute mental powers” of manipulation, take the upper hand.

The powers which qualify as social powers according to the Second Intentionality Constraint are the deontic and telic powers which are gone, due to the shipwreck. The powers that come

¹⁷ Some may point out that one reason to demur from calling “brute power” a social power is that it refers to a capacity that can be used for various purposes, not just social ones. However, even if the given capacity can only be used over other subjects, on Burman’s view, it still would not necessarily qualify as social power. So, for instance, being a good martial arts fighter would also not qualify as a social power – only the status having a black belt in karate would. While being a good martial arts fighter is a capacity that can only be exercised over other people and – in some cases – it might determine important power relations, it does not qualify as a social power as, again, its existence or exercise does not depend on a status.

into play in the aftermath of the shipwreck and define the new power relations, however, do not qualify as social powers according to the Second Intentionality Constraint: they do not depend on pre-existing statuses but rather on the relevant people's "brute powers" (fighting, fishing) and "brute mental powers" (manipulation).

This is how the Second Intentionality Constraint aims to delineate social powers from just any forms of capacity through which one may have an effect on the actions of other subjects. Let us, now, see the challenges this view faces. I will highlight two in particular.

First, one may wonder when exactly a power relation between two subjects – or, zooming back into our example, two shipwrecked individuals – solidifies into their having their respective *social statuses*? In fact, this worry leads to further problems. The static time-slice ontological picture of presupposing that social power depends on existing statuses makes it difficult to explain the emergence of new statuses, let alone the first ones. Surely, the power of bestowing the first statuses does not depend on a pre-existing status. This insight implies that, *even if* we were to accept the Second Intentionality Constraint, we would have to broaden the concept of social power, once again, to hold that social power either depends on existing statuses or it is the power of bestowing statuses.

The second issue is to settle whether, even in the case of those social powers whose existence depends on social statuses, the social statuses themselves depend on collective recognition. As we have seen, Intentionalism holds that all social statuses depend on some form of collective recognition as social statuses are associated with membership in social kinds and social kinds depend on collective recognition. The Second Intentionality constraint, along these lines, holds the same, adding that the existence and exercise of all social powers depend on social statuses.

However, as we have seen in Chapter 2, in recent years, in various spheres of our social lives such as in social media, recommendation systems of streaming platforms, online marketplaces, as well as in databases in healthcare, HR-departments, criminal justice, education, or public administration, data-driven algorithmic systems have constructed new kinds of social kinds, MAKs, and their associated social statuses. These social statuses, as all statuses, depend on some form of representation, but the relevant sense of representation is representation through digital data and algorithmic codes and not collective recognition. In other words, an important consequence of data-driven algorithmic systems is that social statuses can exist without collective recognition.

If some social statuses do not require collective recognition for their existence, then this shows that *even if* we were to narrow the concept of social power and only include those forms of social power whose existence depends on a pre-existing social status, in line with the Second Intentionality Constraint, it would not hold that they necessarily require collective

recognition for their existence. Therefore, the Second Intentionality Constraint is not a necessary condition for social power.

I will, in turn, not only not subscribe to the Second Intentionality Constraint but I will also not assume that it would be the case that *all* forms of social power either depend on existing statuses or is the power of bestowing statuses (even if these statuses do not depend on collective recognition for their existence). Nonetheless, these are the forms of social power which I will focus on.

*

Let us, then, look at a few examples of forms of algorithmic power whose exercise depend on statuses or are the power of bestowing statuses which are, in turn, not necessarily dependent on collective recognition. The relevant statuses, once again, are encoded into digital data and algorithmic codes which are forms of representation which do not depend on collective recognition. The difference, once again, between deontic and telic powers exercised via data-driven algorithmic systems and their “traditional” counterparts lies in whether they require collective recognition for their existence.

In most cases, deontic and telic powers are intertwined in data-driven algorithmic systems. For instance, one may be given certain deontic powers based on one’s telic power. A tangible example is the capacity of companies to use MAKs to track and evaluate the performance of their employees. For a specific example, take Burman’s example of a traditional source of telic power: one’s perceived distance from being an “excellent professor” (2023, 180); now, increasingly, universities employ digitalized teaching excellence frameworks where algorithms analyze data from sources such as student satisfaction questionnaires or peer observations to automatically assess the performance of professors, which may then be used for promotions.¹⁸

Deleuze, in his famous essay, *Postscript on the Societies of Control* (1992), also considers a similar example: allocating bonuses to employees based on performance metrics as opposed to having a fixed salary as characteristic of what he calls “society of control”. This, using our terminology, is another example of the intertwining of deontic powers (having the right to receive a salary of a given amount) and telic powers (measuring how well the employee lives up to the ideal of the best performing employee) through data-driven algorithmic systems.

¹⁸ For instance, Kate Vredenburg (2021), in an essay where she argues that those over whom power is exercised through algorithmic systems should have the right to an explanation of how these systems work, refers to a 2009 case in Washington, DC, where a teaching excellence framework called 'IMPACT' was implemented to evaluate teacher performance. However, the model proved to be overly complex and opaque, leading to a lack of explanation for teachers regarding their ranking (Vredenburg, 2021, 1).

As another example, consider a digital surveillance-based “social credit system”, which assesses the score of “trustworthiness” of citizens and companies (e.g., Zhang, 2020). In such systems, sub-jects are classified into MAKs and bestowed an associated status which, in turn, is a gradable point measured up against the ideal of the “trustworthy citizen”. This, then, determines various constraints and enablements: for instance, one’s deontic powers such as the right to travel, rent or purchase property, or getting a loan.

Finally, take the example of moderating content on social media platform through data-driven algorithmic systems programmed to identify and remove content, such as hate speech, which violates its policies. Here, the telic power lies in measuring up the graded deviation from the “norm” of acceptable speech. Once someone has been identified to have deviated from it, the status of those who are categorized as having published hate speech carries the constraints of the (negative) deontic power of being deprived of the right to publish content.

Now, as we have seen in Chapter 2, through data-driven algorithmic systems, one can bring about social kinds, MAKs, whose associated statuses and *the existence of the forms of power relations enabled by them* do not depend on collective recognition. However, it may be pointed out that the exercise of bestowing algorithmic forms of deontic and telic power are *enhanced* when there is, in fact, collective recognition of the relevant statuses encoded in these systems, accompanied by intentional action of the relevant subjects. In more concrete terms, if one recognizes – perhaps even consciously becomes aware of – the statuses at play in the workings of the given data-driven algorithmic system, then algorithmic forms of deontic and telic power might be more efficient. Forms of algorithmic power may drive the intentional actions of compliance of its sub-jects out of fear of punishment, prudential considerations or internalized intentional states such as beliefs and desires driving the endorsement of the statuses at hand.

In the case of data-driven algorithmic systems of “teaching excellence frameworks”, if both students and professors are collectively aware that this system bestows statuses with, for instance, deontic enablements (e.g., promotions), it may drive students to provide feedback more thoroughly and professors may be driven to meet the criteria set by the framework. In Deleuze’s example of allocating bonuses to employees based on performance metrics rather than a fixed salary, collective recognition of the status constructed by these metrics may drive employees to intentionally meet these metrics. In the case of a social credit systems, if a citizen is aware that their access to traveling or financial services depends on their status derived from their social credit score, they may be driven to change their actions intentionally in order to have a high social credit score. Finally, collective recognition of social media platforms’ algorithmic power of content moderation, which may include taking away one’s right to publish content, may drive users to be cautious and self-regulating.

Thus, while algorithmic forms of deontic and telic powers do not require collective recognition for their existence, they may be more efficient when there is indeed collective recognition of the relevant statuses. This collective recognition, then, allows for individual subjects to intentionally comply, willingly or unwillingly – or, for that matter, do not comply – with the intentions of the powerholders encoded in the data-driven algorithmic systems.

In the final section of this chapter, I will discuss a form of power I will label “nudging” which relies on constant monitoring, predicting and steering its sub-jects’ behaviour. Although well-discussed in social theory and political science literature, nudging is often overlooked in discussions of power in analytic social ontology. Nudging power, as I will show, can also be exercised through data-driven algorithmic systems endowing statuses, thereby making possible a unique form of status-dependent power relation. Forms of algorithmic nudging are, then, distinct from the forms of deontic and/or telic powers, even if they are employed in algorithmic systems, as they are not only not dependent on, but also not enhanced by, collective recognition of its sub-jects.

3.3 Soft but powerful: Modulatory predictive nudging

In this final section, I will discuss a third form of power besides deontic and telic power – expanding our second taxonomy of social which addresses the types of statuses which various forms of social power can construct and depend on.

This third form of power – and, accordingly, third type of status – I aim to discuss below is related to what in behavioral economics and public policy is often labelled *nudges* (cf. Sunstein and Thaler 2009). Nudging assumes that people often don’t engage in conscious intentional action preceded by strategic decision-making but go with the status quo or default option. Therefore, designing the default option in a way that is beneficial for them (or in a way the engineers want them to behave) is a way to exert power over people without coercing or penalizing them if they do otherwise. According to legal scholar Cass Sunstein and economist Richard Thaler, who popularized the term nudge, in the case of nudges, “choices are not blocked, fenced off, or significantly burdened” (Sunstein and Thaler 2009, 5). An often-mentioned and tangible example of the “subtle influence” or “gentle power” of nudging is a cafeteria where healthy food options such as salads and fruits are displayed in the first row while less healthy alternatives are more hidden in the back (Sunstein and Thaler 2009, 1–8).

In our terminology, then, in the case of power through nudges, or nudging power, one gets another subject or other subjects to do what they want them to do by steering their behavior through eliciting certain behaviour without coercing or penalizing their sub-jects if they do otherwise. Nudges, then, are different from deontic and telic powers as they do not rely on

social kinds and their respective statuses which would entail coercing or penalizing sub-jects if they do not align with them.¹⁹

Given that nudges operate without coercing or penalizing, nudging is especially prone to being hidden in two of the three senses of hiddenness delineated above: nudging power is often exercised without resistance on the part of the sub-ject either on the level of behaviour or intentional states, and it is often exercised without sub-jects being aware of its existence. I will have more to say on this topic in the next chapter.

Now, as I will show below, nudging power can be exercised through data-driven algorithmic systems. In fact, sub-jects of nudging power through data-driven algorithmic systems may be even more prone to not becoming aware of the relevant exercise of power. This is because, as we have seen in Chapter 2, data-driven algorithmic systems classify subjects and objects based on large scale, granular and real-time data, which is continuously processed through algorithmic analysis which is often opaque and difficult to understand.

Now, sub-jects of nudging power through data-driven algorithmic systems are endowed with a status: a constraint or enablement based on sub-jects' *predicted* behaviour which is dependent on representation through digital data and algorithmic code. These predictions, then, are adaptable and dynamic as data-driven algorithmic systems adapt to incoming data and thus are updated constantly. To refer to the nudging power exercised through the prediction-driven statuses bestowed by data-driven algorithmic systems, I will use the term, *algorithmic form of nudging power*.

Many scholars in the social sciences write about the new forms of social powers enabled by various types of algorithmic systems, and some of them also provide a taxonomy thereof. For example, legal scholar Marc Schuilenburg and public administration expert Rik Peeters talk about, on the one hand, “algorithmic decision” which is about “status determination” or “determining rights and obligations” as well as various forms of algorithmic “policing” and “risk assessment” – overlapping, largely, with what I called deontic and telic powers so far. On the other hand, they also delineate a form of power they call “algorithmic prediction”, arguing that the latter is “a more recent addition to the tools of government”, defining it as follows: “algorithms are used to generate behavioural insights on specific target populations

¹⁹ This form of power echoes Foucault’s old insight that freedom is not always an antithesis to power, as sometimes it is a precondition thereof:

Power is exercised only over free subjects, and only insofar as they are “free.” By this we mean individual or collective subjects who are faced with a field of possibilities in which several kinds of conduct, several ways of reacting and modes of behavior are available. (...) Consequently, there is not a face-to-face confrontation of power and freedom as mutually exclusive facts (freedom disappearing everywhere power is exercised) but a much more complicated interplay. In this game, freedom may well appear as the condition for the exercise of power... (Foucault 2000, 342, cf. 342–347; cf. also Lukes 2005, 57)

for the purpose of actively influencing their future behaviour” (2021, 2–6). Shoshana Zuboff, in turn, writes of a “new species of coercion” which she calls “instrumentarian power” (2019, 360). As Zuboff puts it, it is a new kind of power aiming at the “instrumentalization of behavior for the purposes of modification, prediction, monetization and control” (Zuboff 2019, 352). Without going into the details of the taxonomies of these theorists, I believe there is a shared intuition that even within forms of algorithmic power, there is a unique form of power which is focused solely on “prediction” and “instrumentalization of behavior” – which, it seems to me, is still largely up for conceptual grabs.

In using the terminology of nudging to describe one of the crucial forms of power exercised through data-driven algorithmic systems and the statuses they bestow through constructing prediction-based MAKs, I am inspired by the insight of Karen Yeung (2017) that the theory of nudges serves as a useful conceptual lens through which to understand this unique form of algorithmic power. The form algorithmic nudging power exercised through data-driven algorithmic systems – which Yeung refers to as a “hypernudge” (2017) – adapts to one’s behaviour and immediately creates an environment where items are highlighted based on one’s predicted preferences as in the cases of social media feed curating algorithms or algorithmic recommendation systems of various online markets or streaming platforms.

To clarify the similarities and differences between algorithmic forms of deontic and telic powers and algorithmic forms of nudging power, I find it helpful to refer, once again, to Deleuze’s discussion of new forms of power enabled by digital technological development. He seems to identify two key elements of a form of power he calls “control”.

First, as we have seen above, the property he calls “modulation”; control, he argues, is a form of power which is “a self-deforming cast that will continuously change from one moment to the other” (1992, 4). As I outlined in Chapter 2, I understand this to mean, in our terms, either that at lower levels of modulation, one’s membership in the given kind or, at higher levels of modulation, that the membership conditions and/or the associated statuses themselves of the given kind are in constant change. This is in line with our understanding of MAKs and deontic and telic forms of algorithmic power such as the above-mentioned examples of forms of performance metrics, social scoring, risk assessment, or content moderation. It seems, then, that all forms of algorithmic power made possible by data-driven algorithmic systems rely on some form of *modulation* driven by constant data collection and its algorithmic analysis in order to assign a social status to individuals and objects.

The second property he associates with control is that it does not “confine” its subjects (2006, 322). This seems to me to be in line with what I have so far referred to as a form of power without coercing or penalizing sub-jects if they act differently. In one of his talks, used as a motto for this chapter, he says,

Control is not discipline. You do not confine people with a highway. But by making highways, you multiply the means of control. I am not saying this is the only aim of highways, but people can travel infinitely and “freely” without being confined while being perfectly controlled. That is our future. (Deleuze 2006, 322)

It seems to me, then, the unique feature of algorithmic form of nudging power is that it combines “modulation” with a lack of “confinement”. It is, then, a predictive and modulatory form of nudging. Thus, while forms of algorithmic nudging power also endow one with a status associated with a MAK, these statuses are not enabling deontic or telic powers. They are neither about rights and obligations, nor about measuring how well someone or something lives up to a norm or ideal. It is, rather, about steering behaviour in a “subtle” but predictable manner without confining or coercing its sub-jects, constantly updating the relevant predictions based on incoming data.

Let’s now go back to some of the most tangible examples of nudging exercised through data-driven algorithmic systems: the profiling activity of social media platforms or the recommendation systems of streaming platforms. The engineers of these systems aim to keep their sub-jects “engaged”, that is, spend as much time as possible on the platform through providing them with “personalized” content. Algorithms analyze data gathered about user’s behavior, both data they have shared as well as the various traces – “digital footprints” or “digital breadcrumbs” – they leave behind through interacting with the platform. Instead of attributing rights and responsibilities or measuring how well users live up to a norm, statuses endowed by nudging power exercised through data-based algorithmic systems, for example, determine the kind of content which becomes more visible for one.

