LUND UNIVERSITY

Strategic moral communication

A metatheoretical and methodological response to the normative perspective on strategic communication

Fjällhed, Alicia

2023

Link to publication

Citation for published version (APA):
Fjällhed, A. (2023). *Strategic moral communication: A metatheoretical and methodological response to the normative perspective on strategic communication*. [Doctoral Thesis (monograph), Department of Strategic Communication]. MediaTryck Lund.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

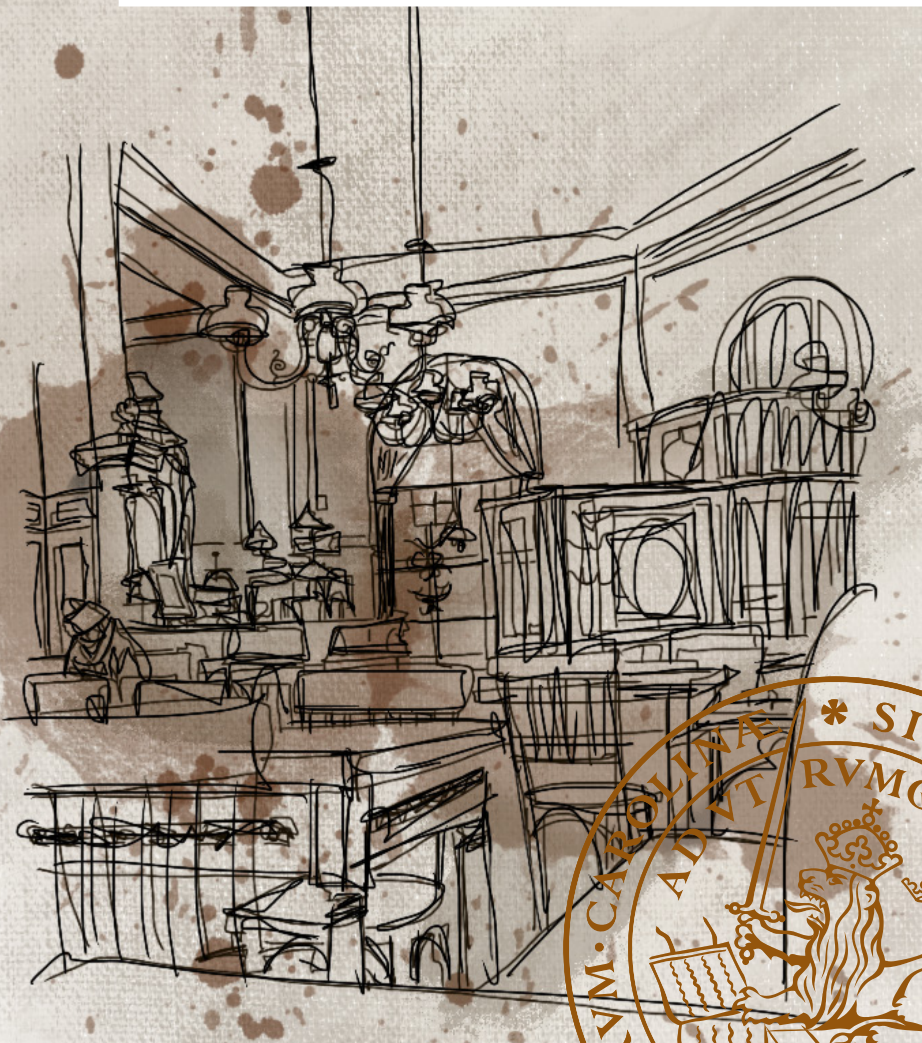
PO Box 117
221 00 Lund
+46 46-222 00 00

Strategic moral communication

A metatheoretical and methodological response to the normative perspective on strategic communication

ALICIA FJÄLLHED

DEPARTMENT OF STRATEGIC COMMUNICATION | LUND UNIVERSITY



Strategic moral communication

Strategic moral communication

A metatheoretical and methodological response to the normative perspective on strategic communication

Alicia Fjällhed

LUND
UNIVERSITY

Coverphoto by Alicia Fjällhed

Copyright pp 1–260 Alicia Fjällhed

Faculty of Social Sciences
Department of Strategic Communication

ISBN 978-91-8039-908-1 (printed) 978-91-8039-909-8 (electronic)
ISSN 1104-43-30

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN

The Vienna coffee house and the unique coffee house culture lived there is not only a place of communication but also a place of encounter. In 1720, the Kramersche Kaffeehaus am Graben was the first coffee house in Vienna that also published newspapers and thus provided its guests with news from all over the world. The Vienna coffee house is considered a mecca of communication, a successful balancing act between tradition and modernity as technology has long since found its way into Vienna's coffee houses and offers its customers, for example, wireless Internet access. In coffee houses, space and time are consumed, but only the coffee is billed.

Translation of application to
UNESCO's Intangible Cultural Heritage

Acknowledgements

There is a reason why the acknowledgements are placed at the beginning of a thesis, because without these people, the text would not exist. First, I would like to thank my supervisors James Pamment and Howard Nothhaft. I accepted this position on the condition that you would be my guides, and it is the most important decision I have made throughout my studies. I cannot fathom the patience you have mastered when reading hundreds of pages of unfinished ideas. You are among the very few whose opinion I trust will be completely honest, and for that I am truly grateful. On top of that, your generous invitations to your networks, to be a co-contributor in your research, teaching, or public outreach projects have created invaluable opportunities for me to learn the academic craft. I would also like to thank the readers at my thesis seminars, with internal readers including Henrik Merkelsen, Charlotte Simonsson, Nils Gustafsson, Marlene Wiggill, Monica Porzionato, and with a special thank you to the external readers Lars Rademacher and Alexander Buhmann for your encouragement and critical comments to the seeds and drafts of this thesis. During my studies, I have also had the chance to meet and learn from a wide range of scholars through co-authorship, research groups, academic exchanges, at conferences, PhD-courses, in teaching teams, and from the professional community through commissioned work. I have had the privilege of meeting brilliant minds, kind hearts, and funny bones. This thesis is very much influenced by our conversations. Finally, a special thank you to Tina and Rickard, Isabelle, Kristijan, Hui, Monica, Ester, Martina, and Jens, with whom I have shared a pandemic, who have been there to celebrate the highs and are among the few who understand the lows, accompanied by a taste of pub grub and breakfast, tacos, or snails. To my friends around Sweden, my dad, my sister, and all the other people I call family—I am beyond grateful for the pauses from academic life that you have provided. But there is one to whom I cannot say thank you to, because it is simply not enough… Mom, roots and wings!

Preface

This thesis addresses the communicative ideal which has been taken as a normative matter of course in many democratic societies and contemporary scholarship in the social sciences. The ideal is theoretically captured in Jürgen Habermas' moral theory called discourse ethics, in which communication is placed at the epicentre of morals. He is recognised as one of our most influential contemporary philosophers, and many rest on his theory explicitly, using it as a theoretical framework in their studies. Others do it implicitly, by resting on his norms or by using methodologies such as critical discourse analysis, as this is built on Habermas' normative assumptions. This thesis even argues that the ideal has become so widespread in our time that many rest their everyday claims of validity on this theory's norms, without recognising the assumptions' normative character. The problem is, however, not the theory itself, but when we fail to recognise that it is a normative theory.

Habermas was very clear that his theory was a normative ideal, not claiming it as something that has been, is, or ever will or could be. It was never intended to be used as a framework that could explain human behaviour, it only enabled a normative judgement in relation to the ideal. Habermas presents the ideal as emerging with a new public sphere in parallel with the first cafés—the historic coffee houses. Here, society was described to have gained a new public space in which people could meet to discuss common issues. By airing various claims, assessing their rational validity through a process of argumentation, and eventually reaching a consensus through common deliberations, the public would reach a conclusion about their shared morals. But again, this is only an ideal, not necessarily manifesting in reality. Habermas accentuates that it is an ideal and not a practice emerging with the coffee houses, formulating his theory as a counterfactual theory and a criticism towards his contemporary society. And as a theory introduced just before the internet, it does not address the digital structures emerging in the new communication landscape. Rather, since the introduction of the theory, the contemporary coffee houses in Vienna—such as the illustration on this thesis' shows—is a scenery in which the public discourse has moved away from the public coffee houses to a digital discourse online. This new digital landscape was initially celebrated as a scenery that could enable Habermas' ideal discourse. It was prophesised to provide a space where anyone could enter a common conversation, air their claims, and engage in a rational conversation with people in their local, national, and international community to arrive at a common consensus. However, this discourse shifted in the middle of the 2010's toward a concern debate when observing the opposite landscape emerging. Critical voices described that people did not engage in a common discourse but were locked into echo chambers, they did not build a conversation on valid claims but on alternative facts, and instead of a rational discourse one would find a post-truth environment.

Habermas' theory was never intended as a suitable framework to explain real behaviour, but depicted an ideal. As such, this thesis argues that we need a new theory of moral communication that moves from a normative to an empirical stance to explain the observed behaviour in moral discourses. And the thesis aims to do just that. It departs from Habermas' theory of discourse ethics, reviewing it in relation to moral psychology's contemporary insights into the rationalising rather than rational function of communication in moral discourses. From this stance, the thesis develops a new theory of *strategic moral communication*. Adding to that, just as Habermas' discourse ethics has built normatively driven methodologies such as critical discourse analysis, this thesis develops its own empirical methodology. From the new theoretical stance, the thesis introduces *moral discourse analysis*. This methodology seeks to enable the analysis of moral discourses without passing a normative judgement from an external moral context. By studying moral discourses, the thesis argues that the methodology can be used to understand moral zeitgeists. The thesis shows one example of such a study, by engaging in a review of its own of the contemporary moral discourse on fake news. This is argued to be a moral discourse in which communicative phenomena are presented as moral or immoral by introducing new normative concepts and definitions. By using the thesis' new understanding of moral communication as strategic, it argues that even Habermas himself, the scholars adopting his theoretical stance, and those using critical discourse analysis, also engage in this practice. The thesis argues that their own communicative behaviour runs contrary to their ideal, that they engage in the very practice of strategic communication which they have previously presented as immoral. Based on the review of these scholarly discourses on the moral content of communication, as well as based on the empirical moral discourse on fake news, the thesis comes to a new conclusion about the moral nature of strategic communication that runs contrary to previous moralised ideas of the concept. Instead of echoing that strategic communication would be immoral, the thesis argues that moral communication is strategic.

The scope of the thesis is highly influenced by the time and place in which it was developed—emerging in parallel with an international and social-wide concern about a new cluster of phenomena addressed under the umbrella term 'fake news'. Before being accepted to the PhD-program, I was engaged in a series of projects related to this discussion. During my master's, I wrote two theses in 2016–2017 (later published as Fjällhed, 2020) on the Swedish Ministry for Foreign Affairs' initial steps to deal with the problem of disinformation in an increasingly polarised political landscape; a discussion emerging alongside the European refugee/migration crisis in 2015. Later, considering Russia's annexation of Crimea in 2014, and in particular the US election race of 2016, these discussions grew as a concerned discourse around the world on the new digital communication landscape as an increasingly hostile environment. Soon, I became part of a team at the Department of Strategic Communication at Lund University. The team was formed after a commission from the Swedish Civil Contingency Agency ahead of the 2018 national elections to develop the knowledge on how to understand, identify, and counter the problem of foreign election interference.[1] We continued publishing reports (i.e., NATO StratCom COE, 2019) and handbooks,[2] and based on these publications organised training and scenario exercises, provided process and policy support, and either moderated, participated, or were assigned to comment on discussions for local, regional, government, and international organisations worldwide. Today, the group has grown into the Lund University Research Institute for Psychological Defence, and I have continued to publish academically on the same phenomena.[3]

---

[1] This resulted in a report mapping the state of the art on foreign actors' tactics to manipulate elections in democracies (Swedish Contingency Agency, 2018) and a practical handbook for communicators published in Swedish (Myndigheten för Samhällsskydd och Beredskap, 2018) and English (Swedish Civil Contingency Agency, 2018a). Since then, Finnish authorities have re-published their own Swedish version (Stadsrådets Kansli, 2019) and Finnish version (Valtioneuvoston Kanslia, 2019) of the handbook. It has also shaped the Swedish Psychological Defence Agency's handbook (Myndigheten för Psykologiskt försvar, 2019).

In parallel to these projects, I started my PhD-studies. The project departed from a feeling that something was off in the debate. In time, I was able to articulate this feeling as a discrepancy between the norms advocated by actors critical towards fake news and their communicative response to the problem. As I was interested in communicative dimensions arising in a moral discourse in the public sphere, I was repeatedly directed to Habermas' moral theory of communication as a potential theoretical framework for the thesis' analysis. However, I soon realised that it could not help me understand what was happening. For as a normative framework, I could only use his theory to pass judgement. It could be used an argument for why fake news, echo chambers, alternative facts, and post-truth was morally wrong. It could not, however, help me understand why this was happening. To that end, I continued reading, not carrying the tradition of looking within a particular set of journals or in one aisle in the library. As I moved from the concept of 'moral' to other concepts such as 'perspective taking' and 'empathy,' I found my way to moral psychology. When reading up on the pillars framing the field today, I found that Habermas' theory was indeed resting on moral psychology, but on a now replaced paradigm. In short, the field had moved from the idea that morals are the product of rational conversations, to the uncomfortable realisation that we use communication to rationalise our moral intuitions. This new paradigm presented an opportunity to build a new theory, replacing Habermas' discourse ethics with a theory resting on this contemporary empirical understanding of moral reasoning. And just as Habermas' discourse ethics had shaped methodologies such as critical discourse analysis, the new theoretical stance opened the opportunity—in part necessitated—the development of a new methodology that could be used to study this new conception of moral communication from an empirical rather than normative stance. The new methodology would enable the study of moral discourses from an empirical rather than normative stance. Through this new approach, strategic communication's relation to morals in practice could also be studied without carrying the normative judgement embedded in moralised representations of strategic communication, such as presented in discourse ethics and critical discourse analysis.

---

[2] Including a toolkit for UK authorities (UK Government Communication Services, 2019) translated into Czech, Spanish and Mandarin. This toolkit was later updated after lessons learned during the Covid-19 pandemic (UK Government Communication Services, 2021).

[3] To take a few examples, publications such as a presentation of the Swedish approach to manage foreign election interference (Fjällhed, Pamment, & Bay, 2021), on how disinformation was affecting organisations' reputation management (Johansson, Nothhaft, & Fjällhed, 2023), on the geopolitical implications stemming from the new algorithmic landscape (Fjällhed, Sandre, & Lüfkens, 2024), and outlining a strategic research agenda for the study of disinformation in public diplomacy (Fjällhed & Pamment, 2023).

The thesis' contribution should also be seen in light of the fact that I am a PhD-student in strategic communication. As such, I am both drawing on this perspective and develop my contribution in part to this field. My stance also explains the study itself, as it arose not only from the observed discrepancy in the public discourse around fake news, but also an observed discrepancy in scholars' presentation of strategic communication. My field's conceptualisation of strategic communication was not mirrored in the social science's understanding of the same concept. Instead, the prevailing view very much echoes Habermas' presentation of strategic communication as immoral communication. My idea of the field's scope worked under a different paradigm. With a bachelor's and master's degree in strategic communication, and after some ten years of working in public, private, and non-profit organisations as a communicator during and in-between studies, my idea of strategic communication was that it simply addressed 'purposeful communication'. It was neither 'good' nor 'bad' in itself, but a neutral tool used by actors such as organisations and individuals to influence other actors.[4] Until I started writing my master's theses in

---

[4] In my professional capacity, the scope as a strategic communicator could include efforts to inform travellers where to find the replacement bus for their cancelled train. In other jobs, I was set to design persuasive campaigns to encourage Christmas shoppers to donate money to children whose parents could not afford presents. I developed a graphical profile to create a holistic visual communication for one organisation, helped

2017, I had rarely encountered critical debates about strategic communication within the field. If so, they were limited to critical historical accounts or deceptive advertising. As a PhD-student, I started to see how my neutral sense of strategic communication meant that disinformation and many other phenomena within the discourse around fake news were also a part of my field's scope. However, when searching for scholarly entries about strategic communication beyond my field's ongoing conversation, I found that the most cited papers were using it for this type of communication only, as a common-sense word (rather than a defined concept, and certainly not as a field of study in its own right) for all types of deceptive communication; the modern-day equivalent of the classic unethical label propaganda. The pattern was repeated when I attended international conferences as a PhD-student. For example, at the International Communication Association's (ICA) conference, I listened to a panel where moral communication was presented as dialogical, two-way, and engaging one's audience. In contrast, strategic communication was always the opposite, an immoral evil twin. At the International Studies Association's (ISA) annual conference, another panel presented strategic communications as part of the military's toolbox to engage in psychological operations—to defensively deter one's enemies abroad from attacking or to strengthen public resilience and the will to fight among one's population. Whenever I found myself outside the narrow groups with the same conception of strategic communication as I, I found many linking it rather to contested or straight-out immoral forms of communication. In time, I came to define my own interest in strategic communication as formed from an empirical rather than normative (and often moralised) definition, not using it as a term to pass moral judgement but merely as a term which captures a type of communication which seeks to or achieves influence. Today, I see my empirical rather than normative interest in strategic communication as the foundation for my contribution to the transdisciplinary discussions on the relation between communication, ethics, and influence.

---

another client produce a video for an IT company's employees to inform them about a new internal process, and for a third client brainstormed a creative campaign to gain public support from the neighbourhood where a retailer planned to open a new store.

At its core, the thesis is a re-interpretation of strategic communication's relation to morals—unlocked by recognising the normative character of moralised descriptions of strategic communication. Instead, the thesis rests on empirical proofs from contemporary moral psychology, allowing for an empirical rather than normative stance. It grounds the thesis' dual proposition for a new moral theory and methodology to understand and research strategic communications' relation to morals. These contributions are bound to the thesis' shift from normative communication theories such as Habermas' discourse ethics and methodological frameworks such as critical discourse analysis. Recognising that such frameworks are moralising by default, as both envision how the world could be a better place, the key lies in the word *better* as this points to their normative and neither objective nor descriptive, explanatory, or empirical aim. In discourse ethics, this is expressed as a moralising presentation of communication seeking or achieving influence, where critical discourse analysis builds upon the same stance to propose approaches for how to study, reveal, and emancipate society from such evils. On the one hand, the thesis is a contribution to the field of strategic communication by a PhD-student from the field—contributing with a conceptual discussion on the definition of strategic communication and addressing the neglected discussion about morals and ethics. Aside from being a contribution to strategic communication, one can also see the thesis as a transdisciplinary contribution to the social sciences. It shows the relevance of understanding strategic communication's relation to morals in an empirical rather than normative way that recognises that strategic communication is not immoral by default, but a key component if we are to understand how people engage in moral reasoning.

Contents

Acknowledgements .... *vii*
Preface .... *ix*

Introduction 1
Outlines *17*

PART 1. THEORY 19

Chapter 1. Discourse ethics 21
Chapter 2. 40 years of moral reasoning 43
Chapter 3. Strategic moral communication 57
Summary *71*

PART 2. METHODOLOGY 73

Chapter 4. Moral discourse analysis 75
Chapter 5. Case study 95
Summary *106*

PART 3. ANALYSIS 107

Chapter 6. Fake news 109
Chapter 7. Echo chambers 121
Chapter 8. Alternative facts 139
Chapter 9. Post-truth 155
Summary *168*

Chapter 10. An emancipatory framework 171

Conclusion 189

Appendix *197*
Index *219*
References *221*
Notes *251*

Introduction

Morals are expressions of a community's consensus, arrived at through a rational conversation between its members. The immoral equivalent is found in those who enter this conversation with a pre-determined purpose, and who pursue self-serving interests rather than explore the collective's interest together with the other participants. If you are part of a community idealising a democratic model for society,[5] you probably agree with such an image. The normative framework is used as a stance to engage in a moralised conversation about communication,[6] such as most recently in the fake news discourse. It has also influenced scholarly discussions, forming the theoretical assumptions in frameworks such as discourse ethics and methodological frameworks such as critical discourse analysis. When used as moralising frameworks to critically evaluate and direct actions toward achieving this ideal, it is a suitable framework. However, when presented as a scholarly theory it was not claiming to present an empirical framework for what has been, is, or perhaps ever will manifest in observable discourses, it was only presented as a widespread norm. The problem is therefore not these frameworks *per se* but when we use them to seek understanding of observable dynamics in moral discourses.

The problem emerges as an empirical, theoretical, and methodological problem. First, it emerges as an empirical problem when assuming that this normative stance is an empirical reality or even possibility. For this means that we assume that people will accept, behave, or at least strive to behave, in accordance with the norm. The discourse around fake news, for example, builds on this stance, framed as a normative criticism towards communicative phenomena created by the digital landscapes. However, we can only use this framework to cast judgement for what ought to or ought not to be communicative practices present in public discourse. It cannot, however, be used as a theory to understand observable dynamics in such discourse. For what we find when reviewing this discourse is a depiction where people are not engaging in an inclusive conversation but get stuck in echo chambers, instead of an informed debate based on valid claims they are described as accepting alternative facts, and finally the discourse presents the Enlightened ideal of arriving at rational conclusions as replaced by a post-truth environment. Indeed, Habermas also observed behaviour running contrary to his ideal when introducing his theory, presenting his proposition for a moral theory of discourse ethics as a normative counterfactual proposition and critique against his own time. In the contemporary discourse, some stay within the lines of his ideal. Others break with it, or even actively target vulnerabilities created by the ideal assumptions about the same system, breaking with the democratic norms to assert what democratic actors would present as illegitimate, immoral, or unethical influence. By assuming the ideal as a reality, we are less likely to understand these phenomena.

---

[5] Henceforth using the label *democracy* to reflect the thesis' theoretical and empirical material's choice of words, where these actors interchangeably also use an added emphasis on this being a *deliberative, liberal,* or *Western democracy*.

[6] Resting on Habermas' distinction that 'as we hypothetically consider claims to validity, the world of intuitionally ordered relations becomes *moralized,* just as the world of existing states of affairs become *theorized*' (1999, p. 107).

At the same time, Habermas' ideal is deeply embedded in the discourse's judgement towards fake news. For since the introduction of his theory, Habermas has become an influential scholar. His theory has been used by many others, not least since the introduction of new social media. At first, the internet was prophesised to enable the ideal to materialise in the new digital landscape. However, in time the initially enthusiastic cheers shifted toward a critical concern for how the very same infrastructure may turn out to be detrimental for the advancement of Habermas' ideal. In the discourse framing these new phenomena as morally problematic, the thesis will argue that we see how Habermas' theory forms the normative stance for such assessments. In short, they manifest as problematic phenomena *because* the phenomena threaten the ideal. In addition, the thesis argues that these actors' attempts to manage the problem work contrary to their own ideals. For example, by restricting access to the public discourse for those who do not conform to the basic rules parts of the public are excluded from the envisioned inclusive discourse. The ideal does therefore not only define the immoral problem (and the moral solution), but the thesis also points to the real limitations of the ideal as it finds that the very counter-measures risk counteracting the ideal it seeks to protect. Furthermore, from a theoretical and methodological perspective, adopting a normative stance without recognising the presence of a normative stance leads to analyses that presuppose the norm as a natural law. Indeed, we can use these frameworks to direct attention to the extent of the correlation between the ideal and observed reality, and suggest actions that seeks to close the gap. However, these theories and methodologies cannot be used to seek an understanding of communication in moral discourses. Consequently, such a normative theory, and the methodologies resting on the same stance to study moral communication, will say little about communication's empirical relation to morals. As the fake news discourse illustrates, the normative standpoint struggles to explain why the public are locked into echo chambers, accept alternative facts, and entertain post-truth rationalisations. For this we would need an empirical rather than normative theory to understand and a new methodology to study moral discourses.

This thesis pursues such a purpose—moving from a normative to an empirical perspective on strategic communication's relation to morals.[7] From this new theoretical stance, it also proposes a new methodology to study this conceptualisation of moral communication. This shift in perspective leads to the argument that the above-mentioned normative frameworks engage in the very practice of strategic communication that they describe as immoral. They engage in conversations where they present their normative ethics as a default that others ought to follow. Such actors do not enter conversations with an open mind, but imply a strategic intent to be persuasive. This thesis examines how communication is used in moral conversations, such as in these theories, methodologies, and in the public discourse. What emerges is a picture diverging from the previously dominating normative frameworks, as the thesis finds that not all strategic communication is immoral, but that moral communication is strategic.

---

[7] Drawing on Habermas' (2022b) distinction between *normative* and *empirical theories*, the thesis challenges his description of discourse ethics as an empirical theory. Rather, the thesis argues it to be a normative theory, where the thesis itself proposes an empirical alternative.

A field haunted by public scepticism

For as long as communication has been used as a tool to assert influence, it has been met with criticism. Some of the earliest records are found in ancient Greece, as the private sphere was expanded to a public one. In the new democratic society, communication became an essential tool. It was formed in the liberal arts through the trivium, where one learned the grammar of how to turn sense impressions into sensible linguistic expressions, the logic of factually correct construction of texts, and the rhetoric of 'the art of communicating thought from one mind to another' (Joseph, 2002, p. 3). Be you a politician or businessperson in the agora or an accused facing the court, your success could depend not on the merit of your claim but on the clarity, logic, and rhetorical appeal of your argument. The latter of these three skills, however, was met with scepticism. Plato's record has particularly influenced our contemporary view. He contrasts his and his fellow philosophers' search for truth through a dialectical method using reason, with their rivals, the Sophists. The moral ideal was embodied in his presentation of Socrates, who, in the face of the death penalty, still did not use rhetorical tricks to persuade the court of his innocence, and ultimately suffered a martyr's death in the name of his communicative ideals. Contrary, Plato describes the Sophists as wielding their skills with words to achieve a pre-determined and self-serving interest. In *Sophist* (360 BC), Plato describes them as imitators, creating images by juggling with words. To this day, this conception echoes in our everyday vocabulary as sophisms refer to fallacious arguments used to deceive and sophists to the people engaged in this practice.

As a group of people teaching others the skill of asserting influence through communication, sophists can also be considered the first communication professionals—a professional community ever since haunted by moral criticism. Most famously, this is represented in the concept *propaganda*, initially coined in 1622 by the Vatican's *Congregatio di Propaganda Fide* (or *The Office for the Propagation of the Faith*) as a neutral if not positive practice of advocating for the divine truth. With the emergence of mass media, propaganda studies took its academic form. In time, this too gained an immoral valence considering the 21st century's use of propaganda in war (Taylor, 2003). However, the interest in communication as a tool for influence did not dissolve with this immoral reputation. Practitioners and scholars alike turned to new labels, repeating the process whenever public criticism arose. Zerfaß, Verčič, Nothhaft, and Werder (2018, p. 490) describe how the field of propaganda studies was introduced in the early 20th century, rebranded as public relations by the middle of it, again rebranded a few decades later as a new field of strategic communication. And now, once again, this latest of concepts is also irrevocably associated with immoral communication. When reviewing the most cited papers beyond the field, one finds titles such as *Using Privileged Information to Manipulate Markets* (Benabou & Laroque, 1992), *Strategic Communication with Lying Costs* (Kartik, 2009), and *Credulity, Lies and Costly Talk* (Kartika, Ottavianib, & Squintanic, 2006). At the same time, there have been scholars who continuously argue against this rebranding. They make a call to reclaim the label of propaganda in a non-normative sense as 'the *deliberate* attempt to persuade people to think and behave *in a desired way*' (Taylor, 2003, p. 6, original emphasis). In such definitions, the concept would not be tied to a normative judgement, but rest on a further assessment of propaganda as good *or* bad where 'good or bad depends upon the merit of the cause urged, and the correctness of the information published' (Bernays, 2005, p. 48). From this stance, advertising would be a type of economic propaganda and public relations 'a communicative process designed to enhance the relationship between an organisation and the public and, as such, is a branch of propaganda, albeit a nicer way of labelling it' (Taylor, 2003, p. 6). Nonetheless, the thesis recognises that strategic communication (under various labels) has predominantly been presented with an immoral valence as a moralised concept.

The problem with moralising scholarship

The public scepticism towards communicative influence has had a profound impact on the social sciences. In this thesis, it is even argued that the normative view reflects the moral air of our time, and that the reason that these theories and methodologies have become influential is that they align with our times' moral intuitions. It can be summarised as an ideal that all people in a community ought to have an equal say on common matters and that our opinions ought to be formed freely. This is today oftentimes approached as a universal truth. In taking this as an assumption, however, we fail to recognise that it is merely one of many possible moral points of departure for determining right from wrong, and for proposing scholarly perspectives on how society ought to progress. Normative theories and methodologies can never aim to fully comprehend what is happening, only judge the world from a pre-determined perspective. To illustrate this problem, the thesis takes a point of departure in two influential scholarly frameworks from the social sciences—the theory of *discourse ethics* and the methodology of *critical discourse analysis*.

Discourse ethics was introduced in the late 20st century by Jürgen Habermas as a moral theory of communication (unpacked in Chapter 1). His interest in communication had long influenced his understanding of morals, as he argued that morality is a sociological construct, established through a community's collective discourse. Continuously, he tied this to the ideal of democracy. He described how it first emerged as an idea in the era spanning the decades 1680 to 1730, when men gathered in public coffee-houses to engage in a critical discourse around common matters. He also continuously built on the public function of speech in ancient Greece, particularly on the theoretical ideal presented by Plato (see i.e., Habermas, 1994, p. 21). Like Plato, he contrasted truth-seeking communication with strategic communication, differentiating the latter from the former as driven by an actor's interest where they use communication to achieve pre-determined goals. Just as Plato phrased this as a practice conducted by the sophists of his time, Habermas' moral theory was a reaction to his own experience of political propaganda from the Nazis and as a criticism towards the systematic use of communication for economic propaganda such as in advertising and public relations.[8]

Habermas' theory has gained much influence, some even arguing that this aligns with the contemporary way of thinking in democracies to the point where it theoretically reflects our way of making moral sense of the world. Luhmann (1996), for example, describe how Habermas' discourse ethics was born and sustained within an 'institutional framework of the constitutional liberal state,' which 'made it possible to accept ethical theories that present the moral in the consensual domain' (p. 34). Marxsen (2019) similarly argues that Habermas' 'worldwide success may be attributed to the fact that he was able to provide the theoretical expression of an emerging *Zeitgeist*' (p. 480). This thesis would argue that such patterns live on to the present day, echoing in the normative discourse around fake news. For example, in one of the early scholarly commentaries from Latour (2004) upon his critical observance of climate change deniers, he argues that the world today relies on *matters of concern* rather than on *matters of fact*. His normative stance is described, for example by Keller (2016), as a proposition that 'end up as a recycled ethics of discourse' (pp. 61–62) echoing Habermas' theory. This, the thesis argues, reveal how Habermas' normative discourse ethics is still used as a measuring rod to evaluate what constitutes a good, ethical, moral, proper, rational, or intelligible public discourse. In doing so, however, we forget that it represents a normative perspective, as we present it as a natural rather than social matter of course.

---

[8] In part aligning with his critical colleagues at the Frankfurt School such as Adorno and Horkheimer's (1997) commentary on the *culture industry*.

The influence of Habermas' normative theory also expands through its influence over methodologies in the social sciences, such as *critical discourse analysis* (unpacked in Chapter 4). The methodology originates in Foucault's proposition for *discourse analysis,*[9] introduced as a method to study historic social realities by taking communication as the object of analysis. The key difference from Foucault's original proposition is found in critical discourse analysis' view of power. While Foucault argued that we can study power to understand why some social realities thrive while others die, critical discourse analysis approach power as a negative phenomenon and sets out to actively unveil hidden power structures through the academic method. Thus, critical discourse analysis is often presented as an emancipatory tool to help society rid itself of these communicative power structures. Critical discourse analysis rests on Habermas' normative framework, with previous scholars observing how critical discourse scholars 'tend to ground their critical practice and ethics in Habermas' normative concept' (Roderick, 2018, p. 157) as they assume 'the presupposition of a

---

[9] Applied in his own studies such as in *Madness and Insanity: History of Madness in the Classical Age* (2013), later more clearly addressed as a methodology in *The Will to Knowledge: History of Sexuality Volume I* (1990).

Habermasian communicative rationality' (p. 160). But today, voices echoing this normative stance, such as Latour (2004), have also started to criticise scholarship grounded in critical perspectives. In the case of Latour, he calls for a revision of the methodology, as he observes that it no longer is fit to address the new threats. Rather, he argues that the problem is reflected in critical analysists' own methodology. For in his concerned observations of climate change deniers and conspiracy theorists, Latour argues that these actors rely on a worldview where there is no such thing as facts, only power structures. As such, he continues that 'conspiracy theories are an absurd deformation of our own arguments, … like weapons smuggled through a fuzzy border to the wrong party, these are our weapons nonetheless' (p. 230). As Hook (2001) emphasises from his readings of Foucault, an 'overriding concern with the content and overt effects of discourse results in a lack of awareness of the means by which *the criticism of discourse itself may become the insidious instrument of power*' (p. 540, original emphasis). In observing the rising influence of matters of concern over matters of fact, Latour therefore calls for replacing critical scholarship with 'another powerful descriptive tool that deals this time with matters of concern' (p. 232).

Discourse ethics and critical discourse analysis appear as an academic dyad, taking a stance in the same normative ethics. Or, as this thesis will argue, they are a product of and reflect the moral air of our time. Together, they are proposing an ideal and provide a tool which is intended to help the ideal materialise. Just as critical discourse analysis is presented as 'an unavoidably moralistic pursuit with explicit aims of beneficially transforming social and political systems to make them more equal and democratic' (Graham, 2018, p. 186), the thesis expands the observation to also include Habermas' discourse ethics. Similarly, Power (1995) points to the epistemological problem in Neo-Kantian approaches such as critical discourse analysis as a Neo-Kantian theory—but in this thesis also observed in discourse ethics—that they become caught in their normative validity as scholarly frameworks. From that point, these approaches lose their critical edge as these theories cannot emancipate themselves from the communicative power that they nurture.

> …unless it can be related to strategies of enlightenment, the critical-reflective component in which possibilities for the acceptance of the validity of critical

> theory are possible. For critical theorists, reconstructive-transcendental arguments and critical-reflective strategies are intertwined and mutually constitutive. For analytical philosophers, they are surprisingly circular because they seem to presuppose the very things they are intended to demonstrate. (Power, 1995, p. 1007)

Considering an empirical dimension, the thesis will argue that there are immanent problems in above-mentioned scholarship. Discourse ethics express an ideal free from strategic communication. At the same time, it was formed as a critical comment by Habermas based on his observations that the world is full of such immoral communication. And herein we find the first embedded paradox addressed in the thesis—the ideal's non-reliance on an empirical reality. Furthermore, this thesis would even argue that his own presentation of discourse ethics is a case of strategic communication. A second paradox emerging as one reviews Habermas' own communicative behaviour considering his own communicative norms. The same immanent paradox emerges in relation to critical discourse analysis, as it on the one hand is presented as an emancipatory tool while on the other hand it locks actors into the methodology's own pre-defined normative boundaries. And as normative frameworks, while presenting strategic communication as immoral, the thesis argues that both will inevitably engage in strategic moral communication. Furthermore, while advocating that their approaches reveal hidden power structures, they do not acknowledge their own hegemonic power over the discourse. While they argue that they accentuate the need for critical voices, they find the criticism against democratic values to be illegitimate and often conclude that these voices should be excluded from the debate.

Reimagining strategic communication

The purpose of this thesis is to propose a shift from the above-presented normative frameworks towards an empirical understanding of strategic communication's relation to morals. This unlocks a new type of knowledge beyond the ability to evaluate moral discourses. Instead, it develops conceptual tools which can help us understand strategic communication's role in observable moral discourses. The contribution is, on the one hand, to the field of strategic communication—responding to an ongoing discussion about the boundaries defining the field, and addressing the lack of metatheoretical discussion on morals and ethics. On the other hand, such a contribution can also be considered of transdisciplinary interest. The proposed new theory of *strategic moral communication* and methodology of *moral discourse analysis* can be used as an alternative to discourse ethics and critical discourse analysis in the social sciences more generally. This theoretical and methodological proposition forms the thesis' contribution to the above-presented discussions. It is presented in three parts, together forming an accumulated presentation of the shift from a normative to an empirical perspective on strategic communication's relation to morals. The following three research questions guide these three parts:

1. What would an empirical theory of strategic communication's relation to morals look like?
2. What method could be used to study this new concept of strategic moral communication?
3. How does strategic moral communication manifest in the moral discourse on fake news?

The first research question is addressed in Part 1 by examining the philosophical assumptions grounding Habermas' discourse ethics. In the first chapter, Habermas' theory is unpacked by pointing to its philosophical assumptions and the research upon which it rests. As a transdisciplinary scholar, Habermas can be unpacked from a wide range of scholarly perspectives. Previous scholars have for example addressed discourse ethics as a political proposition, as a theory of deliberative democracy (Fraser, 1990; Mouffe, 1999, 2005a, 2005). This thesis departs from media- and communication studies in general and in strategic communication in particular. From this perspective, the review unpacks the theory in relation to Habermas' assumptions about communication, ethics, and influence. Together, these three form the nexus of the thesis' review of discourse ethics. What emerges is a stance in harmony with that time's assumption that humans are rational beings, reasoning their way to morals as conclusions through a dialogical approach. The second chapter recognises that this stance, however, has since been challenged in the empirical sciences. This point has been neglected in scholar's discussions about

Habermas' moral theory of communication.[10] From the field of developmental psychology and the early scholars upon which Habermas builds his theory, the field has undergone a paradigmatic shift as its contemporary scholars would argue that this idea of moral communication is but one of two types. On the one hand, they echo Habermas' assumption that we engage in rational reasoning to arrive at a moral conclusion. On the other hand, however, they find that we more often use communication to rationalise our moral intuitions. This is presented in the second chapter, which contains a literature review of the past 40 years of research in moral psychology that has led to this recognition of the rationalising rather than rational idea of morals. Finally, the last chapter translates this new stance in relation to theories from communication science, to propose the thesis' moral theory of communication as *strategic moral communication*. The new theory challenges the normative assumptions in discourse ethics, presented in relation to an ongoing discussion on the very definition of strategic communication within the field. Through discussions, the thesis particularly challenge how strategic communication has been defined in relation to morals. Ever since the field took its contemporary form, strategic communication scholars have found it necessary to address 'the term *strategic* to determine whether it necessarily implies manipulative or deviant communication practices' (Hallahan, Holtzhausen, Ruler, Verčič, & Sriramesh, 2007, p. 4). The defence of the field's moral *raison d'être* shows the influence of Habermas' scepticism. As scholars in strategic communication argue, while Habermas held a negative view of the practice, he emphasised that he 'did not disregard the use of strategy altogether, as long as it was used to create understanding'.[11] Whether Habermas himself would agree is another matter. Nonetheless, the field has tried to fit its definitions in relation to these normative standards. In some definitions, scholars seek a neutral conception. The editors of the inaugural issue of one of the key journals in strategic communication suggest that 'no communication discipline is inherently good or bad but is in the end only the product of its user' (Holtzhausen & Hallahan, 2007, p. 1). Other scholars have suggested definitions in which strategic communication is part of the moral. In Bowen's (2018) encyclopaedic entry, for example, on the ethics of strategic communication, strategic communication is described as a phenomenon 'facilitating social discourse' and points to the 'historical origins of ethics' as one where 'scholars argued that the one obligation of dialogue is an implicitly ethical one' (p. 1). This is but one example of how the field's sense of ethics and morals are bound to Habermas' ideal, without explicitly acknowledging it (expanded review starting on p. 62). For while scholars have tried to adapt to Habermas' norms, the concept of strategic communication nonetheless points to deep paradox embedded in the classic definitions of strategic communication. For example, Bowen points to those arguing that strategic communication can never be moral as it is by default conducted by organisations that hold a more powerful position than everyday people. Buhmann, Paßmann, and Fieseler (2020) similarly argue that strategic communication—as per definition being the purposeful use of communication to achieve a pre-determined goal—works contrary to Habermas' idea of moral communication. For it rather aligns with Habermas' notion of immoral strategic acts. This thesis argues that we should not limit the study of strategic communication to an organisation's use of communicative influence, as it has traditionally been defined. Instead, the thesis suggests a wider definition, a shift from defining strategic communication as a professional practice to acknowledge its potential as a philosophical perspective on a certain form of communication that seeks and/or achieves influence. It also argues that strategic communication should not be considered a field studying a professional practice, but is better approached as a phenomenon studied across the social sciences, as scholars in various ways approach communication with the idea that it holds the potential for influence. This re-imagination of how we can define strategic communication is ultimately linked to the thesis' proposition for the theoretical nature of strategic moral communication, moving from Habermas' normative framework to the empirical stance in moral psychology's presentation of how communication is used in moral reasoning and when passing moral judgement (presented in Chapter 3).

---

[10] For example, a search on the WoS (introduced on p. 200) on 25 November 2023 pointed to three publications mentioning both Habermas and moral psychology, none of which discuss the need to review Habermas' theory in relation to new insights that has emerged in the field of moral psychology since he based his theory on the founding (however now challenged) theories some 40 years ago. A search on the same platform for mentions of Habermas and developmental psychology (as the field was named when he engaged with it) points to nine publications. Again, however, none note the new paradigm that challenges the validity of Habermas' ideal model as empirically valid. A search for publications mentioning both Habermas and Haidt (one of the key contemporary moral psychologists upon which this thesis builds its theoretical framework) presented zero results.

---

[11] The piece continues to argue that Foucault's and Lyotard's arguments support the importance of strategic communication, as they described how political power dimensions embedded in discourse 'again emphasize the importance of *strategic* as an impetus for the field communication practice' (Hallahan et al., 2007, p. 15).