This form of power, then, is not about determining whether someone has the right to view a kind of content or not. Sub-jects are not endowed the right or obligation to view, say, a certain movie on Netflix if it is recommended to them; it might just take more time and energy to view other kinds of content. Sub-jects are also not penalized for viewing these recommended movies. They are not being measured up against a norm or ideal. Rouvroy makes a similar observation when she talks about an “anormative” form of power she calls “algorithmic governmentality” which is not based on a “general norm” but on “a system of eminently evolving relations between various measurements” (Rouvroy–Berns 2013, iv, x). Sub-jects, again, are free to watch whatever movie they please; free to roam the “highways” of social media or streaming platforms without the confinements of deontic or telic powers. The emphasis is on sub-jects being tracked and their behaviour being predictably steered through constantly adjusted nudges in the form of the online content they encounter.

While nudging power is also about getting sub-jects to do something the powerholder wants them to do, there is another important difference to highlight between algorithmic forms of deontic and telic powers and some forms of algorithmic nudging power. It is not just that

forms of algorithmic nudging power lack confinements as defined above – sub-jects are often not coerced or penalized if they do not act in alignment with what the powerholder wants – but the very goals of (at least some of) the powerholders can also be less strictly defined.

Of course, those using data-driven algorithmic systems for posting their advertisement indeed have a specific goal to get their sub-jects to click on, or view, their ads. However, the goal of powerholders, such as the engineers of the given data-driven algorithmic systems, are often *more general*: the goal is no longer for sub-jects to engage in a specific behaviour, merely to steer them in *predictable* ways. Social media algorithms and recommendation systems are designed, as we have seen, to engage users, however – as opposed to, say, a social credit system with an encoded norm of “trustworthiness” or a teaching excellence framework with an encoded ideal of “the ideal teacher” – it is often the case that there is no *specific* behaviour how this engagement should be channeled.

While the distinction between the powers where the powerholders have more and less specific goals might not always be clear or easy to draw, I believe a rough distinction will do. There indeed seems to be an intuitive difference in terms of specificity of the goals of algorithmic forms of telic power with more specific goals used as a standard for evaluation such as a social credit system or an algorithmic teaching excellence system, and ones where there is no specific predefined goal beyond maximizing user engagement.

As we have seen earlier, while deontic and telic powers can be exercised through a content moderating algorithmic system or an algorithmic social credit system even if people do not recognize its existence, their power may be enhanced by getting their subjects to intentionally comply – either willingly or unwillingly. Given the lack of specificity in the goals or wants of the powerholder, however, forms of algorithmic nudging power are not only not dependent upon collective recognition but – in contrast to algorithmic deontic and telic power – is *a form of power which is not even enhanced by collective recognition*.

This new subject – as a sub-ject – is predictable, not intentionally compliant. Unpacking this idea leads us to the last chapter.

Concluding remarks: Towards a subjectless power

In this chapter, I have delineated a minimal conception of social power and have unpacked two versions of what I called the Intentionality Constraint. I have argued that social power is a capacity of the powerholder to have an effect over the actions of another subject or other subjects – who thereby become sub-jects.

The First Intentionality Constraint established that the exercise of social power requires intentional action on the part of the powerholder in the sense of getting another subject or other subjects to do what the powerholder wants them to do. As I have shown, the relevant intentional states of the powerholder may be unconscious and they may have unintended consequences, both side effects and perverse results.

The Second Intentionality Constraint, however, which holds that the existence of social power depends on collective recognition, was shown not to be true of all forms of social power. Even if we accept, for the sake of argument, that all forms of social power depend on the existence of social statuses, it still does not follow that the existence of all forms of social power depend on collective recognition, as our preceding analysis in Chapter 2 has shown that in the algorithmic society, there are social kinds and their associated social statuses which exist without collective recognition.

I have then discussed three forms of algorithmic power: deontic, telic and nudging. I have shown that none of them are dependent on collective recognition and all of them are “modulatory” as they rely on constant data collection and its algorithmic analysis bestowing a social status to subjects and objects. Forms of algorithmic nudging power, however, are unique in the sense that they are exercised without their sub-jects facing confinements – coercion or penalty if they act not in alignment with how the powerholder wants them to – and often (from the perspective of certain powerholders) without specific goals.

Sub-jects of algorithmic nudging power, then, are often free to engage in a wide array of behaviour; however, they are always tracked and their behaviour is being steered through constantly adjusted nudges. Finally, I have highlighted an important implication of the lack of confinement and lack of specific goals in certain algorithmic forms of nudging power: that these powers are not enhanced by collective recognition. Sub-jects, here, are expected to be predictable, not intentionally compliant.

The fact that that certain forms of algorithmic power are not enhanced by shaping sub-jects’ intentional states – whether it is recognition, awareness, belief, emotions or desires – has recently led social theorists to pronounce those forms of algorithmic power are not linked to, using the Foucauldian term, “the production of subjects”. Most famously, Antoinette Rouvroy has argued that “algorithmic governmentality produces no subjectification” (Rouvroy –Berns 2013, x). How are we to understand the metaphor of “subject production”? What kind of intentional states need to be produced in cases of “subject production”, and what kind of intentional states are, allegedly, not produced when exercising certain forms of algorithmic power?

These questions guide the next chapter. I will provide an analysis of the sense in which power can be “productive” as well as the sense in which certain algorithmic forms of power can be uniquely “non-productive” as their sub-jects lack certain intentional states.

CHAPTER 4: ALGORITHMIC POWER WITHOUT SUBJECT-PRODUCTION OR INTIMIDATION: THREE NOTIONS OF REFLECTION

*What kind of a power is it that wants to take my honour and my pride from me,
and does it in such a meaningless way?*

Kierkegaard: *Repetition* (2009, 61)

Introduction: Foucault, subject production and algorithmic power

In this chapter, I will examine forms of algorithmic power through Foucault's concept of "productive power". As we have seen in the previous chapter, on Foucault's view, some forms of power do more than just intimidate: they are "productive" – with several products. Foucault's examples of such products, combined with our terminology set out so far, include the *pleasure* sub-jects derive from complying with the given rules or norms, the *knowledge* powerholders gain through surveillance, data gathering, and classifications of subjects and objects; as well as the central theme of this chapter, the sub-jects *as subjects* themselves. Inspiring countless works – including social ontology classics such as Ian Hacking's works on "making up people" (1995, 1997) – Foucault famously argued that a crucial form of social power, which he called "productive power" lies in changing people's beliefs, emotions and desires constituting their sense of self. While various products of "productive power" are closely intertwined, the specific question this chapter addresses focuses on subjects and goes as follows: *in what sense do forms of algorithmic power produce – or do not produce – subjects?*

A brief note for clarity. We have so far defined "subjects" as conscious beings with minds capable of intentionality and a sense of self. In the context of social power and power relations, we have also distinguished between subjects as powerholders and subjects over whom some form of power is exercised, that is, sub-jects. Now, as I will explain further below, Foucault's idea regarding the "production of subjects", expressed in our terms, is that the sub-jects of the exercise of "productive" power are "produced" *as subjects*, meaning their sense of self is being changed. Thus, the expression of "the production of subjects", in our terms, should be understood as the production of sub-jects as subjects. With this in mind, I will keep using Foucault's expressions of "the production of subjects" or "subject production" and will only specify the distinction between the emphasizing *subjects as sub-jects* (the property of being subjected to a form of power, or "*sub-jectivity*") and *sub-jects as subjects* (the property of having minds capable of intentionality and a sense of self, or "*subjectivity*"), when necessary.

Many social theorists argue that with algorithmic power comes a crucial change in how power may produce subjects (e.g., Coeckelbergh 2022; Fisher 2022; Petersmann–Van Den Meerssche, 2024). Some have put forth the claim that algorithmic power, unlike productive forms of power, is not linked to the production of subjects (e.g., Rouvroy 2013; Rouvroy–Berns 2013; Lukes 2021; Fisher 2022). It is then argued that the distinctive feature to set forms of algorithmic power apart from other forms of power is a lack of certain intentional states – “lack of reflection” – on the part of its sub-ject. This intriguing argument, however, is often not sufficiently explored. Thus, I will aim to clarify both the conclusion that some forms of algorithmic power are, in some sense, not linked to the production of subjects as well as the premise that there is unique “lack of reflection” on the part of the sub-ject in the case of the exercising of algorithmic power.

The main thesis of this chapter is this: the claim circulating in contemporary social theory that there are forms of algorithmic power which are not linked to subject production is to be unpacked as the claim that exercising a specific form of algorithmic power does not require intentional action on the part of their sub-ject. I will proceed as follows. I will first unpack what I take “production” and “subject” to refer to in the Foucauldian claim that some forms of power produce subjects. Then I will explore a premise underpinning the general claim that there are forms of algorithmic power that are not linked to subject production. This premise holds that forms of algorithmic power which are not linked to subject production are *unique* as their sub-jects lack certain intentional states which are often referred to as “reflection” and are taken to be necessary for subject production. I will provide three possible ways to unpack the relevant sense of reflection, labelling these Awareness, Address, and Assessment. My argument is that, in itself, only the lack of Assessment – the intentional states required and accompanied by intentional action, whether willing or unwilling – implies a form of power whose exercise does not, as I will explain, *require* the production of subjects. Thinking through the implications of combinations of lacking senses of reflection, or even the lack of all three senses of reflection, is a task for future philosophical work.

A few caveats are in order.

First, I am not trying to do justice to the richness of Foucault’s thought or offer a contribution to Foucault-scholarship. For instance, I am not considering the genealogical analyses about the specific historical circumstances that gave rise to “productive” forms of power as opposed to other, non-productive forms of power, for instance, those we have previously labelled as “brute power” or “intimidation”. I will also not dwell on Foucault’s arguments about the increased efficiency and cost-effectiveness of productive power. Instead, I am merely extracting from Foucault’s oeuvre the relevant concepts of “production” and “subject” (as sub-ject) which have inspired numerous social theorists to theorize about social power, including forms of algorithmic power.

Second, I am not trying to answer any empirical-causal question on how exactly various forms of algorithmic power, in fact, effect the specific intentional states as the constituents of one's sense of self or one's mental capacities or mental health (e.g., does being subjected to infinite scroll- and notification-based systems lead to shorter attention span? How do content recommendation systems affect cognitive development and critical reasoning skills? Does algorithmic content-filtering make one less willing to engage in dialogue with opposing viewpoints? etc.). Instead, I am asking a conceptual question about the relevant sense of the kind of intentional state – “reflection” – whose lack would imply that the form of power at hand is – as we will see, uniquely – non-productive.

Third, within the scope of this chapter, I cannot do justice to the richness of the historically and culturally diverse concepts of subjectivity. I do not claim that the notion of “subject” and “subjectivity” I relied upon so far, and which I will use to unpack the claim that algorithmic power is not linked to subject production, is historically or culturally neutral. It may well turn out to be close to what social theorists in the relevant literature on algorithmic power label “modernist notions of liberal subjectivity” (Petersmann, Van Den Meerssche 2024) or the “Enlightenment project of subjectivity” (Fisher 2022, 12). I will refrain from making such historical claims.

Finally, as I outlined in the *Introduction*, I will also not make normative claims about whether algorithmic forms of social power are to be “embraced” or not. For some in the literature, “the algorithmic reconfiguration of sociality” is an opportunity to move away from thinking about subjectivity based on a “liberal lexicon” (Petersmann, Van Den Meerssche 2024). Others mourn the fact that forms of algorithmic power leads to the unfulfillment of a “promise born in the Enlightenment” (Fisher 2022, 11). I will remain neutral on the normative implications of the emergence of various forms of algorithmic power.

With these limitations in mind, let us now explore the relationship between forms of algorithmic power and the “production of subjects”.

4.1 In what sense does power produce subjects?

Let us first unpack the famous Foucauldian claim that there is a form of power whose exercise produces subjects. Borrowing his term, I will refer to the form of power which produces subjects as “productive power”. In Chapter 3, we have defined social power as *the capacity to get sub-jects to do what one wants them to do*.²⁰ We have considered several ways in which

²⁰ This seems to be compatible with Foucault's rather abstract definition of power, according to which exercising power is “a way of acting upon one or more acting subjects by virtue of their acting or being capable of action. A set of actions upon other actions” (2000, 341).

the concept can be broadened (e.g., remaining neutral on whether the relevant needs to be exercised to elicit the desired behaviour on the part of the sub-ject, whether the given power relation is characterized by some form of conflict, and whether there is awareness on the part of the powerholder and/or the sub-ject). We have also seen that the relevant intentions of the powerholder can be unconscious and can have unintended consequences. How, then, does the holder of *productive power* get its sub-jects to do what they want them to do?

Before unpacking productive power in detail, here are a few preliminary ideas. Foucault shows that “[t]he exercise of [productive] power can produce as much *acceptance* as may be wished for [...] it incites, it induces, it *seduces*” (2000, 341, my emphases). How exactly does this form of power produce “acceptance”? How does it “seduce”? The key, according to Foucault, is to understand “the different modes by which, in our culture, human beings are made subjects” (2000, 326). Here is an initial idea of productive power:

Productive Power (initial): A form of power whose exercise produces subjects.

Foucault’s famous examples of productive power are taken from “institutions that would enclose individuals” (1982, 57). Along these lines, take the “production” of the obedient student or the performance-driven employee. To see how exactly these subjects are produced, let’s first clarify the relevant concepts of “produce” and “subject”. In the following excerpt, Foucault contrasts productive power with “repression” or “prohibition” which merely “says no” (1980, 119).

In defining the effects of power as repression, one adopts a purely juridical conception of such power, one identifies power with a law which says no, power is taken above all as carrying the force of a prohibition. Now I believe that this is a wholly negative, narrow, skeletal conception of power, one which has been curiously widespread. If power were never anything but repressive, if it never did anything but to say no, do you really think one would be brought to obey it? What makes power hold good, what makes it accepted, is simply the fact that it doesn’t only weigh on us as a force that says no, but that it traverses and produces things... (Foucault 1980, 119, cf. 1995, 23)

For Foucault, power as “a law which says no” prohibits something which people otherwise would do; in these cases, the ability to elicit fear of punishment is sufficient to ensure compliance. From the forms of power identified in Chapter 3, this is in line with cases of *intimidation* where power may not be directly exercised but there is conflict at least on the level of intentional states. In the case of the obedient student or the performance-driven employee, the relevant intentional states could be, say, being afraid of the negative consequences of bad grades or a bad performance report which would drive these subjects to comply. Intimidation, then, indeed changes the intentional states – emotions, beliefs, desires – of its subjects. However, and this is key, this change of intentional states does not

necessarily entail *willing* compliance – merely securing compliance of some form, even if it is unwilling.

Productive power, however, is a form of power which is directly exercised and aims to change its subjects in more fundamental ways than simply intimidating them. In the case of productive power, the obedient student and the performance-driven employee are not intimidated into compliance but comply *willingly*. The relevant form of willing compliance, then, is not merely achieved through alluding to the *prudential* considerations of the given subject, but via an *internalized* willingness – which Foucault above refers to as “traversing and producing” subjects. I will, then, unpack “traversing” below as follows: productive power changes its subjects’ intentional states to change their sense of self to avoid conflict also on the level of intentional states and thereby elicit internalized willing compliance. This is, in brief, how “acceptance” and “seduction” is achieved in cases of exercising productive power.

But, first, a few words on Foucault’s notion of “subject”.

*

What does Foucault mean by “subject”? Let’s consider the following excerpt from one of his later works, titled “The Subject and Power”:

There are two meanings of the word ‘subject’: subject to someone else by control and dependence; and tied to his own identity by a conscience or self-knowledge. Both meanings suggest a form of power which subjugates and makes subject to. (Foucault 2000, 331)

It seems to me that we can extract the following propositions from this excerpt:

- (1) “being a subject” means having a conscience or self-knowledge of one’s own identity;
- (2) “being a subject” means being subjected to someone;
- (3) having a conscience or self-knowledge of one’s own identity means being subjected to one’s conscience or self-knowledge of one’s own identity.

My definition of subject, built on the prevalent binary ontology of subjects and objects assumed throughout much of the humanistic tradition of Western philosophy, defined subjects as those entities having minds, intentionality and a sense of self. As such, my definition seems to be in line with proposition (1); taking “a conscience or self-knowledge of one’s own identity” to be what I have called a “sense of self”. Thus, I will stick with my definition for the terms “subject” and “subjectivity”. Proposition (2) seems to be a relatively uncontroversial use of the term “subject” and one which aligns with its original etymology. However, to avoid confusion and equivocation, I have introduced the term “sub-ject” to refer

to the subset of subjects, those who are “subjected to someone”, or, in our terms, those subjects who are in a power relationship where power is exercised over them.

The Foucauldian twist lies in proposition (3). While the humanistic tradition of Western philosophy tends to hold that being a subject, having mind, intentionality, and, in particular, having a sense of self, implies some form of unique freedom, for Foucault, being a subject, having a sense of self is a unique site for exercising power, and as such, implies *being a subject to oneself*. As such, Foucault points out an important tension stemming from whether one emphasizes the freedom or the self-subjugation which having a sense of self implies. The main reason he emphasizes the latter, it seems to me, is because he holds that the sense of self of sub-jects (as subjects) is “produced” by others. While this might be less surprising for sociologists and social scientists, it is indeed a provocative statement in light of the conception of subjectivity in the humanistic tradition of Western philosophy.

However, philosophers could, and did, criticize Foucault’s notion of subjectivity – one which readily implies sub-jectivity – from several angles. It could be argued that philosophically relevant distinctions are to be drawn along the following questions. Does indeed *every* sense of self imply being “subjugated” to one’s sense of self? Are only those senses of selves subjugating which are produced by others than the relevant subject themselves? If so, is every sense of self “produced” by others than the relevant subject themselves equally subjugating? Say, even if one’s sense of self is always produced by others, the production of the given sense of self might be result of manipulation or, by contrast, that of empowerment, and there might also be rather “innocent” influences on one’s self-constituting beliefs and desires (cf. Taylor 1984).

Some, in turn, may be less interested in who has played a role in the causal history of the production of the given sense of self, and are more focused on the nature of the given sense of self. Existentialists, for instance, have famously argued that there is a kind of sense of self which is “authentic” as it “will[s] freedom within himself” (Beauvoir 1948). Take Sartre’s famous example of the Parisian waiter (2003, 82–83) whose sense of self seems to be wholly constituted by his sense of “being a waiter” or Simone de Beauvoir’s “serious man” (Beauvoir 1948) who blindly subjugates themselves to the beliefs, emotions and desires constituting their sense of self; both authors contrast these examples with the person of “authenticity” whose sense of self is about believing themselves to be free.

While I cannot elaborate in detail on these questions here, I believe they are getting at important distinctions between how one’s sense of self can be subjugating – distinctions which a charitable reading of Foucault should employ in interpreting his claims. Accordingly, in the following, I will focus broadly on cases where the relevant subject’s sense of self is significantly produced by others, and where this sense of self has more concrete content than merely the “authentic” belief in one’s freedom.

*

Let's, now, unpack the relevant notion of "production" to further clarify the Foucauldian picture of productive power. When explaining how the subject's sense of self is produced, Foucault provides several descriptions of the relevant sense of "production". In the longer quote above, he gave a metaphorical but helpful description of production as "traverse". By "traverse", I take Foucault to mean that power does not simply remain an external force intimidating its sub-ject "from above" but, in some sense, becomes *part of* them. In another text, he states that "subjects are gradually, progressively really and materially constituted through a multiplicity of organisms, forces, energies, materials, desires, thoughts, etc." (1980, 97). Foucault also refers to the "capillary form of existence" of productive power, that is, "the point where power reaches into the very grain of individuals, touches their bodies and inserts itself into their actions and attitudes, their discourses, learning processes and everyday lives" (1980, 39).

Now, to clarify these metaphors a bit, given our concept of a subjectivity, we can take this "traversing" of individuals to refer to changing the intentional states which constitute their sense of self. One's sense of self is constituted by *some of* one's intentional states, such as beliefs, emotions, and desires. The distinction between just any intentional state one may have and ones which constitute one's sense of self is a difficult one. Here, it suffices to say that the latter are those which one, in some sense, *internalized*; I will have more to say on internalization in the next sections of this chapter. Assuming, then, that internalized emotions, beliefs and desires constitute one's sense of self, changing these brings about a change in one's sense of self.

Now, as we have seen in the previous chapter, one's sense of self may drive their intentional actions – whether consciously or not. More precisely, then, in the context of productive power, producing subjects by "traversing" them means that the relevant exercise of power is changing the internalized emotions, beliefs and desires which constitute one's sense of self, which, in turn, may change one's intentional actions.

To illustrate, let's return to the obedient student and the performance-driven employee. As we have seen, the intentional states which constitute one's sense of self are the ones they have internalized. Productive power produces intentional states such as their believing or desiring to be – more or less consciously – someone aspiring to be a good student or a good employee. These internalized intentional states lead them to internalized willing compliance: to willingly follow the prescribed curricula or strive to achieve performance metrics.

Furthermore, as we have already noted, subject production does not need to change the relevant intentional states in the sub-jects in a way that the sub-ject is *aware* of the process.

Requiring these would rule out paradigmatic examples of subject production; indeed, even obedient students or performance-driven employees may not be aware of the relevant beliefs and desires that they have internalized. In cases of productive power, then, such as in the case of the obedient student or the performance-driven employee, one's sense of self is changed such that the power relation becomes, as Foucault puts it, "accepted" and "seductive".

The point, in sum, is changing the subject's intentional states which constitute their sense of self, *to avoid conflict also on the level of intentional states so as to elicit internalized willing compliance.*

From what we have learnt thus far, we can already draw a distinction between intimidation and our definition of productive power so far as follows. The power relation of intimidation changes some of the sub-jects intentional states in order to avoid conflict on the level of behaviour. In other words, intimidation is about getting subjects to do what one wants them to do through changing some of their intentional states – e.g., eliciting fear – but not necessarily those which constitute their sense of self. In these cases, sub-jects need to comply either willingly or unwillingly – regardless of how they think of themselves.

Productive power, by contrast, seems to change those intentional states which constitute the sub-jects' sense of self in a way to avoid conflict not only on the level of behaviour but also on the level of intentional states. In these cases, it is not enough to only specify that the given power relationship changes those intentional states which constitute the sub-jects sense of self. This is because, unintentionally, the exercise of several forms of power may lead to resistance on the level of intentional states, that is, change in the sub-jects' sense of self such that sub-jects may begin to harbour hostility against those exercising power over them. Accordingly, once again, productive power is the form of power which, as Foucault puts it, becomes "accepted" and "seduces", that is, changes one's intentional states which constitute one's sense of self to avoid conflict not only on the level of behaviour but also on the level of intentional states.

However, we need to further refine the definition of productive power. This is because it may also be possible that intimidation leads not merely to compliance driven by fear but also, maybe after a while, willing compliance of the sort driven not only, or perhaps not at all, by prudential considerations but by internalized intentional states. Sub-jects of intimidation indeed sometimes give up all resistance even on the level of intentional states and comply willingly, driven by their internalized intentional states.

Even in these cases, however, there is an important conceptual difference between intimidation and productive forms of power. This is because productive power, or, more precisely, the successful exercise thereof, *requires* the production of willing subjects driven

by internalized intentional states. Intimidation simply requires compliance – regardless of where it happens to fall on the spectrum of willing and unwilling forms of compliance. Thus, the difference between intimidation and productive forms of power is not that one may produce subjects while the other never does. Instead, the difference is that while intimidation may lead to subject production, the exercise of intimidation does not *require* the production of subjects. On the other hand, the exercise of productive forms of power requires the production of willing subjects driven by internalized intentional states.

Accordingly, we need to update our definition.

PRODUCTIVE POWER (updated): A form of power whose successful exercise *requires* producing willing subjects driven by internalized intentional states.

*

Now, with the definition yielded by the analysis of the notion of productive power, let us revisit the taxonomies of social power we have outlined in the previous chapter. Specifically, I have developed two taxonomies of power so far. First, one which addressed *the ways in which power can be hidden*. This taxonomy was a result of having asked three questions: whether power must always be exercised to exist, whether power necessarily involves some form of conflict between subjects and whether the relevant subjects are aware of the power relation at hand. The second taxonomy I outlined – inspired by Burman’s work on power (2007, 2023) – addressed *the types of statuses which social power can construct and depend on*.

How does our notion of productive power fit into these taxonomies?

Let’s consider, first, the three questions which yielded the taxonomy of hidden forms of power in relation to productive power. First, as we have seen, productive power is a form of power which gets other subjects to do what one wants them to do by being exercised. Thus, productive power is not hidden in the sense that it would not require the relevant powers to be exercised. What makes productive power hidden, then, becomes evident when we consider how it fares with the last two questions. On the one hand, as we have seen, productive power aims to avoid conflict, not only on the level of behaviour but also on the level of intentional states. On the other, productive power is a form of power about which both the powerholder and the sub-ject are often unaware as both its exercise and compliance often happen unconsciously.

Moving on to placing productive power in the second taxonomy – the one addressing the types of statuses which social power can construct and depend on – leads to the following insights. As I will explain more in detail below, forms of power which construct or depend on

any type of statuses can be productive. That is to say, deontic, telic and nudging forms of power – both algorithmic and traditional forms thereof – can equally be productive. This is because the sub-ject of any of these forms of power may come to have internalized intentional states about what the powerholder wants them to do in a way as to comply willingly. It follows, then, that the types of status the given social power can construct and depend on is independent from whether it is productive or not.

I propose, therefore, that productive power as well as its contrast case considered so far, intimidation, may (also) be put into a third type of taxonomy. This is a taxonomy addressing *the ways in which a given form of power relates to the intentional states of its sub-jects*. As we have seen, productive power changes its sub-ject's intentional states which constitute their sense of self to avoid conflict not only on the level of behaviour but also on the level of intentional states. Intimidation, in turn, changes some of its sub-jects intentional states – e.g., eliciting fear – but not necessarily those which constitute their sense of self. However, the successful exercise of both productive power and intimidation depends on some form of *intentional* action of compliance – whether internalized or otherwise. (Once again, I consider *none* of these taxonomies exhaustive and welcome any additions.)

In the next section, I will consider how forms of algorithmic power fit into this third taxonomy. I will do so by examining a claim made, in various forms, in the contemporary social scientific literature about sub-jects of algorithmic power lacking the intentional state of “reflection” which, allegedly, shows that algorithmic power and productive power are distinct. Given that algorithmic power and productive power are not mutually exclusive, our task, more precisely, will be to pinpoint the precise form of algorithmic power which is different from productive power, and, as I will show, from intimidation. I will then identify both the relevant sense of the lacking reflection on the part of the sub-ject of as well as the specific form of algorithmic power which is distinct from productive power as well as intimidation.

This specific form of algorithmic power, then, will be an addition to our third taxonomy of power which answers the question of how power can relate to its sub-jects intentional states. Preliminarily, this form of power will be a subset of what I have earlier called algorithmic nudging power, and I will call it *behavioural* nudging to emphasize that the relevant compliance is achieved through changing one's *behaviour* without necessarily driving their *intentional actions*. Thus, in the remainder of this chapter, I am not focusing on the nature of the statuses endowed onto sub-jects through algorithmic nudging power but on how certain cases of algorithmic nudging power relate to the intentional states of their sub-jects.

4.2 In what sense does algorithmic power not produce subjects?

There is, as we have seen, an unfolding discourse in media studies and social theory as well as social philosophy, which aims to describe the growing influence of quantifying and classifying data-driven algorithmic systems on our social life – and on our subjectivity. For instance, media scholar Eran Fisher argues that “algorithms create a new way of knowing, which, in turn, changes our fundamental sense of self and our concept of subjectivity” (Fisher 2022, 1); sociologist Rogers Brubaker suggests that we are experiencing an “entirely new techno-social infrastructure of selfhood, an entirely new ecology within which selves are formed and reformed” (2020, 771–773, 779); legal scholars Marie Petersmann and Dimitri Van Den Meerssche claim that “with the rise of ‘algorithmic governmentality’ (...) practices of classification and subjectivation have taken novel forms” (Petersmann–Van Den Meerssche, 2024; cf. Coeckelbergh 2022, 159).

More interesting perhaps are the interpretations of the specific nature of this historical shift towards an algorithmic society and the new forms of power which thus emerge. To explain the exact effect of algorithmic power over its subjects, some put forth the more provocative claim that, as opposed to forms of productive power, algorithmic power – in some sense, at least – is not linked to subject production. I will call this claim “Subjectless”; here is an initial definition:

SUBJECTLESS initial: As opposed to productive power, (some) forms of algorithmic power are not linked to subject production.

The relevant sense of “not being linked to subject production”, however, allows for several interpretations of Subjectless, and these are often not clearly differentiated.

In some cases, Subjectless seems to refer to the negation of the *initial* definition of productive power, that is, algorithmic power *does not* produce subjects. Perhaps most famously, Antoinette Rouvroy have argued that “algorithmic governmentality produces no subjectification”, that it “does not produce any kind of subject” (Rouvroy–Berns, 2013, x; Rouvroy 2013, 153). Some scholars, however, go a step further. They discuss the “dissolution”, “displacement”, “subversion”, “undermining” of subjectivity (in this order: Petersmann- Van Den Meerssche, 2024, 15; Fisher 2022, 2). These formulations suggest a reading of Subjectless according to which algorithmic power has the reverse causal effect compared to productive power as it *actively hinders* subject production. Along these lines, some go as far as to propose that the effects of algorithmic power do not only “weaken,” but “at the extreme, disable” subjectivity (Lukes 2021, 181, 186), that forms of algorithmic power “turn, them [subjects] into objects” (Fisher 2022, 12); that they “make us into machines” (Bowker, 2020, ix). Thus, versions of Subjectless range from simply denying the production of subjects in the case of some forms of algorithmic power, through claiming that some forms

of algorithmic power actively hinder the process of subject production to the idea that some forms of algorithmic power disable the subjectivity of its sub-jects.

Evaluating these claims would, first, require unpacking what “non-production”, “hindering” and “disabling” exactly means in this context. Furthermore, even after clarifying these concepts, answering these questions would be a matter of empirical research: we would need to see how exactly various forms of algorithmic power, in fact, affect one’s intentional states – and mental capacities in general. Therefore, while I believe these are intriguing empirical questions, I will, as mentioned earlier, bracket them as they go beyond the scope and methods of this chapter.

I will, then, focus on unpacking Subjectless as the negation of the updated definition of productive power. In other words, I take Subjectless to hold that the exercise of forms of algorithmic power, or at least some forms thereof, *does not require* subject production. This version seems to be what the following formulations refer to. Some argue that for algorithmic power, “human subject is superfluous” (Coeckelbergh 2022, 116); that it “leaves subjectivity redundant”, “requires no subjects at all”, or “sidestep(s) the role of subjectivity” (Fisher 2022, 2, 3, 12); is characterized by a “radical indifference” (Zuboff 2019) to one’s sense of self and thus by “the avoidance of subjects” (Rouvroy–Berns, 2013, xx). Here, again, the claim is not that algorithmic power does not produce, hinders or disables subjectivity but that some forms of algorithmic power can be exercised without subject production happening in the relevant sense.²¹

As we have seen, intimidation also does not require the production of subjects. This, then, raises the following question. In case some forms of algorithmic power are indeed non-productive forms of power, what, if anything, distinguishes these forms of algorithmic power *uniquely* – where uniqueness is understood as being different from intimidation –from productive power? I believe Subjectless should provide an answer to this question; let us, then, update Subjectless based on these considerations.