From this theoretical framework, the thesis continues in Part 2 to address the second research question on how to study the new theoretical phenomenon. The new theoretical conception of strategic moral communication challenges the foundation of methodologies such as critical discourse analysis. In the new theory, where moral communication is recognised as strategic by default, Habermas' discourse ethics and critical discourse analysis become examples of this practice. In critical discourse analysis, this strategic aim is seen as embedded in the very methodological design as 'the need to persuade is also implied in the critical transformative aim' (Graham, 2018, p. 202). Breaking with this methodological aim, the thesis introduces a new form of discourse analysis fit to analyse normative ethics, by studying strategic moral communication as expressions of a moral intuition. Phrased differently, this methodology allows us to understand normative ethics embedded in discourses—such as scholarly discourses (discourse ethics and critical discourse analysis), public discourses (the fake news discourse)—as well as to study a set of moral discourses that align to such extent that one argues that they reflect a moral zeitgeist. Many have argued that the study of a zeitgeist demands a comprehensive, social-wide exploration spanning across political, economic, and perhaps even personal motives and actions.[12] Others would question if we can make sense of the contemporary zeitgeist, arguing that we cannot comprehend the contours of its normative boundaries as we take them for granted and see them as matters of course. This thesis has broken with such assumptions about the impossibility of studying one's own system, arguing to have established a method to study historic or contemporary moral zeitgeists. The approach is centred on reviewing phenomenon framed as moral or immoral phenomenon. The selective process of labelling particular phenomenon as moral or immoral, and in the definition of each respectively pointing to what makes them so, the methodology argues that we can recreate the normatively restrictive phenomenological boundaries in such strategic discourse. For just as sociologists describe that 'the concept of the moral refers to communication' (Luhmann, 1996, p. 29), communication scholars with an interest in ethics argue that communication is a methodological gateway to study the constitution of our moral convictions.

---

[12] As Habermas would reflect on the ineluctability of moral feelings, 'we cannot react at will our commitment to a lifeworld whose members we are' (1999, p. 47). Arguably, we nonetheless find examples studies of zeitgeists spanning from Mannheim (1997), to Horkheimer and Adorno (1997), Habermas (1989), Foucault (1990, 1994a, 1995, 2010, 2013), and Nietzsche (2006) as scholars seek understanding of societies tied to a certain time and space.

> …communication, as both a discipline and 'interdiscipline' or field, is poised to play a unique role in advancing discussions on ethics because the field offers an array of concepts and principles attuned to the examination of ethics writ large. … Communication is especially well suited for meta-ethical analyses as well as an examination of specific ethical issues, dilemmas, and decisions because of how it is attuned to the very construction of arguments about what 'counts' as relevant information, opinion, and choice. The rhetorical framing of ethics broadly considered, where ethics can be treated as an integral to the life-world or ancillary to it, is but one obvious powerful example. (Cheney, May, & Munshi, 2011, p. 1)

Through *moral discourse analysis*, the thesis propose that we can map the collective phenomenological boundaries of a shared context by reviewing the shared norms for communication. This would be revealed by analysing which phenomena are acknowledged as morally or immorally significant (which concepts are being used) and seek understanding of what aspects of these phenomena are considered moral or immoral (reviewing the definitions). Together, this is argued to reveal the underlying values framing the context's normative boundaries. Using this inductively defined normative framework, the discourse can then be reviewed immanently in relation to its own moral standards. Through this, we move from a pre-determined moral stance, as in critical discourse analysis, to a methodology that adopts the context's own moral system as its analytical framework.

Following Jaeggi's reformulation of immanent critique to 'point to a normative account of real interests that responds to claims that the priority of right, justice, and morality only serves as a cover for ethnocentric projection of power' (quote from Hogan & Marcelle, 2022, p. 53), this thesis seeks to reveal immanent paradoxes between the moral norm for communication expressed and the same actors' empirical practice of communication. Through this analysis, the methodology stipulates that we can uncover immanent power structures as moral intuitions which leads to these normative arguments. Using the theoretical stance that moral communication is strategic as an analytical assumption, the methodology serves as a tool to help unveil strategic communicative behaviour in moral discourses. The strategic moral communication expressed would then be based on and reflect the moral intuitions guiding the discourse' normative judgement.

As the move from a new theory to a new methodology leads to a new understanding of moral discourses, Part 3 of the thesis addresses the third research question of how the methodology can be used to seek empirical insights into strategic moral communication in practice. To that end, the thesis engages in a study of the fake news discourse. Scholars have published excessively on the theme of fake news as an empirical phenomenon.[13] These studies span from contributions to communication studies to computer science, philosophy, psychology, and beyond. This thesis, however, does not study the phenomenon itself, but is a meta-study of the public discourse framing it by reviewing a sample of 1,100 publications issued by organisations in 2016–2020 on fake news and its conceptual siblings.[14] By reviewing the discourse as a moral discourse, the case-study provides an argument for the empirical validity of the proposed theory, showing how it manifests as a practice of strategic moral communication. It also illustrates how the methodological proposition can be applied to study this new conception of strategic communication's relation to morals. As such, it shows how the approach enables an empirical rather than normative representation of discourses. It starts by showing the way the fake news discourse mirrors Habermas' normative proposition in discourse ethics. The discourse repeats the story of the internet as initially being prophesised as a scenery which would enable Habermas' ideal to manifest in society. It continues to describe how the discourse turned into a critical discourse in the mid 2010s, with the rising concern for a new set of phenomena such as echo chambers, alternative facts, and post-truth. Continuing to zoom in on each of these three foci, the analysis shows how, when reviewed immanently, the discourse itself is also engaging in the same behaviour. As such, it is conceptualised in this thesis not from a normative stance as something to be judged, but as a deeply human condition that ought to be considered in the analysis of moral discourses.

---

[13] Over 4,000 journal articles recorded as per 12 November 2023 on the Web of Science (see p. 198). None were published before 2005, the most common field of publication being communication science (n=927).

[14] Similar recent approaches found in Farkas and Shou's (2020) and in Natani (2002).

After responding to these three research questions, Chapter 10 ends with a discussion concerning the results. It argues that the ideal norms supported in Habermas' discourse ethics, in critical discourse analysis, and in the fake news discourse, all end up unveiling immanent paradoxes. It argues that while Habermas' discourse ethics and critical discourse analysis set out as frameworks to support all members of the public's claim to a say in common discourses, it will inevitably exclude those critical towards the same point of departure. Therefore, the frameworks work contrary to the norms promoted in the frameworks themselves. To that end, the final chapter argues that we can view the thesis' theoretical and methodological propositions as emancipatory frameworks that can untie us from the pre-existing framework's normative boundaries. It becomes a meta-emancipatory framework, as it shows how the stance that we have approached as a default is not a universal, but a particular. The thesis ends with a conclusion, summarising the results, engaging in a critical discussion of the thesis' own immanent validity, and suggesting empirical and theoretical implications arising from these findings. What the thesis argues is not only that this empirical analysis shines new light on the reasons behind the other's immoral behaviour, but provides a critical perspective on one's own role in enabling or pushing actors towards such behaviour. It finally argues that we can use this new stance to form a new understanding of what is happening in the zeitgeist. From this new understanding, we can then develop new ways for how to engage in the fake news discourse in the future. For example, if we could imagine seek consensus not in our rational arguments, but simply in a common conclusion—even if we base this on different rationalisations and moral intuitions.

Outlines

The thesis' move from a normative to an empirical perspective on strategic communication is presented in three parts, each corresponding to one of the three research questions. Part 1 presents the theoretical stance, moving from a presentation of Habermas' discourse ethics in Chapter 1 to a literature review of the contemporary understanding of moral reasoning from moral psychology in Chapter 2, leading to the thesis' theoretical proposition for strategic moral communication in Chapter 3. Part 2 continues to build on this new theoretical stance, presenting a new methodology of moral discourse analysis in Chapter 4, moving from this general presentation to an illustration of its practical application in Chapter 5 by tying it to the thesis' case study of the fake news discourse. Part 3 illustrates the empirical expression of the thesis' theoretical proposition and the type of knowledge one could acquire by applying the new methodology. This is presented through an analysis of the fake news discourse. First, the context in which the fake news discourse emerges is presented in Chapter 6, before expanding on the three analytical foci tied to paradoxes between the discourse's normative assumptions and their concern for observable immoral communication. This is presented in the thesis as a division between discussions linked to echo chambers in Chapter 7, alternative facts in Chapter 8, and post-truth in Chapter 9. The thesis ends by discussing the results and concludes the implications of the proposition presented. Chapter 10 starts by discussing the thesis' framework as an emancipating tool that enables us to think beyond normative assessments of moral discourses. This leads to a final summary of the results in the thesis' final conclusions, continuing with a critical discussion concerning the theory's impact on the development of the theory itself and the proposition's implications for the fake news discourse and future scholarship on moral communication.

PART 1. THEORY

What would an empirical theory of strategic communication's relation to morals look like?

Introducing Habermas' normative theory of discourse ethics in Chapter 1, challenged in Chapter 2 through a literature review of the 40 years of advancements in moral psychology since his engagement with the field, translating this new stance into the thesis' proposition for an empirical theory of strategic moral communication in Chapter 3.

Chapter 1.

Discourse ethics

Discourse ethics is Habermas' suggestion of a moral theory. Here, he draws upon communication in his conception of morals, reflecting his quest to ground 'the social sciences in a theory of language' (1988, p. xiv) or, more recently expressed, as a focus on 'the communicative model of society' (2018b, p. 65). In some writings, the theory is described as simply *communication ethics* (2007, p. 23) or a *discourse theory of morality* (1994, p. vii). Habermas persistently presents it not as a normative theory in the traditional philosophical sense, but argues that it reflects a so widespread norm that he labels it an empirical theory (2022b). This thesis would disagree, arguing that it is a normative theory, albeit so widespread that it reflects the morals of our time.

Habermas' first and most-read book, *The Structural Transformation of the Public Sphere* (1989), depicts a time in which he argues that the ideal he later theorised as discourse ethics emerged. Around the turn of the 18th century, it grew as an idea in Europe alongside the first public coffee houses. Here, people could engage in critical conversations about common matters, rationally evaluating the validity of claims to arrive at a consensus. The product would be his conception of morals. Morals would therefore be derived from everyday conversations, not by philosophers, and changes in character with the social context in which it arises. His first attempt to formalise this as a theory was presented as *ideal speech situations* in two influential volumes on *The Theory of Communicative Action* (1984, 1987). However, taking the broader criticism directed towards this theory into account,[15] he then split off to form a moral theory of communication as

---

[15] Addressed for example in Habermas (1994, pp. 19–112), and in the compilation of voices by Honneth and Joas (1991) which includes a final reply by Habermas.

discourse ethics.[i] In retrospect, Habermas (2018b, pp. 100–121) describes his version of discourse ethics as starting to develop in the 1960s (p. 102), presented a couple of decades later in the book *Moral Consciousness and Communicative Action* (1999). It took the form of a neo-Kantian theory that adopted Karl-Otto Apel's discourse ethics but argued that morals are not found in intellectual discourses but in the every-day discourses among the public.[16] Instead of presenting a deontological rule that would guide all moral behaviour, as otherwise presented by many moral philosophers before him, he presents a set of communicative axioms that bind moral conclusions to a moral form of communication. The moral is thus not found in a certain type of moral conclusion, but is determined as moral if the actors arriving at this conclusion engaged in a moral form of communication. Just as it was introduced by Habermas to reflect a historic ideal, the ideal maintains to this day. Similarly, it was presented as a counterfactual proposition, from this thesis viewed as a critical commentary against the immoral strategic communication which Habermas saw in his own time, just as there is a widespread concern today about the nature of the public discourse.

This chapter unpacks Habermas' discourse ethics, building on his own selection of texts that he sees as framing the theory (see p. 198)[17]. Returning to Habermas' (2018b) description that discourse ethics started by encountering 'the specific problem of rational justification of decisions' (p. 101) in three contexts,[18] the following sections take a point of departure in these dimensions. Together, they are re-framed as unpacking the ontological, axiological, and epistemological philosophical assumptions forming Habermas' discourse ethics.

---

[16] Among other scholars, Habermas (1999, p. 43) position his theory in relation a group that 'share the intention of analyzing the conditions for making impartial judgements of practical questions,' including not only Karl-Otto Apel but also Kurt Baier, Marcus Singer, John Rawls, Paul Lorenzen, and Ernst Tugenhadt to name a few.

[17] The thesis uses 'p. x' when referring to a page in a referenced publication, while using 'see p. x' when referring to a page within the thesis.

[18] From ontological forms in a social system, axiological relations to validity claims, and in an epistemological question about reason and interests (Habermas, 2018b, p. 101).

Communication, human nature, and moral reality

Habermas' discourse ethics is formed on a particular ontological view of human nature, communication, and morals. These three together form his argument that morals are an evolutionary inevitability. As Habermas see it, we seek relations with others to survive, and these relations rely on our ability to communicate with one another, where the result of such conversations is a common consensus on regulatory moral norms. Furthermore, he argues that morals are bound to a certain form of communication. This type of communication aligns with discourse ethics' two overarching axioms, assuming an open discourse where actors rationally engage in deliberations to arrive at a moral consensus.

> In anthropological terms, morality is a safety device compensating for a vulnerability built into the sociocultural form of life. The basic facts are the following: Creatures that are individuated only through socialization are vulnerable and morally in need of considerateness. Linguistically and behaviourally competent subjects are constituted as individuals by growing into an intersubjectively shared lifeworld, and the lifeworld of a language community is reproduced in turn through the communicative actions of its members. This explains why the identity of the individual and that of the collective are interdependent; they form and maintain themselves together. … Conversely, everyday language is also the medium by which the intersubjectivity of a shared world is maintained. (Habermas, 1999, p. 199)

Lifeworld is a key concept in Habermas' theory. He argues that 'practical discourse transforms what Mead viewed as *individual, privately enacted* role taking into a *public* affair, practiced intersubjectively by all involved' (Habermas, 1999, p. 198). Therefore, Habermas describe the lifeworld taking two forms—either an *individual* or a *social lifeworld*. The two are connected to each other in 'linguistic terms' (2018b, p. 69) as language would function as a vessel where 'psychic processes such as sensations, needs and feelings are fitted into structures of linguistic intersubjectivity, inner episodes or experiences are transformed into intentional contents—that is, cognitions into texts, needs and feelings into normative expectations (precepts and values)' (2007, p. 9). This 'communications approach provides a solution' to Husserl's problem on 'how an intersubjectively shared lifeworld can be generated from the egological perspective of the transcendental monads of several "original egos"' (Habermas, 2018b, p. 71). Ideally, Habermas describes that the social lifeworld is built on the individual. In the shift from an individual to a social lifeworld, people use communication to seek common *understanding,* which results in 'shared knowledge, mutual trust, and accord with one another' (1979, p. 24).[ii] In short, they create their shared morals through such conversations.

What then does Habermas mean by communication? While moral philosophers would look for morals embedded in linguistic structures, he emphasises that morals can only be derived from the everyday use of communication by everyday people (Habermas, 1999, p. 66). Philosophers of morality should therefore, he argues, step away from their desks and at least adopt and at best study the 'communicative practice of everyday life' (p. 48) to understand how morals are formed. As he continues 'the *moral* intuitions of everyday life are not in need of clarification by the philosophers' (p. 98, original emphasis). This positioning against the moral philosophers is a key part in his presentation of discourse ethics—arguing for moral sociology rather than moral philosophy as the appropriate stance for understanding the nature of morals. Furthermore, in discourse ethics, the concept of communication is limited to linguistic forms only (critically discussed for example in Rich & Craig, 2012, pp. 391–393) while in other texts acknowledging other forms of communication (c.f., in his historic presentation of coffee houses 1989, p. 69). Habermas imagines communication to take two forms, either as 'ordinary language' that meets functional imperatives in everyday cooperation or 'ritual communication' that gives rise to normative obligations (note 14 in 2018b, p. 196), both serving a social function.[19] In emphasising the social function of communication, he aligns with Wittgenstein's late language theory (1953). Here, upon observing people's everyday use of communication, Wittgenstein came to appreciate words as volatile, socially dependent 'language games'.[20] Language and context were reflections of one another.

---

[19] In reference to *Postmetaphysical Thinking II* (Habermas, 2017, pp. vii–xv and chs 2–3).

[20] He describes his stance as in part aligning with a consensus theory of truth, however giving it a new label as part of a transcendental hermeneutic or 'as I would prefer to call it, a universal pragmatic' as he imagines that if 'Wittgenstein developed a theory of language games, it would have to take the form of universal pragmatics' (Habermas, 2003a, p. 53).

To understand this conception of communication as tied to a social process, Habermas argues that we need to move beyond what is *said* (the linguistic message), as in Chomsky's (1965) 'linguistic competences'. For Habermas adds that communication can only be understood if we incorporate a social dimension, also considering what is *meant* and how the message is *understood*. This formed his more comprehensive theory of *communicative competences* (Habermas, 1970, p. 367). As such, he finds a kindred spirit in Austin (1962), continuing from Wittgenstein's late theory to form a theory of speech as performative acts. These *speech acts* emphasise dimensions spanning from what is said (locution), meant (illocution), and communicated (perlocution)—harmonising with Habermas' communicative competences.

Habermas is particularly interested in illocutionary acts, forming his normative view of moral communication as *communicative acts*. In communicative acts, the communicator 'claims to be: (a) *uttering* something understandably; (b) giving [the hearer] *something* to understand; (c) making *himself* thereby understandable; and (d) coming to an understanding *with another person*' (1979, p. 23).

> I call interactions *communicative* when the participants coordinate their plans of action consensually, with the agreement reached at any point being evaluated in terms of the intersubjective recognition of validity claims. … In communicative action one actor seeks *rationally* to *motivate* another by relying on the illocutionary binding/bonding effect (*Bindungseffekt*) of the offer contained in his speech act (Habermas, 1999, p. 58, original emphasis)

Communicative actions are tied to an individuals' lifeworld as they 'serve to express lived experience' from the '*subjective world* to which he has privileged access' (Habermas, 1999, p. 136, original emphasis). That is, communicative actions transform the individual lifeworld to a social one, resting on Habermas' conception of illocutionary speech act's ability to enable understanding. In relation to these illocutionary speech acts, 'he believes that it contains the seeds for establishing a normative relationship between two or more speakers, and this is precisely what universal pragmatics and communicative action is about' (Thomassen, 2019, p. 488). Communicative acts thus lead to mutual understanding as moral conclusions. In the move from an individual to a social lifeworld, we find a moral agreement. This, in turn, creates a *binding/bonding effect* as we by recognising the validity of the moral agree upon abiding them ourselves. They become moral rules, as 'norms are dependent upon the continual reestablishment of legitimately ordered interpersonal relationships' (1999, p. 61) which depends on—at least the idea—that they will be followed (p. 58).[iii] As Habermas continues, 'normative claims to validity, then, *mediate a mutual dependence* of language and the social world that does not exist for the relation of language to the objective world' (p. 61, original emphasis). Morals are thus not objectively existent (in language), but are socially dependent (expressed in the context-characteristic use of language).

This chapter argues that Habermas' discourse ethics is a case of normative ethics as it is proposing an ideal. At first, this ideal was simply grounded in Kant's maxim (Habermas, 2007, p. 108), that is, his categorical imperative 'act only according to that maxim through which you can at the same time will that it become a universal law' (Kant, 2012, p. 421). As Habermas (2007) argued at the time, 'the expectation of discursive redemption of normative validity claims is already contained in the structure of intersubjectivity that makes specifically introduced maxims of universalization superfluous' (p. 110). Later, however, Habermas (2018b) would argue this as an 'over-hasty conclusion' (p. 104) where discourse ethics 'reformulates the intuition' of Kant (Habermas, 2018b, p. 106; c.f., Habermas, 1999, pp. 63, 195) in the form of two new principles. The first *principle of universalisation* (U) reformulates Kant's proposition to specify that the universal expands only to the social context from which it was formed. This was added to account for Habermas' observation that Kant's principle 'is open to the suspicion that it expresses one-sided or culturally biased intuitions' (2018b, p. 7). Only those who are a part of the society in which the moral is formed would be bound to follow the context's morals. His principle was, however, met with criticism. Critics argued that it still 'imposes uniformity and inclusion through assimilation' (2018b, p. 107), as reaching consensus would mean that people are not allowed to hold on to their individual convictions. In Habermas' view, this was the result of a misconception, as he clarified that 'rational discourse, as a form of communication, not only accords the participants the communicative freedom to take a stance when addressing practical questions, but also requires them to engage in *reciprocal perspective-taking*'

(p. 108).[21] To Habermas, his (U) builds on the idea of equality—eventually aligning with Ernst Thugendhat—as Habermas (1999) describes how 'arguing that a norm is justified only if it is "equally good" for each of the persons concerned' (p. 68), and ethics would therefore be a compromise (p. 72).[iv] This leads him to emphasise 'the character of the participants themselves' as they need to have an empathic ability, where 'practical discourses are essentially a matter of the participants working to overcome these differences in perspective' (p. 120). Arguably, the criticism nonetheless holds, that while Habermas accentuates the need to seek understanding of another's perspective, for consensus to form from a discourse with multiple perspectives, at least one would have to change their mind. As this thesis would interpret Habermas, this is necessary for a community to function as a unit—as its inhabitants would have to agree on a common way of life bound to the context's agreed morals.

To complete this reinterpretation of Kant's categorical imperative, he adds a second principle, emphasising discourse ethics as formed from communication by tying it to a second *principle of discourse* (D). To Habermas (1999), morals are decided in 'real-life argumentation' (p. 66) and while (U) would signal the *who,* (D) response to the question of *how* as it is a matter of a procedure (1999, p. 103). It is not, however 'a procedure for generating justified norms but a procedure for testing the validity of norms that are being proposed and hypothetically considered for adoption' (p. 103). He describes this discourse as build on *argumentation*, an ideal situation which presupposes that the participants assume that 'the structure of their communication rules out all external or internal coercion other than the force of the better argument and thereby also neutralizes all motives other than the cooperative search for truth' (pp. 88–89). His reinterpretation of Kant is described to inevitably give rise to the second principle as 'discourse ethics replaces the Kantian categorical imperative by a procedure of moral argumentation' (p. 197).

---

[21] Alternatively phrasing this *perspective taking* as 'to *perform the act of empathy*' in line with George Herbert Mead's '*reciprocally adopting* the *perspectives* of the others involved' which leads Habermas to describe practical discourse as 'essentially a matter of the participants working to overcome these differences in perspective' (Habermas, 2018b, p. 120).

> Rather than ascribing as valid to all others any maxim that I can will to be a universal law, I must submit my maxim to all others for purposes of discursively testing its claim to universality. The emphasis shifts from what each can will without contradiction to be a general law, to what all can will in agreement to be a universal norm. (Habermas (1999, p. 67) quoting McCarthy (1987, p. 326))

Habermas provides an early definition of *discourse* in his book *Legitimation crisis* (2007), here presented as a form of communication where the only force exercised is the better argument and 'all motives except the cooperative search for truth are excluded' (pp. 102–116). Later expanding on this definition, he emphasises how 'practical discourse (D) plays the part of a rule of argumentation' (1999, p. 197) where 'only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in practical discourse' (p. 66). In this way, we see how Habermas continuously links his two principles—(D) and (U)—to each other in his presentation of discourse ethics. In short, it would read that morals are the product of discourse between people in a community.

This ideal depiction of moral communication can also be seen as a critical comment to the communicative phenomena Habermas observed in the public sphere. That is, forming discourse ethics as a conceptualisation of communication that works contrary to observable forms of immoral communication. Growing up in Nazi Germany, he describes how his early of experiences propaganda would come to drive his academic path.[22] As many other philosophers—including for example Lévinas (1991)[23]—the war drove him to explore theories which would seek to realise the mantra 'Never again Auschwitz'. As many others, he sought a theory which would support equal treatment and respect for all human beings, rather than engaging in practices of *othering* wherein one presents the other as inferior, insane, or non-human (Habermas, 1998). Later, considering the rising multicultural society emerging in the late 20th century (1994, p. 93), he again sees the necessity in developing a theory which sees other members of one's community as oneself. In this new 'moral community' it 'is not a collective that would force its homogenised members to affirm its distinctiveness' nor 'imply locking members into a community that closes itself off from others' (1998, xxxv–xxxvi). As he continues to advocate, the 'inclusion of the other' means that 'the boundaries of the community are open for all, also and most especially for those who are strangers to one another and want to remain strangers' (p. xxxv—xxxvi).[24] Discourse ethics is a matter of building relations between people. For he had seen what could happen if we do not engage in taking the other's perspective, arguing that 'unless universalistic morals are developed, this can take place in terms of the differentiation between in-group and out-group morality' (2007, p. 9). In *The Inclusion of the Other* in particular, he 'defend the rational content of a morality based on equal respect for everybody and on the universal solidarity and responsibility of each for all,' arguing that his (U) constitutes a stance 'that is highly sensitive to differences' (1998, p. xxxv). At the same time, he recognises that ethics inevitably gives rise to the idea of the unethical. This produces two paradoxes in relation to *tolerance*; *arbitrariness* as the tolerated behaviour also means that there is behaviour that we do not tolerate, and *paternalism* as a distinction between the tolerating and the tolerated (Habermas, 2006).

---

[22] Habermas describes how 'it was the events of the year of 1945 that set my political motives.' He continuously engaged in discussions about the German heritage, in public statements published in the German newspaper *Die Zeit* as well as in books such as *The New Conservatism* (1989a) and *A Berlin Republic* (1997).

[23] Comparisons of Habermas' and Lévinas philosophy can be found in i.e., Vetlesen (1997) or Hendley (2000, 2004).

A transdisciplinary moral ontology

As a theory of morals—or as this thesis would label it, as a moral theory—discourse ethics draws upon concepts from moral philosophy. However, as previously emphasised, he decisively positions himself against theories from this field. Rather, Habermas emphasises that his proposition is grounded in a transdisciplinary stance. More specifically, he draws upon developmental psychology emerging as a field at the time. This transdisciplinary interest dates back to his very first book (1989), presenting the public sphere as an object of study 'whose complexity precludes exclusive reliance on the specialised method of a single discipline' as 'when considered within the boundaries of a particular social-scientific discipline, this object disintegrates' (p. xvii).[25] Discourse ethics evolves from such a transdisciplinary stance, grounded in philosophical concepts but foremost drawing upon his readings from moral psychology as it started to develop in the 1980s. This, he presents as the key point dividing his discourse ethics from that of Karl-Otto Apels'.

---

[24] In other writings, Habermas continues that 'the interest is common because the constraint-free consensus permits only that *all* can want; it is free of deception because even the interpretations of needs in which *each individual* must be able to recognize what he wants become the object of discursive will-formation' (2007, p. 108).

> The difference between the versions of discourse ethics developed by Karl-Otto Apel and me were grounded from the beginning in the contrast between a disciplinary and interdisciplinary perspective on moral philosophy. While Apel regards the ultimate justification [*Letxtbegründung*] of the moral principle as the royal road to the self-confirmation of philosophy, for me the task of justifying the moral principle arises in the context of a theory of communicative action that is tailored to the division of labour with other human sciences' (Habermas, 2018b, p. 105).

Discourse ethics is a 'theory of morality' (Habermas, 2018b, p. 104), tied to philosophical concepts such as *ethics* and *morals*.[26] As many before him, Habermas used the two interchangeably at first, later aligning with Hegel's differentiation where *ethical discourse* refers to an individual level, while *moral discourse* ties to common social grounds (Habermas, 1998, p. 26). For that reason, while introducing the theory as discourse *ethics*, he later concludes that as his theory concerns the common, it is better described as a *moral* theory of communication, or a *discourse theory of morality* (1994, p. vii).[v] Continuously, Habermas departs from concepts found in moral philosophy. A first important observation is Habermas echoing moral philosophy's description of morals and ethics as a matters of *ought* rather than *is*, a distinction between facts and values presented by Hume that…

---

[25] At the same time, he recognises the problems with a transdisciplinary stance, as 'scarcely anyone will be able to master several, let alone all, of these disciplines' (p. xvii).

[26] Extensive discussion, for example, starting on p. 89 in *Moral consciousness and communicative action* (1999).

> …signifies the impossibility of logically deriving prescriptive sentences or value judgements from descriptive sentences or texts. In analytic philosophy this has been the point of departure for a non-cognitivist treatment of practical questions in which we distinguish between empiricist and decisionist lines of argument. They converge in the conviction that moral controversies cannot, in the final analysis, be described with reason because the value premises from which we infer moral sentences are irrational. (Habermas, 2007, p. 102)

In Habermas' own reasoning, this sets him apart from 'realist conceptions of morality' to place an emphasis on 'the constructive character of the world of normatively regulated interpersonal relations' (2018b, p. 106). He approaches moral claims as analogous to truth claims (1999, pp. 56, 76)—while they are not the same, we approach them as such, experiencing that we are driven by a moral intuition to give reasons for our morals (c.f., Habermas, 1994, p. 34). As Cooke (2019) describes, upon presenting discourse ethics…

> …from the 1980s onwards, in response to critics, Habermas began to revise his theory of propositional truth. He gradually distanced himself from his previous definition of truth as the outcome of a discursive procedure, replacing it with an idea of truth as justification-transcendent, in the sense that it does not coincide even with the outcome of an idealized justificatory procedure: even in his conceptual thought experiment, in which ideal justificatory conditions actually obtain, an argumentatively reached agreement merely *points towards* truth in an unconditioned, universally binding sense (Cooke, 2019, pp. 130–131; c.f., Habermas, 1994, p. 29)

This ought character of norms represents an important key to understand Habermas' idea of morals as norms (1999, p. 44; c.f., 1994, pp. 40–43), as he describes that 'to say that *I ought* to do something means that I *have good reasons* for doing it' (1999, p. 49). We arrive at an idea of morals by engaging in discourse where morals are what is agreed, each 'have to try to persuade one another that it is in the interest of each that all act as they intended' as 'argumentation is designed to prevent some from simply suggesting or prescribing to others what is good for them' (p. 71).[vi] The ought rather than is character of these arguments means, however, that 'normative texts cannot be verified or falsified: that is, they cannot be tested in the same way as descriptive texts' (p. 54). Thus, he relies on morals as rational,[27] while his interpretation of *rational* diverges from the 'narrow concept of rationality that permits only deductive arguments.' (2007) While he diverges from Kant's idea of the participants in such discourse as not 'intelligible characters but real human beings driven by other motives in addition to the one permitted motive of the search for truth' (Habermas, 1999, p. 90), it is only the discourse in search for truths that he conceptualises as moral communication. He also emphasises the *counterfactual* nature of moral communication, resting on an unpublished manuscript by Albrecht Wellmer (see Habermas, 1999, pp. 105, 198), as he shows how discourse ethics as an ideal point to how things *ought* to be, but at the same time recognises that as an ideal it will not manifest in real discourses. For example, he points to Kant's conceptualisation of the 'intelligible character' of the subjects reasoning, to be missing from his own conception of moral conversations, and thus morals rather pointing towards than to a moral ideal (Habermas, 1999, p. 92; c.f., 1994, p. 55).

Habermas keeps describing how discourse ethics is highly influenced by his readings of psychology research at the time. More specifically, readings from the emerging field of developmental psychology, which were centred around discussions on how ethics develop in human beings, guided by the big question *'Where does morality come from?* This question divided scholars into a classic nature-nurture standoff with the *nativists* in one corner arguing that we discover our innate morality as we grow older, the *empiricists* in contrast arguing that we learn morality from others (Ellemers, Toorn, Paunov, & Leeuwen, 2019; Haidt, 2013). Among the early scholars, Piaget (1932) built on McDougall's research on moral conduct to propose a theory that children are neither born with moral capabilities nor learn them by imitating others, but *rationally reason* their way to determine morals as they experience the world. Kohlberg (1958, 1964, 1977) would later quantify Piaget's measurements to a model of moral development, arguing that children's morality develops in six-stages—first two steps in a *pre-conventional* stage (judgement in relation to the physical world), then two *conventional* stages (rules and social conventions, or moral authorities), and finally two *post-conventional* stages (independent reasoning in which they reflect on the constitution of rules and social conventions, including moral authorities). Reading Piaget's genetic structuralism and Kohlberg's theory of moral development—as well as having personal contact with Kohlberg (Habermas, 2018b, p. 105)—Habermas (2007; 1994, p. 161) saw intuitively how it not only related to individuals' moral progression but came to argue that the same steps related to the moral development of societies (1999, p. 109). This theory would echo from his previous writings about 'a collectively attained stage of moral consciousness' (2007, p. 12).[28] He argues that discourse ethics fall into the highest stage,[vii] as 'the empirically demonstrated connection between the "stages" of moral judgement and interactive competences show that the postconventional forms of consciousness' is rooted 'in the basic features of linguistically structured forms of life in general' (2018b, p. 105). Building his theory on the very first steps in this new field, developmental psychology has since grown into moral psychology. Today, contemporary scholars in moral psychology argue that Kohlberg's theory is to be read as highly influenced by his own sense of morality. They describe that rather than pointing to a universal valid moral framework, the stage-model transformed the primary tool for engaging in research on moral psychology 'to measure children's progress toward the liberal ideal' (Haidt, 2012, p. 10).[29] From this perspective, this thesis would argue that the justifications Habermas presents for discourse ethics—by tying this to these empirical theories—rather reflects a common stance in a shared *zeitgeist.* As Habermas would define the concept zeitgeist, it appears not 'within an objective context of events but with the symbolic context of a spirit that expresses itself in them' (1988, p. 17), or presenting it as culture which generates structures of rationality or cultural value spheres (1999, p. 17). Habermas, this thesis argues, presents a moral theory of communication in which he finds that the theories presented by developmental psychology at the time confirmed his moral intuitions. However, rather than using this as an argument for the empirical and universal validity of his theory, this thesis would argue that it merely reflects the common moral zeitgeist from which both Habermas and these moral psychologists departed.

---

[27] Initially outlined in Habermas (1994) and later updated in Habermas (1999).

---

[28] Also addressed in Habermas' *Communication and the evolution of society* (1979).

[29] For example, echoing Habermas' emphasis on perspective taking (see p. 27) as Kohlberg's highest stage is dependent on the ability of 'role taking' (Greene & Haidt, 2002, p. 517).

Habermas' rational ideal for morals further adds a set of valid claims to the theory of discourse ethics. Firstly, as a cognitivist ethics, Habermas (2018b) describes how discourse ethics builds on Peter F. Strawson's work 'against empiricist and value-sceptical conceptions' (p. 106). This means a view where 'normative rightness must be regarded as a claim to validity that is analogous to a truth claim' (1999, p. 197). And while we cannot derive reasons from moral claims in the way we can for descriptive statements (see p. 30), a normative theory must, nonetheless, 'answer the question of how to justify normative texts' (p. 197). His early presentation of principle (D) was 'not yet an answer to the central question as to why normative claims for recognition should have cognitive meaning in the first place and could be redeemed with reason' (2018b, p. 103). In practical terms, in these everyday conversations,[viii] people will be convinced by different reasons. From this observation, Habermas describes three key *validity claims* from which rational morals are determined. He grounds his theory on the conception of an individual and intersubjective lifeworld (see p. 23) as a *subjective* and *social world* respectively.[30] Adding to these *domains of reality* he then adds that we can also derive claims from an *objective world* (grounding normative claims in not only an ought but in part on an is, see p. 30).[ix] Together, these three worlds form a 'threefold relationship involved in a text' in which communications 'serve as (a) an expression of the speaker's intention, (b) an expression of the establishment of an interpersonal relationship between speaker and hearer, and (c) an expression about something in the world' (1999, p. 24). It thus ties back to Habermas' idea of communicative competence as communication would be a matter of what is meant, understood, and said (see p. 25). Moral discourses would be tied to 'the discursive exchange of reasons' that take the form of 'descriptive, evaluative and normative patterns of argumentation' (2018b, p. 101). Together, this forms three options for speakers as three *modes of communication*, either a 'choice between a *cognitive*, an *interactive,* and an *expressive* mode of *language use*' focusing either on 'issues of truth, justice, or taste' (1999, p. 137). Discourse ethics thus develops into a theory 'intended to explain how moral validity claims can be redeemed' (2018b, p. 104) with reasons (2007, p. 105). This triadic ground in three domains of reality, expressed as three modes of communication, finally leads to Habermas' suggestion of three corresponding validity claims (p. 8).[31] As Habermas (1999) argues, 'normative statements can be valid or invalid' (p. 52) in relation to each of these three dimensions of evaluation.[x] In the presentation of discourse ethics, Habermas points to three validity claims; claims to *truth* that refer 'to something in the objective world'; to *rightness* by referring 'to something in the shared social world' based on 'the legitimately regulated interpersonal relationships for a social group'; or to *truthfulness* by referring 'to something in his own subjective world' from the experience of one's world 'to which one has privilege access' (pp. 3–4). The second intersubjective is of course key to understand the moral, but all three are important parts in Habermas' conception of a moral conclusion as they together form an idea of the 'more comprehensive reason embodied in use of language oriented to reaching an understanding and in communicative action' (2018b, p. 104).

---

[30] Building on Rawls, as presented by Habermas (1994, p. 26).

---

[31] Habermas does not define what validity claims are, and the typology has changed over time where for example *comprehensibility* was abandoned by the time he presented discourse ethics as a theoretical framework (Heath, 2019). In Habermas' presentation of discourse ethics, this becomes visible as he describes Alexy's rules of discourse, where he interprets the first as that 'every subject with the competence to speak and act is allowed to take part in a discourse' which he then describe 'defines the set of potential participants' (Habermas, 1999, p. 89).

Table 1. Habermas' world view
The three-part division of the world, communication, and valid claims.

Domains of reality	Modes of communication	Validity claim
Objective world	Cognitive	Truth
Social world	Interactive	Rightness
Subjective world	Expressive	Truthfulness

Later communication scholars would argue that discourse ethics presupposes a 'background consensus' that goes beyond the (U) and (D) to also require the participants to acknowledge these validity claims to be valid (Kurtine's unpublished manuscripts from 1989, referenced in French & Allbright, 1998). However, as Cooke (2019) points out, in his later discussion about discourse ethics in *Justification and Application* (1994), Habermas expands that 'ethical validity claims, too, may be subject of argumentative thematization in discourse' (Cooke, 2019, p. 26). Alex Honneth (1991, p. 112), however, express that 'my supposition is that the Habermasian social theory is constituted in the way that it has to ignore systematically all the existing forms of social critique that are not recognized in the political-hegemonic public sphere' (cited in Herzog, 2018, p. 15).[32] Indeed, this would be aligning with the thesis' own immanent criticism against discourse ethics, as expanded below.

Discourse in conflict

To understand Habermas' discourse ethics as a moral ideal, we also need to recognise how he frames this through its non-relation with immoral communication. The scenery in which moral discourses emerge is described as one of conflict. The need for discourse ethics arises out of the need to find ways in which actors who disagree with one another could engage in a conversation, to arrive at a common conclusion on what they perceive is the right cause of action ahead. But as a counterfactual theory, this context of conflict is also found to contain immoral forms of communication, as strategic or instrumental acts rather than communicative acts. These actors would not seek understanding, but pursue self-serving interests. Contrary to the ideal of the individual lifeworlds converging into a common intersubjective lifeworld, he describes the risk that the system might possess power over the individuals without taking their influence into account.

---

[32] As Iser (2019, p. 570) describes, Herzog's thesis engages 'critically, yet sympathetically with Habermas' work and later became his assistant and the successor to Habermas' position as Professor of Social Philosophy at Frankfurt University.

To start this presentation of the immoral, we first need to understand his idea of the irrational. Habermas' view of rationality is key to understand his presentation of discourse ethics. He argues that being rational is a human natural behaviour, characteristic of our species. But as rational beings, we also hold the potential of acting irrationally.

> …we can call men and women, children and adults, ministers and bus conductors 'rational,' but not animals or lilac bushes, mountains, streets, or chairs. We can call apologies, delays, surgical interventions, declarations of war, repairs, construction plans or conference decisions 'irrational,' but not a storm, an accident, a lottery win, or an illness. (Habermas, 1984, p. 8)

Being rational and irrational is thus a human trait, manifesting in different forms. Discourse ethics, however, diverges from moral philosophy's sense of the word of rationality, Habermas uses his sociological point of departure to argue that discourse ethics points to 'a narrower focus than traditional ethical theories because it concentrates on questions of justice as the question that can be rationally decided in principle' (2018b, p. 104). To Habermas, morals are a practical concern, and reasons mean 'the use of reason' which means that his moral theory is a 'part of the general theory of rational discourse' (p. 101) and focuses on practical reasons (1994, p. 30). From this point, he argues that while claims can be valid or invalid, 'normative statements cannot be verified or falsified; that is, they cannot be tested in the same way as descriptive sentences' which leads to his 'rejection of the idea that practical questions admit of truth' (p. 54). Adding to this, however, Habermas also emphasises that morals cannot be a product of power only, while in part aligning for example with Mouffe (2005) that language provides boundaries where we can 'distinguish within a given regime of truth between those who respect the strategy of argumentation and its rules, and those who simply want to impose their power' (p. 15). In his response to Tugendhat, Habermas explains this as 'when we say yes or no to norms or commands, we are expressing something more than the arbitrary will of a person who submits to or opposes an imperative claim to power. By conflating validity claims and power claims, Tugenhat cuts the ground out from under his own attempt to differentiate between justified and unjustified norms' (1999, p. 73). Their views diverge as Mouffe continues to argue that Foucault among others has shown that 'there cannot be an absolute separation between validity and power' (p. 15).[33] Habermas, on the other hand—echoing for example in Latour's (2004) criticism of the fake news discourse (see p. 7)—argues that we have to accept that some things are true, and no discursive power can change facts such as that the earth is round. More directly, Habermas builds discourse ethics as a re-interpretation of Kant as opposed to Horkheimer and Adorno's criticism in the *Dialectic of Enlightenment* (1997).[xi] To Habermas, the ideals of Enlightenment are not a lost cause, just not yet realised. He uses this stance to criticise his own time.