SUBJECTLESS_{updated}: The (successful) exercise of (some) forms of algorithmic power does not require subject production in a unique way.

²¹ Rouvroy herself can be interpreted as arguing for any of the above-outlined versions of the Subjectless. At one point she argues that people are “governed by preventing or at least complicating the very possibility of subjectification processes”, and usually alternates between “the idea that algorithmic power “prevents” or “complicates” becoming a subject. (2013a). While sometimes she argues that the subject of algorithmic power “does not manage to become a subject” (2013a xx), in other places, she shows that there is a “problem [...] which is] more the result of a rarefaction of subjectification processes and opportunities, of a difficulty to become subjects, than the product of a “de-subjectification””, (2013a, xvi). However, she also talks about “the avoidance of the subject”.

For the sake of clarity, by Subjectless, I will refer to this updated version unless I note otherwise. My aim below, then, is to clarify the distinction between productive power, intimidation, and (some) forms of algorithmic power.

*

There seems to be a consensus that Subjectless (regardless of the specific version the given author may have in mind) is underpinned by some form of a *lack of a specific kind of intentional states which characterizes the sub-jects of algorithmic power*. Specifically, several authors employ the term “reflection” as well as related terms like “reflexivity”, “self-reflection”, “reflective relation/practice” etc., to refer the lacking intentional state. For example, some argue that algorithmic power “undermines a key process in the constructing of subjectivity: self-reflection” (Fisher 2022, 2); or that “reflective relation to oneself is missing, with no space to be with oneself” (Coeckelbergh 2022 116; cf. Couldry-Mejias 2019, 171); while others contrast “reflexive practices” of subjectivization with “data- and algorithm-driven processes” (Brubaker 2020, 771, 773). How to unpack, then, the relevant notion of *reflection*?

As I will show in the sections that follows, in the extant literature, the term “reflection” encompasses many concepts. Given that the meaning of reflection is often left unanalyzed, these concepts can, and often do, get conflated. In fact, the following two lengthier excerpts from Rouvroy seem to me to contain (at least) *three* concepts of reflection – the lack of which is said to characterize sub-jects of (some) forms of algorithmic power.

The algorithmic government thus contrasts with what we know about a neoliberal mode of government which produces the subjects it needs. Through the ubiquitous injunction – and its internalization by subjects - of maximization of performance (production) and enjoyment (consumption), neoliberalism produces ‘hyper-subjects’ having, as their normative horizon, the continuously reiterated project of ‘becoming themselves’, and passionately engaged in ‘self-control’, ‘self-entrepreneurship’, ‘self-evaluation’. Algorithmic governmentality does not produce any kind of subject. It affects, without addressing them, people in all situations of possible criminality, fraud, deception and consumption, which are situations where they are not requested to ‘produce’ anything, and certainly not subjectivation. Rather, algorithmic governmentality *bypasses consciousness and reflexivity*, and operates on the mode of alerts and reflexes. (Rouvroy 2013, 153, cf. 144–145, my emphasis)

Algorithmic governmentality produces no subjectification, it *circumvents and avoids reflexive human subjects*, feeding on infra-individual data which are meaningless on their own, to build supra-individual models of behaviours or profiles without ever involving the individual, and without ever asking them to themselves describe what

they are or what they could become. *The moment of reflexivity, critique and recalcitrance necessary for subjectification to form seems to constantly become more complicated or to be postponed*” (Rouvroy–Berns, 2013, x, my emphasis)

The first sense of reflection I will identify is based on the idea I take to be expressed in the phrase, “the moment of reflexivity, critique and recalcitrance [... is] postponed”. In our terms, I will take this sense of reflection to refer to the intentional state of the subject’s awareness or critical awareness of the existence and workings of the power relation. I will call it Awareness.

The second sense of reflection echoes the idea I take to be implicit in the phrases, “affects, without addressing them”, “without ever asking them”. I take this second sense of reflection, in our terms, to refer to the intentional state for which stems from the subject’s being addressed as a subject, by another subject. I will call it Address.

The third sense of reflection is derived from what I take to be the idea behind phrases such as “operates on the mode of alerts and reflexes”). These phrases, I take it, refer, *as a contrast*, to what I have called the intentional states which are needed for, and accompanied by, an intentional action. I will call these intentional states Assessment.

The three senses of reflection – three kinds of intentional states – I will identify, then, are following: reflection as Awareness, reflection as Address, and reflection as Assessment. Depending, then, on the specific notion of reflection at issue in the argument that Subjectless stems from a lack of reflection, the resulting conclusion can differ in crucial ways.

In the following, I will unpack each meaning of reflection in detail. In each case, my analysis will culminate in posing the following question: does the lack of the given kind of reflection imply Subjectless? To see if the lack of a given kind of reflection implies that the given algorithmic form of power is *uniquely* non-productive (i.e. in contrasting it to forms of productive power as well as intimidation), I will assess if the given kind of reflection (which is said to be lacking in certain forms of algorithmic power) is indeed required for productive power and, if so, whether it is also required for intimidation as a non-productive power.

4.2.1 Reflection as Awareness

Let us begin with reflection as Awareness. In part I, regarding the existence of social kinds, I talked about the intentional state of “recognition”, a form of intentional state representing something as existing, whether consciously or not. Then, in Chapter 3, to pinpoint a specific kind of recognition which involves *conscious* intentional states representing something as existing – in our case, a power relation – I have used the term *awareness*. So, what does this capitalized version, Awareness, refer to?

Following Rouvroy's point that "the moment of reflexivity, *critique* and *recalcitrance* [are] necessary for subjectification to form" (my emphases), I will first take Awareness to refer to a form of *critical* awareness (on the part of the sub-ject) of a power relationship. The idea, for the sake of argument, is that there is a form of reflection which stems from such critical awareness which is necessary for the exercise of productive power. This critical awareness, then, is taken to be missing in (some) forms of algorithmic power. Now, the sub-ject may be critical of the power relationship or show recalcitrance because they are aware of the existence of the power relationship at hand *and* is aware of the workings of the power relationship at hand, or one may be critical of the power relationship and show recalcitrance because one is only aware of the existence of the power relationship but is *not* aware of the workings of the power relationship at hand.

A brief note on what amounts to being aware of the workings of a power relationship at hand: this, of course, is a difficult question. While I will not provide a deeper analysis, I note that we already have the conceptual resources to say a bit more on what such awareness may amount to, at least in cases of social powers which are the power to construct social statuses or those whose existence depends on social statuses. Here, using the concepts yielded by our analysis of social kinds and their mind-dependency in Part I, being aware of the workings of the relevant power relation amounts to saying that one should be aware of the relevant constraints and enablements bestowed by the statues at hand as well as the membership conditions for the given statuses.

A good literary example for being critical of the power relationship or showing recalcitrance because one is aware of the existence of the power relationship at hand *and* is aware of the workings of the power relationship at hand, is Kleist's character, Michael Kohlhaas, a Brandenburg horse dealer who has his horses seized when he aims to cross the border into Saxony. He soon discovers that this was an arbitrary exercise of power as there was no such deontic status in place which would have granted the right to seize the horses, realizing that the legal system is corrupt and biased in favor of the nobility. Thus, he becomes fully aware of the real workings of the power relation, while in an act of recalcitrance – after trying to appeal to no avail – he launches a campaign against the nobleman who illegitimately ordered the seizure of his horses.

An example of being critical of a power relationship while only being aware of its existence, but not its workings, is Kafka's Joseph K. in "The Trial." Joseph K. faces an opaque legal system, unable to understand his own status (or its membership conditions) bestowed upon him as he is also unable to understand the charges against him. Similarly, the Biblical character Job faces immense suffering as having the status of a candidate for a test of faith by God, without being aware of why. Both characters, then, are caught up in a power relation whose relevant statuses are opaque to them *and* towards which they show "recalcitrance"; thus, while Joseph

K. and Job are aware of the existence of the relevant power relation, they are not aware of its workings.

However, none of these are cases of productive power. Productive power, by definition, produces subjects showing *internalized willing* compliance, “accepting” or “being seduced by” the power relationship. Thus, simply from our definition, it follows that “critique” and “recalcitrance” are not necessary for subject production.

So, let us take Awareness to mean being aware of the existence of the power relationship at hand and being aware of the workings of the power relationship at hand – without having *critical* awareness of it, that is, without in any way viewing the given power relationship as illegitimate or showing “recalcitrance”.

So, do sub-ject’s need to be aware of the workings of the power relationship at hand in the case of forms of algorithmic power – and are they, in fact? On the one hand, as we have seen in the previous chapter, forms of algorithmic power do not require collective recognition (and, therefore, collective awareness) for their existence. On the other, given that data-driven algorithmic systems, and the Modulatory Algorithmic Kinds they construct, are analyzing vast amount of data and thus are constantly modulating (on various levels, as we have seen), being aware of the workings of the power relationship at hand can indeed be a challenge. For instance, as I showed in Part I, Google’s category of “Married”, as a MAK, has membership conditions and associated statuses which are determined by constantly incoming behavioural data and complex algorithmic codes, making it difficult for its sub-jects be fully aware of these. As Rouvroy argues, forms of algorithmic power are often characterized by a form of opacity: “our behaviours have never been so processed – observed, recorded, classified, evaluated – underpinned by codes of intelligibility and criteria that are completely opaque to human understanding” (Rouvroy–Berns 2013, xix; the same point is made by several other authors as well, cf., Couldry–Mejias 2019, 124–125; Brubaker 2020, 780; Fisher 2022, 24 etc.).

But, the question remains, is such awareness of the workings of the power relation at hand necessary for the exercise of *productive* power?

It seems to me that such awareness is not necessary. There are indeed cases of exercises of productive power where sub-jects are not aware of the workings of the power relationship at hand, all the while being aware that they are in fact in a power relationship. Consider the examples of the student and the employee again. A student may be aware that a given adult is a teacher and so has (some form of) power over them but may not be aware that they are, in fact, the head of school and so may not be aware of the workings of their actual power relation. In a similar way, a new employee at a large company may recognize that an older colleague has power over them but may not be aware of that colleague’s role in determining their promotions.

Or take an example about a form of power which, while it arguably depends on the existence of statuses, is not itself about bestowing statuses: I have in mind the example of the sub-jects of the productive power of a media outlet. These sub-jects recognize the products of the media outlet as being products of a media outlet. Now, while they may be aware that *as* the audience of media outlet, they are involved in a power relationship, they may also be unaware, or not fully aware, of the relevant workings of the power relationship, e.g., the ways in which information is selected and presented.

Consider the following excerpt from an interview with Chomsky:

I found out about the arms flow to Iran by reading transcripts of the BBC and by reading an interview somewhere with an Israeli ambassador in one city and reading something else in the Israeli press. OK, the information is there. But it's there to a fanatic. You know, somebody who wants to spend a substantial part of their time and energy exploring it and comparing today's lies with yesterday's leaks, and so on. That's a research job. And it just simply doesn't make any sense to ask the general population to dedicate themselves to this task on every issue. (Chomsky in: Achbar–Wintonick 1992)

Here, Chomsky suggests that sub-jects may be aware that they are in a power relationship when reading the media but may not have the time, resources, or expertise to be aware of relevant workings of this power relationship.

There are also examples of forms of algorithmic power which are productive *and* whose sub-jects may be aware that they are in a power relationship *but* may not be aware of the relevant workings of the given data-driven algorithmic system. Take the use of the nudging power through social media feed curating algorithms for the aims of political propaganda, aiming to produce willing subjects driven by internalized intentional states who would support the given political project. Here, sub-jects – users of the given social media – may be aware that they are in a power relationship, agree with it, but not be aware of its workings. That is to say, they may not be aware of the existence of the relevant MAKs constructed through the given algorithmic systems and thus not be aware of their associated statuses and membership conditions.

Finally, let's take Awareness to mean simply the sub-ject's being aware of being in an existing power relationship. Some excerpts from Rouvroy's argument indeed seem to point at this meaning of Awareness; for instance, when she highlights "the difficulty to produce algorithmic subjects who conceive of or think about themselves as such" (Rouvroy–Berns 2013, xvi).

It seems to me, however, that even in cases of productive power, one may not even be aware of being in a power relationship at all. In other words, there are cases of exercises of

productive power (whether algorithmic or not) without sub-jects being aware of the existence of the power relationship at hand.

Above, I have already mentioned that, in some cases, even obedient students or performance-driven employees may not be aware of the (productive) power-relation at hand. Or take again an example of an algorithmic form of power. The nudging power exercised through social media feed curating algorithms can *also* be an example of productive power without even this weaker sense of Awareness. Social media often represents itself as a platform that enables human connection, using Facebook’s motto, “sharing and connecting with your friends, family, and people you know”. That is, social media *does not* represent itself as a platform that collects data about its users, sells it to advertisers, and then nudges the users in the directions which serve the interest of the advertisers. Some may, then, not even know the business model of social media platforms – which involves using data-driven algorithmic systems to categorize subjects and objects into MAKs – and may not even be aware of there being a power relationship at all.²²

In sum, none of the senses of Awareness delineated above are necessary for productive power. Sub-jects of productive power do not need to have a critical awareness or show recalcitrance, do not need to be aware of the workings of the power relationship and do not even need to be aware of the existence of the power relationship. Thus, the sheer fact that one is in a power relationship where one does not have a critical awareness or does not show recalcitrance, or is not aware of the workings of the power relationship or is not even aware of the existence of the power relationship does not imply that the given power is non-productive. In other words, in this section, I have shown that lack of Awareness can also characterize sub-jects of productive power, and so, regardless of which of the above delineated sense it is taken, does not imply Subjectless.

4.2.2 *Reflection as Address*

In this section, I will identify the second meaning of reflection which I will call Address. Preliminarily, I will take Address to refer to the intentional state which stems from the sub-ject’s being addressed as a subject, by another subject. Those who hold power, for instance, in schools and corporations, it seems, address their sub-jects as subjects to produce the

²² For a more specific example, take the mobile application Pokémon Go, a game where players – its sub-jects – use their smartphones to find and capture “Pokémon” in real-world locations. The sub-jects of Pokémon Go – many of them – may think they are merely playing a fun game, having no awareness of the existence of a power relation at hand and no knowledge of its workings, and, thus, also showing no recalcitrance. Nonetheless, they are also contributing to the data collection, analysis and nudging, that is, in Shoshana Zuboff’s words, “the gentle herding of innocent Pokémon Go players to eat, drink, and purchase in the restaurants, bars, fast-food joints, and shops that pay to play in its behavioral futures markets” (Zuboff 2019). One could argue that the sub-jects of Pokémon Go are nonetheless produced as subjects, as the aim of the game was to change how people – especially children – may think of themselves as Pokémon masters, wanting to “catch them all”.

willing subjects they want who are often driven by internalized intentional states, not merely prudential considerations. It is, however, a recurring theme in the literature that in the case of some forms of algorithmic power, the relevant sub-jects are, in some sense, *not addressed*. For example, in the quote above by Rouvroy, she makes the case that what she calls “algorithmic governance” is characterized by the fact that it “affects, without addressing” (2013, 153). Is some form of Address indeed necessary for the exercise of productive power?