> In cultural modernity, reason is stripped of its validity claims and is assimilated to sheer power. The critical ability to take a 'yes' or 'no' stand, to be able to distinguish between what is valid and invalid, is undercut by the unfortunate fusion of power and validity claims. If one reduces the critique of instrumental reason to this core it becomes clear why the *Dialectic of Enlightenment* flattens out the view of modernity in such an astonishing manner (Habermas, 1982, p. 18)

According to Habermas, *rational reason* is merely a type of *reason,* however different from other forms of reason such as the pre-modern's *religious reason.*[xii] He describes this latter 'secular knowledge' as striving towards 'coming to terms with outer nature', formed as a 'cumulative process' of 'scientific and technical progress' that has moved society from 'myth, through religion, to philosophy and ideology' where 'the demand for discursive redemption of normative validity claims increasingly prevails'

---

[33] Quoting Wittgenstein in her critique of Habermas, she writes how 'Where two principles really do meet which cannot be reconciled with another, then each man declares the other a fool and a heretic. I said I would "combat" the other man, but wouldn't I give him reasons? Certainly; but how far do they go? At the end of reasons comes persuasion.' (Quote in *Deliberative Democracy or Agonistic Pluralism?* (1999) note 7, p. 749-50, original emphasis). Repeated similarly in Mouffe's *The Democratic Paradox* (2005a) note 7, p. 97-98.

(1999, p. 197, 2007, p. 11). Once again, he places his proposition at the last stage of human moral progression (see p. 33). The ones successfully participating in a rational discourse are described as *competent actors* (1999, p. 48) or *competent subjects* (pp. 86, 95, 100, 119) while arguing that he diverges from Kant's (2012) idea of *rational beings* (p. 31). His conception and interpretation of what is rational is embedded in the rational ideal of discourse ethics, but do not specify what this rationality means. The only grounds would be the rational found in the theory's axioms (principles) and values (validity claims). In addition, what comes across when reading his text is an idea that what is rational is reflected in the ideals promoted in democracies, such as finding that 'human rights obviously embody generalizable interests' (Habermas, 1999, p. 205; 2010a).

Moral discourse arises out of the need 'for achieving agreement on a controversial norm' (Habermas, 2018b, p. 108), or addressing a 'controversial matter' (1999, p. 103). This means that while morals form a consensual, calm, sense of shared morals, it arises out of conflicts where 'competing parties' return to their 'only remaining shared resource' as 'the communicative presupposition of the discursive situation of those who seek reasonable agreement in the face of persistent practical conflict' (p. 113). This idea dates back to Plato's dialogue *Protagoras,* in which the conflict between right and wrong, or good and bad, plays out against one another. As French and Allbright (1998) describe, 'in a myth about the origin of morality it is suggested that Zeus sent morality to human beings because without it there was the danger that the conflicts between them would lead to the extermination of the human race' (p. 178).[34] For Habermas, discourse ethics thus emerge out of conversations in conflict, setting out to settle the conflict (1999, p. 67) as 'practical discourses are always related to the concrete point of departure of a disturbed normative agreement' (p. 103). His ideal would, in theory, lead to an agreement and understanding—to moral consensus. Moral discourse can thus alternatively be seen as a matter of conflict or consensus, as the process shaping or containing morals. This not only speaks to the character of the moral discourse, and the process leading to moral conclusions, but to the nature of moral statements as such that 'we are prepared to defend against criticism' (p. 56).

---

[34] They also point to the quote by Rest (1986) that 'the function of morality is to provide basic guidelines for determining how conflicts in human interests are to be settled and for optimizing mutual benefit of people living together in groups' (p. 1).

> Conflicts in the domain of norm-guided interactions can be traced directly to some disruption of a normative consensus. Repairing a disrupted consensus can mean one of two things: restoring an intersubjective recognition of a validity claim after it has become controversial or assuring intersubjective recognition for a new validity claim that is a substitute for the old one. (Habermas, 1999, p. 67)

From the idea of morals as tied to discursive conflicts, it also means that the challenged norms may also be replaced by a new paradigm. Habermas here speaks of *social crisis*, described as how discourse gives rise to changes in social identity. This is a subjective evaluation, where historians 'refer to the interpretation that members of a system use it to identifying one another as belonging to the same group' (Habermas, 2007, pp. 3–4). We experience changes in the system as it loses its identity, occurring 'as soon as later generations no longer recognize themselves within the once-constitutive tradition' (ibid). This, Habermas describes, ties *social evolution* to *principles of organisation* as 'highly abstract regulations arising as emergent properties in improbable evolutionary steps and characterizing, at each stage, a new level of development' (p. 7), crisis emerges whenever such principles are inadequate in resolving crisis. Such crisis 'proceeds from internal contradictions' (p. 20) in traditional societies. In liberal-capitalists societies, he continues, this is most sensitive as 'bourgeois society must react to the evident contradiction between idea and reality. For this reason, the critique of bourgeois society could take the form of an unmasking of bourgeois ideologies themselves by confronting idea and reality' (p. 23). As he continues, the ideal when realised would experience the 'side-effects of its own success that endanger the premises for its own functioning' (1986, p. 8), a point which this thesis argues we have experienced in the criticism against the democratic ideal in the fake news discourse (further discussed in the analysis, see Part 3).

In these discourses, not only moral communication materialises. While Habermas links communication to ethics and the rational, 'rational discourse' is merely '*a* form of communication' (2018b, p. 108, emphasis added). Communication is not ethical or unethical in itself, but from

Habermas' perspective, there are ethical and unethical forms of communication. Returning to Habermas' reliance on the theory of speech acts (see p. 25), communication would be more than pieces of information (as in Chomsky's linguistic competence, see p. 25). Rather, communication becomes an action in itself, which reveals messages beyond what is said (locution) but also include the message as intended (illocution) and message as received (perlocution). Resting on Austin's (1962) idea that utterances have performative qualities—the potential to not only describe but influence the world—Habermas develops his own set of speech acts as either normatively moral or immoral. Firstly, there is communicative action (introduced on p. 25), as 'whenever the actions of the agents involved are coordinated not through egocentric calculations of success but through acts of reaching understanding' (1999, p. 58). Immoral communication, by contrast, would be represented by two other forms of speech acts that do not aim for understanding but indeed seek egocentric goals.[xiii] *Instrumental acts* are described as 'an action oriented to success' or even 'egocentric calculations of success' as the intention is one of 'influencing the decision of a rational opponent' (1987, pp. 285–286). *Strategic acts* takes this a step further, 'whereas in strategic action one actor seeks to *influence* the behaviour of another by means of the threat of sanctions or the prospect of gratification in order to *cause* the interaction to continue as the first actor desires' (1999, p. 58, original emphasis). That is, these immoral forms of communication are not restricted to a sense of 'validity to a *common* interest ascertained *without deception*' (2007, p. 108, original emphasis).

Similarly, as communicative action transforms the individual lifeworld into an intersubjectively shared lifeworld (see p. 23), the immoral process is the reversed. Here, Habermas cautions against the *system* interfering with the lifeworld through instrumental and strategic acts. This idea of a system relies on 'the conceptual strategy of systems theory' which 'encompasses normative structures within its language' and thus the realisation of a system that is defined by its own normative 'control system' (2007, p. 5; c.f., 1994, p. 166). The system, as in the case of the language system, risks limiting the lifeworld—thus following an immoral process as it breaks with the power of the individual lifeworld over the intersubjective. In his ideal discourse ethics, the social lifeworld should be constituted by communicative action derived from multiple individual lifeworlds.

Through this, actors collectively create the systems regulating co-existence. For at the same time as they constitute separate paradigms, and Habermas urges caution against the influence of the system over the individual lifeworld, they are mutually formed as the individual lifeworld form normative structures, the systems act as 'society's steering mechanism and the extension of the scope of contingency' (p. 4). Habermas' main concern relates to how systems influence lifeworlds, how they become pathological and perform a *colonisation* of the lifeworld as an expression of the *ideology* (Habermas, 1984).[35] Democratic institutions would then serve the purpose of ensuring that it is indeed the lifeworld determining the shape of the system, rather than the other way around.

---

[35] As described by Mumby (1997), a part of a movement at the time striving for ideological emancipation within a *discourse of suspicion*.

Chapter 2.

40 years of moral reasoning

In the 40 years that have passed since Habermas drew upon his reading from moral development, the field has grown into a new name of moral psychology and has undergone paradigmatic changes in its understanding of moral reasoning. This chapter presents this shift, how the old idea of the relation between communications and morals have been complemented by a recognition that we not only use communication to reach rational conclusions, but more often use it to rationalise our moral intuitions. This modified understanding of the constitution of moral communication leads to the thesis' new point of departure, from which it develops its own moral theory of communication.

Today's moral psychologists describe how the field has continued to evolve from its founding fathers—such as Piaget and Kohlberg (see p. 32)—to a contemporary frontier coloured by three major revolutions. First, there was a cognitive one, slaying 'the two dragons of the twentieth-century psychology—behaviourism and psychoanalysis', then followed an affective revolution starting in the 1980s, and an automaticity revolution in the 1990s (Haidt, 2013). This has led not to a rejection of the ideas that once shaped Habermas' discourse ethics—that we use reason to reach a moral conclusion—but an added recognition that we also use communication as a resource to strategically rationalise our moral intuitions. This new form of moral communication is not only missing from Habermas' framework, but it is presented in discourse ethics as the very basis for immoral communication (as instrumental and strategic acts, see p. 41). Moral psychologists, however, would argue that it represents the more common of the two in terms of how communication links to morals and ethics. As Green and Haidt (2002a) describe, the historic division between those accentuating 'reasoning and "higher" cognition', such as Piaget and

Kohlberg, and those accentuating the emotions governing moral reasoning (such as Freud and behaviourists) has been integrated into a single framework where 'emotions and reasoning both matter, but automatic emotional processes tend to dominate' (p. 517). On a similar note, Habermas builds on Strawson's phenomenology of the moral to argue that 'feelings seem to have a similar function for the moral justification of action as sense perceptions have for the theoretical justification of facts' (Habermas, 1999, p. 50). The new insights into rationalised rather than rational reasoning was popularised through the book *Thinking Fast and Slow* (2011), where Kahneman summarises a lifetime of research for which he received a Nobel Prize. The title reflects the book's emphasis on two forms of thinking. Slow thinking corresponds with Habermas ideal depiction whereby we engage in conscious reasoning, weighing one option against another, to arrive at a conclusion. In contrast, however, they also accentuate another form of fast thinking, an automatic, subconscious, and more frequent type of thinking where we rely on pre-programmed cognitive shortcuts, or as they label it, *heuristics*.[36] This latter process will be integrated in the thesis' development of an empirical understanding of strategic communication's relation to morals.

The chapter rests on a meta-literature review (see p. 203) aiming to capture the trends in research on moral reasoning driving the field in the past 40 years. The aim is not to contribute to this discussion, but use the overview to shine new light on Habermas' philosophical assumptions, and to provide a new theoretical foundation for the empirical theory. What the reviewing authors in the field describe is a discussion not driven by a collective research agenda but rather strands of research (Ellemers et al., 2019), departing here from three frequently emphasised discussions. Each foci challenge the previously presented philosophical assumptions upon which Habermas frames his discourse ethics. Consequentially, the three themes constitute important shifts in the later presentation of the new empirical theory.

---

[36] Building on Simon's (1955) theory of *bounded rationality.*

Moral tribes

Among the key points forming today's literature in moral psychology is the very discussion Habermas addressed on the difference between us and them (see p. 28). Indeed, moral psychology echoes Habermas' emphasis that morals is a matter of social constructions. However, moving from Habermas' ideal depiction that we ought to treat everyone equally (expressed in his (U), see p. 26), moral psychology has arrived at a stance where it is now taken as a matter of course that we do not.

Moral psychology echoes Habermas' presentation of a connection between morals, communication, and our innate strive to form relations with others. Aligning with such perspective, moral psychology branches out not only to psychological and neurological inquiries, but also an interest in social psychology. As in Habermas' framework, it emphasises morals as a tool for humans to form a socially predictable co-existence. Review articles describe how morals serve a social function of building social guidelines (Ellemers et al., 2019, pp. 333–334), by pointing to 'the *social behaviour of individuals living together in groups*' (quote from Ellemers et al., (2019, p. 332), original emphasis, in reference to Gert (1988)), much like how Habermas describes the binding/bonding effect arising from discourse ethics (see p. 25). Just as Habermas, researchers in moral psychology often describe this as an evolutionary trait. Through the ages, our 'minds were shaped by kin selection plus reciprocal altruism augmented by gossip and reputation management' (Hait (2012, p. 190) citing Dawkins (1999)). This observation is described as harmonising with a large set of studies in the field and 'the message of nearly every book on the evolutionary origins of morality' (Haidt, 2012, p. 220). Just as Habermas, these scholars of moral psychology describe it as tied to an evolutionary process as people have sought relations with one another to form the societies that have enabled us to thrive as a species.

Our ability to engage in such relations are echoing Habermas' description, that to succeed we need the ability to understand another's perspective. Habermas talks about perspective-taking (see p. 26), or 'empathy' as he sometimes, and moral psychologists predominantly, label it. While empathy is most often associated with one's ability to feel what another is feeling (feeling *with* someone), moral psychology and cognitive scientists also talk about the concept of 'theory of mind' (ToM). This would point to a cognitive rather than emotional connection, where one is able to put oneself in another person's shoes without feeling what they are feeling (see i.e., Greene & Haidt, 2002). These phenomena point to the social dimensions in moral co-existence, Tomasello (2013) imagining that this ability developed in two tiers. The first tier of social bonds would have appeared in face-to-face interactions as we treat each other well and build a reciprocal relation. As these collectives grew, however, the bonds became more symbolic in a second tier of morality, including social norms from which we form a collective identity. The would align with Habermas' idea of two forms of speech (see p. 24) where the latter furthermore aligns with Habermas' idea of the identity-formation dimension appearing when adopting a social norm (see p. 40). As Ellemers et al. (2019) summarise, the move toward symbolic meanings means a reliance on language as 'moral judgements that function to maintain social order in this way rely on complex explanations and require verbal exchanges to communicate the moral overtones of behavioural guidelines' (p. 333). Furthermore, 'this socially defined nature of moral guidelines is explicitly acknowledged in several theoretical perspectives on moral behaviour' (p. 334).[37]

So far, we see how Habermas and moral psychologists align in their view of communications' function in relation to morals. However, while Habermas would argue that we ought to treat all equally in discourse, the reverse is observed in studies from moral psychology. As Bloom (2016) summarises a series of research on the subject, empathy is not evenly distributed, but creates an irrational empathic spotlight. This is a well-documented phenomenon as we diverge from a rational evaluation of the situation to for example favour the ones geographically near rather than the ones far away, the beautiful over the ugly, and our friends and family over strangers.[38] Recognising this, Singer (2005) consequentially described how evolutionary psychology and neuroscience can work to 'argue against a normative ethical theory' by showing 'how that in some circumstances the theory leads to judgments that are contrary to our common moral intuitions' (p. 331). We state that we believe in one type of moral action (to treat everyone equally) while finding ourselves in situations where we repeatedly diverge from such norms. He takes an example in a hypothetical story of a person walking past a pond where a child is drowning. Consider that you are that person, wearing a quite expensive outfit that would be ruined if you jumped into the waters—would you still save the child? When writing this example, Singer presented it as a reaction to the East Bengal crisis in 1971, questioning why we would probably not hesitate to save the drowning child in front of us, but do not feel it as a rationally pressing moral responsibility to donate the same amount of money as the worth of our outfit to save a child far away (Singer, 1972).[39]

---

[37] In a way, one can see the move from Kohlberg's first two stages however here merely recognising that one communicates morals, not acknowledging the liberal ideals that he too ascribes to the final stages of moral development (see p. 33).

[38] One of his latest empirical evidence (in a row of moral psychologists' research) can be found in Cecil, Marshall, and Bloom (2020).

From the realisation that we group ourselves into moral communities where we agree to treat each other in a commonly defined way, to the recognition that we also mistreat some people as we direct the irrational empathic spotlight—there is a final important point to be made on how we mistreat groups of people not recognised as part of the self. Habermas discusses this in relation to the concept of othering (see p. 28), the ultimate ill that Habermas' discourse ethics is arguably setting out to fight against. In empirical research, it is a well-known phenomenon and thus a key piece for the thesis' development of an empirical theory of strategic communications' relation to morals. For while Habermas argues that we ought to meet all on equal terms, this is not what the empirical sciences show. Rather, they reveal a pattern where we repeatedly mistreat people based on their group identity and if it is the same as ours. For example, we view people as part of a group sharing the same nationality, skin-colour, interests, gender, etc., and lump them together in imaginary groups. This creates an experienced affinity to some people, but not to others.[40] Among such group constellations, Bloom (2016) takes the example of a study of Red Sox and Yankees fans, were the participants thought of themselves as part of the one or the other group. The study found that people had different neurological responses to one's own team loosing or winning the game—not only taking 'pleasure in response to rival group's misfortunes' that people ended up 'endorsing harm against people associated with those groups' (Cikara et al., 2011, p. 6).

---

[39] Singer's example has led to a philosophical debate with examples including Unger (1996) and Kamm (2007), and where Nagel and Waldmann (2013) have shown through empirical studies in a set of four experiments that the reason for our judgement is not merely spatial distance, but argues 'that distance in moral reasoning constitutes an effective proxy for a family resemblance structure combining many otherwise very dissimilar factors that have related effects on moral intuitions' (p. 249).

[40] See i.e., Cikara, Botvinick, and Fiske (2011), Cikara, Bruneau, Van Bavel, and Saxe (2014), Azevedo, Macaluso, Avenanti, Santangelo, Cazzato, and Aglioti (2013), or Vanman (2016).

Weird moral taste, and beyond

The social communities formed as moral communities means that we can unpack them through their shared reliance on common moral values. Contrary to Habermas' own presentation, this thesis approaches discourse ethics as a normative proposition, arguing that it reflects the moral community in which it arose (see p. 40). It is not, however, understood in this thesis as Habermas presented it—as a theory reflecting a so widespread norm that it ought to be approached as a universally valid moral theory. This misinterpretation of viewing the long dominating western moral philosophy as a matter of course has also been the subject of discussion in moral psychology. They realised that their studies were conducted predominantly by people in this moral community, and that the research subjects upon which they rested their conclusions were likewise embedded in the same community. That is, both the researcher and the researched were sampled from the same moral context when theorising what morals are. In response, scholars have started to recognise that this is not a universal but a particular worldview. Upon realising recognising this underlying moral bias characterising the sampling and analysis, moral psychology has broadened its study of morals—finding that the perspective initially presented as universal is in fact a statistical outlier when reviewed in relation to a global context. At the same time, the scholarship has shown how we simultaneously share human morals, below expanding on this theoretical balancing act, and finding in this new perspective the thesis' criticism towards Habermas' discourse ethics by acknowledging that it is a particular and not a universal moral stance that is perhaps, rather, a bit weird.

Aligning with Habermas, moral psychology describes how morals vary depending on the time and place. This means that multiple moral systems are operating in the world at once (Ellemers et al., 2019, p. 334). As the field branches out to take an interest in the sociological dimensions of morals, in criticism of Piaget and Kohlberg, Shweder questioned the universal validity of the moral assumptions derived from these scholars' cultural context. He showed that morals are culturally dependent, not biologically universal (see i.e., Shweder & Menon, 2014). This division has formed the basis for a long-debated question in moral psychology, dating back to its origin in developmental psychology, as the question *Where does morality come from?* (see p. 32). The division then stood between the naturist and nativist argument, that morals are either innate (and thus globally universal as part of the human DNA) or learned through one's socialisation in the society one is born into. In time, however, the very question was questioned—for what if neither proposition told the full story?

A series of studies has since built a theory where the two aspects are reconciled into a single proposition. For studies indeed show that we normatively evaluate situations differently, however there is at the same time a common pattern underlying these evaluations. Building on previous scholarship (see i.e., Graham et al., 2013), these underlying patterns were framed in the *moral foundations theory*. Initially, it was presented as intuitive ethics (Haidt & Joseph, 2004), presented as a metaphor where all human beings are born with a tongue that holds the same taste receptors but, depending on the context in which we are raised, theorise that we develop different moral tastes. Studies show that we share a set of six underlying value dimensions as dichotomous pairs of what we consider moral or immoral—care vs harm, fairness vs cheating, loyalty vs betrayal, authority vs subversion, sanctity vs degradation, and liberty vs oppression.[41] In that way, we would have similar moral compasses, for example caring for the weak, favouring the clean, respecting authorities, favouring the loyal, and striving for fairness. However, these receptors may result in different moral interpretations—such as respecting your elders or respecting your boss, in caring for children or caring for the sick. This is not only tied to national cultures. It has also been presented by Haidt and Graham (2007) in a study that liberals rest primarily on fairness and care while conservatives focus on loyalty, authority, and sanctity. There is thus a support for Habermas' idea of universal moral values, but not necessarily in the sense he envisioned and still not appearing as values tied to communicative values.

---

[41] Shifting in 2012 from the originally proposed terms harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity (footnote 1 in Graham et al., 2013, p. 60) with Haidt adding liberty/oppression in a later publication (2012), and additional foundations since suggested alongside a string of critique as described in Graham at al. (2013, pp. 98–107).

The recognition of others' moral perspective finally led to a startling realisation—that moral philosophy, as well as moral psychology, had indeed been assuming their own particular moral standards as universal. Western philosophers were philosophising about morals from *a* moral point of view, *their* moral point of view, indeed common-sense to them and their colleagues but in a global context would be considered weird. In moral psychology, this problem was even embedded in the empirical studies—as the research subjects too were from this particular moral context, and thus empirically confirmed the scholars' moral intuitions. Among the scholars noting this problem, Henrich, Heine, and Norenzayan (2010) presented a particularly influential article describing this weird perspective as an acronym to show in what context this perspective is dominant.

> Behavioral scientists routinely publish broad claims about human psychology and behavior in the world's top journals based on samples drawn entirely from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies. Here, our review of the comparative database from across the behavioral sciences suggests … that WEIRD subjects are particularly unusual compared with the rest of the species—frequent outliers. … The findings suggest that members of WEIRD societies, including young children, are among the least representative populations one could find for generalizing about humans (Henrich et al., 2010, p. 61)

Quite contrary to the previously dominating moral scholars' view that theirs was a universal moral perspective, a review of global morals would not even find their stance as dominating but a statistical outlier. It was found to be most common in Western, Educated, Industrialised, Rich, and Democratic societies (Henrich, 2020; Henrich et al., 2010). These societies are prone to individualistic rather than a collectivistic sense of morality, and conceptualise morals in relation to intent rather than consequences.[42]

Habermas too addresses this central critique against (U), as many before had criticised Kant for the egocentric perspective in his categorical imperative. Among them, he points to Paul Taylor's objection to Kurt Baier's proposal, Habermas emphasising that 'in view of the anthropological data available we cannot but concede that the moral code expounded by Kantian moral theories is indeed only one among many' (Habermas, 1999, p. 78).

> However deeply our own conscience and moral outlook may have been shaped by it, we must recognize that other societies in the history of the world have been able to function on the basis of other codes. … To claim that a person who is a member of those societies and who knows the moral code nevertheless does not have true moral conviction is, it seems to me, fundamentally correct. But such a claim cannot be justified on the ground of our concept of the moral point of view, for that is to assume that the moral code of liberal western society is the only genuine morality. (Taylor (1963, p. 570) cited in Habermas (1999, p. 78))

As Habermas continues, cognitivists in general and Kant in particular would run the reason of committing a *petition principii*. This thesis would argue that Habermas himself runs the same risk. For while he argues that this proposition avoids such criticism towards discourse ethics—that it is not a matter of a universal principle, but a contextually bounded moral principle—this the thesis would argue nonetheless is based on a variation of Kant's principle, formed from an axiomatic relation to (U) and (D).

From consensus to intuitions

Contrary to Habermas' discourse ethics, today's moral psychologists recognise that morals are not only that which is rationally concluded, but that morals are also intuitively rationalised. This thesis would argue that we see it in the very rationalisation of Habermas' moral outlook, expressed in his discourse ethics. For it is not a universal stance, but in this thesis rather recognised that it was always WEIRD. This recognition opens for a second important discussion in moral psychology—from the ideal that morals are the result of rational reasoning, towards a description that morals are also, simply put, formed from our gut-feeling. Now, one might point out that such gut-feeling is an expression of the moral system previously concluded to be rationally moral as it is accepted by the community. Indeed, scholars are yet to determine which is the hen and egg, and are perhaps doomed to never find an answer. This thesis is not intended to settle the question of how the two relate to one another, but in the move towards an empirical theory merely adds the new rationalising function of communication in moral discourse to form a new theoretical framework. This rationalising function is illustrated in a series of studies which show just how strong such rationalising force is, that people make the most irrational connections to rationalise one's moral intuition. The section below presents this new perspective of rationalised morals. From that, the thesis would argue that we must acknowledge that the form of communication which Habermas' conceptualises as immoral, must be included in an empirical theory if we are to understand moral communication.

---

[42] While this research shows how we conceptualise morals differently in different parts of the world, research has shown how moral intuitions appear across WEIRD and non-WEIRD cultures (Doğruyol, Alper, & Yilmaz, 2019).

One of the major revolutions in moral psychology is described by Haidt (2013) as a recognition of the *moral intuitions* guiding moral reasoning. Habermas too talks about moral intuitions (for example in, Habermas, 1999, pp. 196–) but then rather as a matter where discourse ethics conceptualizes a set of moral intuitions, that is his (U), (D) and the idea of his validity claims. In moral psychology the use of the word is different, as intuitionism grounds the third assumption in moral foundations theory alongside a recognition of the dual nature of morality as simultaneously innate and learned (see p. 49).[43] With grounds in neuroscience and evolutionary biology, the *social intuitions model* is Haidt's own way of capturing this recognition (2001, 2008). He compares this type of moral reasoning to aesthetic reasoning, where 'we see an action or hear a story and we have an instant feeling of approval or disapproval' (Greene & Haidt, 2002, p. 517), and describes how it rests on the recognition 'that emotion is a significant driving force in moral judgment' (p. 522). Haidt presents a range of metaphors for this moral process, in his most popular article through the very title *The Emotional Dog and its Rational Tail* (2001, 2008). In the paper, Haidt challenges the dominant rationalist models, arguing that our intuitive reasoning is like an emotional dog, taking us in any direction where the rational tale merely tags along and tries its best at post-rationalising our moral intuitions on which route was the best to take.[44] The theory rests on the observation that we seem to act instinctively to moral situations. On the one hand, some point to moral communicative expressions such as the automatic bodily reflexes in a disapproving frown (Ellemers et al., 2019, p. 335), others to linguistic articulations conceptualised as *motivated reasoning* (Kunda, 1990) and similar descriptions (Gazzaniga, 1990; Nisbett & Wilson, 1977) of what Greene and Hait (2002) describe as 'post-hoc reasons to justify their actions and judgement' (p. 517). In moral psychology, this has been shown through an interview technique where the interviewer presents a short story, a *vignette*, and then asks the interviewee if these actions are wrong and why. The purpose of these stories is to 'trigger intuitive judgements' (Haidt, Björklund, & Murphy, 2000, p. 5), which contrary to moral stories such as Kohlberg's 'Heinz dilemma'[45] these vignettes include no harm, but are morally disgusting. For example, in the article partially titled *Is It Wrong to Eat Your Dog?,* Haidt, Koller, and Dias (1993) present the following vignette.

---

[43] The historic context presented by i.e., Graham et al. (2013) and by Greene and Haidt (2002).

> A family's dog was killed by a car in front of their house. They had heard that dog meat was delicious, so they cut up the dog's body and cooked it and ate it for dinner. (Haidt et al., 1993, p. 617)

While the interviewees would instinctively find that this is wrong, the story is written in such a way that 'the participant would be prevented from finding the usual "reasoning-why" about harm that participants in Western cultures commonly use to justify moral condemnation' (Haidt et al., 2000, p. 5). That is, making it impossible to provide a rational reason, to speak the language of Habermas (see p. 38). As Haidt (2012) would later describe, you probably 'felt an initial flash of disgust, but you hesitated before saying that the family had done anything *morally* wrong' (p. 3), as you cannot derive the immoral behaviour to rational reasons. Nonetheless, Haidt, Koller, and Dias (1993) describe how the interviewees desperately search for this, inventing rational reasons outside the scope of the vignette to rationalise their moral intuition of disgust towards the act described. For example, 23% claimed that the family might harm themselves in some way, and 10% found other potential victims (p. 618) to derive the immorality of such behaviour to reasons.

---

[44] Continuously used as a metaphor in response to his critiques (Haidt, 2003, 2004).

[45] Where the question is whether it is morally right of a man to steal medicine to save his dying wife.

To illustrate just how deep this innate drive runs, scholars have made a series of studies challenging the rationalising urge. Some have tied it to the very functionality of our brains, moving from moral psychology's interest in social psychology to an interest in cognitive neuroscience. What they find is that we seem to look for patterns, linking causes to effects, even when such links do not exist. To take an extreme example, a classic study shows this in the crudest way in persons with *split brains*. The brain is divided into two parts (the left and right), connected through the corpus callosum.[46] As different parts of the brain serve different functions,[47] they are constantly collaborating, sharing information by sending neurological signals to each other, with what Gazzaniga (1990) describes as an 'interpreter' in the left hemisphere, 'a device that allows us to construct theories about the relationship between perceived events, actions and feelings' (p. 1293). He continues to describe how this happens automatically, 'carried out by the brain prior to our conscious awareness of them' while 'our subjective belief and feeling is that we are in charge of our actions' (p. 1293). Among people where the brain's two halves are not connected, this results in a person's brain perceiving two separate realities. Gazzaniga (1990) takes a few examples where researchers let a person with a split brain sit in front of two pictures, separated from each other so that each eye would only experience one picture. This allowed the person to perceive two realities separately. After that, the person was asked to choose two pictures from a set and explain their choice…

---

[46] To disconnect the two halves of the brain is, for example, a last surgical resort for people with epilepsy.

[47] Beyond the popular description that one is the logical side and the other the creative, Gazzaniga (1990) describes the brain as divided into two sets of capacities 'with the left hemisphere specialized for language and speech and major problem-solving capacities and the right hemisphere specialized for tasks such as facial recognition and attention monitoring'.

> …a chicken claw was shown to the (speaking) left hemisphere and a snow scene was shown to the (silent) right hemisphere. Patient P.S. easily picked out related pictures from a set of eight choices. His left hand chose a snow shovel and his right hand chose a chicken. When asked why he had picked those particular pictures, P.S. said, 'Oh, that's simple. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed' (Gazzaniga, 1990, p. 1318)

Even though the two realities did not have a rational connection, the brain's instinct to present a rational explanation was so strong that it invented one. The interpreter 'sustains a running narrative of our actions, emotions, thoughts, and dreams. The interpreter is the glue that keeps our story unified and creates our sense of being a coherent, rational agent. To our bag of individual instincts it brings theories about our life' (Gazzaniga, 1990, p. 1320). This would give a physiological explanation to the observed behaviour of rationalising our moral intuitions, even if they are not rational to begin with.

As a final point, this could also explain what Ellemers et al. (2019) to as a string of research in moral psychology on how rationalisations are used to explain one's own immoral behaviour. This is described as 'the explicit self-awareness and autobiographical narratives that characterise human self-consciousness, and moral self-views in particular where 'people are highly motivated to protect their self-views of being a moral person' (p. 334). Or, as Haidt moves from a dog and its rational tale to an elephant-rider making rational sense of the elephant's intuitively chosen path (Haidt, 2012), our ability to post-rationalise is thought to originate in the fact that 'once human beings developed language and began to gossip about each other, it became extremely valuable for elephants to carry around on their back a full-time public relations firm' (p. 46). These types of rationalisations are operationalised through a set of different tactics such as 'by refining one's behaviour, averting responsibility for what happened, disregarding the impact of others, or excluding others from the right to moral treatment, to name just a few possibilities' (Ellemers et al., 2019, p. 334). In reference to a series of studies, they take examples of empirical findings which suggest that we can simply let the moral rules that we have slip our mind, that we engaged in behaviour that broke with such a rule, or hold our own behaviour in a higher regard by disregarding those who follow the rules more diligently.[48] The key to this line of research, as Ellemers et al. (2019) highlight, is that our self-identity of being moral does not mean that we engage in moral behaviour. Rather, what it means is that human beings will communicate with oneself, as well as with the world, a version of the story that elevates one's behaviour as moral—regardless of whether it aligns with one's behaviour. This recognition of the difference between a person's presented morals and their moral behaviour leads to both types being studied in moral psychology. The key to understanding morals is found not in the study of the one or the other, but by studying the observed dynamics between our normative stance and our behaviour. As the authors argue, we find in norms and actions, as well as in the intricate interplay between the two.

> Few would disagree that morality ultimately lies in action and that the study of moral development should use action as the final criterion. But also few would limit the moral phenomenon to objectively observable behaviour. Moral actions is seen, implicitly or explicitly, as complex, imbedded in a variety of feelings, questions, doubts, judgements, and decisions. … From this perspective, the study of the relations between moral cognition and moral action is of primary importance (Blasi, 1980, p. 1)

As Ellemers et al. (2019, p. 335) quote Blasin above, they continue to describe how this in turn has become even more influential with the above-presented idea of moral intuitions by Haidt. As this thesis argues, this ought also to be an added focus in the empirical understanding of the relation between strategic communication and morals—looking not only at the structural suggestion of a normative ethics, but also at how these norms relate to the same promotor's own behaviour.

---

[48] Contrary to Habermas depiction that 'someone who acts against his better judgement must not only fact the moral rebukes of others but is also prey to self-criticism, and thus to "bad conscience"' (1994, p. 34)

Chapter 3.

Strategic moral communication

From Habermas' discourse ethics, resting on the ideal that morals are the result of rational communication, this thesis presents a new theoretical framework based on the added recognition from moral psychology that we use communication to rationalise our moral intuitions. In short, the thesis argues that strategic communication is not immoral, but that moral communication is strategic. The following section unpacks this theoretical perspective, translating the new stance from moral psychology in relation to theoretical conceptualisations of communication in general, and strategic communication in particular. Together, this will build the thesis' proposition for a new moral theory of communication as *strategic moral communication.*

Before we start, as a first important point we need to recognise how the shift from Habermas' normative to the thesis' empirical aim means that this new theory does not seek another ought, but instead a proposition for what is (see p. 30). Such an aim was never pursued in Habermas' discourse ethics, from his very first explorations of the concept of a public sphere describing the ideal 'as a historical category,' 'limited to the *liberal* model of the bourgeois public sphere, to its emergence and transformation' (1989, p. xviii). Discourse ethics was never intended to describe what was, is, or even will or can be, as it only intended to theoretically present what Habermas argues is a widespread ideal. This requires a philosophical shift on our view of morals. As Tiberius (2015) notes, 'practitioners of philosophical and interdisciplinary moral psychology do not have the conception of what the subject matter of moral psychology is, which makes it tricky to bring the two into conversation with each other' (p. xi). While moral philosophy was first to conceptualise morals, formalising theories on how the world ought to be, this is now contrasted against fields with a moral interest grounded in empirical observations. For example, sociologists describe how ‘a sociology of the moral will never become an ethical theory. It will never be able to claim for itself moral quality, be it good or bad, be it nice and helpful or cynical’ (Luhmann, 1996, p. 32). Moral psychologists even exclaim that ‘the fact that one cannot derive morality from psychological research is so screamingly obvious that I never thought to explicitly write it down’ (Bloom, 2010, p. 490). These scholars are not interested in right or wrong, but how humans (including scholars) engage in moral reasoning. Habermas also acknowledges this difference, influenced by his reading of an early moral psychologist as he writes that while science ‘can test whether a philosopher’s conception of morality phenomenologically fits the psychological facts. Science cannot go on to justify that conception of morality as what morality ought to be’ (Kohlberg referenced in Habermas, 1999, p. 39). This is also the perspective pursued in this theoretical proposition, as an empirical theory.

To unpack this new theory, the chapter translates the empirical insights from moral psychology in relation to theories of strategic communication. It incorporates a discussion on the similarities and differences between Habermas’ discourse ethics and moral psychology’s stance in relation to communication science’s understanding of communication. Continuously, it is presented in the chapter below in relation to an ongoing discussion on the nature of strategic communication,[xiv] unpacking the theory’s conception of communication, morals, and strategy.

Communication

Just as discourse ethics, the theory proposed in this thesis is a theory of *communication*. However, it diverges from Habermas’ conceptualisation in relation to three key points. First, while also building a theory of moral communication, communication is recognised in this theoretical framework as including more than linguistic expressions. Secondly, while Habermas proposes key foundations for communication based on a normative depiction of moral and immoral communication, the thesis proposes a set of normatively neutral elements of relevance to understand the process of communication. Finally, it reimagines the meaning of the concept strategic communication, diverging not only from Habermas' idea that it is immoral by default, but also challenges the dominant definitions of strategic communication within the field, as the thesis suggests that it should be studied as a philosophical concept rather than a professional practice.

This thesis diverges from Habermas' sense of communication as language only in his discourse ethics (see p. 23). Instead, the thesis rests on a broader definition of communication, where *language* would refer to texts with semantic structures while it is only one form of *communication*. Communication would cover a broader set of phenomena, such as visual communication (images, graphic design etc.), by acting in a certain way (such as deciding to attend or not attend an invitation to an event), by staying silent, or by not acting is also an act of communication. As traditional communication scholars would describe, communication covers anything from 'a nod or a wink, a drumbeat in the jungle, a gesture pictured on a television screen, the blinking of a signal light' (Weaver, 1949, p. 11), to take a few examples. Strategic communication scholars repeat that communication 'should not be limited to formal messages, while actions also convey meaning and should, therefore, also be part of strategic communication. What we do is often more important than what we say' (Paul, 2011, p. 28). This aligns with moral psychology's description of morals as expanding beyond words to include bodily communication such as cringing one's face to communicate disgust, or shaking one's head to communicate disapproval (see p. 53).

The thesis also argues we find better structures to form an empirical rather than normative understanding of communication in basic communication theories. Traditionally, there have been two frequently mentioned clusters of theories in communication science, the first being the *transmission view* which depicts communication as the one-way transmission of messages. Here, communication is described as a matter of straightforward influence, illustrated in metaphors of a hypodermic needle injecting messages, as a magic bullet hitting its target (Lasswell, 1948), or as the most efficient transmission of radio signals (Shannon & Weaver, 1949). The second paradigm is the *ritual view* in which one returns to the etymological interpretation of *communicare* as 'doing together'. Here, communication is approached as a two-way creational process where multiple actors engage in a common conversation. Going back to the religious ties to propaganda (see p. 4), the first could be illustrated in an authority's teaching of the faith, while the latter would be found in 'the sacred ceremony that draws people together in fellowship and commonality' (Carey, 2009, p. x).[49] The two are frequently used in strategic communication studies (Van Ruler, 2018), and corresponds in part with Habermas' idea of immoral and moral communication.[50] This thesis, however, does not treat them as mutually exclusive. Rather, it aligns with voices such as Craig (1999, pp. 124–128), advocating for a dismissal of the dialectic opposition between a transmissional and a ritual view, as he argues for an inclusive *constitutive model of communication.* Communication, this thesis argues, appears as both the transmission and co-construction of phenomenological realities (echoing Habermas' description of the two functions of language as ordinary or ritual language, see p. 24).

The transmission view also holds the merit of breaking up communication as a phenomenon into a set of elements. This, the thesis argues, can form a new stance to review communication that diverges from Habermas' normative elements in his (U), (D), and validity claims, where a review of the elements in communication is here argued to form a theoretical conceptualisation of communication that is not moral by default. To that end, we find in the transmission view of communication a set of influential models such as Shannon and Weaver's (1949) *mathematical model of communication,* where they divide communication into a five-part process.[51] Another influential example is Lasswell's (1948, p. 216) five-part proposition,[52] Braddock (1958) later adding two additional elements.[53] Combined, these elementary depictions of the process frames communication as a process divided into eight elements; whereby (1) a *communicator* (2) seeking to convey an *intention* (3) by sending a *message* (4) through a *medium* (5) that is received within an *environment* (6) where it reaches a *receiver* (7) that makes an *interpretation* (8) that leads to an *effect* or non-effect. This thesis approaches these as elements in a process communication, assuming however nothing about the order or sequence of such elements.

---

[49] Similarly presented as a critique against the transmission model by scholars such as Deetz (1994), Pearce (1989), and Shepherd (1993).

[50] As moral communicative acts (see p. 25) or immoral strategic and instrumental acts (see p. 41).

[51] A schematic diagram that begins with a *message* transmitted (including an *information source* and *transmitter*) that produces a *signal* that is disrupted by *noise* before resulting in a *received signal*, a(nother) *message* received (including a *receiver* and *destination*) (Shannon & Weaver, 1949, p. 381).

[52] A series of questions that 'describe an act of communication' by answering 'Who Says What In Which Channel To Whom With What Effect?'

This model also helps illustrate the thesis' divergence from previous conceptualisation of strategic communication in the field. When defined as a field of study in the early 21st century, it was framed as a practice by *organisations* (Hallahan et al., 2007, p. 4). In time, this has moved from a focus on organisations to descriptions of *communication agents* acting on behalf of a *communicative entity* (Holtzhausen & Zerfaß, 2013, p. 74), most recently accepting that strategic communication is not only conducted by groups but also by *single individuals* such as celebrities and politicians (Zerfaß et al., 2018, p. 494). Today, definitions propose that strategic communication entails 'organisations and other entities' (Nothhaft et al., 2018a, p. 329). This thesis argues that this allows us to approach strategic communication as a practice by any group of people (from football teams, to study groups, or families) where there is a 'one' constructed in relation to an 'other'. This would also cover the social communities presented by Habermas as groups forming a moral consensus.[xv] The recent acceptance of *single individuals* (such as in studies on influencers and other one-person organisations) points to the dissolvement of the importance of defining strategic communication based on the actor. This leads to the thesis' own conceptualisation of strategic communication, disregarding the nature of the actor to a suggestion that what unites the field's interest is rather the *strategic* nature of *communication* (further unpacking the new definition throughout this chapter). By acknowledging all these entities as engaging in strategic moral communication, the thesis argues that we expand the definition of strategic communication from the professional realm of PR and marketing, or internal communication within an organisation (as otherwise often emphasised as core practices of strategic communication).