Let’s focus on the part of Address which highlights that the sub-ject is being addressed as a subject *by another subject*. This formulation, for many, may evoke the phenomenological tradition of a second-personal, or “I-You” relation (e.g., Benton 2017, Zahavi 2023). The idea that being addressed as a subject by another subject is somehow crucial for becoming a subject has been explored by thinkers such as Fichte (“No Thou, no I: no I, no Thou”, 1982, 172–3) Martin Buber (“Man becomes an I through a You” 1970, 80, cited by: Zahavi 2023, 95), or, taking an example from poetry, the early 20th-century Hungarian poet, Attila József writes, “in vain you bathe your own face in your self, / it can be cleansed only in that of others” (József 1999).²³ Given that in-person interactions are the paradigmatic examples of “I-Thou” relation, I will take “by another subject” to refer to a *direct* subject-to-subject relation such as an in-person conversation. Is it the case that productive power requires that sub-jects are addressed as a subject by another subject?

Think of exercises of productive power such as advertisement or, once again, propaganda through mass communication. While these are exercises of productive power which require the production of subjects needed, these power relations seem not to be characterized by a *direct* “I-Thou” relationship. Thus, while an advertisement or political propaganda with an explicit message (“*You* need to buy X”, “*We* cannot let X happen!”) clearly addresses sub-jects as subjects, as beings with minds and intentionality capable of forming a sense of self, they are not addressed directly by another subject.

An even clearer case of a lack of being (directly) addressed by another subject (direct subject-to-subject relation) where sub-jects are nevertheless being addressed *as* subjects are modulatory chatbot algorithms. Not only do these chatbots formulate meaningful written replies, but they also consider the sub-ject’s replies in formulating. However, even if the given chatbot is interactive, on a widely shared view, it does not qualify as a subject in the relevant sense. Yet there are indeed chatbots whose goal, e.g., through intentionally being trained on biased data, is to exercise productive power. Again, assuming that chatbots are not subjects, this is an example of an exercise of productive power where there is no subject-to-subject

²³ In the literature on second-personal relation, Dan Zahavi, for instance, thinks of such relations as requiring two conditions: a “particular kind of communicative connectedness” and “mutual address” (2023, 86); in a similar vein, Matthew A. Benton breaks down the second-personal relation to two main components: “two-way causal interactions” and treating each other “as subjects” (2017, 15–18). Without unpacking these ideas in detail, I think these building blocks correspond to the “as a subject” and “by a subject” parts of Address separately.

relation, even though sub-jects are addressed as subjects, *that is*, as beings with a mind and intentionality, capable of forming a sense of self.²⁴

If being (directly) addressed “by another subject” is not necessary for the exercise of productive power, then let’s focus on being addressed “as a subject”. Given our definition of subjectivity, if one is addressed as a subject, then one is addressed as a being with a mind and intentionality, capable of forming a sense of self. It may seem intuitive to suppose that such addressing is done via a communicative relation which would require the sub-jects’ capacities as subjects.

Some social scientists, then, seem to argue that the relevant communicative relation – which is said to be lacking in the case of some forms of algorithmic power – is *dialogical* in the sense that the sub-ject’s sense of self is somehow considered. Eran Fisher, for instance, argues that algorithmic power “undermines [...] the active participation of the self in creating knowledge about the self” (Fisher 2022, 2). Or, as Rouvroy puts it in an interview: “Does what we might say about ourselves not become redundant, if not suspect, in the face of the efficiency and mechanical objectivity of the automatic profiling to which we are subjected?” (Rouvroy 2016). In the case of social media, for instance, while there is a form of interaction between one’s tracked behaviour and the personalized content which is then fed back onto one through one’s membership in MAKs, it is *not* a dialogical communicative relation as the data gathering and its algorithmic analysis focuses on tracked behaviour, not the sub-jects’ own sense of self.

However, a dialogical communicative relation seems not to be necessary for productive power either. For instance, algorithmic forms of propaganda on social media which do not engage with their sub-jects in a dialogical communicative relation may still lead to subject production. Or take the above-mentioned example of a student or employee. While they are addressed as subjects in the sense that their performance and ranking is communicated to them, it could be the case that they are *not* asked about, say, their own “self-assessment”. Nevertheless, as we have seen, this process *may* still serve as an exercise of productive power: the grades and performance reports could be meant to drive the student or the employee to inform their sense of self and strive for better performance.

²⁴ A beautiful portrayal of the importance of being addressed by another subject is the movie “Her”. While the virtual assistant, Samantha, with whom (or which?) the protagonist, Theodore, falls in love, behaves as a kind and empathetic person would, he discovers the profound limitation in this relationship when he realizes that Samantha simultaneously converses with thousands of individuals and virtual assistants. Could their relationship be an “I-Thou” relation, the movie prompts one to ask, even if Samantha only conversed with Theodore? Whether subject-to-subject relation of Address is ultimately fulfilled in a scenario like a chatbot or “Her”, of course, depends on whether we take the given algorithmic system as a subject. This is perhaps one of the most interesting philosophical questions about algorithms and artificial intelligence in general – and one I cannot, *and do not need to*, settle here. I am merely stipulating that our definition above “beings with minds capable of intentionality and having a sense of self” excludes these algorithmic systems.

Another example of an algorithmic power which is productive but does not involve dialogical communicative relation in the relevant sense is the increased use of fitness and health data tracking devices which provide people with quantified biometric information. Again, subjects are addressed as subjects but not through a dialogical communicative relation: the often-repeated, fashionable motto, “algorithms know you better than you know yourself” captures the trend that behavioural and biometrical data tracking and its algorithmic analysis are regarded as more effective methods for gaining knowledge about oneself than considering one’s intentional states or sense of self. People can, then, internalize the health data numbers and biometrics. This can be viewed as an example of algorithmic telic power, which endows individuals with a status based on their perceived distance from the norms and ideals of a healthy life implied by these analyses. As a result, subjects change their sense of self accordingly, without involving their own sense of self in the data-gathering process. Thus, data-driven algorithmic insights about one’s health is another way of producing one’s sense of self.²⁵

If a dialogical communicative relation in addressing the subject as a subject is not necessary for productive power, then let us go back to our broader definition of being addressed as a subject, that is, the power relation being a communicative relation (not necessarily dialogical) which would require the subjects’ capacities as subjects, that is, beings with minds and intentionality capable of having a sense of self.

Without undertaking the heroic task of defining what constitutes a communicative relation, our examples from the previous section, Kohlhaas, Josef K. and Job, may help us illuminate the difference to some extent. Consider that Kohlhaas and Josef K. are explicitly told about the relevant power relation. Kohlhaas is told that his horses will be seized as collateral and Josef K is informed that “You are not allowed to go from here. You are after all under arrest” (Kafka 2019). On the other hand, Job is not informed, he is only affected by the suffering brought upon him by God; he only seems to *assume*, at first, that it is the case that “*The Lord gave and the Lord has taken away*” (BSB, Job 1:21, my emphasis).

Are these exercises of power, then, productive? While Kohlhaas and Josef K. are addressed as subjects, as we know from their story, are not willingly compliant subjects. Is God’s exercise of power, in turn, productive? As we have seen, for a while, Job shows resistance –

²⁵ I am not trying to settle the issue of when the first-person perspective becomes relevant or unpack what kind of epistemic relevance the first-person perspective and the often associated “privileged access” hold. I am also not trying to make normative judgements about using algorithmic methods to guide one’s practice (as opposed to much of the social scientific literature, e.g., Eran Fisher who argues for “the inherent inability of algorithms to underlie critical knowledge” [2022, 3]). I am merely pointing out that the various self-tracking methods which use algorithmic processes to generate knowledge about themselves, without considering their own sense of self or self-assessment, exemplify productive power (Neff & Nafus, 2016; Coeckelbergh 2022, 112).

the trait which earned him the awe of Kierkegaard. However, he eventually submits and repents. But this only happens after being *addressed* by God “Then the Lord answered Job out of the whirlwind and said: [...] Brace yourself like a man; *I will question you*” (BSB, Job 38:3, my emphasis). So, is addressing subjects through a communicative relation (not necessarily dialogical) which would require the subjects’ capacities as subjects, that is, as beings with a mind and intentionality and a sense of self, necessary (even if, as we have seen, not sufficient) for productive power?

It may help to see what it would mean not to, using Rouvroy’s terms “address” subjects as subjects but only “affect” them. She does not draw a clear distinction between the two forms of power relation, and, indeed, formulating the distinction between power relations which “address” and those which merely “affect” is no easy endeavour. To get off the ground, consider a patient visiting a neurologist for a patellar reflex test. Hitting one’s knee with a reflex hammer does not address one as a subject as it is not a communicative relation which would require the one’s capacities as subjects, that is, as beings with a mind and intentionality, capable of forming a sense of self. It only affects one’s body as a reactive system susceptible to stimuli – completely bypassing one’s consciousness, intentionality and sense of self.

Indeed, social theorists such as Rouvroy argue that, at times, the exercise of what we have labelled nudging algorithmic power – such as recommendation algorithms compiling one’s news feed, “personalized” playlist, watchlist etc. – may count as less direct forms of addressing – or even mere affecting. Sub-jects of (some cases of) nudging algorithmic power are often not informed, say, about the result of their behavioural analysis process – only affected by it.

However, between affecting and communicative addressing, there is a rich spectrum of ways of addressing using more subtle manners. Take propaganda again. While in many cases of exercising productive power via propaganda, sub-jects are in a communicative relation, for instance, through billboards or video ads, there are examples of propaganda where the addressing of sub-jects as subjects, and thereby the exercise of productive power, is more subtle. Take, as an example, the planning and staging of fake attacks as part of false flags operations as pre-texts for war. For instance, consider Sweden’s planned fake attacks in the 18th century to incite hatred against Russians as a *casus belli*. Swedes were not directly addressed, yet the staging was intended as a sign to be interpreted by subjects capable of having a sense of self and whose sense of self might be altered, driving them to think of themselves primarily as Swedes and think of the Russians as a threat, and act accordingly. Thus, in this case we have subject production, some subtle form of communicative relation, but no direct address.

Telling apart various – direct and indirect – forms of communicative addressing from affecting, thus, is a difficult task; one for another philosophical inquiry. However, it is safe to say that the kind of intentional states that the subject as a subject forms is often independent of whether they were addressed through a direct communicative relation or not. Therefore, the assumption that being addressed through some form of direct communicative relation is necessary for a specific kind of intentional state – reflection as Address – is false. Nonetheless, as I will show in the next section, focusing on this borderland between addressing and affecting drives us close to locating to the kind of intentional states – the precise sense of “reflection” – that is said to be lacking in some cases of forms of algorithmic power.

In sum, none of the senses of Address delineated above are necessary for productive power. I have shown that to exercise of productive power, subjects do not need to be addressed as a subject, directly by another subject. We have seen that no in-person direct addressing is needed, no dialogical communicative address is needed where the subject’s sense of self is considered, and, finally, neither dialogical nor any form of direct communicative address is needed. The lack of reflection as Address, then, does not tell non-productive forms of algorithmic power apart from productive forms of powers.

Now, two tasks are still ahead of us. First, pinpointing the precise sense of “reflection” that is said to be *uniquely* lacking in some cases of successful exercise of algorithmic power. Second, pinpointing the precise sense of the form of algorithmic power whose successful exercise is uniquely characterized by *this* lack of reflection.

4.2.3 *Reflection as Assessment*

The idea circulating in contemporary social theory that forms of algorithmic power do not address subjects as subjects in the sense that they do not address them as beings with minds and intentionality, capable of having a sense of self leads us to the last meaning of reflection I will explore. I will call it Assessment. Assessment refers to the intentional states necessary for, and accompanied by, an intentional action which the subject performs willingly, driven by internalized intentional states, *or* unwillingly, driven by prudential considerations or fear of punishment. The lack of Assessment, then, is a state where conjunction of the negation of these disjuncts are true: not performing an intentional action willingly and not performing an intentional action unwillingly.

The lack of both willing (internalized) and unwilling intentional action is often characterized on the model of reflex- or impulse-based behaviour which is, then, held to characterize subjects of some forms of algorithmic power. For example, Rouvroy mentions that what we have termed data-driven algorithmic systems are “designed to accelerate flows – avoiding any form of ‘detour’ or subjective ‘reflexive suspension’ between ‘stimuli’ and their ‘reflex responses’” (Rouvroy–Berns 2013, xiv). In another paper, she observes that a certain form of

algorithmic power “operates on the mode of alerts and reflexes” (2013, 153). As I will argue, it is this third sense of lack of reflection which makes what I will call behavioural nudging algorithmic power – a subset of nudging algorithmic power – uniquely non-productive.

As we have seen, intentional action is a special form of behaviour where one needs to have certain intentional states which represent the action to perform it. One, for example, needs to recognize the existence of money (grasping the concepts making up the relevant statuses) in order to be able to *pay*. Intentional action, however, comes in various forms and degrees which I cannot sufficiently analyze here. Suffice it to note that paradigmatic cases of intentional action involve slow, rational or strategic decision making, which prioritizes with long-term benefits, while paradigmatic cases of non-intentional action, in turn, involve quick, emotion- or impulse-driven behaviour which prioritizes short-term benefits or even reflex-driven behaviour.

Without an elaborate taxonomy of intentional vs non-intentional actions which would encompass the spectrum of behaviour from reflex-responses to strategic planning, I will focus on distinguishing between two kinds of intentional actions. On the one hand, intentional actions performed willingly, more precisely, in a specific sense of willingness. I have mentioned earlier that telling apart various forms of willing and unwilling actions is no easy endeavor; in the following, whenever I mention willing intentional actions, I will mean those intentional actions one does and desires to do, driven by *internalized* intentional states which make the action align with their sense of self. By contrast, there are intentional actions which are performed unwillingly, that is, those one does not desire to do as they do not align with their sense of self.

In the following, I will link these two types of intentional actions to two types of compliance subjects may perform when – either productive or intimidating – power is successfully exercised over them. Then I will explain how the form of power I will call behavioural nudging does not require intentional action (or at least no paradigmatic examples thereof) for its successful exercise. To do so, I will rely on Harry Frankfurt’s (1971) famous distinctions between two kinds of desires and their role in constituting what he calls “personhood” and what I have referred to as a sense of self.

*

Frankfurt (1971) differentiates between “first-order” and “second-order” desires. He defines first-order desires as desires about a particular action, as in “S wants to do X/doesn’t want to do X”. To differentiate between first-order desires of various strength, he introduces the term “effective first-order desire” which refers to a desire which one acts upon. Second-order desires, in turn, are desires about a first-order desire, as in “S wants to want to do X/doesn’t want to want to do X”. Now, to have a sense of self, using Frankfurt’s terms, one needs to be

capable of forming second-order desires. As he puts it, “the capacity for reflective self-evaluation that is manifested in the formation of second-order desires” (1971, 7). Even more importantly, one needs to be capable of forming what he calls “effective second-order desires”. These are desires about one’s effective first-order desire, that is, the desire that a certain desire would drive one to action (1971, 10–11). Let us, now, apply Frankfurt’s distinctions to explain the two kinds of intentional actions.

Take the intentional actions of compliance the sub-jects of productive power which the sub-ject performs willingly as it corresponds to their sense of self. Remember that we have stipulated that not all intentional states which one may experience constitute one’s sense of self. How can we tell, then, which intentional states constitute one’s sense of self? I have mentioned above that they are, in some sense, internalized. Now, borrowing Frankfurt’s terms, internalization can be unpacked as having not only first-order desire and effective first-order desire but also an effective second-order desire which is not in conflict with one’s second order desires. In other words, the desires (or other intentional states) which one desires to have are the ones which one desires to be one’s effective first-order desire, that is, those driving one’s intentional action.

Imagine a student who has the first-order desire to rest as well as the first-order desire to get good grades. They, however, also have a second order desire: the desire to desire to get good grades, which in turn, becomes an effective second-order desire once they act upon it. There is, in these cases, no conflict between what the powerholder (say, the teacher) expects the sub-ject (student) to do and what the sub-ject, in fact, desires or wants to do; and, in turn, there is no conflict between their effective second-order desire and their second order desires.