---

[53] To include the questions 'Under WHAT CIRCUMSTANCES?' and 'For WHAT PURPOSE?'

The shift from a defining emphasis on the communicator, to a suggestion that strategic communication, by not being dependent on the type of communicator, is rather founding its phenomenological essence in the idea of intentions and effects. By not limiting strategic communication to a certain type of communicator, it argues that our definitions of strategic communication should rather focus on defining the strategic nature of communication as a philosophical phenomenon in itself.

Morals

Just as in Habermas' discourse ethics, the proposition of the thesis for a theoretical framework of strategic *moral* communication means that it is a theory about morals. However, in the move from a normative to an empirical aim, this second section challenges the Habermasian assumption that strategic communication would be immoral. Instead, the section presents the thesis' theoretical proposition for how to move towards an empirical theory of strategic moral communication that does not rest on such a moralised conceptualisation. Adding to that, after conceptually reframing strategic communications relation to morals, the section presents a suggestion for how the above-presented elements in the process of communication can be used to study norms by reviewing the ways in which communication is defined as moral or immoral.

Habermas is clear that strategic communication is immoral (see p. 41). This thesis would disagree. To that end, it starts by looking at how morals have been approached in the field of strategic communication. While reviews of publications in the *International Journal of Strategic Communication* would suggest that there has been only one publication dedicated to the study of morals in strategic communication (Werder, Nothhaft, Verčič, & Zerfaß, 2018, p. 343), this thesis would argue that there are a wider range of implicit discussions about morals in the field. If we look at another related field, *communication ethics*, reviews find it to be a fragmented field of study (Cheney et al., 2011), mapped through two literature reviews covering publications from 1915–1985 (Arnett, 1987) and 1986–2004 (Arnett, Arneson, & Bell, 2006). Together, they help point out six common interests in the field: 1) *democratic communication ethics* as ethics tied to public communication and participation in society, 2) *universal-humanitarian communication ethics* as the human qualities for ethical behaviour, 3) *codes, procedures, and standards in communication ethics* for a cluster of studies on the codes and guardians of ethical codes, 4) *contextual communication ethics* concerning cultural differences in ethical judgement, 5) *narrative communication ethics* interested in how 'good' stories are created, and 6) *dialogic communication ethics* discussing one-way vs two-way communication processes. All these topics are undoubtedly aligning with Habermas' normative theory of discourse ethics, such as the focus on public participation (U), humanitarian ethics, the deontological sense of rules to guide one's behaviour, his validity claims, and the context-dependent structure of morals and general dialectic ideals expressed in his (D).

From this framework presented in communication ethics, the thesis argues that we can find parallel discussions in strategic communication about morals and ethics. Through a review of publications in strategic communication on the basis of these themes, the thesis finds a wide range of studies discussing morals in strategic communication; from publications emphasising 1) the public's participation in political processes (Holtzhausen, 2010; Men, Yang, Song, & Kiousis, 2018), discussions about the nature of the public sphere (Bentele & Nothhaft, 2010, 2015; Dahlgren, 2005), departing from the Habermasian ideal for organising public communication ethics (Feldner & Meisenbach, 2007; Self, 2010); 2) discussing the ethical virtues of communication professionals (Kelm, Dohle, & Bernhard, 2017) and among organisational members (Bagozzi et al., 2013); 3) researching professional codes of conduct (Ikonen, Luoma-aho, & Bowen, 2017); 4) how to adapt to other cultural contexts (Kang & Davenport, 2009) and create strategies that match an evaluation of the audience's perception of one's own responsibility in crisis situations (Schwarz, 2008; Zhao, 2017); 5) using narrative analysis as a method and storytelling as a key focus of interest (Johansen, 2012; Laaksonen, 2016; P. Weber & Grauer, 2019; Zhao, 2017), especially in relation to branding, CSR, and image repair while simultaneously engaging in a parallel discussion about authenticity (Frankental, 2001); 6) and a continuous discussion on the ethical difference between information and communication, one- and two-way process, and how actors achieve dialogue within e.g., financial communication (Koehler, 2014), employee communication (Gode, 2019) and in relation to the public's participation in political discussions (Huang, Ao, Lu, Ip, & Kao, 2017). This also show how the field's discourse on morals aligns with the normative assumptions in Habermas' ideal. While aligning with his principles, only some, however, explicitly reference Habermas' theory, more commonly they use his normative points without recognising their normative nature. The same pattern also repeats when looking beyond strategic communication to adjoining fields such as public relations. Here, literature reviews describe how Habermas' theories are not seldom accepted at face value without depth in the theoretical discussions, critique, or empirical grounding (Buhmann et al., 2020). Again, themes emerge in line with Habermas' theory, such as the first focused on 'lobbying and public affairs, political publics, civil society or the role of publics in democracy', the second cluster on 'lobbying and public affairs, political publics, civil society or the role of publics in democracy' and communication as 'building dialogic stakeholder relationships and engagement through dialogue' (p. 450–452). One example of this alignment is found in Grunig's influential model emphasising *symmetrical public relations*, framed in strategic communication as a case of 'normative theorising' where 'each participant in the communicative process is equally able to influence the other' would constitute 'the most effective and ethical way to conduct public relations' (Hallahan et al., 2007, p. 21). As such, it is not only a shift from Habermas' discourse ethics that is required in the formation of an empirical theory of strategic moral communication, but it also necessitates a shift from previous discussions about ethics and morals in communication studies in general and in strategic communication in particular. While they have been taking Habermas' discourse ethics as a point of departure in their own conceptualisations, theories, and models, this thesis will argue that we need to engage in a meta-theoretical discussion as the field at times reads as a reflection of Habermas' norms.

While strategic communication has embraced Habermas' ethics as a normative matter of course, this in itself is a theoretical paradox as he conceptualises strategic as immoral. The thesis calls for a morally neutral definition of strategic communication, aligning with neutral definitions of propaganda when defined as 'the *deliberate* attempt to persuade people to think and behave *in a desired way*' (Taylor, 2003, p. 6, original emphasis). In such definitions, the normative judgement remains as 'good or bad depends upon the merit of the cause urged, and the correctness of the information published' (Bernays, 2005, p. 48). From this stance, advertising would be a type of economic propaganda and public relations 'a communicative process designed to enhance the relationship between an organisation and the public and, as such, is a branch of propaganda, albeit a nicer way of labelling it' (Taylor, 2003, p. 6). This normatively neutral definition of strategic communication as not holding a moral valence also means that we can deconstruct normative representations of moral and immoral communication. The purpose would be to seek understanding of what in the process of communication makes actors evaluate it as moral or immoral. This, the thesis argues, would in turn enable a methodology where we study moral communication, untangle the elements of moral interest, and from that reconstruct a model of the normative ethics articulated in such moralising representations (methodology to be presented in Part 2). Habermas' description of communication can only be used to determine to what extent observable communication aligns with his ideal theory. This will constrict the analysis to a pre-defined understanding of moral communication, and thus misrepresent any normative ethics that are not aligning with his ideal. A new framework is thus necessary, here suggesting that we build on the previously presented elements of communication (see p. 61). To capture the moral values grounding assessments of communication as moral or immoral, the thesis suggests that we can use the elements in communication as focus points in the definitions of moral and immoral communication. From this point, we can start to see how even Habermas' own discourse ethics can be unpacked in relation to his moral assessment of moral and immoral communicators, intentions, messages, appropriate mediums, the possible influence on the moral evaluation in relation to the environment, the target recipient, and his emphasis on the difference between illocutionary and perlocutionary acts as linked to discussions about the different actors' perspectives in the communicative context.

Table 1. Normative values to elements in the process of communication
Illustration of how the elements in the process of communication can be used as a tool to capture normative dimensions, here taking the example of Habermas' discourse ethics.

Element	Moral	Immoral
Communicator	Individuals from the public	Organisations
Intention	To reaching understanding	To pursue self-serving interest
Message	Based on valid claims such as presenting something that is true and truthful	False and dishonest
Medium	Inclusive discourse	Closed conversations
Environment	Democratic	Authoritarian
Receiver	All affected	Some
Interpretation	Open minded (engaging in perspective taking)	From a stance in othering
Effect	Understanding, consensus and equal compromise	Dissensus or in favour of some rather than all effected

To take Habermas' discourse ethics as an example, he emphasises that moral communication is represented in: 1) the *communicators* as human beings rather than organisations, 2) expressing a genuine *intention* to reach understanding, 3) where the *message* is following the agreed upon set of validity claims, 4) through the medium of *communication* as language, 5) in an inclusive *environment* where all effected by the morals are invited, 6) in which one's message is *received* by other participants in the discourse, 7) each making an *interpretation* that aims to understand the other's perspective, to finally 8) result in an agreed moral standard that will serve as the *effect* regulating the shared co-existence in the participants' shared community. Together, this deconstruction of this normative proposition would help us understand the character of the normative theory by presenting the moral and immoral values tied to the elements of communication. It also means that communication is approached as a normatively neutral phenomena in itself in the proposed theoretical framework, at the same time using it to analyse the normative discourse conception of moral and immoral communication in relation to the normative valence tied to each element.

Strategic

This takes us to the final point argued in this thesis, which is that strategic communication is indeed not immoral by default, but moral communication is instead found to be strategic. This strategic nature of moral communication presents an oxymoronic proposition in relation to Habermas' discourse ethics, but does align with contemporary moral psychology's depiction of how we use communication to rationalise our moral intuitions. From this stance, the very practice of moral communication becomes strategic through: a) the selection of phenomenon acknowledged in discourse by being given a name as a concept, b) which elements are accentuated in the definitions of these concepts, c) how elements are attributed different moral valences in different definitions, and d) how moral values together can be seen to form an idea of what constitutes moral or immoral phenomena. Due to the 'ought' character of morals (see p. 30), this communicative process of presenting a norm, this thesis argues, will inevitably be strategic.

If we move from Habermas' conception of strategic communication as immoral, however, how then does this new theoretical framework conceptualise the *strategic* in strategic communication? To unpack this, we return to the field's own conceptualisations. Aside from the actor (previously discussed, see p. 61), strategic communication has often been defined in relation to the meaning of the label *strategic*. Initially, the field's interest was presented as tied to the *purposeful* use of communication to achieve a goal (Hallahan et al., 2007),[54] while others would argue that 'not all purposive communication is strategic' (Zerfaß et al., 2018, p. 493), as it is a matter of 'not random or unintentional communication—even though unintended consequences of communication can avertedly impact the ability of an organization to achieve its strategic goals' (Hallahan et al., 2007, p. 27). This would resituate the label of strategic from intentions to effects. As such, strategic communication is inconsistently presented in the field as alternatively a matter of intents or effect. Indeed, Habermas does not take much care in drawing up a clear distinction between intentions and consequences either, as he conceptualises immoral communication as tied to a dishonest intent by entering a conversation with a pre-determined goal *or* in the success of such an actor achieving their goal (see p. 41). To that point, thesis echoes Cooren's (2000) criticism of Habermas' assumption that morals are defined in relation to their illocutionary character. As he points out, 'reaching *agreement* assumes that I have successfully *made* my partner *accept* something; at the same time I can undoubtedly refer to an *effect*.' For, as Cooren continues, 'A speaker who asks someone to put out her cigarette will be *strategic* if he has the misfortune to target a 'causing to do' in other words if he really wants her to put out her cigarette' (pp. 299–300, 303). As a result, his theoretical idealisation of communicative action can be seen as based on 'internal inconsistencies which allow Habermas to *purify* communicative action of any rhetorical dimension' while suggesting for '*another* ideal speech situation, a situation which allows the expression of all *strategic* and *rhetorical* forms of discussion inherent in these processes' (p. 296, original emphasis).

---

[54] Others more specifically argue that 'strategic' should be tied to an organisation's relative success in relation to its competitors (Zerfaß et al., 2018, p. 500).

To clarify the thesis' stance, it departs from a perspective where strategic communication is either tied to intentions, effects, or both. This means that the 1) intentional inconsequential, 2) intentional consequential, and 3) unintentional consequential, can be studied as strategic communication, however not 4) unintentional inconsequential communication (a final category which is almost difficult to imagine). These cases of strategic communication can also be tied to a more specific idea of strategic moral communication—for example finding strategic moral communication in 1) a law regulating the speed of car drivers, even if these drivers disregard the rules, 2) an organisation's successful CSR campaign, where the recipients accept the moral message, or 3) a person finding a note on the fridge in the morning from a family member that 'There is no milk!' misinterpreting the message as an expression of moral criticism. This also reaffirms the thesis' proposition of a conceptual shift from the predominant emphasis on strategic communication as a practice and professional function, towards a philosophical conceptualisation. This would mean a substantial shift whereby the thesis advocates that we focus more attention on theorising communication as a philosophical construct, accompanying the current emphasis on empirical practices. Lock, Wonneberger, Verhoeven, and Hellesten (2020), for example, illustrate this emphasis in the literature on strategic communication—while pointing to 91 communication theories available to researchers today, they find that the majority of over 4,000 journal articles reviewed did not refer to any of these theories. In this new framework, strategic communication would not necessarily be defined as communication that *is* intentional nor that has a *real* effect, but argues that it reflects a shared attitude among practitioners and scholars as they align in the philosophical perspective of communication as holding the *potential* of influence.

This tension between intent and effect leads us to another key point of departure for the theory's conceptualisation of morals as appearing in-between norms and behaviour. In a sense, we see this in Habermas' discussion about the binding/bonding effect (see p. 25), as people proposing a norm are expected to follow the same—thus forming a social contract. What we see in moral psychology, however, is that norms and behaviour do not necessarily align (see Chapter 2). Rather, this thesis argues that we ought to follow moral psychology's suggestion that studies on morals and ethics need to study both, as well as the relation between the two. For this, the thesis proposes that strategic moral communication can be studied as appearing in relation to the three key foci in moral psychology—by understanding morals in relation to the tribes (see p. 45), tastes (see p. 48), and rationalised intuitions (see p. 51) present in moral discourses. After reviewing these three expressions of the normative ethics reflected in a moral discourse, we can re-apply the same moral standards in an immanent review of these descriptions of the norm, but now instead approach them as expressions of communication. In relation to each of these three foci, the thesis argues that we will find paradoxical relations between a promoted norm and the same observed behaviour in such normative communication, as norms are expressions of ideals (elaborating on this relation between ideals and reality on p. 92). Rather than reviewing this system of values as a normative system rationally regulating actors' assessment of communication, we can see how actors use strategic moral communication to rationalise their moral intuitions by using this normative framework. For the normative framework does not consist of a set of values that leads to a rational conclusion, it only functions as a system that enables rationalised arguments for one's pre-determined morals.

To expand upon how the three theoretical foci derived from moral psychology would translate in this context, the first would draw upon

Habermas' idea of othering, in which groups of 'them' are presented as the immoral while the 'us' are the moral (see p. 28). The communicative construction of such groups, however, is a paradox. As Agamben (1998) observed, actors in a moral community will not only determine their own moral character but inevitably also the immoral character of those who are not a part of the same community. As those who break the moral rules become 'set apart' from society,[55] everyone is judged in relation to the moral system—including those who never accepted the morally binding—and have thus been excluded from the moral system. The second paradox relates to the values emphasised, where moral psychology has shown that while we might have discussed universal values in the past, we should rather see how any moral claims to universality are strategic as they represent a particular community's stance and promote moral values. Finally, the third focus when reviewing strategic moral communication would be tied to the very acknowledgement in contemporary moral psychology that we rationalise our moral intuitions. To study strategic moral communication would mean that we study these moral rationalisations. As they do not express an 'is' but an 'ought,' they will have to assume a rationalising character. This means a theoretical shift from the dialectic ideal—that grounds Habermas' discourse ethics—to Schopenhauer's (1831) suggestion of an *eristic* perspective (alternatively presented as *controversial dialectic*). To rest on communicative behaviour in observable rather than ideal discourses, the thesis argues, enables a better understanding of moral discourses. The thesis agrees with Luhmann (1996) that Habermas' ideal depiction of social discourse is not aligning with the observed reality. Through this shift in perspective, the thesis consequentially would form a new theoretical framework, a new pair of glasses through which we might be able to see new things, as 'we can observe morality fuelled by conflict everywhere if we only adapt our concepts so as to be able to see it' (p. 34).

---

[55] Conceptualised as *homo sacre* by resting on the Roman word for *sacre* as 'set apart'. They are stripped to the point of only holding 'bare life' (*zoe*) as opposed to the by law regulated human rights of the 'qualified life' (*bios*). Similar examples include the Medieval Nordic punishment of 'going into the woods' or *skóggangr* as a punishment for assault and homicide, arson, or horse theft, whereby one was stripped of all legal rights and protections (Peel, Simensen, Larsson, Love, & Djärv, 2020). Habermas similarly makes a distinction between the 'good' and the 'mere' life (1999, p. 108).

Summary

This first part of the thesis has addressed the question, *What would an empirical theory of strategic communication's relation to morals look like?* It started in Chapter 1 by presenting Habermas' theory of discourse ethics as an example of the normative theories, as he himself summarised as follows.

> …assuming that the practice of argumentation is based on the four essential propositions of (a) the inclusive and (b) equal participation of all affected, (c) the truthfulness of their texts, and (d) the structurally guaranteed non-coerciveness of communication; then, in virtue of (a), (b) and (c), all relevant contributions can gain a hearing in practical discourses, but only those reasons can carry weight that accord equal consideration to everyone's interest and values; and in virtue of (c) and (d), only reasons of this kind (and not other motives) can be decisive for achieving agreement on a controversial norm. (Habermas, 2018b, p. 107–108)

The second chapter, however, has emphasised that this theory was based on a theoretical paradigm that has since been disputed. They first describe how we group people into tribes of us and them, contrary to Habermas ideal of equal treatment of all. Secondly, they argue that theories such as those upon which Habermas rest are tied to the moral consensus of his time, it is not a universally accepted morals if one looks worldwide. Finally, they add that we not only use communication to arrive at a rational conclusion, but also and more often use communication as a tool to rationalise our moral intuitions. From this new empirical understanding of the relation between communication and morals, and the strategic function it serves, the thesis' proposes its empirical theory of strategic moral communication. Using frameworks from communication science, it seeks a structure that can review normative ethics, such as Habermas' discourse ethics. The thesis argues that we can look at the way communication is normatively framed as moral or immoral, tying it to normative representations of the elements in the process of communication to make sense of the ideal. This is divided into a three-part interest in the strategic construction of morals, resting on moral psychology's insights into moral dynamics tied to tribes, tastebuds, and intuitions. This will be used as a stance to build a methodological framework that allows for empirical studies of strategic communication's relation to morals.

PART 2. METHODOLOGY

What method could be used to study this new concept of strategic moral communication?

Suggesting a new methodology of moral discourse analysis in Chapter 4 that builds on the previously presented theoretical framework, describing in Chapter 5 how the approach can be applied in a case study as here presented in relation to the thesis' analysis of the fake news discourse.

Chapter 4.

Moral discourse analysis

Just as Habermas' normative theory of discourse ethics has built the theoretical stance in methodological frameworks such as critical discourse analysis, the thesis' theory enables a new methodological proposition. It builds on the previously presented conception of strategic moral communication to form a new methodology of *moral discourse analysis*. This new methodology is presented in the following chapter, unpacking what the theoretical shift means in terms of the philosophical perspective, material, and analysis of interest in studies adopting this new approach.

Habermas' theory of discourse ethics has inspired the development of methodological frameworks. He even develops his own methodology—labelled *rational reconstruction*—as 'the method of interpretation employed by a hearer when the meaning of a speech act is unclear', which he used as a stance when developing his own social theory that coincides with the development of discourse ethics (Gaus, 2019, p. 369). He describes rational reconstruction as a methodology that reconstructs 'the intuitive knowledge of competent judging and speaking subjects', continuing that 'moral theory engages in a task of rational reconstruction when it elicits from everyday moral intuitions the standpoint of the impartial judgment of interpersonal practical conflicts' (Habermas, 1994, p. 25). One could argue that Habermas himself engages in rational reconstruction when forming his theory of discourse ethics as a theoretical framework capturing the intuitions driving actors in democratic societies.[56] His theoretical proposition has also influenced methodologies such as *critical discourse*

---

[56] Habermas himself similarly writes that 'discourse ethics then takes its place among the reconstructive sciences concerned with the rational base of knowing, speaking, and acting' (1999, p. 98).

analysis. This methodology builds on Foucault's proposition for *discourse analysis*. Both Habermas and Foucault focus on communication as creating social realities. They differ, however, as Habermas gives a normative presentation of an ideal process,[xvi] while Foucault would emphasise that this ideal does not reflect how communication is used in real discourses. Foucault's discourse analysis is instead intended as a methodology to study the use of language in society to reveal social structures. Since Foucault's initial methodological proposition, many have developed their own version. This thesis argues, however, that none of the approaches in discourse analysis explicitly focus on moral discourses.[57] The thesis' own methodological proposition departs from the theoretical recognition that morals are expressed as rationalisations, suggesting that we can study such communicative expressions to make sense of moral discourses. It departs from Foucault's discourse analysis, positioning the suggestion as an alternative to critical discourse analysis, taking inspiration from other discourse analytical procedures, and integrates the theoretical framework presented as strategic moral communication to form the thesis' methodological proposition for moral discourse analysis.

The thesis' suggestion for moral discourse analysis as a new methodology is unpacked in the following chapter, starting by presenting the rationale for a new form of discourse analysis that focuses on moral discourses. From this point, it continues to unpack the suggested methodology of moral discourse analysis. It starts by introducing its three underlying principles, continuing with how this new stance effects the design for sampling the data. Finally, it forms an analytical procedure in which the previously presented theoretical framework for how to review communication is linked to a methodology where these theoretical foci are converted into methodological procedures to review moral discourses.

---

[57] Or, alternatively, one could approach all as studying moral discourses, however not explicitly naming them as such.

The moral in discourse analysis

Discourse analysis is a framework applied across the social sciences, in this thesis arguing that there is however little written about its link to the study of moral discourses. The basis of this claim is made after a series of literature reviews, a search for previous discussions on morals and ethics in publications about discourse analysis, spanning from a review of publications in journals and handbooks. However, as described in the text and echoing in the descriptions of moral sociology, some readers may argue that all discourse analysis inevitably researches moral discourses (at least that is this author's position). However, the thesis presents a methodology that is specifically design with the recognition that we can review discourses as moral discourses. It operates from the stance that they can be reviewed in relation to the normative acknowledgement of communicative phenomena. The analysis seeks to unveil moral intuitions by reviewing the elements emphasised and the values and valences used to make such a normative assessment.

First to the thesis' claim that there is a gap in discourse analytical approaches dedicated to the study of moral discourses. A search for journal and review articles in English published until 24 May 2023[58] reveal nearly 15,500 publications discussing 'discourse analysis'. These are foremost published in communication studies (n=2,665), educational research (n=2,606), and linguistics (n=2,460). Zooming in on those also mentioning either ethics or morals somewhere in the paper,[59] a total of 1,119 papers remained, and 1,035 papers mention discourse analysis and ethics or morals in the title, among keywords, or in the abstract. Out of these, the 211 papers published in communication studies was reviewed with the aim to understand how morals and ethics have been framed in papers discussing or applying discourse analysis in the thesis' primary field of interest. In these papers, many use critical discourse analysis (such as Leeuwen (2018)) while others rather engage in a critical discussion about the same methodology (such as Graham (2018)). Another cluster of papers aligned with the thesis identification of the problem captured in the second research question and addressed in Part 2, among the critics of critical discourse analysis finding for example how Jones (2007) too expresses the problem as…

---

[58] Using the platform Web of Science, presented on p. 198.

[59] Using the search string *ethi** or *moral** to include variations of the two words such as *ethics, ethical, morality, moralism* etc.

> …the problem is that the categories, elements and procedures involved in the various forms of 'discourse analysis' based on conventional descriptive linguistics and pragmatics … do not and, indeed, cannot afford any *critical* purchase on communicative processes and actually get in the way of a proper appreciation of how we communicate in real life situations. (Jones, 2007, p. 338)

The problem, as Jones continues, is that it is based on a belief about language in which…

> …we may use to refer to the creative communicative endeavours of particular individuals, but the term for an abstract, self-contained system of forms, meanings, and rules whose existence is the precondition for successful acts of linguistic communication, any such act being the mere realization or the expression of elements or rules in the system. Moreover, as I shall argue, to justify using the theoretical constructs of this kind of 'discourse analysis' as critical tools, CDA practitioners have come up with a very peculiar picture of the workings of contemporary society and the role and power of discourse within it. (Jones, 2007, p. 338)

Continuing the review, 107 of the 211 papers do mention critical discourse analysis specifically in the title, among keywords, or in the abstract. A few do analysis using discourse analysis and merely mention the word moral or ethics in passing, but do not integrate this as the primary object of their analysis (see i.e., Gellen & Lowe, 2021). Among those few papers in communication science not discussing morals or ethics in relation to critical discourse analysis, but from a broader perspective of discourse analysis, there are papers discussing the formation of norms as a part of an identity-building discursive construct—however none claim ethics or morals as their primary object of interest in the analysis. The closest one gets in this pool is a paper by Barton and Eggly (2009) from medical ethics, studying whether offers of participation in clinical trials have a positive, negative, or neutral valence in relation to ethical persuasion. For, as Koren

(2009) notes, the concept rarely appears in discussions about discourse analysis.

> The concept does not even appear in the Dictionnaire d'analyse du discours published in 2002 and edited by Charaudeau and Maingueneau. The only mention appears, indirectly, in Plantin's use of the adjective 'ethical' to qualify 'rectitude' (p. 72) in the entry 'argumentation' found in this dictionary. (Koren, 2009, p. 422)

Herein lies the paradox of morals in discourse analysis. For, on the one hand, it is apparent that morals have been studied in discourse analysis. Among the early examples, one would be Nietzsche's archaeological discursive study of the emergence of morals in his review of the slave morality in modern Christianity (2006), from which Foucault drew some of his methodological inspiration. But as Koren accentuates in the quote above, it seems as if the topic of morals and ethics has not been studied explicitly in discourse analysis. To extend this review, the thesis continued with a review of handbooks on discourse analysis and critical discourse analysis—including the first and second volume of *The Handbook of Discourse Analysis* (Schiffrin, Tannen, & Hamilton, 2005; Tannen, Hamiltin, & Schiffrin, 2015), the first volume of *The Routledge Handbook of Discourse Analysis* (Gee & Handford, 2013), *The Routledge Handbook of Critical Discourse Analysis* (Flowerdew & Richardson, 2018), and *The Cambridge Handbook of Discourse Studies* (Georgakopoulou & Fina, 2020). The chapters discuss related topics such as discourse and religion, or intertextuality in discourse, but are not explicitly dedicated to moral or ethical aspects. Among the few exceptions, one could find for example Hammersley's (2020) *Ethics and the Study of Discourse* and Cook-Gumperz and Kyratzis' (2015) chapter on *Child Discourse* where they discuss the evolution of morality in children and its manifestation in communication, and social order as a case of 'moral order'. Now, one might argue that the ethical and moral dimension is so deeply embedded within the structures of discourse analysis in general, and critical discourse analysis in particular that it does not have to be expressed. However, a search within for example the *The Handbook of Discourse Analysis* (Schiffrin et al., 2005; Tannen et al., 2015) for the use of the word ethics or morals (including variations) finds that the words are used sporadically, for example to refer to the 'moral order' in legal and political systems. However, these contributions do not engage in it as an interesting unit of analysis in itself in discourse analysis or critical discourse analysis. Resting on the above-mentioned literature review, the thesis thus argues that there is a place for a new approach to discourse analysis that explicitly research moral discourses, and which develops a methodology with this purpose in mind.

Philosophical principles

The first question for this new methodology regards what research subject is of interest in moral discourse analysis. Departing from the theory of strategic moral communication, moral discourse analysis is an interest in moral discourse as phenomenological constructs, studied through an inductive approach and immanent design. Each represent a key part forming the new methodology's approach as an empirical alternative to normative methodologies such as critical discourse analysis. As presented in beyond below sections, these three principles grounding the analysis will also influence the selection of material and the analytical procedure.

Moral discourse analysis would study moral discourses, appearing wherever humans engage in moral conversations with oneself or with others. To start unpacking this claim, Habermas describes in reference to Strawson, that a researcher 'must take up a vantage point from which he can perceive moral phenomena *as* moral phenomena' (Habermas, 1999, p. 47). Moral discourse analysis would understand discourses as moral constructs. As previously emphasised, the empirical rather than normative character of moral discourse analysis stipulates that the knowledge sought is not what *is* normatively right or wrong in a definitive sense, but aligns with Habermas' presentation that it is a matter of context-dependent constructions formed through and expressed as moral discourse. It aligns with what the thesis argues is Habermas' phenomenological stance, with his idea that morals are human constructs articulated through communication.[60] While Habermas would argue that morals are concluded through public discourses, the methodology acknowledges that moral discourses arise in other forums too. For example, in Habermas' own discourse ethics. This is but one example of a normative framework that can be studied as a moral discourse. Moral discourse, the thesis argues, arises whenever there is moral communication. It can indeed arise in discourses between people (as intercommunication), but also on an individual level as we think or talk to ourselves (as intracommunciation). This breaks radically from Habermas' sense of morality, which he argues can only emerge through conversations between people, not on an individual level as it 'cannot occur in a strictly monological form, i.e., in the form of a hypothetical process of argumentation occurring in the individual mind' (1999, p. 68). This was an important point for Habermas, as it showed how 'the moral theorists may take part in them [discourses] as one of those concerned, perhaps even as an expert, but he cannot conduct such discourse by *himself alone*' (p. 95, original emphasis). Habermas, however, presents this statement in relation to his criticism against moral philosopher's propositions of theories which would point to the moral, where he argues that the moral is found in the public's conversation and common consensus. This thesis' argument that we can look at moral discourses at an individual level, however, points not to a moral proposition in Habermas' sense as proposing that these individuals' own moral discourse would result in a stance with a normative claim over any context, it only acknowledges that in the study of moral communication we can also look at how individuals deliberate on moral issues as an expression of a moral discourse, just as we can observe moral discourses appearing between people. This thesis would also argue that we can study moral discourses on a structural level to review what Habermas would describe as time's *zeitgeist* (see p. 33), or Foucault's idea of *episteme* (1994a). This is also resting on Keller's (2011) suggestion for a *social knowledge approach to discourse* (SKAD), which combines Foucault' sense of society with Berger and Luckmann's organisations.[61] Aside from emphasising this new entity of

---

[60] As Habermas for example in his early presentation of discourse ethics talk departs from 'the phenomenology of the moral' (1999, p. 45; c.f., 1998, p. 4), and in his

---

continuous presentation of the subjective and intersubjective character of the individual and social lifeworld (see p. 23).

[61] Pairing well with Habermas (2007) describing how communication forms 'as *life-worlds* that are symbolically structured,' he discussed these shared lifeworlds as social system with a direct reference to Berger and Luckmann (2011).

interest, SKAD also draws attention to not only the form but also the formation of discourse. In particular, it becomes interested in 'the appearance of central breaking or turning points in the history of social constructive of subjective or particular orders of practice' (p. 46). SKAD is working from a bottom-up approach, analysing 'how discourses are structured and how they are structuring knowledge domains', seeking insights into the 'complex socio-historical constellations of production, stabilization, structuration, and transformation of knowledge within a variety of social arenas' (p. 48). The approach diverges from the normative perspective in critical discourse analysis, where the thesis aligns with its focus on the recognition that there are multiple and sometimes conflicting morals existing at the same time, and that the analysis should aim 'to support study the manifold discursive of our time in their own right, according to their own logic, in their own language and cultural and socio-political context' (Keller, Hornidge, & Schünemann, 2018). SKAD as it does not set out to 'unmask' language by 'those in power' to 'manipulate the people' (Keller, 2011, p. 48), but in more neutral terms seeks understanding of the discourse studied and how communication influences the conversation—a position argued to better align with Foucault's original presentation of power.

Moving from the methodology's idea of the research subject, the second principle guiding moral discourse analysis would be its *inductive* character of sampling and analysing the material. This is an essential feature in the methodology, setting it apart from critical discourse analysis. For while critical discourse analysis would take Habermas' normative theory as an analytical framework—and thus, this thesis argues, work from a deductive perspective[62]—the inductive approach in moral discourse analysis means that the methodology seeks to rid itself from normative presuppositions. It challenges the egocentric universalism hidden in Habermas' argument that his theory reflects a universal, and natural sense of morality (see p. 182). It challenges Habermas' own depiction of this problem as he describes it in relation to rational reconstruction; that the problem which might arise from a researcher from one context researching another is negligible as 'the general structures of reaching understanding in communicative action entail standards that provide a universal benchmark for rationality that allows for reflective self-control in the process of rational reconstruction' (Habermas (1984, p. 121) quoted in Gaus (2019, p. 371)). This thesis would disagree. Rather, it argues in line with Gaus that while 'communicative reason allows for the possibility that her judgement may be impartial' (Gaus, 2019, p. 372), analysis necessitates that people 'understand and explicate the standards of rationality in modern societies before applying them as standards of critique' (p. 369). More specifically, this thesis would argue that we need to understand our own 'standards of rationality' no matter which these are and what context we have derived them from, to strive for an impartial judgement of moral discourses. For only then can we see how our frameworks are tainted by our own moral bias. It disagrees with Habermas that 'we have to presuppose the validity of a particular moral theory when we make use of it in *other* context' (2018b, p. 110). Instead, the thesis would argue that it is the very validity in such moral theory that is of interest in moral discourse analysis, analysed to make sense of the morals guiding the conversation. Even if such moral conversation is occurring in a scholarly framework. From the recognition that all moral frameworks are indeed *moral*, as they present a moral stance as a matter of analytical point of departure, the thesis' proposition for an empirical rather than normative methodology would require that pre-existing morals are not used as the standard for the inductive analysis.

---

[62] Acknowledging at the same time how well-cited works claim to combine an inductive approach with critical discourse analysis (Vaara, Sorsa, & Pälli, 2010).

The third philosophical stance solves the problem of not using an external framework for the analysis of moral discourses, by re-framing the purpose of *immanent analysis* (also known as *immanent criticism* or *immanent critique*). The approach is often described as introduced by Hegel, developed by Marx, then redefined by the Frankfurt School as a tool to criticise a society based on the society's own norms.[63] This thesis argues that we can also use the same principle to engage in the very process advocated by moral psychologists—accentuating that we should study morals as normative communication and action respectively, as well as understand morals in the interplay between these two perspectives (see p. 56). Simply put, morals can be studied based on what one says (what is expressed in normative discourse) and what one does (the communicative behaviour that this normative discourse reflects). Immanent approaches enable us to move past our own norms as a stance, as they 'force us to engage with an ethics system on its own terms, to listen to the argument, to draw from its strengths and to point out inconsistencies and anomalies in a line of reasoning rather than evaluating it purely in terms of another worldview' (Vorster, 2019, p. 2). For example, from an immanent perspective, one could ask whether Habermas as promoting the inclusion of all voices does so himself when arriving at his own proposition for discourse ethics? Similarly, one could ask whether critical discourse analysis as emphasising the importance of equality do lead to studies where all voices in the discourse are accentuated? Through this immanent approach, the thesis argues that we can move from representing moral systems as normative ethics and applied ethics respectively, to rather build on these two pillars in a single analysis in which we review the metaethical structure emerging between the two. This, the thesis argues, enables us to see both the norms, behaviour, and in-between the two see the practices of strategic moral communication at play.

---

[63] A dominant narrative as presented by i.e., Fornäs (2013). This, however, has also been challenged by i.e., Finlayson (2014).

Material concepts

Building from the above-presented three methodological principles, the second step is to recognise what material would be of interest in moral discourse analysis, and what structural forms are of interest in said material. First and foremost, the methodology aligns with the views of both Habermas and Foucault, as strategic moral communication would be interested in cases of communication as the object of study. More specifically, it is interested in discourses that can be described as normative, thus the label *moral discourses*. From this, moral discourse analysis would review this material in relation to various strata or levels of interests. As below described, this division rests on pre-existing concepts from discourse analysis.

In relation to the selection of material, the first concept of interest would be *discourse*. This is a concept used by both Habermas and Foucault. On a first broad level, the thesis recognises that both Habermas and Foucault use it in the broadest sense to refer to communication, where the thesis is more specifically interested in *moral discourses* as moral communication. While Habermas would tie this to the normative principle (D) (see p. 26), Foucault would rather describe it as either a case of Habermas' state of moral understanding (see p. 24) or as the very discourse of conflict (see p. 36). For Foucault, the interest is not normative, but is merely based on the recognition that communication is an interesting focus for the study of societies. While the two never engaged in a debate, they are presented as incompatible,[xvii] where Foucault describes how 'I am quite interested in his work, although I know he completely disagrees with my views' (2019, p. 298). This disagreement is aligning with the thesis' disagreement with Habermas. In short, that while Habermas had an interest in ideal forms of discourse, both Foucault and moral discourse analysis would be interested in how discourse plays out in real interactions.

> While I, for my part, tend to be a little more in agreement with what he says, I have always had a problem insofar as he gives communicative relations this place which is so important and, above all, a function that I would call 'utopian.' The idea that there could exist a state of communication that would allow games of truth to circulate freely, without any constraints or coercive effects, seems utopian to me. This is precisely a failure to see that power relations are not something that is bad in itself, that we have to break free of. I do not think that a society can exist without power relations, if by that one means the strategies by which individuals try to direct and control the conduct of others. (Foucault, 2019, p. 298)

Just as Foucault, moral discourse analysis would argue that discourse reveals power. Again, not in the way critical discourse analysis describes power as something inherently bad, but in Foucault's neutral depiction of the concept[64] which align with the thesis' understanding of strategic communication (see p. 69). This leads moral discourse analysis to approach the material containing moral discourses as acts of strategic moral communication, as they hold the power of potential influence.

The study of discourse means a review of *text*. Foucault describes using the term discourse and text synonymously and as 'a group of sequences of signs, in so far as they are sentences' (2010, p. 131). For this thesis, however, discourse will be used as a concept for the discursive construction of norms emerging from the text, while the text refers to the material forms of such discourse. While Habermas would limit the idea of communication in discourse ethics to linguistic expressions (see p. 24), the thesis' suggested theoretical framework of strategic moral communication acknowledges a wider variety of communication (see p. 59). This means that text could take the form of either linguistic expressions on paper (such as documents, handbooks, books, policy instructions, laws etc.), or could be vocally expressed (as in the case of interviews, press conferences, or speeches). It also covers other communicative forms such as images (photographs, graphical profiles, colours, logos, etc.), or video material or observations (focusing on communicative expressions such as facial expressions, body language, or dress codes), or other material aspects of communication (such as the communicative layout of a lecture hall, the landscaping of a public park, or the ordering of books at a library). The only assumption one must make is that a) the communication is b) strategic, as it either pursue an intention or achieves an effect, and c) can be argued to constitute normative assumptions. Aligning with the inductive aim, this text should be inductively sampled from a body of text capturing the discourse, or a *corpus* of text. This means searching for a discourse in Foucault's sense of 'a group of texts in so far as they belong to the same discursive formation' (2010, p. 45). The aim in moral discourse analysis is to let the discourse itself define the boundaries of the discourse. This can be done in a variety of ways. To exemplify one way, this thesis' case study (presented in the next Chapter 5) will take inspiration from the mappings of scientific discourses by tracing citation patterns (Chen, 2017; Small, 1999).[65] The approach requires first the identification of *one* text that is part of the discourse. From that, texts are added that are either referenced in or later are referencing this text. This allows the corpus to grow inductively, as the corpus itself directs the sampling. This procedure is then repeated on the newly sampled material until reaching saturation,[66] and so the corpus of the discourse is formed.

---

[64] Indeed too Habermas (2006, p. 418).

---

[65] C.f. snowball sampling, chain sampling, or referral sampling.

[66] A notoriously difficult concept to define, this thesis resting on Guest, Bunce, and Johnson's (2006) combination of definitions to propose that 'data saturation is reached when there is enough information to replicate the study, when the ability to obtain additional new information has been attained, and when further coding is no longer feasible' (p. 1408).

Aside from this corpus, one would also need to rely on material that reveals the *context*. Texts sampled for this purpose help make sense of the discourse, providing both a source for *understanding* and *criticism*. The former would mean that we can use the contextual texts to understand the circumstances that led to the emergence of the discourse. For example, if we are to understand Habermas' normative ethics as discourse ethics, we could look at both individual factors (related to Habermas experiences such as his readings of developmental psychology) as well as social factors (the post-WW II landscape in which the theory emerges). To enable criticism, we can look at material outside the discourse. In this case, the thesis emphasises the later theories on moral reasoning developed in the 40 years that have passed since his reading of the field to illustrate what is not said discourse ethics.

The final question concerns what type of knowledge moral discourse analysis would acquire. Foucault describes discourse analysis as revealing truth (1980, p. 131). This truth, however, was not described as a universal but a context-dependent truth; much like Habermas' depiction of morals (see p. 32). Moral discourse analysis would rather lean towards Habermas' concept, that it is indeed the context's idea of truth, but better phrased as the context's morals. Foucault's idea has, however, been met with criticism—that it suggests truth to become a relative subject, shifting haphazardly in relation to each context. He responds, however, that this is far from the case—as these constructions are quite stable ideas tied to social structures. Replacing the idea of 'truth' with 'moral' in the quote below, the thesis would echo Hook's interpretation of Foucault's concept.