Let’s move onto cases where sub-jects perform intentional actions of compliance unwillingly. That is, cases of intentional actions which sub-jects do not desire to do as they do not align with their sense of self. This is often the case with intimidation. Here, just like in the case of willing performing of intentional actions of compliance, sub-jects also need not only first-order desires but also second order desires and effective second-order desires. However, in these cases, as we have established, there may be conflict on the level of intentional states.

With Frankfurt’s vocabulary, we can now unpack the relevant conflict on the level of intentional states as follows: given the looming possibility of punishment, sub-jects’ second-order desires (both what they desire to desire to and what they do not desire to desire to do) are in conflict with their effective second-order desires (the desires which they desire to be their effective first-order desire, that is, those driving their action. For instance, the student or the employee may not internalize the desire to be a good student or good employee but simply perform out of fear of punishment. In this case, their intentional action choosing to study or work instead of resting is in conflict with their sense of self – as they think of

themselves as someone who values rest more than hard work – but they still do so to avoid being punished.

Now, given that Assessment is a disjunction of either willing or unwilling intentional action, it is true even if only one of its disjuncts is true. Willing intentional action, as we have seen, is necessary for the successful exercise of productive power. Some form of intentional action (willing or unwilling), then, is necessary for the successful exercise of intimidation. Assessment, then, seems to be necessary for both productive power *and* intimidation. Finally, given that Assessment seems to be necessary for both intimidation and productive power, the lack of Assessment indeed seems to imply Subjectless.

But let us not stop here: let us see how exactly sub-jects over whom some forms of algorithmic power is exercised lack Assessment.

*

As we have seen, in *some* cases, algorithmic power can indeed be productive or intimidating. The use of propaganda through the MAKs constructed via data-driven algorithmic systems in the case of social media is an example of productive algorithmic power. Through feed compilations and targeted advertisements, the aim is not merely to want sub-jects to believe X and desire X but also to make them internalize these intentional states (want to believe X and desire X) so that these intentional states change their sense of self which would drive them to intentional action. An example of intimidating algorithmic power, as we have seen, is a data-driven algorithmic social credit system which only aims to elicit compliance, be it willing or unwilling. Such a system still requires second order desires and second order volition, even if they are in conflict with one another because although the sub-ject may *disagree* with how the system defines “trustworthiness”, formulating the second-order desire of not wanting to want to comply, they still formulate the effective second-order desire which leads them, eventually, to comply.

By contrast, the unique cases of non-productive and non-intimidating algorithmic power which we are after are the cases where Assessment is lacking. We have now reached the point where I can introduce *behavioural* nudging power. Take, once again, the example of scrolling on one’s social media feed and being shown “personalized” content based on one’s status stemming from membership in various MAKs which then drives one’s – impulse-based – behaviour to keep mindlessly scrolling.

This form of power is an example of nudging as there is no penalty or coercion employed in eliciting compliance. More precisely, as I mentioned in Chapter 3, algorithmic nudging is modeled after traditional nudging where one gets another subject or other subjects to do what they want them to do by steering their behavior through eliciting certain behaviour

without coercing or penalizing their sub-jects if they do otherwise. The peculiar nature of some forms of *algorithmic* nudging as opposed to traditional forms of nudging, however, is that its sub-jects are endowed with a status characterized by real-time adaptability. Placing low-cost snacks near the checkout line, as an example of traditional nudging power, which does not necessarily rely on an endowed statuses and only needs a fixed assumption about what customers waiting to pay would be susceptible to. By contrast, algorithmic nudging power, for instance, through algorithmic recommendations, is a form of power which depends on endowing prediction-based statuses and is more adaptable with its assumptions being based on real-time data deriving from various forms of tracking, and thereby being able to update the status accordingly.

Further, the relevant nudging here is “behavioural” given the lack of intentional action on the part of the sub-ject. Sub-jects are required to formulate first-order desires and effective-first order desires but not second-order desires and effective second-order desires. Thus, when Rouvroy argues that the aim of algorithmic power of social media is “to prompt individuals to act without forming or formulating a desire” (2013, xiii), I take her to refer not to first-order desires or effective first-order desires but to effective second-order desire. The relevant difference here, then, say, between the desire to be a good student or a good employee and the desire to continue mindless scrolling is *the kind of desire* that leads to the given behaviour.

The aim of the given nudging power, as we have seen, is no longer “I want you to desire X and desire desiring X, that is, internalize your desire about X” (productive power) or “whatever you desire, I want you to form the effective second-order desire about X” (intimidation), but “I want to *predictably* elicit first-order desires X, Y, Z in you and make them your will driving your behaviour” (algorithmic behavioural nudging). This means that for some cases of exercising algorithmic nudging power, no effective second-order desire (whether it is internalized or in conflict with one’s second-order desires) is needed.

What is needed, instead, are sub-jects who are susceptible to the given nudging stimuli, eliciting the relevant (close-to) reflex- and impulse-based responses and, therefore, being predictable. Predictability – behaving in the way the powerholder wants the sub-ject to – in these cases stems not from a produced willing subject with an aligned sense of self or intimidation but from their susceptibility to (close-to) reflex- and impulse-inducing nudging stimuli. In other words, the work of “subjugating” is delegated from the sub-ject as a subject to an ever-modulating data-driven algorithmic system surveilling the sub-ject.

Thus, for the sake of clarity, three points are in order. First, behavioural nudging power is a subset of forms of nudging power. Second, behavioural nudging power, just like any form of nudging power, can be exercised through data-driven algorithmic systems or through, other, non-algorithmic tools. Third, behavioural nudging power is another category, next to

productive power and intimidation, in our taxonomy outlining how various forms of power relates to the intentional states of its sub-jects.

*

Now, it seems to me that Frankfurt's analysis (1971) of various forms of desires (which, for him, culminates in an analyses of freedom) is also helpful in outlining two ways in which – using our terms – behavioural nudging power, and especially behavioural nudging *algorithmic* power can lack Assessment. In Frankfurt's terms, these are two ways in which the successful exercise of the given power does not require effective second-order desires, whether they are internalized (productive power) or in conflict with one's second-order desires (intimidation).

The first is when algorithmic power produces, as opposed to subjects, what Frankfurt calls "wantons". A wanton, in Frankfurt's interpretation, is someone whose "desires move him to do certain things, without its being true of him either that he wants to be moved by those desires or that he prefers to be moved by other desires", thus the wanton is "not concerned with the desirability of his desires themselves" (1971, 11). The wanton is characterized by "his lack of the capacity for reflection or to his mindless indifference" (1971, 13). The wanton thus is neither a willing (produced) nor an unwilling (intimidated) subject: they simply do not care. It seems to me that we do have our wanton moments when engaging in social media or other recommendation-based algorithmic systems, mindlessly scrolling for hours. One simply does not care what content one is consuming, simply following their first-order desires. There is no tension between one's sense of self and what one is made to do because one simply does not form effective second-order desires.

The other case is when there is a unique distance between our sense of self and what the subject is made to do. For such a unique distance, the sub-ject needs to internalize some of their desires. In other words, they require effective second-order desires which may be uniquely distant from the desires which drive one's behaviour. This distance is unique as it is not the distance of the unwilling intimidated sub-ject: it is, instead, the distance of a sub-ject who is being bombarded with hard-to-resist stimuli inducing first-order desire-driven behaviour. As technology ethicist Tristan Harris puts it in an interview:

I think a good example of this is YouTube. You open up that YouTube video your friend sends you after your lunch break. You come back to your computer and you think OK, I know those other times I end up watching two or three videos and I end up getting sucked into it, but this time it's going to be really different. I'm just going to watch this one video and then somehow, that's not what happens. You wake up from a trance three hours later and you say, "What the hell just happened?" And it's because you didn't realize you had a supercomputer pointed at your brain. (Thomson 2018)

This, I take it, shows that the experience of distance between one’s sense of self and the reflex- or impulse-driven behavior of non-Assessment can be in the form of the shame that one feels when one has scrolled for hours when one did not want to. In other cases, however, the subject may have what we have called (one version of) Awareness, being aware of a relation as a power relation and even knowing about the nature of the power relation, while being subjected to algorithmic forms of behavioural nudging. Assessment, in these cases, is still lacking: the given action is not only not a willing intentional action, but, given its impulse-based nature, arguably not even an intentional action of the sort required for cases of intimidation.

Consider again, as an analogy, the patient visiting a neurologist for a patellar reflex test. The patient might be another neurologist, being aware of the existence and knowing about the workings of that reflex-power, that is, being aware that the neurologist will test their reflexes as well as knowing the workings of the power that accompanies doctor-patient relationships. Nevertheless, how their leg reacts to the tap of the reflex hammer is still not a result of Assessment. In a similar vein, revisiting the example of impulse buying while waiting in line to reach the checkout counter, even if many are aware of the nudging power of strategically placed products, this, again, does not imply that they will resist these – very likely superfluous – purchases.

When one is scrolling on social media, one may be aware of the existence of a power relationship, even know its workings but not be able to resist on the level of behaviour. One may become aware that these systems are exercising power on them through the MAKs they construct, and which nudge their behaviour in subtle ways, without their behaviour of scrolling qualifying as being driven by Assessment. The sentiment the goes with experience of this distance between lacking Assessment and having Awareness in an algorithmic context is aptly shown by a New Yorker cartoon depicting two people sitting side by side, one of them checking their phone and saying something. The caption reads “I hate the person my targeted ads think I am” (Ehlers 2023). This, of course, does not imply that these ads will not have behavioural nudging algorithmic form of power over one.

Incidentally, Frankfurt’s example of the “unwilling addict” is an eerily apt analogy to the subject of behavioural nudging algorithmic power who may have Awareness but no Assessment:

[He] hates his addiction and always struggles desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful for him to withstand, and invariably, in the end, they conquer him. He is an unwilling addict, helplessly violated by his own desires. (1971, 11)

The unwilling addict maintains their unique distance from the power over them, the first order desire which is planted in them. Given these conflicting desires in the person, they

“make one of them more truly his own and, in so doing, he withdraws himself from the other” (1971, 13). Frankfurt’s “identification” and “withdrawal” are, then, in line with how I have used the terms “internalization” and “distance”.

In sum, the second way in which some forms of algorithmic nudging power can be non-productive in a unique way is by eliciting first-order desires and wills driving the subject to action without effective second-order desires. These desires are not internalized (as opposed to productive power) and the resulting behaviour is not an unwilling intentional action (as opposed to intimidation) but a form of (close-to) reflex- or impulse-driven behaviour.

In this section, following up the idea that subjects of some forms of algorithmic power are not addressed as subjects, I considered the last sense of reflection which is said to be lacking in subjects of some form of algorithmic power. I called it Assessment and used it to refer to the intentional states necessary for, and accompanied by, an intentional action which the subject performs willingly or the intentional states necessary for, and accompanied by, an intentional action which the subject performs unwillingly. I have shown that that Assessment is necessary both for forms of productive power and forms of intimidation.

Then I have shown that Assessment is lacking in cases of what I have called behavioural nudging algorithmic power which drive subjects to non-intentional forms of action. Finally, I explained, relying on Frankfurt’s ideas, the two types of examples of lack of Assessment in behavioural nudging algorithmic power. These are, on the one hand, cases of “wantons”, that is, subjects who do not form desires about their own desires, and, on the other, unwilling addicts, that is, subjects who feel a distance between their will and their sense of self.

Concluding remarks: Lack of reflection on all fronts

In this chapter, I have first unpacked Foucault’s notion of productive power in order to evaluate the claim that some forms of algorithmic power are uniquely non-productive. Producing subjects, then, was taken to mean changing those intentional states of the subject which constitute their sense of self in a way that they willingly comply with what the powerholder wants. In brief, then, productive power is a form of power whose exercise requires producing *willing* subjects driven by internalized intentional states. I contrasted productive power with intimidation which, while it may produce either willing or unwilling subject, only requires changing some of the subjects’ intentional states but not those which constitute their sense of self.

Then I have shown that there is an interesting claim circulating in contemporary social theory literature that some forms of algorithmic power do not require the production of subjects in a unique way – where uniqueness is understood as being different from intimidation. I have

pointed out that this claim is said to be supported by the premise that the subjects of these forms of algorithmic power lack certain kinds of intentional states, often referred to as “reflection”. I looked at three ways to unpack the relevant sense of reflection. First, as Awareness, that is, whether the subject is aware or critically aware that they are in power relationship and/or knows the workings of this power relationship. Second, as Address, that is, whether the subject is addressed by another subject and/or as a subject. And, third, as Assessment, that is, whether the subject has intentional states necessary for, and accompanied by, an intentional action which the subject performs willingly or the intentional states necessary for, and accompanied by, an intentional action which they perform unwillingly.

I argued that in and of itself only the last version of reflection, Assessment, is necessary for productive power. Given that it is also necessary for intimidation, the lack of Assessment implies the claim I called Subjectless: that at least some forms of algorithmic power are non-productive in a unique way. I have, then, identified these forms of algorithmic power as behavioural nudging power. Behavioural nudging power – as a subset of forms of nudging power – is the third category I identified, next to productive power and intimidation, in our taxonomy outlining how various forms of power relates to the intentional states of its subjects. Behavioural nudging power is a form of power – which can also be exercised through data-driven algorithmic systems such as a social media feed curating system – which does not require the production of subjects for their successful exercise (as productive power does) and they also do not require any (paradigmatic) form of intentional action (as intimidation does). The work of controlling the subject is delegated from the subject capable of intentional action to an ever-modulating data-driven algorithmic system which is surveilling and nudging the behaviour of the subject.

This leads us to several new questions. First, one can address the empirical issues I have set aside. In what sense does the lack of one or a combination of the above identified senses of reflection lead to a hindering, undermining or disabling of our subjectivity? Second, I have only analysed what each of the possible interpretations of reflection in themselves might imply. Upcoming research may tackle the question of what the implications of the combination or simultaneous lack of all the above-mentioned senses of reflection are. For example, given our examples above, it seems to me that our most often cited form of algorithmic power, social media curating algorithms, at times, may qualify for a lack of all three forms of identified reflection. What are, then, the philosophical implications of spending our time in an environment with which may exhibit all three senses of lack of reflection? Third, as I mentioned, I am not trying to formulate a normative claim. I am neither trying to mourn some sense of a loss of subject production, nor unpack the opportunities that the lack of either one of these senses of reflection, or a combination of them might bring about. Thus, another set of philosophical questions involve the normative assessment of the new forms of power which are born out of the historical shift towards algorithmic societies.

I aimed to have paved the path towards answering such questions by clarifying how some algorithmic forms of power – behavioural nudging forms of algorithmic power – are non-productive in a unique way.

CONCLUSION

This dissertation aimed to provide an explanation of a contemporary shift concerning the conditions of existence of social reality. I argued that there is an ongoing historical shift towards what I called algorithmic societies. I aimed to show that the conceptual framework of social ontology – specifically, its focus on explaining the existence of social reality through the notion of mind-dependence – can be used to uncover the profoundness of this historical shift.

The historical shift, in brief, can be outlined as follows. I differentiated between two basic ontological levels of social reality. First, building blocks, which constitute the most fundamental level. Second, elements, which are partly built out of building blocks and for which I used “kinds” and “forms of power” as my units of analysis. The traditional picture of social ontology, which I later labelled Intentionalism, holds that the necessary building block of social reality, making up elements of social reality such as social kinds and forms of social power, is representation through collective intentionality, more specifically, collective recognition. By contrast, as I showed, in an algorithmic society, there are now also elements which do not depend on collective recognition. Instead, these elements of the algorithmic society, which include kinds constructed through, and forms of power enabled by, what I called data-driven algorithmic systems, depend on the following two building blocks: representation through human intentionality (although not necessarily *collective* intentionality as in the case of traditional elements of social reality) *and* representation through digital data and algorithmic code. My broader goal was to show that elements of the algorithmic society are dependent on our minds – our collective and individual intentionality – in a weaker sense than traditional elements of societies.