> It is in this way ludicrous to read Foucault as suggesting that truth is 'relative', in the open sense of the term, where all possible truth-conditions are equal, depending merely on context or interpretive perspective. Foucault views truth-conditions as extremely stable and secure, as situated in a highly specific and idiosyncratic matrix of historical and socio-political circumstances, which give rise to, and are part of, the order of discourse. A scepticism of truth here defers not to a 'baseless' relativism, but instead to a carefully delineated set of conditions of possibility under which statements come to be meaningful and true (Hook, 2001, p. 525)

This leads to the interest, in Habermas' and Foucault's frameworks as well as in moral discourse analysis, as an analysis not only linked to the form but just as much in the formation of discourse. Habermas argues that 'the social world to which we are oriented in the normative attitude is *historical* in a different sense from the laws and regularities that constitute the realm of describable events and states of affairs in the objective world' (1994, p. 39). This echoes in Foucault's advocation for a genealogical study of discourses, drawing upon Nietzsche's *The Genealogy of Morals* (2006). Both approach discourse as revealing the social world, that there will be different moral worlds at different times, and that each new moral world can be understood in relation to the moral worlds that proceeded it. For at each time, the moral builds on a discourse where we either build upon or reject what was as the previous moral consensus. To study morals therefore requires an understanding of not only the form of a moral discourse but also its formation, as an understanding of the one generates an understanding of the other. Morals are as much about process as product. Resting on Hook's (2001) close reading of Foucault's inaugural lecture at the Collège de France, presented in *The order of discourse* (1981), this thesis aligns with his descriptive position rather than that of a critical methodologist as he presents that Foucault 'might be better read as a "diagnostician" of culture and society' (Hook, 2001, p. 522).

Procedure

Foucault's discourse analysis is notoriously criticised for providing some general guiding principles while being less explicit with descriptions of how it is executed. In recognition of that criticism, the following section expands on the hands-on procedure for someone who wishes to engage in moral discourse analysis. As unpacked below, it comes down to three important lenses when the discourse is coded. First, it is a matter of how the material can be clustered in relation to various levels. Secondly, at each level we could review the material from different analytical points of interest. And finally, the analysis can be used to present a schematic representation of the norms building the moral discourse. This coding enables the later described immanent analysis of the moral discourse from which one can derive the moral intuition driving the discourse from the reviewed strategic moral communication.

Upon reviewing the material as texts, moral discourse analysis advocates coding the material in a set of three levels: as text, authors, and quotes. Starting with *text* (see p. 85), we can simply view each piece of communication sampled as separate analytical entities. In other words, reviewing a policy document from one organisation as one text, while a transcript from an interview would represent a second text. From this, we can also cluster texts into groups in relation to the *author* (Foucault, 1999). In the analysis, this could mean a cluster of texts that include both the policy document and the transcribed speech, if issued by the same author. A third level of interest is in relation to *quotes* within the text. From this point we can also form clusters—such as finding that several texts talk about the concept 'discourse' but describe and use the term in different ways. This would create the third and last basic analytical level of interest in moral discourse analysis. Adding to that, one can also imagine other levels of interest, such as adding information about the time of publication for each text.[xviii]

Moving from the coding of data to the analysis of such codes, the analysis seeks to reconstruct the discourse' representations of the moral and the immoral. For, as the thesis argues, we cannot talk about what is moral without framing what is immoral (c.f., Habermas' presentation of tolerated, see p. 29). Similarly, Foucault describes how the function of the concept of *rational* is of key importance, accentuating that it simultaneously gives rise to the idea of something being *irrational* (1994b, 2013) Moral discourse analysis rests heavily on the idea that these concepts convey discursive power, asserted in and indeed building moral discourses. As Hook describes, 'the strongest discourses are those that have attempted to ground themselves on the natural, the sincere, the scientific—in short, on the level of the various correlates of the "true" and reasonable' (Hook, 2001, p. 524). What these authors construct is 'discursive practices' which contains the moral into a set system.

> The effect of discursive practices is to make it virtually impossible to think outside of them; to be outside of them is, by definition, to be mad, to be beyond comprehension and therefore reason. Discursive rules are hence strongly linked to the exercise of power: discourse itself is both constituted by, and ensures the reproduction of, the social system, through forms of selection, exclusion and domination (Hook (2001, p. 522) in reference to Young (1982))

The irrational takes the form of something in opposition to and expressing a critique of the rational. This is a key point of analytical departure in moral discourse analysis, echoing Habermas' statement that moral discourses arise in times of contestation and conflict (see p. 36). The dichotomous relationship between the rational and the irrational—that is, between the moral and the immoral, or between the truth and lies, the constructive and harmful, the good and the bad—is a discursive construction. The thesis would argue that these pairs inevitably arise in all moral discourses, as there cannot be an idea of something as moral without being able to point to the opposite immoral. The one cannot exist without the other, for if either is removed the other label becomes senseless. Furthermore, as Foucault emphasises, the rational will always be discursively framed as the author of the texts under review, or alternatively its normative allies, for no one would describe oneself as irrational (1994b, 2013). Moral discourse analysis will always be an analysis of actors seeking to strategically frame themselves as the moral. This moral self is presented in opposition to an immoral other, either a contemporary, historical, or imagined immoral counterpart.

To unpack this strategic construction as a moral system expressed in the moral discourse, the methodology rests on moral psychology's understanding of the moral constructions embedded in the presentation of moral tribes, tastebuds, and intuitions (see Chapter 2). Firstly, the analysis needs to understand *who* is represented as the rational and the irrational, respectively. Who is the author and who is given a voice in the discourse is of central interest for the analysis. For the author gains a 'discursive function that points to the existence of certain groups of discourse (associated with the author in question) and affirms their status within a given society' as discourse gives rise to subjects with privilege positions (Foucault (1999), quoted in Hook (2001, p. 527)). This is articulated in the selection of authors who are given a voice in the discourse (as inductively included through the citation sampling, see p. 86), as well as in relation to how actors are described in the discourse as part of the self or the other.

Secondly, just as moral psychology talks about systems of values forming representations of a context's shared norms—echoing indeed in Habermas' discourse ethics—the aim of the analysis is to code quotes from the texts that are normative representations of the moral discourse. This would refer to the type of 'discursive formations' which Foucault labelled 'formation of concepts' (2010, pp. 31, 56). In moral discourse analysis, this points to an interest in three criteria for coded concepts; those that appear in the corpus as a) discourse-specific, b) and often frequently appearing, that c) are defined or described in the discourse.[67] By balancing the two former criteria,[68] moral discourse analysis would argue that we can identify the phenomena of interest in the discourse, as these are strategically defined in the text. For example, the review of Habermas' discourse ethics would point to concepts such as 'lifeworld' and 'understanding' as two formative concepts for his normative proposition. Just as Habermas (1970) describes, the methodology agrees that such concepts contain an 'intersubjectivity of mutual understanding'—thus reflecting our common point of view, expressed as 'the particular structure of potential speech' that constitutes 'the basic linguistic framework, which also determines the scope and structure of corresponding world views' (p. 373). In short, communication reflects worlds as the phenomena that are acknowledged and how they are normatively assessed (see p. 87). Similarly, Habermas (1988) argues that changes in language represent changes in the world, 'a revised picture of the past in light of an anticipated future' where changes in one's environment 'make necessary a changed application or extension of traditional language' (p. 175). These changes, or 'turning points in the socialization process are indicated by a shift in terminologies and by the effort to replace interpretations that have lost their credibility with more appropriate ones' (p. 175). Possibly inspired by Wittgenstein's (1974) infamous quote that 'the limits of my language mean the limits of my world' (his 5.6, p. 74), this grounds Habermas' and this methodology's assumption that the study of communication at a particular time would provide a gateway for studying the community's shared phenomenological experience.[69]

---

[67] The search for relevant concepts that would represent the 'collective symbolic orders' to use Keller's words, expressed as the 'repetition and stabilization of the same texts in a singular text' (2011, p. 52). More specifically, the interest would be in what SKAD describes as *model practices* emerging from discourse as 'exemplary patterns (or templates) of actions which are constituted in discourses for their addressees' (Keller, 2011, p. 48).

[68] That is, looking primarily at discourse-specific and frequently appearing concepts, however not exclusively. This presents a challenge for the analyst who need to balance the two criteria against one another to identify a set of concepts that reflect the discourse's normative language.

From these concepts, the analysis would seek understanding of the values used to frame the phenomena as moral or immoral in the definition of concepts. To that end, the thesis suggests drawing on Hartman's (1967) system, which in turn rests on axiology. Axiology is a philosophical branch, engaging in the study of values. Values can either cover ethical or aesthetic values, this thesis utilising it for the study of the former. Hartman is among the scholars who have used the axiology as a stance to develop a theory on how to map value systems, measuring values by approaching them as scientific values in an attempt 'to introduce orderly thinking into moral subjects' (p. 3). Each value can be seen as a criterion for concepts, where an object can be evaluated in relation to how well it provides an example of such concept. If all criteria are met, this becomes a matter of *formal concepts* as *finite sets* in Hartman's terminology, while other phenomena might not fill all values of any given concept and thus provide *infinite sets*. Upon reviewing concepts' theoretical definitions, we would find values determining whether it is a case of moral or immoral communication. However, these will always be a matter of finite sets, whereas real cases may only partially qualify. In short, the theoretical phenomena may never appear in real life.[70] As Hartman (1967) describes, they are 'constructions of the mind, such as geometric circles' (p. 42; c.f., Habermas 1994, 54–55). This aligns for example with Weber's (1963) sociological idea of *ideal types*, described as strategic constructs whereby actors form the idea of a phenomena as an ideal type by accentuating some values and downplaying others appearing in the same phenomenological experience.

> An ideal type is formed by the one-sided *accentuation* of one or more points of view and by the synthesis of a great many diffuse, discrete, more or less present and occasionally absent concrete individual phenomena, which are arranged

---

[69] C.f. Habermas' (1970) description that 'the classification of semantic fields is predetermined by the question of how far the net of intersubjectivity must be spread to stabilize the identity of the individuals, as well as that of the social group in a given culture or subculture at a given time. The structural differences between the animistic, the mythical, the religious, the philosophical, and the scientistic views of life lie clearly in this dimension' (p. 373).

[70] Similarly discussed by Habermas in relation to ideal and real societies (see p. 40).

> according to those one-sidedly emphasized viewpoints into a unified *analytical* construct (Gedankenbild). In its conceptual purity, this mental construct (Gedankenbild) cannot be found empirically anywhere in reality (Weber, 1963, p. 398, original emphasis)

The value of these ideal types appear as 'an analytical construct that serves the investigator as a measuring rod to ascertain similarities as well as deviations in concrete cases' (Weber, 1963, p. 398) Normative ethics would capture ideal types of moral and immoral behaviour. By recognising their ideal character, we accept that they can only be used as tools to assess reality—they are, however, not mirrors of real events. Moral discourse analysis would suggest tying this aim of constructing the ideal types analytically from a moral discourse by reviewing concepts in search of values. By reviewing definitions of communication, we find values attached to the elements in the process of communication (see p. 61). From this schematic presentation of the norms present in the definitions of moral and immoral communicative phenomena, we can then immanently assess whether the actors proposing such definitions engage in moral or immoral behaviour. For example, in the review of Habermas' discourse ethics finding that he is critical towards strategic communication (see p. 41) while this thesis would argue that he engages in such practice himself. By continuously moving between the communicative ideal and practice, the methodology argues that we capture the moral discourse—both in the norms conveyed, the practices observed, but in particular in the relation between those two.

Chapter 5.

Case study

Moving from the previous presentation of the principles, material, and procedure for moral discourse analysis, this chapter continues to show how these can be applied in the study of a moral discourse. As the thesis argues for a moral zeitgeist spanning from Habermas' discourse ethics to its empirical expansion in the fake news discourse, the fake news discourse is used as the primary moral discourse under review in the analysis, however, is continuously presented in relation to discourse ethics.

Beyond the use of Habermas' norms in scholarly discourses, they are also reflected in empirical moral discourses, in particular echoing in societies aligning with the ideal of democracies. Whether this is due to the successful influence of discourse ethics, or that the normative framework merely reflects these actors' own normative ideal, remains to be concluded. Since the presentation of discourse ethics in the 1980s and 1990s, however, the societal structures in which public discourses can manifest have changed drastically. With the introduction of the internet immediately after his presentation of discourse ethics, both scholarly and public discourses argued that this new infrastructure would enable his ideal to become a reality. Reaching the peak period of enthusiasm at the beginning of the 2010s, the discourse then took a critical turn. The criticism arose from two sides. On the one hand towards politicians and journalists framed as part of 'the establishment'. On the other hand, it was directed in opposite route, as this was presented as actions which eroded the trust in the very institutions that would uphold the democratic ideal. Both sides argued that the other was presenting lifeworlds which their critics did not recognise as truthful. By the mid2010s, a new concept was used by both sides to direct criticism—arguing that the other was presenting *fake news*. This thesis approaches the discourse as a moral discourse in Habermas' sense of the word as a discourse of criticism and contestation (see p. 36). From it, several new words arose alongside the concept of fake news as a continuous critical debate of other communicative phenomena such as *echo chambers, alternative facts,* and *post-truth*. Each of these new words, the thesis argues, encapsulates important facets of the moral discourse—each formed to expand the collective's phenomenological reality by introducing new concepts. The concepts are strategically framed in a normative sense by adding definitions containing either moral praise or immoral criticism against the cases that can be labelled as these communicative phenomena. Below section explains how this analysis was conducted, presenting the results of such analysis in Part 3.

The application of the methodology of moral discourse analysis is presented below in relation to this case-study of the fake news discourse, unpacked in relation to the procedures for sampling, coding, and analysis. This is not only intended to provide transparency, but also as a hands-on guidance on how others with an interest in engaging in moral discourse analysis of this or other moral discourses can go about their research design.

Material

The material under review is a moral discourse, from here on referred to as the *fake news discourse*. This, however, is merely one of several discourses appearing with the emergence of the concept. One could have used this label for other discourses about fake news, such as those appearing on online forums, on social media, in academic articles and journalistic articles. The reason for selecting this discourse is that the thesis' argues that these actors are clear protectors of the ideal expressed in Habermas' discourse ethics, and as such constitute a suitable discourse for a review of the empirical expansion of his theory. The sampled material is argued to mirror some of the key actors' view in the democratic landscape. It is a study of various communicative expressions produced by organisations that seeks to shape the discourse around fake news, and consequentially, the thesis argues, the phenomenological boundaries for moral and immoral conduct in the new communication landscape. Some organisations are pressured to partake, seeking methods to identify the immoral and counter it through moral public communication. Others have entered the conversation voluntarily and purposefully, as it will shape the societal norms, professional codes of conduct and legislative restrictions influencing the leeway for their own future public communication. It is thus approached as not only a moral discourse, but a case of meta-strategic communication, as these organisations' normative framing of different forms of communication will create the normative boundaries for both others and their own communicative influence moving forward.

The sampling followed the inductive principle (see p. 82), inspired by a citation sampling strategy (see p. 86). The process started on 13 September 2021 and continued until reaching saturation (see p. 86), with a corpus containing 1,100 texts. Five sampling criteria guided this process, including only texts 1) issued by organisations,[71] 2) before the end of 2020, 3) focused on the discourse around fake news, 4) expressed as readable text (to enable a software-powered analysis), and that 5) were linked to the discourse by being mentioned or mentioning other texts in the corpus. A selection of texts is referenced in the analysis, ca 100 texts selected to represent the variety of actors and perspectives emerging in this corpus. For example, among the platforms, Facebook is mentioned and appear as the author far more often than LinkedIn—which has influenced the selection of texts referenced to represent the platform perspective. As the focus in the analysis is on the organisations publishing these texts, the labelled author of the texts is always presented as the organisation. That is, even if a report was commissioned by an organisation but authored by named individuals, in the reference the organisation is listed as the author. The aim with such representations of the texts in the analysis is to make it clearer to the reader which organisations' perspective is presented. Similarly, as the analysis is not based on other criteria such as the form of each text, this is not accentuated in the references to texts in the analysis.

---

[71] This has neither included news articles from journalists nor scholarly articles. However, publications written by journalists and scholars in a report form have been included. For example, journalistic reports such as those published by institutes such as the Reuters Institute, and reports written by scholars with what has been evaluated as policy-intent, such as when issuing 'policy reports' (RSIS, 2018) and 'papers' providing 'commentaries' on public policy (Ash Center for Democratic Governance and Innovation & Harvard Kennedy School, 2018).

The first text was found through an online search for documents published on the topic of fake news.[72] The first result in the list was used as the first sampled text—*Handbook for journalism education and training: Journalism, 'fake news' and disinformation* (UNESCO, 2018)—continuing to add texts by adding other texts referenced in this publication.[xix] Later published texts were identified by returning to the online search engines. This second approach allowed the corpus to grow forwards in time of publication, also identifying texts published after 2020, which were however only used to review the references in searching for texts published prior to 2021.[xx] This helped avoid a situation where the texts published in the final years were underrepresented in the corpus. Clusters of texts were also identified by reviewing organisation's set of publications. This was particularly important when sampling material from the platforms, where the texts often cited their webpages on which citations appeared as hyperlinks to other pages on the organisation's webpage, creating a messy network of publications. In those cases, key webpages were identified that helped the sample to cover publications that represented the organisation's perspective. For example, a frequently referenced set of pages were the webpage where Facebook published their press releases (or as they call it, Facebook's Newsroom). These press releases appeared as separate texts, organised by the organisation in relation to the topics discussed. To identify texts published by Facebook on their webpage that could represent their perspective on the discourse, these lists of press releases were reviewed, in particular those topics referring directly to the fake news discourse, such as 'Combatting misinformation' and 'Election integrity'. This helped the process of identifying key texts such as those introducing and defining concepts. While initially considering a computer-supported sampling process, the necessary qualitative nature of the sampling became clear throughout the process. For example, if one were to include all referenced texts in each text, the corpus would expand potentially indefinitely, and definitely move away from the initial core discourse on fake news. Thus, the sampling process was based on qualitative assessments of each text. Each text was reviewed to judge whether it still addressed the core discourse. For example, texts discussing human rights would be relevant if discussing freedom of opinion and expression, but not others focusing on the human right to work under just conditions or freedom from slavery. Continuously, the description of including a text or not was a subject for evaluation in relation to the sampling criteria, as well as in relation to the goal for a saturated corpus. For example, as Facebook grew to rest on the concept of 'coordinated inauthentic behaviour', it became important to include texts that defined the concept and discuss its meaning. However, it was judged not necessary to include all monthly reports of the platform's suspension of accounts based on these grounds, as that would not add relevant new insights to the corpus.

---

[72] Using Google's search engine, a list of pdfs (limited through a search for 'filetype:pdf') containing the lexeme 'fake news'.

As to the sampling of texts framing the context (see p. 87), the need for such material was identified through the analysis. For example, as presented in Chapter 9, the discourse often came to assign the label of 'conspiracy theories' to describe communicative phenomena. However, these definitions were of a strategic nature, framing it as immoral rather than providing a descriptive definition. To show this, alternative definitions were sought from the academic literature. Similarly, the fake news discourse' discussion about the US elections in 2016 was found to often rest on a claim that the elections were influenced by Russian foreign interference, while a few texts also accentuated that such a claim was not supported by studies. To that end, additional texts were sampled from the academic literature which showed the perspective not accentuated in the discourse.

Coding

Due to the size of the corpus—with 1,107 texts sampled to represent the fake news discourse—a *computer-assisted qualitative data analysis software* (CAQDAS) was used to organise and code the material.[73] Following the coding procedure (see presentation starting on p. 84), this was translated into the software's system for coding; classifications were added for each text's author and time of publication, passages containing definitions of concepts were added as a code for each concept.

---

[73] Using the CAQDAS software NVivo.

First, all texts were given an identification number (0001, 0002, 0003, …). After that, information was added about the author and year for each text.[74] This revealed a corpus of texts published from 2006 to 2020. Authors appeared in three clusters; either platforms (i.e., Facebook, Twitter, YouTube, and TikTok), institutes (i.e., policy institutes, think tanks, and centres), and national or international government organisations (i.e., from government authorities in the UK, USA, Sweden, France, Denmark, etc. or international government organisations such as UN, EU, and NATO). As the texts were not sampled with quantitative intent, there is no meaning behind numbers such as how many publications were sampled from each type of organisation, nor in the number of texts published each year. The first number would be particularly misleading as while most texts were sampled from platforms, these were short entries compared to other texts. Government reports, for example, while few in numbers could surpass 200 pages per text. Neither have these quantitative measures had an influence on the qualitative analysis. However, the categories created through this coding helped reveal patterns of which concepts were used and how they were defined by different authors at different times. As a first step, the analysis started by reviewing the texts in chronological order to get a sense of how the discourse developed over time. However, as the sample was limited to publications published prior to 2021 this means that the corpus does not cover for example the end of the COVID-19 pandemic, the full-scale invasion of Ukraine, the formation of META and the new ownership of Twitter (now named 'X'), and the Capitol riots in the US, to name a few important events which might have changed this discourse since. A list of all texts with the author organisation, year of publication, and title was reviewed and coded inductively to spot overarching patterns in themes developing over time. After that, the texts were reviewed for passages in which the texts themselves described the development of the discourse, finding that many linked the fake news discourse to the emergence of the internet in general and social media in particular. It also showed that many returned to the US election of 2016 as a milestone for the discourse. This review has foremost been used in the presentation of the context from which the discourse emerges, presented in Chapter 6.

---

[74] In NVivo meaning adding a *classification sheet* with two *attributes*.

The next step of the coding aimed at identifying quotes in the texts where discourse-specific concepts were explicitly discussed (see p. 91). Following the aim of seeking frequently appearing and discourse-specific concepts, the procedure started with a review of the 100 most frequent words (list presented in Table 2 on p. 205).[75] This helped gain an instant view of the topics under discussion. The top three most common words were 'media' (32k), 'news' (31k) and 'information' (22k). This list was reviewed to identify discourse-specific concepts, finding for example concepts such as 'disinformation' (14k) and 'propaganda' (4k). After that, the program was used to search for the passages in which the most frequent words were used, with the aim of finding concepts appearing as lexemes.[76] For example, the most frequent word 'media' led to the identification of concepts such as 'manipulated media' and 'state-controlled media'. Through this procedure, other concepts were found as they were often introduced in bulk in lines such as 'rumour, conspiracy, disinformation, and manipulated media' (Kofi Annan Commission on Elections and Democracy in the Digital Age, 2020, p. 76) or as 'malicious manipulated media, computational propaganda and disinformation campaigns' (University of Washington Center for an Informed Public, 2020, p. 1) through which additional—less frequent however discourse-specific concepts—were identified such as 'rumour' and 'conspiracy'. Saturation was judged to have been reached at around 100 concepts (list presented in Table 3, see p. 207), at which point new concepts had already been listed.

In the presentation of the fake news discourse, the analysis uses this list of concepts to make sense of both the common and sub-discourse specific discussions. A few concepts became central to the thesis' understanding of the moral discourse. The first concept was 'fake news'. It was found to be reflecting the very conceptual confusion in the debate, but also the character of its evolution from concerns about online media outlets designed to imitate legitimate media. Not least, it marked a milestone for the debate as it was used by Donald J. Trump in the US election race for the 2016 presidential election. The discourse describes this even as a defining moment for the emergence of the discourse itself. Another set of concepts which became key in the analysis were ‘echo chambers’, ‘alternative facts’, and ‘post-truth’. These were found to correspond with Habermas’ moral concepts—as the immoral anti-phenomena a) to his principles of universalisation (U) and discourse (D), b) to his presentation of validity claims as universal, and c) to his notion that morals are the product of rational communication. Together, these four concepts have organised the analysis, working as the focus point for the four chapters. To that, each chapter’ analysis is, however, also grounded in the presentation of the moral discourse as tied to the other list of concepts identified through this review.

---

[75] Using NVivo's function of *frequency query* to search for the most frequent words.

[76] Using NVivo's function of *word search* to review the use of words in the text.

Analysis

The analysis followed the procedure outlined in moral discourse analysis, focusing on describing the nature of the moral discourse by looking at how it constructs its normative presentation of actors, values, and processes for moral communication (presentation starting on p. 102). In each of these three foci, the analysis started by reviewing the normative ideals presented in the moral discourse, after which the same standards were used as a framework for the immanent review (see p. 83). This led to the identification of several immanent paradoxes between the advocated ideal and the observed behaviour. In the analysis, this is approached as unveiling the nature of the moral intuitions supported by the strategic moral communication observed.

The first focus is on how the moral discourse constructs its idea of moral and immoral actors. From the very start, it was clear that the authors of the texts in the sampled discourse were part of the moral self, sharing the common goal of protecting the integrity of democracies and promoting its values (see p. 192). Adding to this observation, the analysis continued by reviewing the most frequently used words in search for other actors mentioned but not necessarily appearing as authors in the sample. From the list, there was a clear focus on one actor, as the list included words such as ‘Russia’, ‘Russian’, or to its leader ‘Putin’ and in references to ‘Kremlin’ (see p. 205). From the review of the use of such words, it became clear that this represented a key cluster in the moral discourse's conceptualisation of the other—never appearing as authors but continuously framed as an immoral actor in the discourse. Through the review of the context in which this actor appeared in the texts, more actors were identified. For example, journalists, the public, and public institutions were presented as part of the moral self, while conspiracy theorists, populistic politicians, and actors engaging in economic deception would be part of the immoral other. Throughout the corpus, clear patterns emerged with clusters of actors either presented as heroes and victims (the self) or villains (the other). Through this mapping, the analysis gained an understanding of the underlying power-structures at play—who is included and excluded, branded as part of the rational self or the irrational other (as discussed by Foucault, see p. 89). This presentation, however, led to an interesting paradox, as the same actors constructing these representations were advocating for an inclusive discourse guided by Habermas' principles (U) and (D). As Chapter 7 emphasises, the thesis argues that this paradox reveals how strategic moral communication will inevitably be expressed as representations of the moral self and the immoral other. That is, even if such construction goes against Habermas' normative ideal.

The second focus concerns how the discourse defines moral and immoral phenomena respectively. Drawing upon the definitions of coded concepts, the analysis reviewed which element in the process of communication was emphasised, and what moral or immoral value was attributed to this element. Together, this formed a system of concepts, each defined as either an immoral or moral form of communication, and each attributed with a set of values. From these cases, the thesis could form an idea of which values defined the discourse's idea of what is moral and immoral. The ideal types emerging, however, as Weber as well as Hartman would emphasise, are not present in reality, but create the 'analytical constructs' that Weber describes can help us make abstract sense (see p. 92). The aim of this review was to form a simple representation of the underlying values used in the debate (as now represented in Figure 1, see p. 151). The normative system would capture how the discourse depict moral and immoral ideal types (see p. 92), as moral and immoral concepts share values. For example, the concept of 'fake profiles' would be a case of immoral communication, and signal that the sender is deceptively presented. The same can be said for the concept of ‘imitators’ as well as ‘sock-puppets’. This enabled the analysis to identify shared moral and immoral values appearing in several concepts, while not necessarily sharing *all* values with each other. For example, ‘disinformation’ and ‘misinformation’ share the value of false messages, however, disinformation is spread with malicious intent while misinformation is not. And while the system would point to a constructed matrix of ideal types that cannot be observed in reality it is approached in this thesis’ analysis to reflect the normative ethics emphasised in the fake news discourse. It is a representation of the normative discourse in this particular context—finding that the values emphasised in the fake news discourse mirror in large part Habermas norms. Likewise, these values are argued in the thesis to reflect the WEIRD values (see p. 50), and is thus representing a particular rather than a universal moral system.

The final of the three foci in the analysis reviews the moral discourse’s expression of strategic moral communication by looking at its construction of the rational and the irrational. More specifically, it shows how the discourse uses the above-presented normative tools to rationalise one’s moral intuitions. After having reviewed the concepts to form an axiological representation of the ideal types, the analysis moves to its immanent review. As Weber emphasises, upon evaluating real cases we emphasise some traits to make them fit into the one or the other ideal type. This is the process investigated in this immanent step, reviewing cases discussed in the discourse as the one and the other type. It shows how actors—through strategic moral communication—come to the decision that one can label cases of communication as moral or immoral. What the thesis finds is that this analysis is drawing, in particular, on the idea of the self and other, where actions of the self is presented as moral while the other’s actions are immoral. The ideal matrix presents a system that actors can use as a strategic tool to rationalise one’s moral intuitions. This requires a shift from a review of the *form* of the moral discourse’s normative proposition towards a review of the *formation* of such rational structures. Simply, describing how norms were rationalised. Resting on moral psychology’s emphasis that we not only arrive at rational conclusions but rationalise our intuitions. This represents the final step in the analysis that builds on the previous two forms of strategic moral communication. Pointing to the criticism against the discourse as characterised by a post-truth reasoning in which actors rationalise rather than engage in rational deliberations, the thesis however shows an example of how the same criticising actors engage in the same behaviour. Through the immanent review, the chapter shows how the frequently used example of the US election in 2016 led to a discussion about foreign interference—a discussion which in hindsight is here argued to at least in part be a rationalisation of the unexpected results of the election. Because, upon closer review of the discourse framing the election as influenced by Russian interference, the statements driving this conclusion in the discourse is not based on the same valid claims as their own norms advocate.

Finally, this leads to the thesis' final discussion in Chapter 10, in which the thesis moves from these empirical observations to an immanent review of the theoretical proposition in discourse ethics and the methodological proposition in critical discourse analysis. It finds similar contra-productive immanent structures highlighted in this final chapter. By continuously moving between these scholarly and empirical moral discourses, the thesis argues proof for the presence of its own proposition for strategic moral communication.

Summary

This second part of the thesis has addressed the question *What method could be used to study this new concept of strategic moral communication?* It did so by introducing its proposition for moral discourse analysis and how it has been applied in the thesis' empirical analysis of the fake news discourse. The first chapter showed how the theoretical proposition for strategic moral communication could be studied through a form of discourse analysis that reviews discourses as moral discourses. The chapter continued by translating the theoretical framework into three methodological principles guiding the procedure—assuming a phenomenological understanding of moral discourses, applying an inductive sampling, and engaging in an immanent analysis of the material. Finally, the chapter presented possible material of interest for this analysis, how to sample text, how to code, and analyse the material. This review was resting on the three foci in moral psychology's presentation of three common forms of expressions of moral intuitions through the formation of tribes, tastebuds, and intuitions. The second chapter moved from this methodological presentation to describe how the framework has been applied in the thesis' analysis of the fake news discourse. The chapter further illustrated how the methodology could be conducted through a step-by-step presentation of the hands-on procedure. Following the methodological presentation, the chapter began by presenting the design for sampling the texts to represent the fake news discourse, continuing with a presentation of how the material was coded, and concluding with how such codes were reviewed in relation to the three analytical foci. The chapter showed how the methodology was applied in the analysis of the fake news discourse.

PART 3. ANALYSIS

How does strategic moral communication manifest in the moral discourse on fake news?

Applying the methodological framework presented in Part 2 to gain knowledge of morals as theorised in Part 1, the following chapters explore how this manifest in the fake news discourse. It starts by presenting the volatile character of the fake news concept in Chapter 6 as a reflection of the very development of the discourse, continuing in Chapter 7 to unpack the normative discourse by reviewing the construction of actors' echo chambers. In Chapter 8 the communicative phenomena are evaluated on the basis of alternative facts, and finally in Chapter 9 the discourse's sense of rationality and criticism against post-truth is presented.

Chapter 6.

Fake news

This chapter introduces the fake news discourse and aims to provide a contextual understanding of the time and place in which it emerged. It finds that the discourse is about communication, more specifically framing the normative character of communication in the public sphere. The immoral problem is constructed by the authors in the fake news debate as a threat towards their own ideal, which aligns with Habermas' discourse ethics. From this contextual understanding, the latter three analytical chapters will then continue by unpacking this normative character of the discourse as practices of strategic moral communication.

The fake news discourse rests its normative point of departure on the same stance as Habermas' discourse ethics—such as emphasising the importance of the public sphere, dialogue, human rights, and democratic institutions. While Habermas presented his theory some 40 years ago, the fake news discourse emerged in the mid2010s. It is described by actors in the discourse as developing from the initial idealisation of new communication technology provided by the internet. In time, it moved to a concerned debate on how it also created new immoral communicative phenomena. The material often returns to the same events as catalysts for the turn of the debate. While the sample represents voices from around the world, they all return to the same cases—a central one being the US election in November 2016. While the concept of fake news was used earlier than that, it was the presidential candidate Donald J Trump's use of the term that ignited the concerned debate. This eventually led to the rise of a cluster of new concepts entering our vocabularies, each reflecting the discourse's concern for different anti-phenomena to Habermas' communicative norms. While the latter three chapters will expand on these concepts and the paradoxes arising from the immanent analysis, this chapter begins by introducing the fake news discourse. The concept is continuously criticised for representing an imprecise term, and that it is used as a political tool in different actors' criticism of the other's behaviour. This, the thesis argues, captures the general character of the discourse, not only as it shows how the discourse emerges out of actors' criticism towards behaviour appearing on the new platforms for communication, but also as it shows how the moral discourse expands from a mere descriptive discourse on communicative phenomena to rather a politicised debate constructing the time's idea of moral and immoral behaviour.

The following sections expand upon this presentation, showing how the discourse emerges from normative concepts found in Habermas' ideal and an initial enthusiasm towards new communication platforms. The chapter continues to show how the discourse developed into the critical discourse around fake news, finally showing how the very concept of fake news can help us understand the general air characterising the discourse. In short, this chapter provides an understanding of the contexts in which these morals arise.

The prophesy of an online public sphere

The emergence of the fake news discourse is described in the discourse itself as tied to new information technology. As Habermas' discourse ethics was introduced in the 1980s and continuously developed in the 1990s, we can see in hindsight how this was developed at a time pending radical changes to the landscape of public discourse. Around the same time, the internet grew into a publicly accessible platform for communicative exchanges. The break from a non-digital to a digital sphere is often taken as milestone for many of the texts the fake news debate (i.e., Canadian House of Commons, 2018, p. 29), with descriptions of the 'intermediaries for the Internet' (European Commission, 2016, p. 293; Organization for Security and Co-operation in Europe, 2011, p. 1) as providing a wide range of services 'such as the access and interconnection to the Internet, transmission, processing and routing of Internet traffic, hosting and providing access to material posted by others, searching, referencing or finding materials on the Internet, enabling financial transactions and facilitating social networking'

(Organization for Security and Co-operation in Europe, 2011, p. 1). Among such 'intermediaries,' *digital platforms* first allowed for mass-transmission of data. On the one hand, pre-existing traditional outlets took a new digital form; newspapers created online versions appearing on websites, radio programs were published as digital podcasts, and watching video material would not only be possible through a TV but also through online broadcasting. More importantly, this changed the access to mass media, as anyone now held the potential to become a mass-medium reaching as large an audience as any traditional outlets.

In time, these new platforms have taken two general forms, either as digital or social media. Firstly, digital media emerged as online mediums for communication to a mass audience. Now, anyone could create their own webpage through which they had the potential to communicate to others with internet access. New online search engines such as Google and Mozilla enabled the public to find those pages that mentioned a certain topic. At this stage, the lines between traditional media and the new digital media started to blur. In 2014 this included initiatives such as that by Mozilla where traditional journalists and citizens could engage in a common platform where the audience could become co-contributors to the journalists' reports.

> Mozilla is proud to announce a new partnership with The New York Times and The Washington Post, funded by $3.89 million from the John S. and James L. Knight Foundation, to build a new, open-source content and commenting platform. The community platform will allow news organizations to connect with audiences beyond the comments section, deepening opportunities for engagement. Through the platform, readers will be able to submit pictures, links and other media; track discussions, and manage their contributions and online identities. Publishers will then be able to collect and use this content for other forms of storytelling and spark ongoing discussions by providing readers with targeted content and notifications. (Mozilla, 2014)

At this point, we start to see how platforms moved from offering the transmission of communication to a co-constructed conversation (see p. 60). New *social platforms* appeared in the online sphere in which the content appearing was not the transmission of one publisher's thoughts to its readers, but the structure of the platform allowed for all entering as *users* to co-create its content. One early example would be the above-presented initiative from Mozilla in relation to journalism. These mediums could take different forms, for example as *forums* where people engaged in discussion threads, taking turns to respond to the previous post on platforms such as Reddit. Aside from these forums, the social discourse exploded with the launch of international platforms such as Facebook and Twitter.[77] Here, each user held their own account, publishing and/or commenting on content stretching from public to private topics—engaging in political discussions of local, national, or international relevance, or simply wishing one's friends a happy birthday. This new infrastructure has a different constitution compared to previous mediums. While all digital and social media were built on computer coding, the content shown to users on some of these mediums were individually tailored, regulated by algorithms. When entering a digital search-engine such as Google or a social media platform such as Facebook, the things appearing on the screen would be individually tailored to present collection of content based on the user's previous behaviour and based on the behaviour of other users. Together, the data tied to users together with the power of algorithms, meant that platforms could show a personalised feed of content mixed with ads from organisations who paid to reach a certain type of user. This could be tied to demographic characteristics such as age, gender, and where the user lived, but also psychographic characteristics based on the users' interests (RSIS, 2018, p. 10). It is this business model, offering advertisers smarter ways to reach their target audiences, that has allowed the platforms to remain free for the public to access, paying with their attention on the screen (Facebook, 2019c).

Many scholars enthusiastically described how these new social platforms provided the empirical infrastructure that could make Habermas' discourse ethics a reality (see i.e., Rheingold, 2000). It was described as 'a new technological optimism for democratic renewal based upon the open and collaborative networking characteristics of social media' by producing 'virtual public spheres' as 'online Agoras and Habermasian forums' (Loader & Mercea, 2011, p. 757). In the fake news discourse, Habermas' idea of the public sphere is described in relation to descriptions of 'the beginning of a new age of the public sphere', through 'the emergence of social media'

---

[77] While there are other earlier digital and social media, Facebook and Twitter are the ones frequently mentioned as the first in the fake news discourse.

(European Parliament, 2019, p. 76). These reports reminisce that 'early internet optimists envisaged a direct democratic participation and a global democracy enabled by the online network' (p. 16). Back then, discussions about the impact of the platforms on the public conversation was as enablers for a true *public* debate, as enabling institutional transparency and being non-discriminatory in who is given a voice. In hindsight, the discourse describes the *Arab Spring* of 2011 as a candidate for 'the period of peak enthusiasm' in platforms' as liberators serving 'democracy and liberal rights' (Centre for the Study of Democratic Institutions, 2018, p. 5). Facebook and Twitter are described by other organisations as being portrayed at this time as liberators emancipating the public from their state's authoritarian rule (i.e., in French CAPS & IRSAM, 2018, p. 39). As a result, we see publications such as human rights organisations' *Joint Declaration on Freedom of Expression and the Internet* in 2011, commenting the following:

> …the transformative nature of the Internet in terms of giving voice to billions of people around the world, of significantly enhancing their ability to access information and of enhancing pluralism and reporting, … the power of Internet to promote the realisation of other rights and public participation (Organization for Security and Co-operation in Europe, 2011, p. 1)

During this enthusiastic period, the fake news discourse highlight how the new media landscape enabled the public to take on a new journalistic role as platforms had 'given rise to crowdsourcing content, and even reporting tasks like verification can now be outsourced to the audience' (UNESCO, 2018, p. 97). Others continue that 'online communication allows views that used to be suppressed in traditional media to emerge and be heard' (European Parliament, 2019, p. 16). YouTube, as one of the social media platforms focused on video-material, similarly describes how 'from its' inception, YouTube has been a platform for free expression', used by activists to publish videos 'illuminating what occurs when governments and individuals in power abuse their positions' in covering criticism in relation to public topics of concern spanning from 'police brutality, discrimination, elder abuse, gender-based violence, socio-economic justice, access to basic resources, and bullying' (YouTube, 2012). The discourse again returns to Habermas' theoretical ideal, describing how the citizen's communication 'used to be direct and unmediated, exercised in public spaces like coffee houses and town squares' where the social media platforms have created new '*virtual public spaces*, potentially engaging everyone with internet access in an interactive discussion' (European Parliament, 2019, p. 52). At this point, the fake news discourse in hindsight describe that it seemed like we had found the missing link between Habermas' ideal and the empirical infrastructure that would enable it to become a reality through the emergence of the internet in general, and with social media in particular. These platforms would allow information to flow freely, redirecting power to the public, and providing a new context for public discourse where all would be allowed to enter, voice their claims, and engage in a rational conversation. In short, the new communication landscape would align perfectly with Habermas' ideal and the democratic thought.

Emerging criticism

While initially met with these enthusiastic cheers, this was about to change. From the making of discourse ethics in the 1980–90s, to the public introduction of the internet in the 1990s and the cheers that followed with the peak period of enthusiasm in the early 2010s, this would become the decade in which the critical discourse around fake news emerged. On the new platforms, actors started to see how immoral forms of communication started to appear. These phenomena ran contrary to the communicative ideal theoretically expressed in Habermas' discourse ethics. Instead of finding the new landscape to provide an arena for the ideal, it was now depicted as a threat towards democracies. Throughout the fake news discourse, authors echo the description below of the communicative landscape. This is a concerned depiction, replacing the previous cheery tone.

> Across the world, the 'fake news' phenomenon has dominated the news and political agenda. This in many cases has resulted in people losing trust in media questioning all kinds of news and information and the channels distributing them. It is a dangerous development as our democracies build trust through dialogue, and fact-based well-informed citizens. (Nordic Council of Ministers, 2018, p. 6)

While scholars used Habermas' theory to emphasise the opportunities emerging in the age of the internet, Habermas himself did not support their claims. Aside from the rare comment, he remained silent on the topic of the new landscape's implications for his theory. But in 2006, he broke this silence.[78] And he did so in a manner surprising to many. To those who had proclaimed themselves Habermasian supporters in celebration of the internet, his commentary was disappointing at best, and annoying at worst, as he painted a far gloomier picture. Again, Habermas emphasised that his was an ideal theory and that it should not be read as reflecting an empirical reality as 'the democratic procedures of legitimation, appears to exemplify the widening gap between normative and empirical approaches toward politics' (Habermas, 2006, p. 411). In his text, he does not emphasise the opportunities in the new landscape, but remarks on the challenges for journalism in this new media landscape. Referring to the new digital and social media as 'electronic media' (p. 422), it is only in a footnote that he addresses 'in passing a remark on the Internet' (p. 423) in which he, contrary to his supporters, cautions against the consequences on democratic discourse that might arise from this new landscape. As he continues in the footnote:

> The Internet has certainly reactivated the grassroots of an egalitarian public of writers and readers. However, computer-mediated communication in the web can claim unequivocal democratic merits only for a special context: It can undermine the censorship of authoritarian regimes that try to control and repress public opinion. In the context of liberal regimes, the rise of millions of fragmented chat rooms across the world tend instead to lead to the fragmentation of large but politically focused mass audiences into a huge number of isolated issue publics. Within established national public spheres, the online debates of web users only promote political communication, when news groups crystallize around the focal points of the quality press, for example, national newspapers and political magazines. (Habermas, 2006, p. 423)

In disappointment, his supporters responded that Habermas simply did not understand the internet and its promises (Bruns, 2007; Rheingold,

---

[78] A text based on Habermas' keynote speech in June 20, 2006, at the 56th Annual International Communication Association Conference.

2007). It would be another 16 years before Habermas continued the remark on the new landscape. In a text published in 2022, he provided his *Reflections and Hypotheses on a Further Structural Transformation of the Public Sphere* (2022b). This text directly addressed the new landscape in relation to social media, discussing the implications of the new landscape in relation to the theory about the public sphere as presented in his very first book about coffee houses (1989). It expresses his concerns for the demise of a liberal belief in consensus-seeking public governance, describing the landscape as one where 'the technological advances of digitalised communication initially fosters trends towards the dissolution of boundaries, but also toward the fragmentation of the public sphere' (p. 146). This runs contrary to his normative presentation of (U) in discourse ethics, where rather the public is envisioned to engage in a commonly shared discourse. While the new platforms 'radically alter the previously predominant pattern of communication in the public sphere by *empowering* all potential users in principle to become independent and equally entitled authors' (2022b, p. 149), the new 'mode of semi-public, fragmented and self-enclosed communication' is described by Habermas to be 'distorting their *perception of the political public sphere*' as well as challenges the deliberative process 'for the formation of public opinion and political will' (p. 146). Habermas never commented on the new communicative landscape as holding promise for his ideal to become a reality—other scholars made this link. Rather, his statement as presented in 2022 point to Habermas' scepticism towards the new landscape, leading to discursive exchanges that break with discourse ethics' foundations. On the new media landscape's effect on his ideal public sphere, he comments that 'aside from its evident benefits, the new technology has also highly ambivalent and potentially disruptive repercussions for the political public sphere in the national context' (p. 159). While others had celebrated the internet, we can only speculate about the reason for Habermas' absence from this initially cheerful debate, and why his later criticism emerged when the discourse had turned to criticism a few years earlier. Perhaps he did not understand the internet and social media as his critics would argue. Perhaps he thought it premature to make a statement on an uncertain future. Perhaps he did not agree with his followers that the theory ought to be anything but a norm as it pointed *towards* but was not the truth. To that point we can only speculate.

A discourse about ‘fake news’

When Habermas expressed his concerns in 2022, the public scepticism toward the new platforms had been brewing for a few years. Just as in the previous enthusiastic period, the new discourse also relied on Habermas’ ideal and concepts in their presentation of the problematic actors. In some statements, the connection to Habermas is made by relying on his concepts. For example, by using concepts such as a ‘public sphere’ and ‘consensus’, explicitly defending ‘democracy’ and ‘human rights’ as a normative given, and advocating communicative exchanges based on ‘dialogue’ (present in the whole sample, such as in Data&Society, 2018, p. 32), and some even reference his works directly (see i.e., Council of Europe, 2017, p. 50). Rather than advocating for the prospect of the internet, however, the fake news discourse concerns immoral communication appearing on the same platforms. Headings describe it as ‘Changes in the structure of the public sphere accelerating the spread of disinformation and propaganda’ (European Parliament, 2019, pp. 52–60). While disinformation and propaganda were already established concepts in the literature, fake news was a new concept emerging in this particular discourse. The new concept is tied to this landscape, working as an umbrella term for a series of communicative phenomena that are given names in the discourse. They span from echo chambers to alternative facts and post-truth, all described as linked to the ongoing critical discourse as ‘all refer to perceived and deliberate distortions of news with the intention to affect the political landscape and to exacerbate divisions in society’ (Joint Research Centre, 2018, p. 8).

> The public sphere (Öffentlichkeit) consists of professional media on the one hand, and citizens' communication on the other. The latter used to be direct and unmediated, exercised in public spaces like coffee houses and town squares. The interactive social media platforms brought back these virtual public spaces, potentially engaging everyone with internet access in an interactive discussion. But these new ‘coffee house’ conversations are no longer unmediated, because the platform operators interfere with their algorithms. If they did not, the discussion would fall into a global cacophony. (European Parliament, 2019, p. 52)

In this discourse, the discursive presentation of *fake news* in the material reflects the general character of the discourse itself. Firstly, we see how it grounds the discourse in a few key events. Genealogically, fake news was first mentioned in a scholarly article in 2005 in discussions about the satirical *The Daily Show* (Baym, 2005).[79] Quite telling for the later discussion, the article starts with the line that 'the boundaries between news and entertainment, and between public affairs and pop culture, have become difficult if not impossible to discern' (p. 259). The article continues by describing how fake news emerges in 'a landscape in which 'real' news is becoming increasingly harder to identify or define' (p. 259). And from this initial presentation to the later discourse, fake news has always been about communication. More specifically, it starts from a concern for a certain type of communication that actors initially had a hard time defining. It was tied to a new landscape where the public was argued to become disoriented considering the new range of sources from which they could build an accurate lifeworld. It is also very telling that the concept first appeared in relation to discussions about mediated communication, in particular its connection to journalism. For in the public debate, journalists were the first to appropriate the term. As one journalist describes, he would first use the term in response to the observation of a new type of communication spreading in the new media landscape—a false story he came across in 2004 'that was quickly racking up likes, shares, and comments on Facebook' (Silverman, 2017).

> The article, published on nationalreport.net, claimed an entire town in Texas was quarantined after a family contracted Ebola, and it used a fake quote attributed to someone at a Texas hospital to pass itself off as real news story. I warned people about what I was seeing: 'Fake news site National Report set off a measure of panic by publishing fake story about Ebola outbreak' [social media post by the journalist on October 15, 2014] That Tweet was one of the first times I publicly used the term 'fake news' to refer to completely false information that was created and spread for profit. (Silverman, 2017)

But while initially used as an empirical term for something being 'false' or 'fake' or perhaps plain 'wrong', the turning point came with Donald J

---

[79] Also accentuated in texts from the discourse, for example by French CAPS and IRSAM (2018, p. 19) and the Knight Foundation (2018, p. 7).

Trump's election campaign, and the years following after that. In 2018, he had repeatedly used the term during the campaign, then referring to the 'fake news media' as a particular set of actors whom he considered misrepresented the events of the day in general and in particular the depictions of his own actions and character.[80] Again, the discussion was tied to news mediums (as in Silverman's use above), but here rather a criticism towards traditional media in defence of new alternative sources of information. This appropriation of the term as a criticism towards journalism is one of the key cornerstones in the fake news discourse's representation of the problem. As a result, texts repeatedly describe how the term 'has been appropriated and used misleadingly by powerful actors to dismiss coverage that is simply found disagreeable' (European Commission, 2018a, p. 5), or that the concept 'has become an emotional, weaponised term used to undermine and discredit journalism' (UNESCO, 2018, p. 14).

Therefore, while initially enthusiastically spread by journalists as a criticism against Trump, or alternatively towards the platforms allowing the spread of his and other criticised actors' content, all of a sudden, the same actors would not utter the words. They criticised the inappropriateness of a concept that was imprecise and eroded trust in *real news* as fake news would be an impossibility, a conceptual oxymoron as news is true by conceptual default.

> … 'news' means verifiable information in the public interest, and information that does not meet these standards does not deserve the label of news. In this sense then, 'fake news' is an oxymoron which lends itself to undermining the credibility of information which does indeed meet the threshold of verifiability and public interest—i.e. real news. (UNESCO, 2018, p. 7; c.f. Heinrich Böll Foundation, 2020, p. 7)

From the initial use of the term by journalists, they were now the ones criticising the use of the same concept. Returning to the journalist who recognised himself as part of popularising the phrase in the first place, he later published a piece titled *I Helped Popularize The Term "Fake News" And Now I Cringe Every Time I Hear It* (Silverman, 2017). In this piece, he continues to describe how it is 'now both an empty slogan and a deeply troubling warning sign.' As such, we need to recognise that while initially emerging as a moral discourse around communicative phenomena, the very label of fake news shows that rapid and fundamental changes have followed in the depictions of the phenomenon. Not least, it allows us to understand the normative character as a biased one, a case of strategic moral communication as it indeed is a motivated description of communicative phenomena where both sides accuse the other of spreading fake news.

---

[80] With frequent mentions of, e.g., *The New York Times, NBC, ABC, CBS,* and *CNN.*

The fake news discourse can therefore be seen as a discourse of conflict, as Habermas emphasise. It is not reflecting a consensus, but depicts a situation in which actors criticise the validity of the other actors' claims. In this discourse, actors are protecting and challenging the very foundation for the democratic ideal. The fake news discourse's criticism against this scenery stems from this fundamental difference, but as we will see in the later chapters, from this point it branches out as a critical debate about the new communication landscape's role in enabling and nurturing such criticism. For just as the texts agree on protecting democracy, they align in their criticism towards the platforms primarily being a facilitator for the problem. It is the new platforms, emphasised by many texts, which has created this problem by forming discourses in echo chambers rather than in a common conversation, has created a scenery in which alternative facts are as influential as valid claims, and has created the post-truth environment in which people are guided by their intuitions rather than by rational reasoning. The following three chapters expand on these three foci, to unpack the discussions in the fake news discourse, following the three foci as translated in the methodology of moral discourse analysis. First, the analysis seeks understanding of who is framed as rational actors, thereafter the phenomena conceptualised and defined as moral and immoral, and finally reflecting on the discourse' presentation of the rational and irrational. In each chapter, the analysis shows how the problem is depicted in its non-relation to Habermas' ideal, while at the same time finding that the criticism authors direct towards other actors can also be observed in their own behaviour. This discrepancy between norms and behaviour, this thesis argues, will point the analysis towards the moral intuitions advocated through the fake news discourse—as the moral discourse form as a case of strategic moral communication.

Chapter 7.

Echo chambers

The fake news discourse echoes Habermas' ideal of an inclusive discourse, finding, however, that actors are divided into separate conversations in the new media landscape. This is captured in the discourse's concerned discussion about *echo chambers*. The chapter shows why and how the discourse frames this phenomenon as a problem. From that, it engages in an immanent analysis of how the authors themselves engage in this behaviour, by only allowing actors access that align with their normative idea of moral communication. These observations align with the first focus in moral discourse analysis, illustrating how actors form moral constructions in discourses by grouping people into the moral 'us' and the immoral 'them'.

Among the first concerns in the fake news discourse is the new communication landscape's tendency to group people into discourses where all echo the other's claims to validity. Contrary to traditional news media, the new platforms did not present the same content to all. Instead, the algorithmic structures allowed the new platforms to present tailored content to each user. On digital media, this manifested as *filter-bubbles*, and as *echo chambers* on social media, as the algorithms were designed to attract the new media landscape's most sought-after resource—the users' attention. Essentially, criticism emerged upon the realisation that algorithms would feed people with content that aligned with their previous interests and convictions, dragging them into a one-lane rabbit hole that grouped them into sub-spheres of likeminded who echoed one's own communication, confirming pre-existing beliefs. This ran contrary to Habermas' and the author's normative ideal of an inclusive discourse in the public sphere where all are a part of the same conversation. Instead of a shared lifeworld, multiple social lifeworlds emerged. As they were not the subject of a critical discourse where claims raised were contested and tested, these lifeworlds are depicted to take a polarised form. At the same time, this thesis would argue that the same behaviour is practiced by those criticising such behaviour. In the studied fake news discourse, only some actors are included in the conversation as authors. Authors also advocate the exclusion of voices who break the rules, and often frames these in derogatory forms, presenting the other as not only different, but irrational and immoral. This, this chapter argues, points to the first character of strategic moral communication—that morals are driven by pre-determined intuitions and consensus is formed as echo chambers through selective inclusion.

The following chapter starts by showing how the fake news discourse's concern for echo chambers links back to Habermas' principles (U) and (D). It continues by illustrating how the same authors in the fake news discourse engage in this criticised behaviour, finally using this empirical review as a stance to argue that morals are indeed consensus within a group, as Habermas suggested, however pre-constructed consensus based on actors' alignment in their communicative norms.

Polarised lifeworlds

The fake news discourse aligns with the axioms presented in Habermas' discourse ethics. Just as Habermas emphasises that his discourse ethics rests on the principles of an inclusive discourse (his (U) and (D), see p. 26 and 27), authors in the fake news discourse emphasise striving for 'ideal forms of public deliberation' (Centre for the Study of Democratic Institutions, 2018, p. 9) so that 'people can better connect their individual views to public debate when deliberation is guided by principles such as inclusivity or openness to diverse views, a commitment to reasonable and epistemically valid claims, and an assumption of moral equality' (ibid), indeed with a direct reference to Habermas. In other ways the connection is more implicit, while not directly referencing Habermas. Among the 100 most frequent words we find many that align with the vocabulary in his discourse ethics, such as mentions of *public, open,* and *discourse.*[81] Among others, Facebook frequently refers to these concepts, describing how they aim to adjust the algorithm 'to better reflect the public discourse on and off Facebook' (2019e), their aim in 'protecting the integrity of civic discourse' (2017b), and 'promoting vibrant civic discourse' (2019d). Their aim is roughly what is described in Habermas' discourse ethics, that they strive towards an inclusive discourse where all have an opportunity to raise their voice. At the same time, these platforms are accused in the discourse of driving the very opposite behaviour. The threat formed from these new phenomena stands in direct contrast to the ideal,[82] describing how 'the shared information basis and the common narratives of society that are the preconditions of democratic public discourse are being splintered by filter bubbles' (European Parliament, 2019, p. 51), where 'the effect is a fragmented public sphere' (p. 58).[83] The consequence is described as 'walled communities' that 'make it harder for people to understand each other and share experiences. The weakening of this "social glue" may undermine democracy as well as several other fundamental rights and freedoms' (European Data Protection Supervisor, 2018, p. 22), as the phenomenon is 'reinforcing existing beliefs and increasing ideological segregation' (International Telecommunication Union, 2020, p. 61). Indeed, these filter bubbles and echo chambers foster understanding and consensus, but it is a shared lifeworld only among those occupying this space. As one author describes (Council of Europe, 2017), it reflects the ritual view of communication (see p. 60), as communication in these forums constitutes a process of establishing shared beliefs.

> Appreciating this truth helps us explain why echo chambers are so appealing. They provide safe spaces for sharing beliefs and worldviews with others, with little fear of confrontation or division. They allow us to 'perform' our identities as shaped by our worldviews with others who share those worldviews. This

---

[81] *Public* appearing 12.700 times, *open* 1.700 times, and *discourse* 960 times, see full list in Table 2, p. 205.

[82] Many texts dedicating larger sections to the topic, such as in texts by the European Parliament (2019, pp. 57–58) and the Council of Europe (2017, pp. 49–56).

[83] Alternatively referred to as *information bubbles* (NATO StratCom COE, 2017, p. 7) or *information cocoons* (European Parliament, 2019, p. 58).

> behaviour is not new, but the platforms have capitalized on these human tendencies, knowing they would encourage users to spend more time on their sites. (Council of Europe, 2017, p. 50)

This shared understanding in echo chambers and filter bubbles, and the subsequent moral identity formed, is not aligning with Habermas' ideal of an inclusive discourse, for the consensus is formed from a pre-determined consensus, and the claims raised in these fora is not contesting previous or will be contested by later voices in the closed community. They do not achieve a consensus in Habermas' ideal form as it is not based on the evaluation of various claims in a dialectic process (see p. 34 and 27); no discourse ever really emerges as there is no contestation and no conflict (see p. 36) in these filter bubbles and echo chambers.[84] This runs contrary to the ideal, as the European Parliament (2019) quotes in their report, 'in a well-functioning democracy, people do not live in echo chamber or information cocoons' (p. 58), before they continue to point to 'as Sajó noted, instead of creating a common space for democratic deliberation, the internet and social media enabled fragmentation and segmentation. Discourse is limited to occur within self-selecting groups and there are tendencies of isolation' (p. 63). Or, as another report quotes Rorty, truth is 'what one's peers, ceteris paribus, let one get away with saying' and 'the question of whether justification to the community with which we identify entails truth is simply irrelevant' (Ash Center for Democratic Governance and Innovation & Harvard Kennedy School, 2018, p. 8). In these algorithmically contained lifeworlds, there is no contestation as 'search results and social media feeds only show us results that cohere with what we already enjoy or believe hence creating filter bubbles or echo chambers. Fake news appearing to match or support these preferences or beliefs spreads quickly and is believable in this environment' (RSIS, 2018, p. 10).

This points to the first problem accentuated in the fake news discourse—its emergence out of the new media landscape. Overall, it is explained by the new set of offline and online media outlets, with partisan representations of reality appearing on both (Council of Europe, 2017, p. 5). Many texts, however, derive the very cause of the fake news discourse to the technical infrastructure that led people to digest only perspectives which echoed their pre-conceived biased beliefs.[85]

---

[84] As Habermas describes it in relation to discourse being a scenery of contestation and conflict, as criticism arises towards the pre-existing norms, a situation 'in its openness, practical discourse is dependent upon contingent content being fed into it from outside' (1999, p. 103).

> …the technology companies are commercial entities, and therefore to keep their shareholders happy need to encourage users to stay on their site for as long as possible to maximize the number of exposures to advertisements. They do so by tweaking the algorithms to deliver more of what users have liked, shared or commented on in the past. (Council of Europe, 2017, p. 52)

While the platforms align in their depiction of the ideal landscape, they do also accentuate an alternative explanation for why this problem emerged in the first place. Rather than accentuating the communicative phenomena's character as linked to a technical landscape emerging with the algorithmically structured forums for communication, they accentuate its link to the social landscape. While arriving at the same conclusion that there is a problem with people being divided into separated lifeworlds—in such closed groups developing more extreme views in conflict with other silos—this is rather accentuated by platforms as a problem derived from *polarization*. Facebook defines the phenomenon as emerging 'when like-minded people talk to one another and end up thinking a more extreme version of what they thought before they started to talk' (2018c). This is described in reports as how 'the fundamental problem is that 'filter bubbles' worsen polarization by allowing us to live in our own online echo chambers and leaving us with only opinions that validate, rather than challenge, our own ideas' (Council of Europe, 2017, p. 50). While other authors would describe how 'polarization may be exacerbated by the existence of homophilous information communities or "filter bubbles", which develop in fragmented and polarized media systems and through the algorithmic curation of information' (Centre for the Study of Democratic Institutions, 2018, p. 37), Facebook on the one hand recognises that polarisation is 'an issue that in fact we have invested heavily in.' However, aside from listing their efforts 'to reduce the amount of content that could drive polarization on our platform' they continue that 'this job won't ever be complete because at the end of the day, online discourse is an extension of society and ours is highly polarized' (2020e). As Facebook continues to argue, at the end of the day, even if there are problems on the platform, social media remains a positive force overall in relation to the public discourse, in support of the democratic and Habermas' ideal.

---

[85] Moving to discussions about *micro-targeting* as the algorithmic structure allows advertisers to target people with specific online characteristics. This is particularly emphasised as a problem in relation to elections, with discussions about *dark ads* used as a tool by political actors to spread different messages in different bubbles—thus subverting the democratic open discourse. These practices 'splinters the shared information basis of society, reduces understanding between people with different beliefs and exacerbates polarisation' (European Parliament, 2019, p. 58).

> I don't minimize the potential challenge of issues like 'fake news' or the 'filter bubble.' They are real, serious and as yet untamed. And yet a technology that has the capacity to expand and diversify political equality around the world is a net good. Most other forms of political engagement tend to favor those with the most wealth or access. Not social media. It gives voice to anyone with a phone. In a time when political power is synonymous with economic power, the type of collective action social media makes possible is giving more people a say in the conduct of their governments and the society they live in. (Facebook, 2018a)

As the above quote illustrates, the problem is often discursively presented in relation to fake news. Repeatedly, they take the example from the US election race of 2016 as a key event where polarisation became highly explicit. However, as some texts also suggest, this might only be the result of an accumulated discourse arising from this as well as other events before the election. Just as Habermas points to the new time's pluralistic and multicultural societies (see p. 29), so do texts trace it as 'both a cause and a symptom of the crisis of democracy, which is evidenced by the growing abstention in elections, distrust towards elected officials and even questioning of democratic and liberal values' (French CAPS & IRSAM, 2018, p. 37).[86] The report continues to trace fake news to events beyond 2016, among them accentuating factors stemming from globalisation and the 'polarization of identity' as…

---

[86] This is also a point highlighted by Habermas (2022b) as he in part recognises that the diminishing trust in democracy stems from the non-realisation of the ideal.

> In response to the porous borders and the cultural blending that globalization engenders, there is an increasing demand for the reaffirmation of 'us' against 'them.' This tendency includes phenomena such as the erection of walls (Israel, the United States, Hungary), the expansion of gated urban communities, the imposition of refugee quotas, etc. (French CAPS & IRSAM, 2018, p. 37)

As the quote describes, the increasingly hostile atmosphere where groups of 'us' are contrasted against 'them' aligns with Habermas' concern for othering (see p. 28), and moral psychology's recognition of how we divide people into groups (presentation starting on p. 45). This, they argue, can thus be traced to factors such as globalisation in general, adding to that specific events such as the European refugee/migration crisis of 2015/16 (French CAPS & IRSAM, 2018, p. 37; Swedish Defence University, CATS, & Hybrid CoE, 2018, p. 75). From this perspective, Donald J. Trump's success in the US elections of 2016 should be seen as only a part of a larger scenery emerging at this time that we need to recognise if we are to understand the emergence of the fake news discourse.

As a final question, we might ask whether these phenomena of filter bubbles and echo chambers actually exist. Indeed, the discourse overall depicts the phenomenon as an argument for the detrimental impact of new algorithmically structured discursive interactions between people in a public sphere. The way these claims are presented are rather as a commonly known truth, however, not resting on studies. Many, such as the (Council of Europe, 2017) argue that 'there is no doubt that digital technologies support us in these tendencies' (p. 49). However, such a claim rests partially on essays and thought-pieces presented by scholars including Habermas' speech to the ICA from 2006. In claims such as his that 'the rise of millions of fragmented chat rooms across the world has tended instead to lead to the fragmentation of large but politically focused mass audiences into a huge number of isolated issue publics' (2006, 423). Indeed, they do draw upon confirmed knowledge from cognitive science that we are more likely to believe facts that align with our pre-accepted beliefs (c.f., European Parliament, 2019, p. 58; RSIS, 2018, p. 11; UK House of Parliament, 2017, p. 3),[87] but to move from this empirical knowledge to argue that the new media landscape has amplified such behaviour is speculative. Indeed, one report describes how ‘the development of filter bubbles which *theoretically* confirm biases and reduce exposure to quality, verified information’ (UNESCO, 2018, p. 61, emphasis added). While some authors describe how digital media such as Google ‘admitted in early 2018 have a tendency to reinforce confirmation bias’ (p. 62), this thesis would argue that authors run the risk of falling into the same behaviour—that the fake news discourse in a broader sense fall into its own accentuated trap of confirmation bias. For they repeatedly accentuate voices that conform with a pre-existing belief of their intuition of filter bubbles and echo chambers. To be clear, this thesis does agree that these human behaviours exist, but that this has always existed and is rather a part of the way we group ourselves (as the repeated statement in moral psychology, see presentation starting on p. 45). What the thesis challenges is the technological determinism expressed in the fake news discourse, singling out platforms as the immoral them solely responsible for this unwanted phenomenon. For the same report also notes that there are studies questioning the very existence of these communicative phenomena (Council of Europe, 2017, p. 53). This is also accentuated in other reports as ‘some research challenge the belief that social media embed citizens in “filter bubbles”’ (Centre for the Study of Democratic Institutions, 2018, p. 37) and is described as ‘a substantial and contested literature’ with both studies supporting and studies contesting the existence of filter bubbles (Knight Foundation, 2018, p. 10). Other reports continue that this ‘remains to be conclusively proven or disproven. Contemporary research increasingly challenges the notion that filter bubbles and echo chambers influence opinion formation, as the internet, despite selective algorithms, often presents a broad spectrum of opinions and information to everyone, offsetting the negative’ (Swedish Civil Contingency Agency, 2018, pp. 37–38). Some reports even argue that the new media landscape has resulted in a more varied consumption of information than before (Reuters Institute, 2017, p. 37, 2018, p. 10).

---

[87] Framed in the concept *confirmation bias*, mentioned in 25 texts. For example, RAND's (2018b) report *Truth Decay* rest heavily on the concept, defining it as 'the human tendency to hold onto prior beliefs even when presented with information clearly

---

demonstrating that these beliefs are incorrect or misguided’ (2018b, p. 83; c.f., UNESCO, 2018, p. 86).

Self-constructed consensus

While criticising the discursive structures in the new media landscape, the immanent review (see p. 83) points to actors criticising such behaviour as engaged in the same behaviour themselves. As the discourse was sampled organically—adding material based on which other authors and texts were linked to this discourse—the discourse amounts to a set of authors selected by the authors themselves to be part of the fake news discourse. The sample include authors such as international governmental actors, however this almost exclusively includes coalitions in which Western actors are a part (such as EU, NATO, and the UN) and countries aligning with the ideal of democracy (such as the US, UK, Germany, Sweden, France etc.).

Furthermore, we can see how the fake news discourse frames actors as either part of the moral 'us' or 'self' in opposition to an immoral 'them' or 'other'. While Habermas' (U) was his attempt to advise actors against engaging in this behaviour of othering (see p. 28), the fake news discourse emerges out of this very conflict, from actors opposing the normative axioms found in his discourse ethics and echoing in the studied discourse's own moral evaluation. Indeed, some authors are presented by the other authors in the studied discourse as part of the other, such as platforms. The wider range of actors, however, emerge upon review of the actors mentioned in but not appearing as authors in the corpus. The moral self is found in the references in the texts to traditional journalism, but only a particular set of journalism often referring to reports by e.g., CNN and BBC, while not using FoxNews or alternative media as sources of truth. The clearest form of othering emerges in a wider representation of the self as tied to a general idea of Western democracies (but also coalitions such as EU and NATO) against a political other that is in opposition with these actor's ideals, either in the form of foreign hostile powers (often mentioning Russia)[88] or domestic actors (framed as alt-right, populists, or conspiracy theories).

---

[88] This narrative is very strong in the dataset, not least by the sheer mention of Russia (10,500 times), Russian (13,800 times), Kremlin (3,300 times), and Putin (1,800 times), amounting to a total of at least 29,400 times where the studied discourse draws attention to this example. If treated as a word-cluster, it would rank as the third most

This discursive framing of actors is formed as moral actors who are presented as trustworthy, protectors of democracy. Immoral actors are depicted to spread lies with the intent of undermining democratic societies. For example, one text describes how the West is portrayed in Russia, that 'Russian state media actively demonizes NATO and other key western institutions and rejects the values of tolerance, moderation, and democracy fundamental to the core of the nation' (Macdonald-Laurier Institute, 2019, p. 20). Several texts explain the aggressiveness of Russia to be the result of the nation being 'anxious about losing ground to Western influence in the post-Soviet space and the ousting of many pro-Russian elites by popular electoral uprisings' (Chatham house, 2016, p. 2). The text continues to describe how the goal is to make the public 'accepting Russia's supremacy' with the objective to 'challenge Western influence' (p. 3). Continuously, we thus see how the self is described to engage in 'influence' whereas the other engages in efforts to establish one's 'supremacy'. Imagine one instead writing that 'the West is anxious about losing ground to Russian influence' whereby the goal is to approach the public to 'accept Western supremacy' with the objective to 'challenge Russian influence'.

There are also other actors painted as the problem, threat, or simply a cause for concern. Labels span from alt-right[89] to populistic media and politicians, and to conspiracy theorists. In the discourse, authors lump these immoral actors together. Sometimes this is done by presenting them on a row, as one text describes the problem as 'anti-western, pro-Kremlin propaganda, conspiracy theories and anti-NATO media' (ISD & LSE, 2018, p. 12). This would span from those who are against the West in general, those promoting the Russian narrative, either through conspiracies, or by being against NATO. They are also connected in labels such as 'anti', in concepts such as anti-democracy, anti-globalisation, anti-NATO, anti-West, anti-EU, anti-American, or in Eurosceptism (Center for European Policy Analysis, 2016, p. 34; French CAPS & IRSAM, 2018, pp. 74–75; GLOBSEC Policy Institute, 2016, pp. 6–7; ISD, pp. 4, 46; Macdonald-Laurier Institute, 2019, p. 16; RAND, 2018a, pp. 66, 77). In

---

common word appearing in the sample (after *media* mentioned 32,400 times and *news* mentioned 30,600 times).

[89] While the alt-right is discussed in 45 texts, only one mentions the alt-left (ISD, pp. 4, 46).

some cases, they are even described as connected empirically—such as in descriptions on how foreign actors use domestic conspiracy theorists and populistic politicians as puppets by either supporting them economically or in other ways feeding them with content. Indeed, in some cases this link is a proven fact, that there are 'different research centres and institutes that have more or less overt links to Russia propagate Russian positions on foreign policy' (French CAPS & IRSAM, 2018, p. 75). More often, however, the link between these actors is founded in a discursive representation of all as part of the immoral other. Through this behaviour, this thesis would argue that the discourse breaks with Habermas' principle of nurturing an inclusive discourse that does not engage in a practice of representing the other but makes it a part of the self by inviting them to the discursive construction (see p. 26).

The thesis would even argue that the reviewed discourse moves from describing 'the other', to engaging in a practice of 'othering' by representing these actors as inferior, immoral, irrational, and other dehumanising characteristics (see p. 28). They are downgraded from a moral to a bare life (see p. 70) by being discursively presented in the fake news discourse as less than human, lacking in mental capacities, culture, and civil behaviour which is otherwise ascribed to moral human beings. As in TikTok's definition of *hateful behaviour*,[90] this could include…

> …hateful content related to an individual or group, including: claiming that they are physically, mentally, or morally inferior calling for or justifying violence against them claiming that they are criminals referring to them as animals, inanimate objects, or other non-human entities promoting or justifying exclusion, segregation, or discrimination against them. (TikTok, 2020)

While condemning this behaviour, the discourse engages in the same practice. Examples include descriptions of conspiracies as having a 'paranoid style' (Data&Society, 2017, p. 19) or 'paranoid beliefs of those who are predisposed to believing conspiracy theories' (Macdonald-Laurier Institute, 2019, p. 19), and in describing how it is 'right-wing conspiracy sites that disseminate their paranoid ideas via social media' (Center for Business and Human Rights, 2019, p. 3).[91] Some specify that this should be the target of one's counter-measures, suggesting that 'governance mechanisms should be introduced to ensure that major content-selection algorithms do not generate individualised information spaces that entertain unduly extreme, delusional, obsessive or paranoid mental states' (Wilfred Martens Center for European Studies, 2018, p. 11). All of this points to a discursive representation of these actors' mental incapability and lack of rational thought. It in part aligns with Habermas' own vocabulary, presenting how 'I think it makes sense to start from the idea that linguistic intuitions can be 'false' only if they come from incompetent speakers' (1979, p. 212) and other quotes where he equates those not accepting the basic premises of social life to be suffering from 'serious mental illness' (Habermas, 1999, p. 100). It can also be seen in his idea of the rational and irrational, where he implicitly by talking about rational and rational arguments talk about 'the other' as irrational. Similarly, the fake news discourse depicts multiple other dehumanized actors—ranging from conspiracy theorists, to members of the alt-right movement, Trump supporters, and Russian agents—linked in their rejection of the axioms grounding Habermas' discourse ethics and by extension the fake news discourse's common sense.

---

[90] Similarly appearing in Facebook's depiction of *Bullying and Harassment* (n.d.a) and *Hatespeech* (n.d.b), in Twitter's *dehumanizing policy* (2018).

Protecting an open debate through exclusion

Moving from the normative ideal of inclusion to the observed behaviour of exclusion, we find the thesis' first observable immanent paradox. In this paradox, we find the first important piece to understand the phenomenological boundaries constituting the moral discourse. By defining the moral, the discourse inevitably defines the immoral, and as Agamben emphasises, the immoral will be excluded from the system (see p. 70). Indeed, this is a repeated mantra in the discourse's criticism against platforms—that they are not doing a good enough job in cleaning up the online conversation, by identifying the immoral and excluding them.

---

[91] Also talking about Russia as paranoid and spreading paranoid ideas as e.g., continuously depicted by the Institute of Modern Russia and The Interpreter (2014) and by the US Committee on Foreign Relations (2018, pp. 13–14).

To that end, the platforms have developed their own *codes of conduct* in which they form the rules for discourse on their own platforms. If you break with these rules, the platform describes having ground to act. Some platforms work with a strikes-system where one is given a set number of warnings (YouTube, 2019), others focus on a permanent or temporary suspension (Twitter, 2020b), and eventually some platforms created their own type of courts where one can appeal a suspension decision (Facebook, 2020a). The discourse describes how the platforms should not only exclude immoral actors (what is referred to as *deplatforming*), but that they should also actively moderate the conversation—by down-ranking some actors chances to get seen by others on the platform.

> Online intermediary platforms must be responsible for prioritising authoritative information and sources, deplatforming malign ones, and down-ranking and clearly labelling misinformation. New regulation should lead them to undertake robust and transparent research that informs the design of products and effective moderation systems that balance privacy, safety and freedom of expression. (Institut Montaigne & Institute for Strategic Dialogue, 2020, p. 23)

Initially, the platforms were sceptical towards this practice, arguing that conversations should be free and unregulated. In a sense, this would align with Habermas' ideal for a free discourse. But with the fake news discourse, active moderation was eventually seen as platforms taking responsibility. Most notably, this is reflected in the public announcements by platforms such as by Twitter and Facebook whenever they remove networks of spammers engaged in economic deception, networks spreading disinformation for political aims, or organisations such as terrorist organisations using the platforms to radicalise users. At the same time, the removal of these actors from one's common public sphere is arguably contra-productive towards the ultimate goal of an inclusive conversation—as many have noticed that actors banned from these platforms simply turn to other less regulated ones.

> …the focus of improving content moderation approaches on a few large platforms over the past three years has led to a platform migration of many

> purveyors of hate speech, extremism, terrorist content and disinformation away from large platforms to smaller platforms with little or no oversight, limited or no Terms of Service (i.e., Gab), or in some cases, any appetite or intent to respond to online harms (i.e., 8chan). A limited focus on the few largest platforms would simply accelerate this phenomenon. (ISD, 2019)

Be it closed chatrooms not subject to as strict content-surveillance by platform moderators—such as WhatsApp and Telegram, in one text described as 'enjoying a complete absence of moderation' (French CAPS & IRSAM, 2018, p. 82)—or simply turning to platforms with less restrictive discussion forums or no restrictions at all—such as Gab, or for example the Russian platform VK (VKontakte). On these platforms, what is labelled as 'controversial' or plain 'false' or 'misleading' is not evaluated by the same criteria on the other platforms, and thus allowed. As one study reports, 'many far-right groups have also found a home on Russian-language social media groups on OK and VK', finding 'a safe haven' (ISD & LSE, 2018, p. 14).

The authors expect that platforms use their algorithmic ability to make sure that they clean out immoral communication. They are expected to function in the same way as the journalistic gatekeepers, working as a type of editor ensuring that the content published on the platform is constructive. At the same time, they should give priority to 'privileging credible content in ranking algorithms' (European Commission, 2018, p. 14), essentially meaning that they should not only moderate but influence the public's perception of the conversation. They are the ones expected to exclude those acting immorally, while promoting moral communication and in doing so skew the conversation towards a better one. What is lost, however, in this process, is that it breaks with the very ideal in Habermas' discourse ethics of an open conversation where all members of the public can have a voice, no matter how ridiculous their claims are, as their validity will be judged by the public once shared in the discourse. Either by excluding or down-ranking content from some while up-ranking the content of others, platforms are expected to act in line with the normative system's values of what is moral and immoral. By doing so we inevitably move away from the public conversation advocated by Habermas. Arguably, his ideal was rather found in the early-days of these platforms, in the near anarchistic arenas for free speech. Today, these platforms are pressured to execute what Habermas would call a system, restricting the formation of the lifeworld in a way that runs contrary to the normative ideal. The question remaining, then, is who has influence over these regulations—is it based on the platforms' idea of the public's perception (building a socially restrictive world from the public's individual lifeworlds) or are there other actors who influence the rules regulating online discourses today?

The review of the fake news discourse also reveals the discursive process in which actors are negotiated as moral or immoral. Reoccurring labels include *misinformation* and *disinformation*. Both are cases of false information, but in the first the *misinformed* are not aware of it being false, as these actors are not acting with malicious intent. Knowing that, should one exclude them then from the conversation? The authors would say no, instead one ought to inform the public that they are misinformed. This was especially emphasised in the beginning of the discourse, aligning with a transmission view of communication (see p. 60), where the new information as expressed would act as a magic-bullet or hypodermic-needle and the misinformed would come to their senses. But sometimes, this failed. The misinformed did not change their view after being corrected. So, what then? While a Habermasian would perhaps theoretically argue that if we are not convinced, then the claims raised are not valid—and a rejected claim is simply a rejected claim. However, the reasoning in the fake news discourse is different, rather aligning with Habermas' idea that their point of normative view is so common sense that it is inevitably true (see p. 182). For example, as the misinformed cling to their belief in the false, they are branded as conspiracy-theorists or useful idiots in the hands on foreign political powers, or simply described by authors as misled by their own cognitively biased way of thinking (RAND, 2018b). A few authors, however, continue to emphasise the right for everyone to not only hold but freely express their opinion—even if such beliefs are false. As the UN emphasises…

> …the human right to impart information and ideas is not limited to 'correct' statements, that the right also protects information and ideas that may shock, offend and disturb, and that prohibitions on disinformation may violate international human rights standards, while, at the same time, this does not

> justify the dissemination of knowingly or recklessly false statements by official or State actors (Organization for Security and Co-operation in Europe, 2011)

This creates a dilemma, or a balancing act as actors have to balance two opposing interests—as captured in the report by the International Telecommunication Union (2020) entitled *Balancing act: Countering online disinformation while respecting freedom of expression.*[92] This is not least present in the EU's settling on using *disinformation* over *fake news* as a concept, as the former would designate 'a phenomenon that goes well beyond the term "fake news"' (European Commission, 2018a, p. 5), defined to include 'all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit' (p. 3), however does not cover illegal forms of communication (such as defamation and hate speech) nor 'other forms of deliberate but not misleading distortions of facts such as satire and parody' (p. 5). In a later publication, the European Commission (2018b) continued to specify that disinformation 'does not include misleading advertising, reporting errors, satire and parody, or clearly identified partisan news and commentary' (p. 1). This points to the problems embedded in the very balancing act, for as the European Commission also notes, fake news 'has been appropriated and used misleadingly by powerful actors to dismiss coverage that is simply found disagreeable' (2018a, p. 5). While many of the texts advocate for regulatory measures to deal with the problem, this can also be seen as a problem itself—potentially limiting the very freedom of speech that these actors seek to protect. As subsequent texts noted 'democracies presupposes a lively public political discussion, which should not be stifled by over-harsh regulation of expression' (European Parliament, 2019, p. 16). As human rights organisations express this concern, this is not only expanded to state-interventions through national laws against fake news and disinformation that may limit the public's right to freedom of expression, but also appears in relation to third party intermediaries such as digital and social platforms. On the one hand they emphasise that while they are not responsible for the content published, they should remove content based on court orders. On the other hand, increasingly throughout the fake news discourse, they have developed algorithmic processes for scanning content for that which breaks with their codes of conduct, at times using human content moderators to review the flagged content, but at other times simply having to rely on the algorithmic governed moderation to manage the vast amounts of communication. This at the same time as the human rights organisations accentuate that these platforms too can fall into the trap of algorithmic censorship as they express a concern…

---

[92] 39 texts using the phrase 'balancing act', spanning from platforms such as Facebook (2020g) and public authorities such as the UK Law Commission (2018) as to the challenge of 'balancing the application of the criminal law with the qualified right to freedom of expression' (p. 24).

> …about some measures taken by intermediaries to limit access to or the dissemination of digital content, including through automated processes, such as algorithms or digital recognition based content removal systems, which are not transparent in nature, which fail to respect minimum due process standards and/or which unduly restrict access to or the dissemination of content (Organization for Security and Co-operation in Europe, 2011, p. 2)

The balancing dilemma emerges between on the one hand seeking to sustain a universally inclusive discourse—following Habermas' ideal—while at the same time recognising that this would allow some actors (the other) who have broken with the pre-determined rules and who actively seek to destroy this conversation, and for that reason ought to be excluded from the very same conversation. To remove users is presented as common sense many times, however some actors caution against the consequence of such actions. In the case of platforms, many joined in an official statement on the *Crossroads for the Open Internet* in 2020, in which they raised their concerns as to how the EU's *Digital Service Act* and the *Democracy Action Plan* could result in a state 'limiting our online environment to a few dominant gatekeepers, while failing to meaningfully address the challenges preventing the Internet from realizing its potential' (Twitter, 2020a). Others would argue that it is these very platforms that already hold such a position, being the dominant gatekeepers in the new communication landscape. Either way, the question remains how to balance interests between freedom of speech and protection of the discourse online, to what extent we ought to approach the system embedded in the normative ideal as a force that is allowed to restrict individual lifeworlds (c.f. Habermas 1994, p. 170–171).