The focus of this dissertation, then, was unpacking this weaker mind-dependency which characterizes the elements of the algorithmic society. I argued, first, that data-driven algorithmic systems bring about social kinds – which I called Modulatory Algorithmic Kinds – which do not require collective recognition (in the relevant sense) for their existence. This, as I showed, challenges Intentionalism, the traditional picture of social ontology. Second, I argued that there are forms of algorithmic social power which do not require collective recognition for their existence. Third, I showed that forms of algorithmic *nudging* power are further removed from human intentionality. This is because their exercise is not even enhanced by collective recognition. Furthermore, a specific subset of algorithmic nudging power, behavioural nudging, also does not require that the relevant sub-jects (subjects upon whom power is exercised) engage in intentional action, which sets this form of power apart from cases of “productive power” as well as cases of intimidation.

Here is, then, a brief review of the argument of each chapter.

In Part I, I have analysed what social kinds are and how they are dependent on minds. My aim was to show that social kinds constructed in an algorithmic society challenge our traditional understanding of the conditions of existence of social kinds.

In Chapter 1, I considered the traditional picture of social ontology, which I have labelled Intentionalism, which holds that social kinds such as money, marriage and gender are mind-dependent in the sense that their existence is dependent on collective intentionality, more specifically, collective recognition. Intentionalism, as I construed it, aims to be a monist theory of social reality, in other words, it aims to formulate a general formula of mind-dependence which applies to all social kinds. This is in contrast with, on the one hand, weaker monist theories as well as pluralistic theories of social reality which hold that social kinds are mind-dependent in a weaker sense, that social kinds are mind-dependent in different ways, or that only some social kinds are mind-dependent.

I aimed to provide a charitable reconstruction of Intentionalism, taking into account the points of criticism it received. I have shown that, given legitimate counterexamples brought in the extant literature, the scope of Intentionalism, the relevant sense of recognition as well as that of dependence needs to be weakened in order for it to remain a general – monist – theory of social reality. I argued that the scope of Intentionalism should be restricted to human social kinds, that the relevant sense of recognition can accommodate being wrong about the membership conditions of the given social kind, being unconscious and being non-language dependent, and that dependence on collective recognition can be either direct or indirect.

However, in Chapter 2, I have argued that with data-driven algorithmic systems, that is, systems which quantify and classify subjects and objects based on large scale, granular and real-time data through algorithmic analysis for various socially consequential purposes, such as those used in social media, automated hiring, and risk assessment, new ways (even if only in degree) of kind-making emerge. Data driven algorithmic systems are characterized by, first, data collection of unprecedented scale and granularity, second, vastly increased computational capacities for algorithmic analysis, and, third, both data collection and processing happening in a dynamic, real-time manner. While the difference from between these new ways of kind-making through data-driven algorithmic systems and earlier quantification and outside-perspective based kind-making may only be one of degree, it is, as I have argued, enough to carry implications for social ontology.

Kind-making through data-driven algorithmic systems leads to the construction of what I have called Modulatory Algorithmic Kinds (MAKs). These are modulatory because, given incoming data and its constant analysis, on a lower level, the membership of their members is in constant change, and on higher levels, the very rules specifying the member's associated

status and their membership conditions are in constant change as well. MAKs, as I have shown, challenge the traditional picture of social ontology captured by Intentionalism. This is because they qualify as social kinds without their existence being dependent on collective recognition or any form of collective intentionality. First, as I argued, given their predictive potential stemming from their multiple shared properties, they qualify as kinds. Second, as I aimed to show, they are *social* kinds as they are representation-dependent, (at least some of them are) causally (not constitutively) mind-dependent and carry group-level significance. Thus, Intentionalism, I argued, should give way to a pluralistic understanding of the mind-dependence of social kinds.

In Part II, I focused on the element of social reality I called social power (or, simply, power). My aim was to highlight how forms of power enabled by data-driven algorithmic systems fit into our understanding of social power.

In Chapter 3, I defined a minimal notion of social power according to which power is a capacity of getting other subjects to do what the powerholder wants them to do. I aimed to capture the insights of both “power-to” and “power-over” views in the extant literature by emphasizing that power is a capacity and that it is relational, respectively. I also considered two versions of a crucial constraint on the minimal concept of social power which I called the Intentionality Constraint. The first version, which I have accepted with caveats, holds that the exercise of social power requires intentional action on the part of the powerholder, even if these intentional states may be unconscious and they may have unintended consequences.

Regarding the second version, according to which the existence of social power depends on collective recognition, I argued that it does not hold in all cases of social power. Relying on my earlier analysis of the mind-dependence of social kinds and their associated social statuses constructed through data-driven algorithmic systems, I argued that even if we accepted that all forms of social power depend on the existence of social statuses, it would still not follow that the existence of all forms of social power depend on collective recognition. This is because, as I argued, exercising any form of power via data-driven algorithmic systems brings about forms of social power – meeting the conditions of the minimal concept thereof – where neither the power of bestowing statuses nor the powers stemming from the statuses themselves depend on collective recognition.

My aim was, first, to clarify the relationship between social status, intentional action, and social power. Second, to begin the project of providing an explanation of the forms of algorithmic powers made possible by data-driven algorithmic systems. To accomplish these aims, I have introduced two – non-exhaustive – taxonomies of social power: one addressing the ways in which power can be hidden, and another addressing the types of statuses which social power can construct and depend on.

Regarding the ways in which power can be hidden, I identified the following forms of power. First, intimidation and inaction where the relevant powers do not need to be exercised in order to elicit compliance. Second, power relations characterized by a lack of conflict, either on the level of behavior or intentional states. And, third, power relations the without awareness of the power relation either by the relevant powerholder or the sub-ject.

Regarding the types of statuses which social power can construct and depend on, relying heavily on Åsa Burman's analyses of power (2007, 2023), I introduced deontic powers and telic powers. The statuses associated with deontic powers are various forms of rights and obligations, which usually come in binaries; the statuses associated with telic powers are one's perceived distance from an ideal or norm, which are therefore graded. I showed that these forms of power can be exercised through data-driven algorithmic systems as well.

Then I expanded this taxonomy with algorithmic forms of nudging power where sub-jects are not confined, coerced or penalized for not complying but their behaviour is constantly steered, "nudged", towards a direction the powerholder deems desirable. I have shown that, as opposed to algorithmic forms of deontic or telic powers, in the case of algorithmic form of nudging power, the power of bestowing statuses is not only not dependent on, but also not enhanced by, collective recognition.

The focus on algorithmic forms of nudging power paved the way towards Chapter 4. Here, I addressed the question of how various forms of power relate to their sub-jects' intentional states. The chapter was inspired, first, by Foucault's idea on how certain forms of power do not intimidate but "seduce" and elicit "acceptance", thereby, in a sense, "producing subjects", a form of power he called "productive power". Second, the inspiration arose from having encountered, in the extant contemporary literature on algorithmic societies in social theory, media studies, sociology, several formulations of the idea that forms of algorithmic power are, in some sense, not linked to subject production. Thus, the specific question this chapter addressed was this: *in what sense do forms of algorithmic power produce – or do not produce – subjects?*

I first unpacked Foucault's notion of productive power. I took "producing subjects" to mean changing their internalized intentional states, that is, those intentional states of the sub-ject which constitute their sense of self in a way that they willingly comply with what the powerholder wants. I, then, following Foucault, contrasted productive power with intimidation. Intimidation, just like productive power, also changes the intentional states of its sub-jects. However, it only requires changing some of the subjects' intentional states but not those which constitute their sense of self. As such, intimidation may produce either willing or unwilling subjects, not necessarily willing ones as in the case of productive power. Productive power and intimidation, then, form two categories in the third taxonomy of power

I developed: a taxonomy addressing the ways in which a given form of power relates to the intentional states of its sub-jects.

Then, I showed that social theorists contrasting productive power with algorithmic power tend to cite the lack of “reflection”, which I took to be a certain kind of intentional state, as a reason for the alleged difference. Given that the concept of reflection is often muddled in the extant debate, I distinguished between three ways to unpack it. My conclusion was that only what I called Assessment, that is, reflection in the sense of the sub-ject having intentional states necessary for, and accompanied by, an intentional action which the subject performs willingly or unwillingly, is necessary for productive power – as well as intimidation.

The specific form of algorithmic power which I identified as lacking Assessment – being thus non-productive in a *unique* way, setting them also apart from intimidation – is what I called behavioural nudging power. It is a form of power which does not require intentional action on the part of its sub-jects. As I argued, it is a subset of forms of nudging power, one which can be exercised through data-driven algorithmic systems – such as a social media feed curating system – and through other means, and, finally, a third category, besides productive power and intimidation, in my third taxonomy.

These are, in sum, the points I argued for to make the case that we are now constructing societies which are dependent on our minds in a weaker sense than traditional societies have been.

Now, numerous crucial questions have not been sufficiently answered. There are many empirical questions which I only considered from the zoomed-out perspective of the philosopher. For instance, questions concerning how exactly data-driven algorithmic systems carry out their quantifying and classificatory work, to what extent and in what sense people are, in fact, aware of the existence and workings of these systems, how these systems impact people’s intentional states, mental capacities as well as larger social inequalities and social reproduction.

However, there are also many philosophical questions left unresolved. Here is a brief, far from exhaustive, list. There are, first, several intriguing legal issues concerning what rights should those over whom power is exercised through data-driven algorithmic systems have – and, as a corollary, what responsibilities are to be bestowed upon those exercising power through these systems. Furthermore, one may ask how exactly a detailed pluralistic – but still not overly convoluted – theory of the mind-dependence of social kinds should look like. I gave a preliminary suggestion based on whether the given kinds are representation-dependent, and if so, whether through collective recognition or digital data and algorithmic code; surely, this taxonomy may be further refined. And, as I have briefly alluded to at the end of Chapter 4, one may also tackle the normative assessment of contemporary ways of kind-making and

forms of power which are enabled by data-driven algorithmic system. How exactly are we to *critically assess* the implications of the historical shift towards algorithmic societies – most notably, collective intentionality no longer being the sole glue of social reality? These are all intriguing questions for future inquiry.

I conclude, then, that this work serves as a small step toward understanding the social ontological implications of the historical shift towards algorithmic societies – how exactly our social reality is becoming “mindless”.

Bibliography

Achbar, M., & Wintonick, P. (Directors). (1992). *Manufacturing Consent: Noam Chomsky and the Media* [Documentary]. Necessary Illusions Productions Inc., National Film Board of Canada (NFB), Téléfilm Canada.

Ananny, M., & Crawford, K. (2018). Seeing Without Knowing: Limitations of the Transparency ideal and Its Application to Algorithmic Accountability. *New Media & Society*, 20(3), 973–989.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. <https://www.wired.com/2008/06/pb-theory/>

Ásta. (2017). Social Kinds. In *The Routledge Handbook of Collective Intentionality*. Routledge.

Ásta. (2018). *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*. Oxford University Press.

Bachrach, P., & Baratz, M. S. (1962). Two Faces of Power. *The American Political Science Review*, 56(4), 947–952.

Baker, L. R. (2019). Just What is Social Ontology? *Journal of Social Ontology*, 5(1), 1–12.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender Identity and Lexical Variation in Social Media. *Journal of Sociolinguistics*, 18(2), 135–160.

Barnes, E. (2014). Going Beyond the Fundamental: Feminism in Contemporary Metaphysics. *Proceedings of the Aristotelian Society*, 114, 335–351.

Bartoletti, I. (2020). *An Artificial Revolution: On Power, Politics and AI*. The Indigo Press.

Baudrillard, J. (1983). *In the Shadow of The Silent Majorities Or, The End of The Social Order and Other Essays*. Semiotext(E).

Baudrillard, J. (with Poster, M.). (1988). *Selected Writings*. Stanford University Press.

Beauvoir, S. de. (1948). *The Ethics of Ambiguity* (B. Frechtman, Trans.). Philosophical Library.

Beer, D. (2017). The Social Power of Algorithms. *Information, Communication & Society*, 20(1), 1–13.

Benn, S., I. (1967). Power. In *The Encyclopedia of philosophy*. Macmillan.

Benton, M. A. (2017). Epistemology Personalized. *The Philosophical Quarterly*, 67(269), 813–834.

Biagioli, M. (Ed.). (1999). *The Science Studies Reader*. Routledge

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. Routledge & Kegan Paul.

Bowker, G. C. (2020). Foreword. In *The Constitution of Algorithms: Ground-truthing, Programming, Formulating* (pp. ix–x). The MIT Press.

Boyd, R. (1991). Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies*, 61(1), 127–148.

Brubaker, R. (2020). Digital Hyperconnectivity and the Self. *Theory and Society*, 49(5), 771–801.

Brusseu, J. (2020). Deleuze's Postscript on the Societies of Control Updated for Big Data and Predictive Analytics. *Theoria: A Journal of Social and Political Theory*, 67(164), 1–25.

BSB (2022). *Holy Bible: Berean Standard Bible*. BSB PUB.

Buber, M. (1970). *I and Thou* (W. Kaufmann, Trans.). Charles Scribner's Sons.

Bucher, T. (2018). *If ... then: Algorithmic Power and Politics*. Oxford University Press.

Burman (Andersson), Å. (2007). *Power and Social Ontology*. Bokbox Publications.

Burman, Å. (2015). *Extending the Scope of Searle's Theory of Social Reality: How to Account for Opaque Social Phenomena Like Power Structures*, *ProtoSociology – Essays on Philosophy*.

Burman, Å. (2019). Categories We Do Not Know We Live By. *Journal of Social Ontology*, 5(2), 235–243.

Burman, Å. (2023). *Nonideal Social Ontology: The Power View*. Oxford University Press.

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1), 2053951715622512.

Burrell, J., & Fourcade, M. (2021). The Society of Algorithms. *Annual Review of Sociology*, 47(Volume 47, 2021), 213–237.

Cakebread, C. (2017). *You're Not Alone, No One Reads Terms of Service Agreements*. Business Insider.

<https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11>

Caruso, C. (2016). *Who's Top Monkey? How Social Status Affects Immune Health*. Scientific American. <https://www.scientificamerican.com/article/whos-top-monkey-how-social-status-affects-immune-health/> (retrieved October 19, 2024)

Cheney-Lippold, J. (2011). A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society*, 28(6), 164–181.

Cheney-Lippold, J. (2017). *We Are Data: Algorithms and the Making of Our Digital Selves*. New York University Press.

Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values* (First edition). W.W. Norton & Company.

Coeckelbergh, M. (2022). *The Political Philosophy of AI: An Introduction*. Polity Press.

Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It For Capitalism*. Stanford University Press.

Crane, T. (1998). Intentionality as the Mark of the Mental. In A. O’Hear (Ed.), *Contemporary Issues in the Philosophy of Mind* (pp. 229–252). Cambridge University Press.

Crane, T. (2009). Intentionalism. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind* (p. 0). Oxford University Press.

Crane, T. (2017). The Unity of Unconsciousness. *Proceedings of the Aristotelian Society*, 117(1), 1–21.

Dahl, R. A. (1957). The Concept of Power. *Behavioral Science*, 2(3), 201–215.

De Landa, M. (2006). *A New Philosophy of Society: Assemblage Theory and Social Complexity*. Continuum.

Deleuze, G. (1992). Postscript on the Societies of Control. *October*, 59, 3–7.

Deleuze, G., & Lapoujade, D. (2006). *Two Regimes of Madness: Texts and Interviews 1975-1995*. Semiotext(e); Distributed by MIT Press.

Diaz-Leon, E. (2012). Social Kinds, Conceptual Analysis, and the Operative Concept: A Reply to Haslanger. *HUMANA.MENTE Journal of Philosophical Studies*, 5(22), Article 22.