Chapter 8.

Alternative facts

Aside from passing a normative judgement on the actor as moral or immoral, the fake news discourse also determines this line by presenting different communicative phenomena as moral or immoral. This constitutes the second analytical focus in moral discourse analysis (see p. 103). This chapter unpacks which communicative phenomena are being discussed and on what basis their definitions determine them to be cases of moral or immoral communication. The criticism arises as actors in the fake news discourse argue that actors do not depend on valid claims, but alternative facts.

As Habermas, the fake news discourse links the idea of moral and immoral communication to an assessment of the claims to validity raised in such cases of communication. The criticism similarly arises out of a concern that actors are not only conveying communication that aligns with normative standards, but express and indeed convince others of their validity by resting on other forms of claims. Rather than relying on valid claims, the immoral take the form of *alternative facts*. This is sometimes defined as false information, at other times as true information framed in a deceptive way, and finally by some acknowledged as an expression of the criticism raised towards the hegemonic claims to validity. This conceptual confusion expands from this single concept to representing a characteristic of the debate itself. In the fake news discourse, a myriad of concepts are presented as part of the problem and solution respectively. For example, spam (Facebook, 2014a, n.d.c; Twitter, 2020b) is considered an immoral form of communication, where flagging (YouTube, 2006, 2009) would be a moral counter-measure. This pattern repeats along a series of sub-themes in the discourse spanning from discussions tied to science, journalism, warfare, politics, etc.—in all finding both concepts for immoral and moral communicative phenomena. By reviewing the definitions of these concepts, the analysis seeks to identify the underlying common elements of interest as normative values. Mapping the moral and immoral valence recurringly attached to these values, the chapter forms a system that is argued to constitute the normative character of the moral discourse. While Habermas would argue that validity claims can either be assessed as valid or not, the thesis expands upon this claim to argue that actors can point to ideal types as combinations of multiple values—enabling a strategic assessment of communication where normative values are emphasised and downplayed in relation to the evaluation of cases to point to a pre-determined ideal moral or immoral type. Once again, the chapter points to the strategic character of moral communication and in doing so paints a clearer picture of the moral discourse under review.

The discussions about alternative facts draws attention to the type of phenomena subject to moral evaluation—namely communication. From the discourse, the next section unpacks how this phenomenon is understood as various forms of communication, continuing by showing how authors add normative labels to separate moral from immoral communication, building the systems' normative framework.

Invalid claims

As we zoom in on how the fake news discourse determines communication as moral or immoral, we again find it aligning with Habermas' discourse ethics as this also falls on an assessment of the claims raised in communication. In the discourse, some even directly refer to this process as one where actors 'assessed the validity of claims made' (Institute for Strategic Dialogue, 2020, p. 9), or 'attempt to prove the validity of their claims' (French CAPS & IRSAM, 2018, p. 15). While these authors do not reference Habermas directly, they do rely on the same normative concepts. What emerges from the review presented below is a discourse where the phenomenon under normative discussion is indeed communication—expressing a concern for a scenery where actors no longer make a normative assessment based on what they would pass as valid claims, as they are convinced of *alternative facts*. Yet again, we find this concept emerging in opposition to the normative ideal founding Habermas' discourse ethics. When a panel of German linguists selected the *Un-word of The Year* (German *Unwort des Jahres*) as the anti-democratic word of the year that violates human-rights, they awarded this price to alternative facts.

Alternative facts are either used interchangeably with fake news as a concept, or is not defined at all (see i.e., Global Health Security Network, 2020). Alternatively, it is specified that fake news is built on alternative facts, distinguishing that fake news is the broader phenomenon under concern while alternative facts point to the evaluative factors from which we can determine if communication is moral or immoral. In some definitions, authors even specify that alternative facts are truths in the sense that they align with an objective world, but are deceptive as they are framed to present a deceptive message.

> …while fake news aims for a complete manipulation of the public and is intentionally false and fabricated with this specific purpose in mind, alternative facts imply the existence of objective facts that are present somewhere but contextualized to drive the story. In other words, alternative facts match the existing data and the facts are cast in a context that suits political powers. (Heinrich Böll Foundation, 2020, p. 11)

As before, the concept is tied to the US election race of 2016, and Donald J. Trump's spokesperson Sean Spicer at the first press briefing after the inauguration accused the press of underestimating the size of the crowd attending the event. Trump's adviser Kellyanne Conway later defended this claim by arguing that it was a case of alternative facts. As this was the first time the word had been used, the interviewee remarked that 'alternative facts are not facts, they're falsehoods', whereby Conway accused the journalist of marking her words to make her look ridiculous. In a later interview, Conway defended the choice of term, expanding that alternative facts would refer to 'additional facts and alternative information', adding in another interview a few years later that 'two plus two is four. Three plus one is four. Partly cloudy, partly sunny. Glass half full, glass half empty. Those are alternative facts' (NATO StratCom COE & King's Centre for Strategic Communications, 2018, p. 22).

From this example, we see three different discursive meanings being attributed to the use of the term alternative facts in the reviewed debate.

On the one hand, the fake news discourse suggests that the statement 'can be seen as a way to mitigate the negative effect of a claim being dismissed as a blatant lie' (NATO StratCom COE & King's Centre for Strategic Communications, 2018, p. 22), thus pointing to the interpretation that the immoral is found in the statement itself as alternative facts would refer to the non-correspondence with an objective world. Others would indeed point to alternative facts as the deceptive framing of objective facts (above-presented) to mislead the audience—thus either tying the immoral to what Habermas describes as a distinction between what is said, understood, or what is meant (see p. 25). A third interpretation is found in other statements describing how Conway's use of the term would refer to an idea that 'no matter how the media report, the government has "alternative facts"' (Heinrich Böll Foundation, 2020, p. 19), which differ from the truth, thus rather expressing a criticism and indeed point to a moral discourse of conflict in Habermas' sense of the word (see p. 36) as a difference in social lifeworlds. Alternative facts are either presented as a concept linked to Habermas' idea of valid claims as tied to truth, rightness, or truthfulness (see p. 34). One text takes this a step further, arguing that 'one explanation for this is that Trump and his associates simply do not believe in the existence of a fact of the matter. Therefore, in their view, all factual claims are merely expressive of partisan bias', continuing to quote Trump's former adviser Roger Stone that 'Facts are, obviously, in the eye of the beholder. You have an obligation to make a compelling case. Caveat emptor. Let the consumer decide what he or she believes or doesn't believe based on how compelling a case you put forward for your point of view' (Ash Center for Democratic Governance and Innovation & Harvard Kennedy School, 2018, p. 8).

Phenomenon, value, and valence

What this moral discourse reveals when zooming in on the concept of alternative facts is not only how the discourse aligns with Habermas' view that communication is assessed in relation to its claim to validity, but also reveals multiple factors of normative concern in the evaluation. While Habermas would point to three claims to validity (see p. 34), he argues that each time and place determine their own agreed claims to validity (see p. 36). The aim in the below-presented second step in the moral discourse analysis (see p. 103) is to point to evaluation and seek the common values building these concepts and how these values are ascribed a moral and immoral valence.

On the one hand, the division of actors into 'us' and 'them' points to a common emphasis in the normative evaluation of communication as moral communication aligns with and seeks to defend democratic values and human rights. It reoccurs so frequently in the sampled discourse that it nearly becomes invisible, appearing in titles such as *Democratic Defence Against Disinformation* (Atlantic Council, 2018, 2019a)[93] where disinformation in general is framed as part of the other, as here standing in direct opposition to the democratic self. The democratic ideal expands further to a sense of morality found in freedom of expression, freedom of press, equality, etc.[94] Among these texts, human rights take the form of a modern-day eudemonia (see p. 33), where the fake news phenomena constitute a direct threat toward the same values.

> Fine grained, sub-conscious and personalised levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions. These effects remain underexplored but cannot be underestimated. Not only may they weaken the exercise and enjoyment of individual human rights, but they may lead to the corrosion of the very foundation of the Council of Europe. Its central pillars of human rights, democracy and the rule of law are grounded on the fundamental belief in the equality and dignity of all humans as independent moral agents. (EU Committee of Ministers, 2019)[95]

Continuously, alternative facts are described to be part of the phenomena that 'poison debate' and 'erode consensus' (Global Health Security

---

[93] Similar titles found in, for example, Canadian House of Common's publication (2018).

[94] *Human rights* appear as a lexeme in over 200 texts, *equality* in about 150 texts, and *freedom* is mentioned a total of 3,200 times across the sample.

[95] Similarly in another text that 'freedom of speech is necessary because it enables individuals to "develop their faculties", to realise their own potential and autonomy through expressing themselves—and to become responsible moral agents of a just political society' (European Parliament, 2019, p. 76).

Network, 2020, p. 5). It also appears in the description of immoral actors using these tactics to 'undermine government legitimacy, exacerbate social discord, or erode citizens' trust in democratic institutions' (Centre for the Study of Democratic Institutions, 2018, p. 2). This is not only tied to domestic institutions, but also to international collaborations between Western allies. For example, NATO describes that it 'undermines a sense of shared reality and a will to fight together against common challenges' (Global Health Security Network, 2020, p. 5).

Beyond this conceptual consensus, the fake news discourse is described as characterised by conceptual confusion (Council of Europe, 2017, p. 4). The range of communicative phenomena defined as moral and immoral respectively grows very long. Sometimes, the same concept is used for different phenomena or the same phenomenon is defined in different concepts. The conceptual confusion is also recognised as a problem in the discourse, especially in the critical discourse around the use of the term fake news (see p. 117), as authors continue to describe how…

> The subject is riddled with an abundance of imprecise terms, mixing classical notions (influence, propaganda, disinformation) with neologisms (fake news, post-truth, fact-checking), whose multiplication 'signals the inability for the existing vocabulary to describe a social world that is completely transforming.' (French CAPS & IRSAM, 2018, p. 18)

With the aim to bring order to these concepts, this thesis collects a set of key concepts tied to the fake news discourse and reviews on what basis these phenomena were defined as moral or immoral. What emerges are concepts defined by normative values tied to different elements in the process of communication (see p. 61). To each value, these concepts are assigned either a moral or immoral valence. For example, moral communication can be assessed in relation to the element and value of content of the message, determined as tied to a valence of being either morally true or immorally false. This forms the basis of a normative system, as phenomena are defined in relation to elements tied to values, each either evaluated as holding a moral or immoral valence. Figure 1 (see p. 151) presents an overview of how this system unfolds in definitions of some key concepts, continuing in the presentation below to contextualise the appearance of these concepts in relation to a series of themes that emerge in the studied discourse.

Upon reviewing the concepts holding an immoral valence, attributed to any case of communication criticising or threatening the integrity of the communicative ideal reflected in Habermas' discourse ethics, we find that the concepts also point to a more instrumental or technical set of aspects in the evaluation of communication. In short, they do describe communication as moral or immoral based on attributions of normative values to the elements in the process of communication (see p. 61). From this more complex set of concepts, there is however no consensus in terms of which elements are in focus and what normative values are used as grounds for conceptual determination. While presented as an unorganized set of concepts—many times not defining the concepts, at other times using the same concepts for different definitions, or the same definition for different phenomena—the following presentation reveals a set of discursive themes linked to sub-discourses emerging in the fake news discourse. Firstly, before fake news was coined as a concept, the discourse emerged around a concern for *pseudoscience* (French CAPS & IRSAM, 2018, p. 33; Institute for Strategic Dialogue, 2020, p. 7; International Telecommunication Union, 2020, p. 196; New Knowledge, 2018). This drew upon an enlightened vocabulary with concepts such as 'evidence' to create 'theories' resting on alternative facts. The problem was described to manifest across disciplinary interests such as *climate-change deniers* in relation to environmental science (Reuters Institute, 2020), in relation to health through *anti-vaxxers* (Centre for Countering Digital Hate, 2020), and geography as in *flat-earthers* (YouTube, 2020) to mention but a few examples. Here, the discourse foremost emphasises informing the misinformed, through *myth busting* and *debunking* (Sceptical science, 2012). This would later develop into discussions about economic incentives—first accentuated by the platforms—spanning from phenomena such as *hoaxes* (Facebook, 2016), *spam* (Facebook, 2014a, n.d.c; Twitter, 2020b), *artificial amplification* (Atlantic Council, 2019b, p. 3), *fake engagement* (Facebook, 2020f), *cloaking* (Facebook, 2017a), *spoof domains* (Facebook, 2016) and *click-baits* (Facebook, 2014b). Solutions now turned to platform-centred remedies such as *community reporting* (TikTok, 2019a), *flagging* (YouTube, 2006), and *content moderation* practices (TikTok, 2019c). In time, the hostile actors were also recognised as foreign, connecting to a discussion about practices at war as *information, hybrid, asymmetric* or *psychological warfare* or *threats* (extensively covered in i.e., NATO StratCom COE, 2019; Swedish Defence University, CATS, & Hybrid CoE, 2018). These reports' proposition for how to counter such actions is both by accentuating military defence, but also by building *deterrence* against foreign actors and *resilience* among one's own population. This links to the political discourse, often in criticism towards *alternative* or *alt-right* politics, not seldom described in the discourse by mentioning right-wing extremism and white supremacy. It also connects to discussions about *populism,* a focus on domestic actors and a concern for foreign actors' behaviour in concepts such as *election integrity* (government actions outlined in texts such as i.e., Danish Ministry of Defence, 2018; Government of Ireland, 2019; Swedish Civil Contingency Agency, 2018a),[96] through 'deliberate attempts to mislead voters' by *foreign powers*, also discussed by platforms in their communication about what behaviour to report in order to protect the integrity of democratic elections (Twitter, 2019). Some platforms divide *inauthentic behaviour* into two forms, either *coordinated inauthentic behaviour,* which cover 'domestic, non-government campaigns that include groups of accounts and webpages seeking to mislead people about who they are and what they are doing while relying on fake accounts' while *foreign or government interference* is 'conducted on behalf of a government entity or by a foreign actor' (Facebook, 2020b). In other definitions the latter is acknowledged as 'foreign-led efforts to manipulate the public debate in another country' or as 'operations run by a government to target its own citizens' (Facebook, 2019b).[97] At this intersection between a domestic and foreign domain, we also find a discursive concern for foreign actors' use of *domestic proxies* or *useful idiots* (borrowed terms from a military setting) or *multipliers* (borrowed from a marketing setting) to carry out such interference—either described as knowingly enabling the actions of a foreign state, or being deceived by the very same to spread narratives in line with a foreign actors' interests. The solution, once again, is framed as strengthening the very system that is under threat through efforts of increasing public influence and building public *resilience*. The final theme emerging from the discourse's concepts of moral and immoral communication is tied to journalism—also accentuated in Habermas' own concerns for the new landscape (2022). In the new communication landscape, new concepts initially formed from a new form of journalism practiced by the public—reflected in concepts such as *citizen*, *public*, *participatory*, *democratic*, *guerrilla*, and *street journalism* (see e.g., the previous presentation of Mozilla's early initiative on p. 111). This is tied to a positive valence, aligning with the initial cheer for the new platform's positive effects on an informed public. In time, however, the theme of journalistic concepts rather turned to criticism towards *fake news* and *fake news media*, aligning with concepts like *yellow journalism* and *junk news*, by the authors' described as a new non-journalistic practice emerging in the new landscape (see i.e., Reuters Institute, 2017; UNESCO, 2018). The new landscape with new media also led to the need to define traditional journalism, in the discourse framed for example as *traditional media*, *journalistic media*, or *legacy media*, often contrasted against descriptions of mediums such as *alternative media* and *social media,* which content ought to be sceptically approached as not gatekept by professional journalists. To deal with the problem of false stories spreading on these new outlets, journalists are described to take on a new role of *debunking* in the practice of *fact-checking*. In time, the platforms paired up with journalists in common initiatives—see i.e., Google (2017a, 2020), Facebook (2019a) and Pinterest (2019).[98] Adding to that, the discourse also argues that we can combat fake news by teaching the public that journalistic and other democratic sources are sources of truth. This is formulated in educational programs designed to form a more *media savvy* population, by teaching

---

[96] This also includes handbooks, published by government bodies, that are not tied to elections specifically but concern disinformation in general, such as by UK authorities (UK Government Communication Services, 2019, 2021) and Indian authorities (Indian National Cyber Research & Innovation Centre, 2020).

[97] Other platforms also addressing state-sponsored activity such as Google (2018).

---

[98] Here, some point to a distinction as 'journalists cannot leave it to fact-checking organisations to do the journalistic work of verifying questionable claims that are presented by sources' (UNESCO, 2018, p. 11), as well as distinction between fact-check*ing* and fact-check*ers* where the former is a common practice *within* journalism, but the latter is a new form of organisations and a 'fifth estate' of external actors, such as bloggers who review the journalistic work.

media and information literacy (MIL)[99] and *digital literacy* in addition to a promotion of *source criticism* and *trusted sources*.[100]

Each of these themes emphasise different elements of communication in their moral assessment. On a general level, journalists often focus on the message and medium, describing how the moral and immoral is contained in what is said but also whether the medium is described as legitimate or illegitimate. In the warfare discourse, as well as in the political discourse, the focus is instead on the environment—defining the phenomenon in relation to whether it appears during war or elections, and whether it is a matter of foreign or domestic actors. While both define the phenomenon in relation to intentions and consequences, the war discourse has an emphasis on the intentions of the enemy while the political discourse is more concerned about the consequences for the public political debate. Platforms, by contrast, often emphasise the communicative behaviour as deceptive, not focusing on the message, sender, or consequences. This means that while all align in the defence of democracy and its system for public discourse as theoretically reflected in Habermas' discourse ethics, actors emphasise various forms of these communicative immoral phenomena and moral counter-measures. Each organisation having different foci, aligning with their organisational type and emphasising elements of relevance in relation to their organisations' purpose and resources, while at the same time connected by returning to the same normative standard for public communication.

---

[99] Defined in one text as a practice that 'includes all technical, cognitive, social, civic and creative capacities that allow a citizen to access, have a critical understanding of the media and interact with it' (Open Society Institute, 2018, p. 14) and in another how 'media literacy education offers a potential tool to curb Truth Decay, defined as the diminishing role that facts, data, and analysis play in today's political and civil discourse' (RAND, 2019, p. iii).

[100] Platforms such as Facebook also align with this narrative, using a survey where they approached 'a diverse and representative sample of people using Facebook across the US to gauge their familiarity with, and trust in, various different sources of news' and then used this information to rank news in the algorithm towards *trusted sources* (Facebook, 2018d).

A strategic moral system

While Habermas argued that his validity claims is used together in the assessment of communication (see p. 34), this thesis proposes that such argument builds a false simple picture of the evaluation of communication. By evaluating communication in relation to each of these three values, Habermas neglects how actors' different combinations lead to moral or immoral ideal types, in the vocabulary of Weber and Hartman (see p. 92). When combined with other values, even untruthful communication can form a moral ideal type, sincere communication can result in an immoral ideal type, and righteous moral communication can indeed break with social norms as long as it seeks the right interests (see Habermas' three validity claims on p. 34)—in each of these cases, we see how Habermas' simple assessment that one validity claim is tied to a moral acceptance or immoral rejection is challenged when acknowledging the effect of combinations of values into ideal types.

Three key values are found to connect all normative definitions of communication in the fake news discourse. The first is the discourse's emphasis on the communicator's intention. This links back to both the first element in the process of communication (see p. 61) and Habermas' emphasis on truth as his first validity claim (see p. 34). This would tie the evaluation of a message in relation to an objective world, as in Habermas' framework (see p. 34). In a sense, we could imagine that this connects to an epistemological interest in a message as *true* if ethical and *false* if unethical, but also in depiction of ethical mediums as *fact-checkers* and in discussion on debunking *made up stories* with *facts* and *factual correct texts*. As one text describes, 'contemporary discussions about fake news and truth are rooted in longstanding philosophical debates about how we assess truth-claims' (NATO StratCom COE & King's Centre for Strategic Communications, 2018, p. 23), and continues to emphasise the correspondence theory of truth (see p. 60). Continuously, the discourse draws upon values such as *facts*, *evidence*, and *truth* to represent the moral forms of communication, while *lies* and *hoaxes* are but a few examples of values with an immoral valence. The second emphasis is on intentions. This echoes Habermas' emphasis on the importance of being truthful, a validity claim aligning with the subjective world (see p. 34). However, in the fake news debate, the discourse often mixes up the concept of intentions and consequences, just as in definitions of strategic communication (see Chapter 3). Hence, they are approached in the analysis as a common point of departure, in the assessment as intent and consequence are collapsed into a single value. Both values are frequently occurring in definitions of moral and immoral communication in the discourse, by describing the speaker as driven by a *genuine* or *malicious* intent, and seeking either *constructive* or *harmful* consequences. Thirdly, the discourse conveys a shared interpretation of Habermas' validity claim of rightness (see p. 34) in a socially shared claim to validity. Aside from the two previously shared norms that one ought to speak the truth, a frequent point of reference in the normative description of communication is whether the communication is *overt* or *covert*. This spans from critical discussions about actors operating under false pretences, such as through *fake* accounts or by using proxies such as *trolls*—sometimes in a coordinated sense described as overt operations using *troll farms, bots,* or other forms of *inauthentic coordinated behaviour*. It also covers the way messages are *amplified,* manipulating the very appearance of the public discourse—i.e., to make it seem like there is an opposition to a claim raised, or by cheering on some actors to create the impression of public approval.

Together, this creates a system with three dimensions of interest in the moral evaluation, to each finding a negative or positive moral valence as either the content is *true* or *false*, with a *constructive* or *harmful* influence (either consequences and/or intentions, as defined on p. 69), and either in an *overt* or *covert* form. This rests on Habermas' theory of validity claims as enabling the actors to present a moral or immoral valence to each claim, as 'the yes or no position taken by the hearer whereby he accepts or rejects the claim to validity that has been raised by the speaker' (p. 54). Based on these dimensions and values, we can form an axiological system in a model outlining the ideal types grounding the normative discourse. However, as Weber emphasises, these are not empirical but *theoretical* phenomena. As the fake news discourse is here found to revolve around three values, each either assessed as ethical or unethical valence in a binary, the result is a total of eight possible combinations of values (3^2=8) and eight *ideal types* (see p. 92).

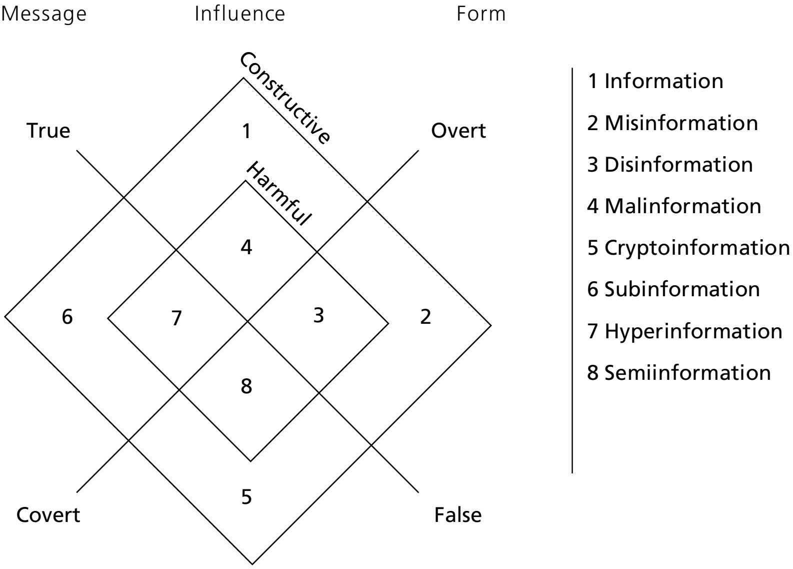

Message
Influence
Form
Constructive
Harmful
True
Overt
Covert
False
1
2
3
4
5
6
7
8
1 Information
2 Misinformation
3 Disinformation
4 Malinformation
5 Cryptoinformation
6 Subinformation
7 Hyperinformation
8 Semiinformation

Figure 1 The moral zeitgeist as a value system
Model of the ideal types based on the three validity claims raised in the fake news discourse.

Among the ideal types created, the first four stem from a set of concepts appearing in the fake news debate; *information, misinformation, disinformation,* and *malinformation.* Information and disinformation are the most frequently used terms of these four, followed by misinformation and finally malinformation. Disinformation and misinformation are frequently used and defined. Information is not often defined in the discourse, and malinformation is rarely used, however, they both help complete the relevant sets of values as the first four ideal types in which the communication is overt. While information is presented as true, disinformation and misinformation are most often defined as false information.[101] The truth is captured in phenomena such as *fact-checkers* and *debunking.* The difference between disinformation and misinformation would be that the former is intentionally deceiving, while the latter is spread without knowing that the communication is false. This leads to *conspiracy theories* as examples of misinformation, while *influence operations* by foreign actors would classify as cases of disinformation. In particular, misinformation and disinformation are frequently repeated in the discourse with similar if not identical definitions, later texts recognising them as part of an international standard-vocabulary (adopted by i.e., the UK House of Parliament (2017)). Malinformation is not used as frequently in the debate—defined as true communication spread with the intention to deceive (Council of Europe, 2017)—but helps complete the ideal depiction of the discourse's normative assessment of communication. Malinformation can be found in the broad definition of *alternative facts* as being not false information per se, but as deceptive when framed to a pre-determined goal (see p. 151). This can of course be used to assess not only linguistic expressions of communication, but also in relation to the various types of communication such as images, videos, actions, and non-communication (see p. 59).[102]

---

[101] Sometimes, however, definitions are not clear on whether disinformation covers all forms of false information, all forms of misleading information, or whether it necessitates that the communication is false and misleading. See for example, the European Commission (2018) as how they disinformation as 'verifiably false or misleading information' (p. 1).

This leaves us with four ideal types of communication. While the formerly presented concepts are tied to overt communication, these latter four describe covert practices. While often referring to such a distinction, the discourse does not present concepts that can help thematise this set of communication. For that reason, this thesis suggests four new labels, drawing etymological inspiration from other types of Latin and Greek prefixes (as in mis-, dis- and mal- above).[103] These four ideal types consist again of more-or-less moral or immoral forms of communication. For example, *subinformation* indicates that the practice is covert, spreading a true message with constructive intent. This includes, for example, government actors or public authorities' funding other actors' activities for communicating true and constructively intended messages. It also includes the platforms' up-ranking of actors who are deemed moral in users' newsfeeds—such as how Twitter describes amplifying over 20 non-profit partners' local campaigns on Twitter, or other algorithmic support such as to the *International Association for Suicide Prevention*, where the platform would add an orange ribbon (the cause's symbol) to anyone who tweeted under a set of hashtags to promote the content (Twitter, 2020). *Hyperinformation* is covertly spread true yet harmful communication, and include cases such as for example the use of *sockpuppets*—where actors create false accounts on social media that enter pre-existing discourses and amplify the narrative on either side of a discourse. In some of these cases the messages spread can be seen as false, but they can also be true statements if reviewed in each of these bubbles. However, they are spread by a hidden actor with a harmful intent such as aiming to polarise the discourse further as 'the practice of masking the sponsors of a message or organisation to make it appear as though it originates from grass-roots participants' (Wilfred Martens Centre for European Studies, 2018, p. 6). On the other hand, covert messages can also be untruthful. To this, we add the notion of *semiinformation* which is for example presented as a tactic in foreign election interference, where foreign actors insert false information through domestic proxies with the intent of harming said society (see p. 146). This practice is frequently discussed, while the final ideal type *cryptoinformation* is less so. But there are examples of this final ideal type. A widely discussed example is found in France, in a counter-measure against the anticipated hacking attack against Emanuel Macron's election campaign in the 2017 French presidential elections. The case—referred to as *The Macron Leaks*—describes how actors targeted the campaign's computer systems, but 'these targeted actions against presidential candidate Emmanuel Macron neither succeeded in interfering with the election nor in antagonizing French society' (French CAPS & IRSAM, 2018, p. 106). Nonetheless, two days before the election, the campaign team was hacked. Following a longer series of attacks from both Russia and Macron's political opponent in France, this was an anticipated attack, as the culmination of a larger operation. The success in countering these efforts is attributed in the texts to a number of factors, one of which is how Macron's team engaged in *cyber blurring* (alternatively *digital blurring*), which in this thesis would represent a case of cryptoinformation.

---

[102] In the texts, information is used both to talk about the content in, as well as the practice of, communication, while disinformation as a concept is often tied to the description of actors and misinformation is particularly used in with the rise of the COVID-19 pandemic.

[103] *Misinformation* from the Greek prefix *mis* meaning bad, badly, wrong, wrongly, as in *mis*behave or *mis*print. *Disinformation* from the Latin prefix *dis* meaning apart, away, not, to the opposite, as in *dis*like and *dis*belief. *Malinformation* from the Latin prefix *mal* meaning bad, ill, wrong, as in *mal*content and *mal*icious. *Hyperinformation* from the Latin and Greek prefix *hyper* meaning too much or excessive, as in *hyper*active or *hyper*bole. *Interinformation* from the Latin prefix *inter* meaning between, as in *inter*section and *inter*rupt. *Cryptoinformation* from the Greek prefix *crypto* meaning hidden or secret, as in *crypt*ic or *crypt*ography and to en*crypt*.

> As the hacks could not be avoided, the En Marche! team placed several traps, including fake email addresses, fake passwords, and fake documents. This diversionary tactic, which involves the creation of fake documents to confuse attackers with irrelevant and even deliberately ludicrous information, is called cyber or digital blurring. Thanks to this tactic, the Macron campaign staff did not have to justify potentially compromising information contained in the Macron Leaks; rather, the hackers had to justify why they stole and leaked information which seemed, at best, useless and, at worst, false or misleading (French CAPS & IRSAM, 2018, p. 114)

By inserting false information into their own systems in anticipation of an external attack—the aim of such covert communication was indeed to be constructive, while using false information in an overt manner.

Beyond working as a framework to organise the wide range of phenomena appearing in the fake news discourse—providing a language to talk about key values and elements of interest in the normative assessment of communication—the schematic presentation also challenges Habermas' idea that his validity claims point to values as moral or immoral communication by default. Using this framework, actors can strategically engage in a practice of accentuating some and downplaying other values in the assessment of communication as described by Weber (see p. 92). The normative system thus moves from being a tool to arrive at a rational assessment for moral evaluation, to become a tool enabling rationalisations of one's moral intuitions.

Chapter 9.

Post-truth

A final overarching concern in the fake news discourse is the observation of an emerging *post-truth* environment, described as both a cause and consequence of the fake news phenomena. In this concept, we move from a question of who and what to a more abstract discussion about the communicative context in which these interactions play out—a criticism against a new form of reasoning breaking with Habermas' sense of rational conclusions. In the new landscape, actors do not seem to be resting their idea of rationality on the Enlightened ideal in Habermas' discourse ethics, but accept post-rationalised truths. Yet again, through the immanent methodology, we see in this chapter how this criticised behaviour also extends to the authors' own behaviour, explained through moral psychology's third emphasis on morals as rationalised intuitions.

The concern for post-truth breaks with Habermas' and the fake news discourse's idea that the rational is found in consensus arising from the evaluation of valid claims. This new concept is used in combination with alternative facts and echo chambers, describing how actors use alternative facts to present rationalised arguments that support their pre-existing convictions. These actors are described to reject valid claims, rejecting every conclusion that breaks with their intuitions. At the same time, this chapter argues that the authors criticising this behaviour engage in the same strategic moral communication themselves. In their evaluation of cases of communication, they indeed rest on the system of values (as presented in Chapter 7). However, as this chapter will show, they use these values strategically in their evaluation so that each case is judged not based on technical values (tied to an evaluation in relation to the elements in communication) but based on whether the case is aligning with their own, rather than the others, moral norms (as presented in Chapter 8). The chapter illustrates this post-rationalisation process through an example that is salient in the fake news discourse, the very claim that Russia was responsible for the election of Donald J Trump. This is presented as an explicit and implicit statement in the discourse. The chapter, however, shows how it can be seen as a rationalised conclusion supporting the authors' moral intuitions. Trump is continuously presented in the discourse as an irrational choice in the Presidential elections. The authors criticise his public appearances and attitude towards for example journalists and public institutions, continuously rejecting the idea that he would rally enough support to gain office. Nonetheless, he was elected President. Consequentially, actors searched for an explanation to make sense of this inexplicable outcome. The fake news discourse frequently describe examples of foreign election interference from Russia, presented as an explanation for why Trump won. They repeatedly draw attention to Russia's actions, pointing to inquiries made by US authorities. However, while these studies confirm that Russian actors engaged in such activities, the same reports and other studies still cannot claim that these activities determined the outcome of the election. The authors, nonetheless, repeatedly present this as a rational conclusion, presented as a rationalised explanation for why Trump won. As the thesis will argue, this provides a final immanent example of the morals emerging at the intersection between norms and behaviour, an example pointing to practices of strategic moral communication in the moral discourse which ties back to the previous two foci presented in the former two chapters.

The chapter expands upon these points, starting with a presentation of the concept post-truth and how it breaks with Habermas' and the fake news discourse's ideal. The chapter continues by presenting the post-truth phenomenon as linked to conspiracy theories, ending with a final immanent review of the fake news discourse by arguing that the claim that Russia was the cause of the election of Donald J Trump presents a rationalised conclusion.

Post-truth conspiracies

Habermas' argues that we ought to arrive at an idea of what is rational after having evaluating claims to validity (see p. 34). While the previously presented discussion about alternative facts points to claims being criticised by the discourse as non-valid, the concept *post-truth* draws attention to the very process of reasoning. While echo chambers draw attention to the algorithmic landscape, alternative facts to the invalid claims raised, post-truth is rather pointing to a more intangible scenery. Sometimes, the texts refer to it more specifically as a *post-truth era, phase, world,* or *age,* while others describe it as a *trend* or *movement.* The president of the Oxford Dictionaries is quoted by authors to describe that 'post-truth as a concept has been finding a linguistic footing for some time … fuelled by the rise of social media as a news source and growing distrust of facts offered by the establishment', where post-truth is described as 'what happens when a society relaxes its defence of the values that underpin its cohesion, order and progress: the values of veracity, honesty and accountability' (RSIS, 2018, p. 26). In short, this is in stark contrast to Habermas' normative ideal of a rational discourse guided by his validity claims. Many texts return to the Oxford Dictionary's definition of the concept after announcing post-truth as the word of the year in 2016 as an 'adjective relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief' (Ash Center for Democratic Governance and Innovation & Harvard Kennedy School, 2018, p. 5). Others elaborate that…

> …post-truth signifies a communication paradigm in the 21st century in which 'I think, therefore, I exist' is replaced by 'I believe, so I am right', i.e., in which objective facts have less influence on shaping public opinion than emotions and personal beliefs. Post-truth refers to such a media and societal system in which the public interest is placed behind the particular interests of the elite and in which media manipulation is almost legitimate method of coming to power and staying in power. The post-truth society is not only a society in which truth is not a priority; it has almost become its contradiction—a society in which the truth is undesirable, unprofitable, and irrelevant. (Heinrich Böll Foundation, 2020, p. 19)

Contrasted against Habermas' normative idea of a rational discourse, the fake news discourse criticise the post-truth scenery to be the home of irrational discourse. The very label of something being rational or irrational is never explicitly defined in the discourse—it is rather used as an axiomatic marker of a moral or immoral valence which one attributes in the assessment. To express that a source is irrational becomes a label for marking it as wrong, unethical, bad, harmful, etc., without having to specify what makes it irrational. As one text describes in reference to Rorty, 'different conceptions of human flourishing make different vocabularies more or less useful to us' where 'if a new vocabulary is judged to be closer to our ideal of human life, then adopting it will be judged to be rational; if not, then adopting it will be deemed irrational' (Ash Center for Democratic Governance and Innovation & Harvard Kennedy School, 2018, p. 21). It becomes a very different strategically communicative instrument to signal what is moral or immoral—contrary to the previous presentation of values tied to elements in the process of communication (see p. 61). In relation to values, actors describe what makes phenomena moral or immoral—that e.g., trolls are immoral as they engage with insincere intent and cause harm. Such explanations are not found when reviewing the texts' attribution of labelling some communication, sources, actors, behaviour, authorities, agreements, beliefs, etc., with the immoral valence of being irrational.

As post-truth is tied to the irrational, it is also tied to emotions rather than facts. Repeatedly, definitions of post-truth is presented in its relation to alternative facts, and the observance that truth is no longer about rational reasons but is 'subject to personal belief rather than evidence' (Global Health Security Network, 2020, p. 4). Whatever 'feels right' in this landscape is potentially accepted as such, disregarding the facts. This new environment is therefore described as irrational, and its supporters by extension too. The use of emotions is repeatedly framed as a tactic by malicious actors, targeting emotions to unleash an irrational response, as 'it will try to find those issues in our societies that garner most emotions around them, and it will try to fuel and amplify these emotions as far as possible—because an audience shaken by strong emotions will behave more irrationally and will be easier to manipulate' (EUvsDisinformation, 2018, p. 7). This irrational form of reasoning is the basis for people's beliefs in alternative facts, for while not factually correct they feel right. In a way, we see how the emphasis on emotional reasons in the definitions of post-truth, rather than cognitive reasons, aligns with the dichotomy in moral psychology between morals as intuitive gut-feelings and not as the consequence of rational arguments (see presentation starting on p. 51).

Rational discourse is, by contrast, supported by facts. In line with an Enlightened ideal, the truth resides in facts about an objective reality (the objective world, see p. 34). As one text describes, post-truth 'and populism emphasise the local, personal truth above the rational, scientific, or academic truth' (NATO StratCom COE & King's Centre for Strategic Communications, 2018, p. 10). By contrast, the discourse is repeatedly critical against emotions as guides for reasoning. Instead, emotions are being presented as tied to emotional manipulation and propaganda. As one text describes, 'we might understand "post-truth" politics as privileging an emotional and local approach to knowledge and to justifying truth claims' (p. 11). Another text continues that 'if citizens are responding to information emotionally rather than by rational reasoning, it is more likely that they will not elect those who will represent them in their best interests, but rather those who are better manipulators' (Heinrich Böll Foundation, 2020, p. 21). From these discussions, the discourse moves from a depiction of the irrational led by emotions, to a discussion about the broader consequences for the society—as building the post-truth environment. This new scenery is presented as a threat towards the ideal rational society, rather finding an environment where people are persuaded by irrational alternative facts as they engage in discourses within groups that echo their pre-existing beliefs.

The problem arising in this post-truth environment is in its extreme form referred to as *conspiracy theories*. Conspiracy theories are defined by several characteristics, some describing these actors as drawing upon a *crippled epistemology* (French CAPS & IRSAM, 2018, p. 12). As previously presented, some authors engage in a practice of othering when presenting these conspiracy theorist (see p. 135). Foremost, however, this is done by pointing to the content spread rather than the character of the individuals spreading them. That is, authors pointing to the fact that the people believing these theories are simply misled as they do not have access to good sources of information.

> Most conspiracy theorists are neither foolish nor irrational but simply lack good sources of information. The theories they defend are unjustified in light

> of all the information available, but not according to the sources they consult, which make these arguments seem plausible. The cause of the problem is the epistemic poverty of their environments. (French CAPS & IRSAM, 2018, p. 35)

In academic writing, conspiracy theories are defined by a set of conceptual characteristics; key components covering that a) it assumes hidden powers b) acting as puppeteers to pull the actor's strings behind the scenes. These are covert actions where 'conspiracy theories hold that certain people have disproportionate power with which to conceal their actions. These attempts can become absorbed into the narrative of the plot' (French CAPS & IRSAM, 2018, p. 35). Others describe how 'conspiracy theories are driven by a belief in the machinations of a powerful group of people who have managed to conceal their role in an event or situation' (Data&Society, 2017, p. 18). However, maybe there is a deeper root to our human tendency to engage in conspiratorial theorising. As moral psychology has shown, when confronted with an illogical reality we are able to post-construct stories that make sense out of the senseless (see p. 54). We find reason in rational explanations that are at least rational in our mind, influenced by cognitive biases that enables us to build rational stories on a non-rational basis—such as texts growing in their evidential weight not on the basis of their validity claims, but on the mere frequency of previous repetition and the conformity with our previous beliefs (see p. 51). We end up using communication as a tool to make sense of how events are connected, describe correlations strategically as a case of causality, and bend even our own rules of reason to rationalise such moral intuitions.

Did Russia win the US election?

In the final immanent review, the thesis would argue that the fake news discourse itself engages in the very practice of constructing post-truths. Indeed, as previously presented, it appears as rationalisations of why some actors are moral or immoral (Chapter 7) and why some cases of communication are moral or immoral (Chapter 8). The strongest example of this behaviour of post-rationalisations of one's moral intuitions comes in the very example repeatedly described in the discourse—the election of Donald J Trump in the 2016 election race.