Diaz-Leon, E. (2015). What Is Social Construction? *European Journal of Philosophy*, 23(4), 1137–1152.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press.

DW Documentary (Director). (2019, September 26). *Artificial Intelligence and Algorithms: Pros and Cons | DW Documentary (AI documentary)* [Video recording].
<https://www.youtube.com/watch?v=s0dMTAQM4cw>

Ehlers, I. (2023). *Cartoons from the April 10, 2023, Issue* [Graphic].
<https://www.newyorker.com/gallery/cartoons-from-the-april-10-2023-issue>

Epstein, B. (2014). How Many Kinds of Glue Hold the Social World Together? In M. Gallotti & J. Michael (Eds.), *Perspectives on Social Ontology and Social Cognition* (pp. 41–55). Springer Netherlands.

Epstein, B. (2016). A Framework for Social Ontology. *Philosophy of the Social Sciences*, 46(2), 147–167.

Epstein, B. (2018). *The Ant Trap: Rebuilding the Foundations of the Social Sciences* (First issued as an Oxford University Press paperback, 2018). Oxford University Press.

Epstein, B. (2019). Anchoring versus Grounding: Reply to Schaffer. *Philosophy and Phenomenological Research*, 99(3), 768–781.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (First edition). St. Martin's Press.

Farkas, K. (2023). The Lives of Others. *Aristotelian Society Supplementary Volume*, 97(1), 104–121.

Fichte, J. G. (1982). *Science of Knowledge: With the First and Second Introductions* (P. Heath & J. Lachs, Eds.). Cambridge University Press.

Fine, K. (1995). Ontological Dependence. *Proceedings of the Aristotelian Society*, 95(1), 269–290.

Fisher, E. (2022). *Algorithms and Subjectivity: The Subversion of Critical Knowledge*. Routledge, Taylor & Francis Group.

Foucault, & Gordon. (1980). *Power/knowledge: Selected Interviews and Other Writing, 1972-1977*. Harvester Wheatsheaf.

Foucault, M. (1982). The Subject and Power. *Critical Inquiry*, 8(4), 777–795.

Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison* (A. Sheridan, Trans.; Second Vintage books edition). Vintage Books.

Foucault, M. (with Faubion, J.). (2000). *The Essential Works of Michel Foucault, 1954-1984. Vol. 3*. Penguin.

Franda, F. (2021). *Social Kinds, Social Objects, and Vague Boundaries*. 2nd International Workshop on Ontology of Social, Legal and Economic Entities (SoLEE).

Franda, F. (2022). *Social Kinds: A User's Manual (1st edition)* [PhD Thesis, University at Buffalo].

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5–20.

Gallotti, M., & Michael, J. (Eds.). (2014). *Perspectives on Social Ontology and Social Cognition*. Springer.

Gilbert, M. (1992). *On Social Facts*. Princeton University Press.

Gilbert, M. (1996). *Living Together: Rationality, Sociality, and Obligation*. Rowman & Littlefield Publishers.

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (p. 0). The MIT Press.

Gillespie, T., Boczkowski, P. J., & Foot, K. A. (Eds.). (2014). *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press.

Godman, M. (2021). *The Epistemology and Morality of Human Kinds*. Routledge, Taylor & Francis Group.

Greenwood, J. D. (1997). *The Mark of the Social: Discovery or Invention?* Rowman & Littlefield Publishers.

Griffith, A. M. (2017). Social Construction and Grounding. *Philosophy and Phenomenological Research*, 97(2), 393–409.

Griffith, A. M. (2018). Social Construction: Big-G Grounding, Small-G Realization. *Philosophical Studies*, 175(1), 241–260.

Guala, F. (2010). Infallibilism and Human Kinds. *Philosophy of the Social Sciences*, 40(2), 244–264.

Guala, F. (2014). On the Nature of Social Kinds. In M. Gallotti & J. Michael (Eds.), *Perspectives on Social Ontology and Social Cognition* (pp. 57–68). Springer Netherlands.

Guala, F., & Hindriks, F. (2023). The Nature and Significance of Social Ontology. *Synthese*, 201(4), 145.

Hacking, I. (1990). *The Taming of Chance*. Cambridge University Press.

Hacking, I. (1995). The Looping Effects of Human Kinds. In *Causal Cognition: A Multidisciplinary Debate* (pp. 351–394). Clarendon Press/Oxford University Press.

Hacking, I. (1997, August 21). Taking Bad Arguments Seriously. *London Review of Books*, 19(16).

Hacking, I. (1999). *The Social Construction of What?* Harvard University Press.

Hacking, I. (2006, August 17). Making Up People. *London Review of Books*, 28(16).

Hacking, I. (2007). Kinds of People: Moving Targets. In H. Ian (Ed.), *Proceedings of the British Academy, Volume 151, 2006 Lectures* (pp. 285–318).

Harré, R. (1997). Crews, Clubs, Crowds, and Classes: “The Social” as a Discursive Category. In *The mark of the social: Discovery or invention?* (pp. 199–212). Rowman & Littlefield Publishers.

Haslanger, S. A. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.

Hosanagar, K. (2019). *A Human's Guide to Machine Intelligence: How Algorithms Are Shaping Our Lives and How We Can Stay in Control*. Viking Press.

Hume, D. (with Nidditch, P. H.). (1978). *A Treatise of Human Nature* (L. A. Selby-Bigge, Ed.; Second edition). Clarendon Press; Oxford University Press.

Ian, H. (1999). Making Up People. In *The Science Studies Reader* (pp. 161–171). Routledge.

Jaton, F. (2020). *The Constitution of Algorithms: Ground-truthing, Programming, Formulating*. The MIT Press.

József, A. (1999). *The Iron-blue Vault: Selected Poems* (Z. Ozsváth & F. Turner, Trans.). Bloodaxe.

Kafka, F. (with Parry, I.). (2019). *The Trial*. Penguin Books.

Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press.

Khalidi, M. A. (2015). Three Kinds of Social Kinds. *Philosophy and Phenomenological Research*, 90(1), 96–112.

Khalidi, M. A. (2016). Mind-Dependent Kinds. *Journal of Social Ontology*, 2(2), 223–246.

Khalidi, M. A. (2022). *Social Ontology and Linguistic Reclassification* [Video recording of conference presentation]. <https://www.youtube.com/watch?v=PPrHwg3g7Z0>

Khalidi, M. A. (2024). Historical Kinds in the Social World. *Philosophy of the Social Sciences*, 00483931241228906.

Kierkegaard, S. A., Piety, M. G., & Mooney, E. F. (2009). *Repetition. Philosophical crumbs*. Oxford University Press.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805.

Koslicki, K., & Massin, O. (2023). A Plea for Descriptive Social Ontology. *Synthese*, 202(3), 60.

Lash, S. (2007). Power after Hegemony: Cultural Studies in Mutation? *Theory, Culture & Society*, 24(3), 55–78.

Lechner, F. J., & Silver, D. (2022). Georg Simmel's "How Is Society Possible?": A Dialogue. *Simmel Studies*, 26(2), 9–9.

Lee, R. L. M. (1997). The Reversible Imaginary: Baudrillard and the End of the Social. In *The mark of the social: Discovery or invention?* (pp. 253–272). Rowman & Littlefield Publishers.

Leibniz, G. W., Remnant, P., & Bennett, J. (1996). *New Essays on Human Understanding*. Cambridge University Press.

- Lippmann, W. (1998). *Public Opinion* (2. print). Transaction Publishers.
- List, C. (2021). Group Agency and Artificial Intelligence. *Philosophy & Technology*, 34(4), 1213–1242.
- Locke, J., & Nidditch, P. H. (1975). *An Essay Concerning Human Understanding*. Clarendon Press.
- Lukes, S. (2005). *Power: A Radical View* (Second edition). Macmillan International Higher Education ; Red Globe Press.
- Lukes, S. (2021). *Power: A Radical View* (Third edition). Bloomsbury Academic, Red Globe Press.
- Lyon, L. (2019). Hyenas Probably Have More Friends Than You: Spotted hyena social hierarchies. *Science in the News*. <https://sitn.hms.harvard.edu/flash/2019/hyenas-probably-have-more-friends-than-you-spotted-hyena-social-hierarchies/> (retrieved: 26th February 2024)
- Madrigal, A. C. (2012, March 1). Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/>
- Madrigal, A. C. (2014, January 2). How Netflix Reverse-Engineered Hollywood. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/>
- Mason, R. (2016). The Metaphysics of Social Kinds. *Philosophy Compass*, 11(12), 841–850.
- Mason, R. (2021). Social Kinds Are Essentially Mind-Dependent. *Philosophical Studies*, 178(12), 3975–3994.
- Mason, R., & Ritchie, K. (2020). Social Ontology. In R. Bliss & J. Miller (Eds.), *The Routledge Handbook of Metametaphysics*. Routledge.
- Mikkola, M. (2015). Doing Ontology and Doing Justice: What Feminist Philosophy Can Teach Us About Meta-Metaphysics. *Inquiry*, 58(7–8), 780–805.
- Mikkola, M. (2017). On the Apparent Antagonism Between Feminist and Mainstream Metaphysics. *Philosophical Studies*, 174(10), 2435–2448.
- Mill, J. S. (1974). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. University of Toronto Press; Routledge & Kegan Paul.
- Morriss, P. (2002). *Power: A Philosophical Analysis* (2nd ed). Manchester University Press; Distributed exclusively in the USA by Palgrave.
- Neff, G., & Nafus, D. (2016). *Self-tracking*. The MIT Press.

Nietzsche, F. W. (1974). *The Gay Science: With a Prelude in Rhymes and an Appendix of Songs* (W. Kaufmann, Trans.; First edition). Vintage Books.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.

Oppenheim, F. E. (1981). *Political Concepts, a Reconstruction*. University of Chicago Press.

Östlund, R. (Director). (2022). *Triangle of Sadness*. Imperative Entertainment, Plattform Produktion, Neon.

Pasquale, F. (2015). *The Black Box Society: The secret Algorithms That Control Money and Information*. Harvard University Press.

Passinsky, A. (2020a). Social Entities. In M. J. Raven (Ed.), *The Routledge Handbook of Metaphysical Grounding* (pp. 510–520). Routledge.

Passinsky, A. (2020b). Social Objects, Response-Dependence, and Realism. *Journal of the American Philosophical Association*, 6(4), 431–443.

Passinsky, A. (2021). Finean Feminist Metaphysics. *Inquiry*, 64(9), 937–954.

Petersmann, M., & Van Den Meerssche, D. (2024). On Phantom Publics, Clusters, and Collectives: Be(com)ing Subject in Algorithmic Times. *AI & SOCIETY*, 39(1), 107–124.

Reeves, R. (2023). The Ezra Klein Show: Opinion | The Men—And Boys—Are Not Alright. *The New York Times*. <https://www.nytimes.com/2023/03/10/opinion/ezra-klein-podcast-richard-reeves.html>

Richardson, K. (2023a). The Social Turn in Analytic Philosophy: Promises and Perils. *Daily Nous*. <https://dailynous.com/2023/08/08/the-social-turn-in-analytic-philosophy-promises-and-perils-guest-post/>,

Richardson, K. (2023b). Critical Social Ontology. *Synthese*, 201(6), 1–19.

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton University Press.

Rouvroy, A. (2013). The End(s) of Critique: Data Behaviourism versus Due Process. In *Privacy, Due Process and the Computational Turn* (pp. 143–167). Routledge.

Rouvroy, A. (2016, January 22). Big data: L'enjeu est moins la donnée personnelle que la disparition de la personne. *Le Monde.Fr*. <https://www.lemonde.fr/blog/binaire/2016/01/22/le-sujet-de-droit-au-peril-de-la-gouvernementalite-algorithmique/>

Rouvroy, A., & Berns, T. (2010). The New Statistical Power: Or When Control Is Exercised on a Standardized, Docile, and Eventless Reality, Made Up of “Digital” Bodies. *Multitudes*, 40(1), 88–103.

Rouvroy, A., & Berns, T. (2013). Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? *Réseaux*, 177(1), 163–196.

Rouvroy, A., Berns, T., & Carey-Libbrecht, L. (2013). Algorithmic Governmentality and Prospects of Emancipation. Disparateness as a Precondition for Individuation Through Relationships? *Réseaux*, 177(1), 163–196.

Sartre, J.-P. (with Warnock, M., & Eyre, R.). (2003). *Being and Nothingness: An Essay on Phenomenological Ontology* (H. E. Barnes, Trans.). Routledge.

Schaffer, J. (2017). Social Construction as Grounding; or: Fundamentality for Feminists, a Reply to Barnes and Mikkola. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 174(10), 2449–2465.

Schaffer, J. (2019). Anchoring as Grounding: On Epstein's the Ant Trap. *Philosophy and Phenomenological Research*, 99(3), 749–767.

Schmitt, F. F. (2003). *Socializing Metaphysics: The Nature of Social Reality*. Rowman & Littlefield Publishers.

Schuilenburg, M., & Peeters, R. (Eds.). (2020). *The Algorithmic Society: Technology, Power, and Knowledge*. Routledge.

Searle, J. (2010). *Making the Social World: The Structure of Human Civilization*. Oxford University Press.

Searle, J. R. (1995). *The construction of social reality*. Penguin Books.

Searle, J. R. (2006). Social Ontology: Some Basic Principles. *Anthropological Theory*, 6(1), 12–29.

Sider, T. (2011). *Writing the Book of the World*. Clarendon Press; Oxford University Press.

Simmel, G. (1910). How is Society Possible? *American Journal of Sociology*, 16(3), 372–391.

Steiner, C. (2012). *Automate This: How Algorithms Came to Rule Our World*. Portfolio/Penguin.

Sveinsdóttir, Á. (2015). Social Construction. *Philosophy Compass*, 10(12), 884–892.

Taylor, C. (1984). Foucault on Freedom and Truth. *Political Theory*, 12(2), 152–183.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Revised and expanded edition). Penguin Books.

Thomasson, A. L. (2003a). Foundations for a Social Ontology. *ProtoSociology*, 18/19, 269–290.

Thomasson, A. L. (2003b). Realism and Human Kinds. *Philosophy and Phenomenological Research*, 67(3), 580–609.

Thompson, N. (2018). When Tech Knows You Better Than You Know Yourself. *Wired*. <https://www.wired.com/story/artificial-intelligence-yuval-noah-harari-tristan-harris/>

Tufekci, Z. (2014). Engineering the public: Big Data, Surveillance and Computational Politics. *First Monday*.

Tuomela, R. (2002). *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge University Press.

Velasco, P. F. (2023). Social Reality Without Language: Enacted Representations, Not Declarations, Create Status Functions. *Journal of Social Ontology*, 9(1), Article 1.

Vico, G. (1974). *La scienza nuova. Giusta l'edizione del 1744* (F. Nicolini, Ed.). Laterza.

Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology* (G. Roth & C. Wittich, Eds.; E. Fischoff, Trans.). University of California Press.

Wennerås, C., & Wold, A. (1997). Nepotism and Sexism in Peer-Review. *Nature*, 387(6631), 341–343.

Wilson, R. A. (2007). Social Reality and Institutional Facts: Sociality Within and Without Intentionality. In S. L. Tsohatzidis (Ed.), *Intentional Acts and Institutional Facts: Essays on John Searle's Social Ontology* (pp. 139–153). Springer Netherlands.

Yeung, K. (2017). 'Hypernudge': Big Data as a Mode of Regulation by Design. *Information, Communication & Society*, 20(1), 118–136.

Zahavi, D. (2023). Observation, Interaction, Communication: The Role of the Second Person. *Aristotelian Society Supplementary Volume*, 97(1), 82–103.

Zhang, C. (2020). Governing (Through) Trustworthiness: Technologies of Power and Subjectification in China's Social Credit System. *Critical Asian Studies*, 52(4), 565–588.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (First edition). PublicAffairs.