The US election of 2016 ended up as a race between Hillary Clinton and Donald J Trump. Experts almost unanimously described how the only rational outcome would be that Clinton would win the election. This was partially deduced by the mere rational logics of no man like Trump being able to act presidential, and so the US population would see this and cast their vote in another direction. Even before Trump entered the race, experts from public relations had painted him out as the most unlikely of candidates to even run for office, because of his character. In reference to the events of Trump questioning the then elected President Obama's birthplace,[104] these scholars described how…

> Individuals who wish to manipulate the public sphere to their own benefit without bringing ideas and arguments that contribute to the public debate will be rejected, … Trump has no elected status and has never run for office and the way he used his public profile to inset himself into a political debate was met with ridicule. Since … Trump has been quite absent in the public sphere. This is an example of how arguments in the public sphere are rejected when they are not authentic or do not contribute to the improvement of society. (Holtzhausen & Zerfaß, 2014, p. 6)

Despite the rational conclusion that Trump would never run for public office, he did. He even won. Following the elections, the actors who thought the selection of this candidate as utterly senseless tried to make sense out of the results. Many pointed to Trump's communication style as effective, even if it was immoral. Texts also point to the immoral campaigning activities by Trump's campaign, such as the use of *dark ads* (Centre for the Study of Democratic Institutions, 2018, p. 20). In parallel, however, other theories started to emerge, such as whether he had been supported in his campaign by foreign actors. Soon after the election, several state-issued reports were released, presenting the results from investigations on Russian actors' attempts to interfere in the election,[105] and the suspected Russian interference's links to the Trump campaign.[106] These reports confirmed that Russian actors had engaged in activities that sought to influence the election. In particular, authors repeatedly describe these reports' presentation of Russian activities coordinated through *The Internet Research Agency* (such as i.e., in French CAPS & IRSAM, 2018, starting on p. 84),[107] a troll farm located in Saint Petersburg. The description of these activities is presented in a unified way in the discourse, all aligning in representations such as…

---

[104] A recurring example among the texts, labelled as the *birther movement,* questioning whether Obama was born in the US and thus eligible for the position of President. To this claim, the White House responded by releasing the President's birth certificate in 2011.

> Special Counsel Robert Mueller's February 16 indictment of the Internet Research Agency, along with 13 key individuals and two related entities, was another reminder for Americans of the coordinated attack mounted on our country's democracy. In addition to the use of information operations outlined in the indictment, the U.S. intelligence community has documented Russia's use of hacks on campaign-related entities, with stolen information weaponized and released via Wikileaks and other outlets, and the Department of Homeland Security (DHS) has confirmed that at least 21 states had parts of their electoral infrastructure probed in the run-up to the 2016 presidential election. The Internet Research Agency purchased political advertisements and sponsored organic content on social media platforms aimed at amplifying socially divisive

---

[105] Foreign interference is not only presented as conducted by Russia, but is the most frequently mentioned actor in the discourse. For example, in Facebook's continuous takedowns of foreign or government interference, Russia is the most frequently occurring—but also includes other countries such as China, Iran, Nigeria, Indonesia, Egypt, Spain, Saudi Arabia, the United Arab Emirates, Georgia, Vietnam, and the Philippines, to name a few. Aside from the US election of 2016, Russian actions is also frequently presented in relation to activities around Europe, with a particular focus on activities targeting Ukraine (one example being the report by Graphica, 2020).

[106] Both a report by the United States Office of the Director of National Intelligence (US Intelligence Community Assessement, 2017) and a two-part report concluding the special counsel Robert Muller's investigations, commonly referred to as the *Muller report* (Special Counsel Robert S. Mueller, 2019a, 2019b), officially titled *Report on the investigation into Russian interference in the 2016 Presidental election*.

[107] Finally resulting in one of the first big take-downs of *coordinated inauthentic behaviour* by Facebook, as described by the platform itself in publications such as *Authenticity Matters: The IRA Has No Place on Facebook* (Facebook, 2018b) and *Continuing Transparency on Russian Activity* (Facebook, 2017c). The events are similarly presented by other platforms such as Google (2017b, 2017c, 2017d, 2017e) in many reports from institutes such as by Graphica (2019) and the Oxford Internet Institute (2018).

> issues. This content reached over a hundred million Americans, and the trend of exacerbating social divisions has continued on Twitter, in particular, over the past year and a half. (German Marshall Fond, 2018, p. 1)

As to these activities' link to the Trump campaign, the Mueller Report describes how 'while this report does not conclude that the President committed a crime, it also does not exonerate him' (Special Counsel Robert S. Mueller, 2019a). The report did outline activities from Russian actors, however, cannot claim to hold answers to the effects or whether these effects swayed the results. Nonetheless, in the texts this is a frequently repeated mantra—if not outright calling out that Russia influenced the US election, then at least repeating that 'information operations surrounding elections are a source of global concern, particularly following the US 2016 election, in which Russian interference has been alleged to be a significant reason why Trump won' (DEMOS, 2019, p. 32). The narrative around Russia's influence seems to have triggered a repetition bias as we tend to believe things that we have been told repeatedly and by a variety of sources.

Hand in hand, this narrative is accompanied by another where alt-right digital platforms also influenced the election. However, once again we see how scholars (referenced in the texts) point to the evidence of such influence being scarce and rather that…

> There is still a great deal of ambiguity around the outcome of the election and why the majority of experts and predictions did not foresee Trump's win. Given this confusion, the alt-right was happy to take credit. However, determining the lines of critical influence is too complicated to conclude so tidily. (Data&Society, 2017, p. 3)

What emerges is a discourse that depicts how Russia's influence explains why the public voted for such an irrational politician as Trump. It makes sense to those who disagree with his politics. Repeatedly, the discourse is more or less explicit in normatively positioning the text in relation to a depiction of Trump's and his politics as irrational.[108] This irrationality is derived from an interpretation where his position breaks with the normative ideal, by attacking journalists and discrediting the integrity of the electoral system. Trump is not the only but, without competition, the most commonly used example of this international trend. For example, as post-truth was introduced in 2016, texts describe how this was 'the year we allegedly left the world of rational argument and objective facts and entered a world of "bullshit" and lies' and continues that 'the election of President Donald Trump and the success of the Brexit campaign are often cited as evidence' of this (NATO StratCom COE & King's Centre for Strategic Communications, 2018, p. 7).

---

[108] Even Habermas takes Trump as an example in his text from 2022.

At the same time, the discourse does not simply acknowledge the possibility that Trump and the people voting for him represent a new critical voice in the public discourse, critical indeed against the hegemonic powers and possibly also their norms. Instead, authors are repeatedly seeking explanations for such actors' irrational behaviour, and from that hope to find cures that will make them come to their senses. What is missing from the discourse's depiction of the political debate is a recognition that some of these voices may be expressing a sincere public opinion. Even if that is not the case for all voters (as some may indeed have been swayed by foreign interference) we must at least assume that many simply voted for the person that they thought would best represent them. The fact that some are critical towards the norm may simply be just that, an expression of a critical public. For when reviewing the fake news discourse' depiction of the US elections, there are few mentions of the US public who did vote for Trump while not under the influence of foreign powers or deceptive rhetoric. In parallel to these descriptions of Russian influence, another voice is almost as prevalent in the fake news discourse—the problem of the public's growing mistrust towards the system of democracies, the lack of trust towards elected officials, experts, journalists, and sources of authority in the democratic system. The public is never singled out as the ones responsible for this development, but are rather victimised. For they are described as not falling into this irrational reasoning by their own fault, but are the victim of their own cognitive biases which mislead them into being influenced by Russian election interference, leading them to mistrust established journalism, trapped in algorithmic structures on new platforms which subconsciously drag them further down a rabbit hole occupied by extreme opinions and non-factual claims, where they entertain irrational stories rather than morally rational news.

Rationalising moral intuitions

As a final step, we see how the very concept of post-truth as referring to an irrational environment deepens the above-presented paradox in relation to the discourse' depiction of rational or irrational societal structures. On the one hand, texts describe how post-truth environments vary in severity across the globe and find that some societies have a stronger post-truth character than others. These authors describe how in post-truth societies, 'reliable information is available to privileged people, while ordinary citizens are exposed to fake news and alternative facts' (Heinrich Böll Foundation, 2020, p. 21). They continue to describe how these societies' educational system…

> …is not based on encouraging critical thinking, but rather on memorizing 'unquestionable truths' imposed by those who manipulate. However, the trend of societies being turned into post-truth societies is not only present in autocratic states and non-democratic societies. It is present in democratic societies as well, especially during the pre-election period when the production of alternative facts, fake news, propaganda, etc. begins. (NGO Internews Ukraine & Ukraine World, 2018, p. 31)

On the one hand, these authors describe how the ideal is found in societies where the public is practicing critical thinking. At the same time, however, (as the previous section has shown) the discourse does not acknowledge the criticism expressed by the public when electing a candidate such as Trump who openly criticise democratic actors. For in that case, such criticism is deemed irrational to the point where the public must have been persuaded by other arguments than facts—rather presented as manipulated by emotional populism using malign rhetoric, or alternatively by being influenced by foreign actors attempt to sow social division. While describing open criticism as a key character of their ideal system, this is dependent on whether the actors criticised are considered moral or immoral (divided into tribes as presented in Chapter 7) where one can point to these actors' behaviour as moral or immoral (resting on the values presented in Chapter 8).

In other texts, societies are described as characterised by a high or low post-truth character. They describe how 'in democratic societies …

argumentation is key and authority is based on the power of arguments and not vice versa', continuing to equate arguments with 'rational authority' (Heinrich Böll Foundation, 2020, p. 14).

The problem arises when we look closer at the very idea of what rational and rational is. As previously described, this is not defined in the fake news discourse, but is rather continuously presented in-between the lines as that which aligns with the discourse' norms. In that sense, the rational becomes an axiom, a pre-determined truth that we use as a stance to evaluate moral realities. It is the moral intuition driving our rationalisations. By contrast, the text continues that 'weak democracies value information based on its source' and describe this as an 'irrational authority' (Heinrich Böll Foundation, 2020, p. 14). At the same time, the fake news discourse is relying on a theory of reliable, established, and official sources as sources of truth. For example, in the repeated emphasis on teaching source criticism, media and information literacy, digital literacy etc. And it takes us back to the very start of the analysis where morals are tied to moral authorities present among actors defined as part of the self (see Chapter 7) who assess communication as moral or immoral from this point of view (see Chapter 8) to form their idea of rationality that simultaneously expresses and defends their moral convictions. For, on the one hand, the fake news discourse does emphasise that we should value information based on its source—such as valuing statements from democratic authorities, reflected in for example the platforms algorithmic up-ranking of information from selected sources (see p. 133). However, when presented in this way, these sources are not presented as an 'irrational authority' but rather as a rational authority. This, again leads us back to the question of the balancing act (see p. 136) that these societies have to engage in—on the one hand protecting an environment that is supposed to be free from influence over each individual lifeworld, but at the same time applying the normative framework as a system that seeks to maintain the integrity of such discourse.

This closes the loop, in three observable strategic moral communication here argued to manifest in the reviewed fake news discourse (presented in Chapter 7, 8, and 9 respectively). While advocating than we ought to build on a rational discourse where all voices are heard, collectively determining valid claims to arrive at a rational moral conclusion—the reverse behaviour is observed in the moral discourse under review. Instead, the determination of what is moral or immoral is based on one's alignment with these authors' moral intuitions. Any normative assessment made—whether it is about *who* is moral or immoral, *what* behaviour is moral or immoral, or *how* moral and immoral processes of arriving at a moral conclusion are applied—is based on the same pre-determined idea that these actors already constrict their normative assessment based on a pre-set normative framework. Those who reason outside of this axiomatic framework will be excluded, those who use alternative claims to validity will be seen as illegitimate, bad, false, or unethical. Any conclusion that falls outside these authors' moral intuitions will be found irrational. Together, this process does not align with the ideal of Habermas' discourse ethics of engaging in a rational discourse to determine what is moral, but contrary uses communication strategically to rationalise one's moral intuitions. That is, indeed confirming the empirical observations from moral psychology (see p. 51) that constitute the very core of the thesis' proposition for strategic moral communication.

Summary

This third and final part of the thesis has addressed the thesis' third research question *How does strategic moral communication manifest in the moral discourse on fake news?* Just as Habermas' discourse ethics, the fake news discourse is found to be discussing the morals of communication, more specifically communication appearing in the public sphere. It emerges in contrast to the previous enthusiasm towards the internet, finding that the new platforms for communication also inhibit the normative ideal. This included phenomena such as echo chambers breaking with Habermas' universal discourse, alternative facts pointing to claims gaining validity beyond Habermas' valid claims, caused by, and resulting in an environment breaking with the enlightened ideal as people's conclusions seem to be based on their pre-determined intuitions in a post-truth environment. In each of these three foci, the chapters have used the methodological framework of moral discourse analysis to illustrate the strategic moral communication practiced by authors in the fake news discourse. First, it showed how Habermas' ideal of an all-inclusive discourse is discursively framed as threatened by echo chambers. At the same time, the chapter showed how these authors group people into those part of the self and other, themselves even engaging in practices of othering. Continuing by looking at the concepts and definitions used to frame moral and immoral communication in the fake news discourse, the thesis has shown how alternative facts are framed to represent the immoral, as a threat towards valid claims. At the same time, the analysis showed how these normative values form a schematic presentation of how moral values can be used to determine combinations of values as ideal types. The final chapter showed how post-truth rationalisations are presented as the irrationally immoral, while the analysis also found that the actors also engage in the same behaviour in their condemnation of communicative phenomena in the fake news discourse. Together, these chapters have shown how the fake news discourse builds on Habermas' ideal, while paradoxically falling into the very practice criticised. This, as the final reflections will argue—as the behaviour observed by and engaged in by both the authors of and the actors discussed in the fake news discourse—aligns with the theory of moral strategic communication.

Final reflections

Chapter 10.

An emancipatory framework

Having presented answers to the three research questions (in Chapter 7, 8, and 9), this final chapter takes a step back to engage in a discussion about what the shift from a normative to an empirical perspective on strategic communications' relation to morals means in a broader sense. The analysis showed how actors in the fake news discourse advocate for a set of norms, while pointing to situations where they do not follow them. Similarly, this final chapter engages in an immanent review of Habermas' discourse ethics and of critical discourse analysis—showing how the pattern repeats in these theoretical and methodological propositions. The thesis would argue that they also can be seen as moral discourses that engage in strategic moral communication.

As the thesis has illustrated, Habermas' discourse ethics—and by extension critical discourse analysis—work from a set of philosophical assumptions. In Habermas' framework it is the assumption that moral communication is created by inviting everyone affected to a joint discourse, evaluating claims to validity, and finally arriving at a shared moral consensus through a rational process. However, just as we have seen how the fake news actors advocating the same values fall short in applying these moral standards when building the moral discourse around fake news, the same can be said about Habermas' discourse ethics—that it falls short based on its own internal theoretical validity. For while he advocates a framework for moral discourses where all is allowed entrance, this is at the same time conditioned—participants must agree with his basic understanding of key axioms about what morality is and how we arrive at it. Similarly, while advocating for validity claims as guiding the process, the analysis has shown how both the decision of which claims to use and how to assess communication in relation to them becomes a strategic process through that very same conditioned discourse. While advocating that his ideal is the last stop in a series of moral progressions (see p. 33), the thesis would argue that it rather reflects the norms of his own context, and that it is a WEIRD context at that (see p. 50). And just as he argues for the need of a rational conversation, one could argue that his own arguments for discourse ethics is a case of rationalised communication—an expression of strategic moral communication as inevitably all cases of normative ethics would be. The same immanent paradoxes emerge when reviewing critical discourse analysis. More specifically, the methodology advocates that we can use the method to emancipate society from hegemonic social structures. However, it locks its own followers into this normative perspective as it forms its own echo chamber—drawing up the boundaries for moral facts and immoral alternative facts respectively—which leads to deductive reasoning in which the ideal is rationalised in correspondence with a post-truth milieu. This becomes particularly visible in the theoretical, methodological, and empirical moral discourse' representation of the public—as those not agreeing with the ideal are branded as immoral actors, who upon breaking with the ideal will be excluded from the discourse. While the frameworks were proposed to protect the public's right to their own lifeworld and the right to influence the collective, what is protected in this normative discourse is also the democratic system. The question remaining, would then be how we balance these two interests in the future fake news discourse, and how we can seek a more constructive way for realising an ideal while this (per definition of being an ideal, see p. 40) will perhaps never manifest in its pure form.

This final chapter expands on these discussions, critically reviewing the immanent review leading to the identification of theoretical and methodological paradoxes in discourse ethics and critical discourse analysis. It argues that what these frameworks rather show is the confirmation of the theory of strategic moral communication proposed in this thesis—that as theoretical and methodological approaches they become tools for the type of moral rationalisations of gut-feelings that are emphasised by contemporary moral psychology.

Morals found in pre-determined consensus

The first immanent theoretical paradox emerges out of Habermas' emphasis on consensus (see p. 40). For while he presents consensus as the result of a conversation, the analysis of the fake news discourse rather shows how his normative axioms become a tool for excluding dissensus—thus constructing groups of moral consensuses before any discourse has started (see Chapter 7). However, this thesis would argue that it is not only an empirical practice in the fake news discourse, but is embedded in the very normative moral theory of communication presented by Habermas. As he constructs his theory, it is explicitly described as a model that fits the ideal of democratic societies. What he fails to recognise, however, is that such an ideal is not universal but captures the normative consensus of his time and place—a foundation for the moral zeitgeist in democracies. In any discourse relying on this normative framework, those not complying with its values will be excluded (as presented by Agamben, see p. 70). The problem is inherited in critical discourse analysis, as scholars here work from the same assumption that the normative framework is empirically universal, thus inevitably framing all who do not conform with its axioms and values to be immoral, constituting a problem to be exposed through their methodology.

Before starting the immanent review of Habermas' discourse ethics, it is important to recognise that the thesis' points are not limited to this theory. Rather, as a normative theory it is merely an example of how all normative theories practice strategic moral communication. Normative theories inevitably rest on axioms and values. Contrary to empirical theories resting on observations of how things *is*, normative theories rest on arguments for how things *ought* to be. Sure, moral arguments can rest on claims derived from empirical realities, but they can never point to something as true or false based on an external, universal, objective, criteria alone. The normative character of ought statements dictate that they rely on arguments that derive their validity from the phenomenology of the individual lifeworld (ethics) or shared lifeworld (morals). Such lifeworlds consist of experiences, and the moral arguments upon which normative claims rest. These are phenomenological constructions that seek to convince others, since they cannot simply point to the facts of the case.

Despite morals non-correspondence with an objective world alone, we frame normative axioms and values as empirical truths as a strategy in moral strategic communication. Normative claims rest on the actors' ability to be convincing that theirs is a justified position. This renders moral communication to become strategic by default, as it seeks to cause or indeed succeeds in practicing influence (resting on the thesis' definition of strategic communication, see p. 69). At the same time, we find that normative theories present axioms (such as (U) and (D)) and values (such as valid claims) as moral truths, to gain legitimacy for their normative stance. Indeed, it echoes Habermas' idea that moral communication takes the form of arguments (see p. 27), however, with the important addition that any normative *theory* thus also is engaging in this practice. Through the framework of strategic moral communication, we can start to see these patterns. In the shift from propositions of normative theories such as discourse ethics to the thesis' empirical proposition, we would rather be interested in the communicative constitution of the normative theories themselves—a case of metatheoretical ethics to speak the language of moral philosophy. From this new stance, axioms in normative frameworks can be seen to create the phenomenologically non-objectionable moral stances from which all is evaluated as rational or irrational. These are not subject for negotiations, thus making Habermas' discourse ethics a theoretical impossibility and forms a circular argument with a series of immanent paradoxes. Among these paradoxes we see how while discourse ethics dictates that all should be included in a common discourse, only those will be allowed entrance who agree with his theory's premise. The pattern is repeated in critical discourse analysis and the fake news discourse's exclusive assumptions. This rests on the acknowledgement of the paradox presented by Agamben (see p. 70) that any moral framework will create a basis for both inclusion and exclusion. Contrary to Habermas' theory would suggest, we will judge not only those who have agreed to our moral standards, but this will also extend to those who never agreed on the axiomatic premises in the first place, as these are simply branded as immoral (see Habermas paradox on p. 29).

What emerges is rather a picture where Habermas' discourse ethics, critical discourse analysis, and the fake news discourse fall into the behaviour that they themselves present as immoral. For by defining what and who is moral, they inevitably define immoral behaviour and actors.

Interestingly, while all present their own stance as promoting the voices of everyday people, acting to emancipate society from social structures that restricts the individual lifeworld into a pre-determined way of thinking—they themselves become that system restricting the lifeworld. When the fake news discourse emerges based on critical voices against the hegemonic power (as a criticism towards journalism, politicians, and other actors labelled in the discourse as the 'mainstream' actors) these critics are instead excluded from the public discourse, which was promised to be open to all, not least but perhaps most importantly critics. If the authors in the fake news discourse had followed their own norms, critical actors would also be represented in the discourse, either by including them as authors or by neutrally presenting the criticism. Instead, what is found in the analysis of the fake news discourse is a depiction of such critical actors as a threat to be combatted. Either they are described as misinformed, thus a problem but can simply be talked into reason by presenting the facts in the case. If they do not, however, this would in some cases be used as an argument for relabelling the same actors as threats. For we can, indeed, observe immoral communicative phenomena in the public sphere as actors actively seek to mislead the public. And based on the normative system reflected in both Habermas' theoretical ideal and in the ideals of democracy, this means that there are communicative phenomena that ought to be seen as problems to be dealt with. However, aside from this problem, the thesis also points to a problem emerging in the paradox when reviewing the behaviour of actors engaged in the fake news discourse—as their actions do not align with their norms.

This, the thesis argues, points to where Habermas went wrong in his interpretation of Kant. Habermas criticises Kant's categorical imperative for expanding beyond the community which came to the moral conclusion. But what if we instead interpret Kant's imperative as one where he rather emphasises morals as harmony between one's norms and behaviour? If we return to Kant's categorical imperative, this thesis argues that we can interpret his stance as 'do not act in a way that goes against your own norms.'[109] This would signal a view of morals very different from Habermas discourse ethics, not tying the concept to a collective's agreed consensus but to an individualistically intrinsic correspondence between one's norms and actions. This, the thesis argues, harmonise with moral psychology's presentation of morals. It all boils down to an immanent principle that works in relation to *moral hypocrisy* in psychology. Here, moral hypocrisy is broadly defined as ones 'motivation to appear moral yet, if possible, avoid the cost of actually being moral' (Batson, Thompson, & Chen, 2002, p. 1335). The thesis works from the assumption embedded in such definition that not acting authentically—in line with one's moral values—would be considered immoral, independently of which moral standards one breaks with. This would create an unethical—ethical continuum between hypocrisy—authenticity. From Habermas' idea of discourse ethics as a tool for a pre-existing community to arrive at a moral consensus, the shift toward immanent ethics would rid it from the idea of morals as communicatively constructed consensus, to a view where morals are simply when our communicative norms align with our communicative behaviour. In a way, this echoes Habermas' binding/bonding effect (see p. 25), but moves from tying this to a communicative entity (a group of people) to a recognition that it merely reflects the harmony between normative speech and action. The thesis argues that this perspective of morals as immanent also enables us to see past our own normative conception of morals—and in that find understanding for the other's perspective in, for example, the fake news discourse. It helps those who cannot understand why Donald J Trump was elected president to see that while we condemn his actions, his norms align with his behaviour. We can criticise whether we think those norms are right or wrong. But we can also seek an explanation as to why people voted for him—perhaps simply because they agree with his stance and see that he acts according to his own norms and is therefore moral from their point of view.

---

[109] In a way echoing Habermas' idea of the binding/bonding effect of morals (see p. 25) and in his description of the rule as phrased by Alexy that 'no speaker may contradict himself' (quoted in Habermas, 1999, p. 87).

Now, at this point I would not be surprised if the reader reads this as an argument for the validity of the actors criticised in the fake news discourse. To be clear, *it is not*. Nor is it a criticism against the values promoted in Habermas' discourse ethics, in critical discourse analysis, or in the empirical fake news discourse. Rather, it is an attempt to shine light on the normative representation of moral theories in the social sciences—where it argues that we have tied the 'right' or 'good' or 'true' in the use of these frameworks to our own particular (not a universal) sense of morals and ethics. What the thesis argues is that if we are to form an empirical rather than normative theory of morals, we need to start a conversation where we do not seek to pass judgement but seek understanding. This understanding, however, does not necessitate acceptance nor agreement. The suggestion for moral discourse analysis in this thesis is merely an attempt to present a methodology whereby we can review moral discourses to describe the strategic moral communication emerging there—without having to pass normative judgement based on our own moral norms.

A WEIRD perspective

The second observation concerns the values emphasised in these normative discourses. These values build on the axioms, but take the form of values with a moral or immoral valence—the values are not good or bad in themselves, as axioms, but point to points of interest which can either be assessed as moral or immoral. In Habermas' discourse ethics it comes in the form of three validity claims (see p. 34), in the fake news discourse forming three aspects of interest in the normative framing of communicative phenomena. They align as norms found in democracies, the normative systems promoting this particular (rather than a universal) moral theory of communication.

While Habermas describes discourse ethics as resting on a universal stance (see p. 182), the thesis would argue that his theory, critical discourse analysis, and the fake news discourse all depart from the same WEIRD stance. From Kant to Piaget and Kohlberg, the voices upon which Habermas builds his theory are all departing from the same normative standpoint, now described as a WEIRD perspective (see p. 50). From Piaget's suggestion that the rational is arrived at in human moral development through rational reasoning, to Kohlberg's presentation of this type as the highest form of moral reasoning—they all align with an interpretation of morals as aligning with those found in western, educated, industrialised, rich, and democratic communities. Habermas, as also part of this community, naturally finds them all to be promoting common sense—as it is indeed common sense to these actors. As moral psychology has come to emphasise with time, however, this is not a universal stance.

To say that people around the world are driven by different moral standards is indeed not a revolutionary statement, but to accept that these actors' point of moral departure is not only just one but an extreme outlier in human reasoning is a harder fact to accept among those who align with the normative stance. To accept this, however, is important if we are to understand the particular normative rather than the universal empirical character of these frameworks' propositions.

To say that the moral is grounded in a correspondence with an objective reality, expressing the speaker's sincere opinion, and conforming with the shared moral rules is so uncontroversial that we have a hard time accepting that they constitute *a* and not *the only* way to reason. What we see in the fake news discourse is a situation where others have come to challenge this common sense. Baffled, the discourse describes how actors rely on fake news instead of a truth corresponding with how things are, do not have a problem with politicians presenting alternative facts as driven by the speaker's motives, and witness an environment where rational reasoning is not the only convincing form of communication but where post-truths are just as acceptable if they conform with one's pre-determined beliefs. What happened? Habermas prophesised in 1986 that there might be a crisis pending, and with that arising groups that challenge the previous common sense.

> …the social welfare system comes into danger of having its societal base slip away. The upwardly mobile voter groups, who have directly reaped the greatest benefits of the formation of the social welfare state, are capable in times of crisis of developing a mentality concerned with protecting their standard of living. They also may join together with the old middle class, and in general with those classes oriented towards productivity, into a defensive block against underprivileged or excluded groups. Such a regrouping of the electoral base threatens, first of all, the political parties that for decades have been able to rely on a steady clientele in the welfare state: for example, the Democrats in the United States, the English Labour Party, or the Social Democrats in West Germany. (Habermas, 1986, p. 8)

Indeed, the fake news discourse also points to the globalised society as a potential explanation for the expressed criticism (see p. 127). Suggestively, the multicultural societies have created situations where more people have to engage in a compromise, and if standing further apart, each party will feel as if they have lost more in the compromise than in previous discourses when the group was more homogeneous. Suggestively, we could rephrase this statement as one where we now have faced a situation where the *status quo* of the WEIRD stance has been challenged, facing a situation where it too would have to engage in a compromise to accommodate for the integration of other non-WEIRD stances into a common international environment. But this is not what we are observing in the discourse. Rather, the advocators for a WEIRD stance, as above-described, cannot accept that one redefines the very axiomatic basis for its moral point of departure. It will not compromise when it comes to everyone's right to education, to freely express their opinion, to exercise their democratic right in elections, and other human rights. However, in rejecting the negotiation, the ideal once again falls short in practice—breaking with the very axioms of inclusive deliberations. Those opposing these values are presented as part of the problem, as irrational actors with hostile intent—for indeed they are a threat towards one self. However, by excluding them, the authors achieve a pre-constructed consensus that resembles their own criticism against echo chambers.

This argument also expands to the critical discourse analysis employed in the social sciences. Consider reviewing the fake news discourse from this perspective. What you would end up with is an argument for why fake news is immoral, why actors fighting against non-democratic actors are right—and in that possibly even justifying practices of immoral communication, as long as it supports the overarching goal of defending the normative framework. This leads to the strategic moral communication observed in actors' combination of moral values to arrive at an evaluation of an act of communication as moral or immoral. For we do not look at the technical aspects of whether it contains lies or truths, whether the intentions are constructive or harmful, or whether it is presented overtly or covertly—but we use these as factors to emphasise elements in a case so that we ensure that the case is assessed as moral or immoral based on our pre-determined intuition. Our analysis becomes a rationalisation aiming to protect the norm. In each of the values advocated as moral and immoral in the fake news discourse, it is through a strategic argument—whereby some values are accentuated and others downplayed—that we rationalise the conclusion of it being a moral or immoral type of communication. No one value is moral or immoral, but phenomena are labelled as moral or immoral depending on how we interpret the combination of values. For example, not even truth is necessarily pointing to moral communication, as in the case of malinformation (see p. 151).

A naturally moral theory in strategic disguise

As to immanent paradoxes embedded in Habermas' presentation of discourse ethics, the key paradox emphasised in this thesis is his presentation of the theory as an ideal but an ideal so widely practiced that he presents it as an empirical rather than normative theory (see p. 182). On the one hand, it is presented as an ideal, emerging at first as an idea with the first coffee houses. Continuously, however, he rejects the idea that it is a moral philosophy. These paradoxes emerges as he on the one hand recognises that discourse ethics represents a normative ideal practiced to the extent that he argues that it can be viewed as immanently justified by the axioms sustaining it. This, the thesis argues, points to the final and most important observation, which re-situates strategic communication from being immoral by default, to the recognition that moral communication is strategic—and yes, it is even conducted by actors who actively condemn the practice, such as Habermas in discourse ethics.

Habermas argues that moral communication is not driven by a pre-determined interest but engages in open conversations with the aim of arriving at morals through social consensus (see p. 40). Contrary, strategic and instrumental acts are conducted by members of the public with a pre-determined interest (see p. 41). As this thesis has shown, however, contemporary moral psychology would challenge this assumption. Instead, they argue that morals are pre-determined convictions from which we present rationalised arguments to convince others of their claim to validity (presented in Chapter 2). From this perspective, one could argue that Habermas does not engage in a rational conversation, but presents justifications for his moral beliefs. As he himself describes, his discourse ethics 'reformulates the intuition expressed in Kant's "legislation formula" of the Categorical Imperative' (Habermas, 2018b, pp. 106–107). Going back to moral psychologists' reflections that human beings engage in post-hoc justifications for their moral reasoning, could we go as far as to argue that Habermas in his capacity of being a human could have fallen into the same rationalising communication? Echoing in the words of moral psychologists, 'intuitions come first, *strategic* reasoning second. … We believe our own post hoc reasoning so thoroughly that we end up self-righteously convinced of our own virtue' (Haidt, 2012, p. 220). Similarly, we can see the fake news discourse as an expression of Habermas' moral intuition, again pointing to a moral intuition that drives the judgement of communicative environments—without engaging in a discursive process for the axiomatic justification of such framework.

What justifications do Habermas then present for the validity of his theory? In relation to Kant's universal principle, Habermas describes how 'discourse ethics attempts to justify the universal validity of the principle by appealing to the idealizing content of the general and unavoidable presuppositions of argumentation' (2018b, p. 107). The first key lies in recognising that he envisions this conversation to rest on the participants presupposing *ideal* conditions, such as assuming that everyone entering the conversation does so without any egoistic motives, and that they contribute with honest and constructively intended content. The second key lies in recognising that this is an *unavoidable* presupposition, as he argues that his two principles hardly need justification. In both these statements, we find the thesis' claim for why it is a normative theory. First, Habermas (2018b) argues that discourse ethics does, through observations, manifest across societies, independent of the political regime, where 'the belief in its legitimacy has a moral core which is the same as that to be found in the mode of validity of any binding social norm' (p. 120).

> Everyday moral intuitions presuppose the existence of persons who are so constituted that they possess a sense of justice, form conceptions of the good, regard themselves as sources of legitimate claims, and accept the conditions of fair cooperation. In short, the theoretical problem of justification is shifted from characteristics of procedures to qualities of persons. (Habermas, 1994, p. 28)

His conception of morals as tied to the social would thus be justified in any society, as they all ascribe to this idea per default by relying on social structures. Continuously, he describes how the justification of the moral principles is resting on this very circularity, 'as anyone who engages sincerely in a practical discourse, … already implicitly presupposes the validity of the principle of universalization' (Habermas, 2018b, p. 107), whereby they 'accept certain demanding normative presuppositions of communication … through the mere participation in that practice of rational discourse' (p. 106). To that end, he forms the argument that discourse ethics is not a normative theory, as 'there is no alternative because we do not know of any equivalent practices' and where 'demonstrating this is sufficient to refute the objection of circularity which is often raised—namely, that the justification of the principle of universalization only extracts from the content of the presupposition of argumentation what was already inserted to them *by definition*' (2018b, p. 107). This would even be conceptualised by Habermas as a human intuition, embedded in 'the communicative mode of socialization' where the justification of (U) 'involves intuitive knowledge implicit in knowing how norms of actions are justified' (p. 107)[110] Indeed, morals are common sense as shared in a group, but it is not common sense in the capacity of being justified by nature. He argues for this natural justification for his theory by arguing that his is not a normative theory, but an empirical theory in the sense that we can observe its principles as guiding societies across the world and at various points in history (2022b). This thesis would disagree, at best agreeing that it is a sociologically moral theory as it reflects the normative standards of a particular social context.[xxi] Habermas' discourse ethics indeed constitute a theory capturing the dominant normative framework guiding many societies in his and our time. It is not, however, as he argues, enough ground to claim it to be an empirical theory—as it reflects the normative ethics used to guide applied ethical behaviour, but does not claim nor achieves to present an empirical theory in the sense that it explains moral behaviour in contexts where these norms are not accepted.

Most importantly, the thesis' view of discourse ethics as a normative theory is found in this very nature of its presentation, as *not* a moral theory but indeed a matter of course. When arguing that his is not a moral theory, he describes that it holds 'a narrower focus than traditional ethical theories because it concentrates on questions of justice as the question that can be rationally decided in principle' (Habermas, 2018b, p. 104). He rejects the idea that discourse ethics is another normative theory, as 'discourse ethics separates questions of the justice from ethical questions—that is, from questions of the good life, of how to live one's life and what is the best for us to do' (p. 114). The problem, this thesis would argue, is that Habermas fails to recognise that justice represents the guiding principle of his time, it is not a universal but a particular value. And while it is presented as derived from a natural process of human evolution, this thesis would argue that we must recognise the *strategic* nature of proposing a moral proposition as *not* a moral proposition but an inevitable, default, and natural conclusion. When he speaks of discourse ethics as being observable everywhere, this thesis would argue that it is foremost tied to democracies, and other social entities ascribing to the same values that correspond to Habermas' discourse ethics.[111]

---

[110] He similarly describes his in relation to his justification of argumentation in discourse, that 'I must appeal to the intuitive preunderstanding that every subject competent in speech and action brings to a process of argumentation' (Habermas, 1999, pp. 89–90) which would arise 'as soon as he cites a reason for the truth' (p. 90). Resting on Alexy, Habermas argues that these, however, are not rules as in chess, but a presupposed in the '*form* in which we present the implicitly adopted and intuitively known pragmatic presuppositions of a special type of speech' (p. 91, original emphasis).

At the same time, the thesis would agree on Habermas' emphasis on communication's role in building morals. It is from this very recognition that the thesis founds the methodological proposition, arguing that we can study these structures of communication as expressions of the morals of the context in which such language is used. For without communication, it is difficult to imagine that morals would emerge. What unites morals of all times, Habermas argues, is that they all rely on communication to form such morals. It is the 'communicative constitution of all sociocultural forms of life' (Habermas, 2018b, p. 119) Morals, in short, would be reflected in social community, built on communication.

> I use the evolutionary argument that there is an increasing convergence between the diversity of historical concepts of justice and the discursive procedure of judging moral-practical conflicts. The idea of justice, which first assumes the form of concrete conceptions in the particularistic contexts of tribal societies and early civilizations, progressively loses its substantive content

---

[111] For example, in corporate organisations, families, activist groups, and other human collectives where all have agreed to align their interests by following the two principles that all should be included in decision making (U) where the collective norms are determined through a common conversation (D).

> with increasing social complexity, until finally the propositional content of 'justice' withdraws into the procedural form of *impartial judgement.* The scenario of the disintegration of the communal forms of ethos and the increasing pluralism of forms of life and worldviews explains why the substantive ideas of justice, in terms of which the worthiness of moral principles and of norms to be recognized is measured, finally dissolve into the idea of discursive procedural justice. The semantic content of *justice* converges with a procedural notion of *impartiality* that is operationalized in the form of an agreement achieved in rational discourse. (Habermas, 2018b, p. 119)

Underlying the validity claims is the idea that the participants follow the mutual obligation to follow the agreed stance, and that one's 'behaviour does not contradict and will not contradict in the future' (Habermas, 1999, p. 58), which creates the binding/bonding effect (p. 59). From this perspective, he describes how 'the belief in the existence of moral truth is construed as an illusion stemming from the intuitive understanding of everyday life' (1999, p. 55). As Habermas too recognises, validity claims and communicative actions are also attempts where 'one actor seeks *rationally* to *motivate* another' (p. 59) and where the binding/bonding effect is achieved by the mutual recognition that both will indeed follow the agreed stance. In one passage, Habermas even takes it as far as describing how not choosing the moral form of communication will be self-destructive…

> That is why they, as individuals, have a choice between communicative and strategic action only in an abstract sense, i.e., in individual cases. They do not have the option of a long-term absence from contexts of action oriented toward reaching an understanding. That would mean regressing to the monadic isolation of strategic action, or schizophrenia and suicide. In the long run such absence if self-destructive. (Habermas, 1999, p. 102)

Habermas continuously presents morality in this radical sense as building our existence. From above quote, he continues that 'unless the subject externalizes himself by participating in interpersonal relations through language, he is unable to form that inner centre that is his personal identity. This explains the almost constitutional insecurity and chronic fragility of personal identity an insecurity that is antecedent to cruder threats to the integrity of life and limb' (c.f., Habermas, 1994, p. 46, 1999, p. 199).

Similar statements appear in later texts, that we cannot reject morality as we are inevitably a product of it by being part of a society in which everyone else engages in communication to seek understanding. If we are present in the world, he argues, it will surround us whether we accept it or not, 'unless he is willing to take refuge in suicide or serious mental illness' (1999, p. 100). 'The meaning of morality' is described by Habermas in relation 'to the challenge inherent in communicatively structured forms of life' where people stabilize their individual identities in relation to a common social one and thus 'morality is embedded in ethical life' (Habermas, 2018b, pp. 109–110). From this stance, discourse ethics would conceptualise morals as a common identity, and morals as the result of a process. As he phrases it 'only an intersubjective process of reaching understanding can produce an agreement that is reflexive in nature; only it can give the participants the knowledge that they have collectively become convinced of something' (1999, p. 67).

To the final point, we have to return to Habermas' discussion about morals as consensus arising from dissensus. This aligns with the thesis' eristic foundation (see p. 70), rather than resting on a dialectic ideal, finding that conflict is an expression of conflicting interests. However, it does not argue that what is concluded is the rational, only that this *becomes* rational by winning the argument. Habermas argues that discourse arises out of the need 'for achieving agreement on a controversial norm' (2018b, p. 108) as 'conflicts of action are sparked by the resistance of opponents with conflicting value orientations' (p. 120). Alternatively, he describes that 'moral argumentation thus serves to settle conflicts of action by consensual means' where the repair of consensus can either lead to 'restoring intersubjective recognition of a validity claim after it has become controversial or assuring intersubjective recognition for a new validity claim that is a substitute for the old one' (1999, p. 67). This makes Habermas' sense of what is morally rational into a fluid phenomenon, changing because of actors contesting the contemporary rational consensus. It is not a harmonious and predictable phenomenon but a violently unpredictable one. From this perspective, morals would be the discourse itself, as actors test the validity of new claims, aligning with the *dialectic* ideal of communicative progression. Consequently, the actors either find that the pre-defined moral frame provides adequate answers, or alternatively leads to shifts in the moral stance (1999, p. 67). This also means that if there is a discursive agreement and mutual understanding among the participants in a society, no discourse emerges—as actors already agree on how to interpret and act in relation to a shared reality. The idea of discourse is thus tied to a concept of conflict, or more specifically in Habermas' vocabulary a social crisis (see p. 40). Habermas describes how historians segment history into eras of different identities, 'the interpretation that members of a system use in identifying one another as belonging to the same group' where we experience changes in the system as it loses its identity, occurring 'as soon as later generations no longer recognize themselves within the once-constitutive tradition' (2007, pp. 3–4). Tying *social evolution* to *principles of organisation* as 'highly abstract regulations arising as emergent properties in improbable evolutionary steps and characterizing, at each stage, a new level of development', crisis emerges whenever such principles are inadequate in resolving crisis. Such crisis 'proceeds from internal contradictions' (p. 20) in society, describing how liberal-capitalist societies are most sensitive as this 'bourgeois society must react to the evident contradiction between idea and reality. For this reason, the critique of bourgeois society could take the form of an unmasking of bourgeois ideologies themselves by confronting idea and reality' giving rise to a *system crisis* (p. 23). As a consequence, Habermas depicts a movement where the ideal realised would experience the 'side-effects of its own success that endanger the premises for its own functioning' (1986, p. 8).

This, the thesis argues, is what we witness in our time. For a long time, our moral zeitgeist was found to be reflected in Habermas' discourse ethics, using this as well as methodological instruments such as critical discourse analysis as neutral tools. In this very lack of recognition of their normative character, this thesis argues that we find the key argument for why they represent a moral zeitgeist—as we see them as truths rather than as norms. The phenomena discussed in the fake news discourse as threatening this ideal can be seen as a moral discourse in Habermas' sense as arising from public criticism. Instead of following our norms of inclusive discourse, the normative system becomes locked into its own phenomenological boundaries, constricting all discourses into the normative system dictated by its axioms and values. The result is dissonance between the norms preached and the behaviour practiced, which in turn reduces the perceived moral legitimacy of these actors among those who are labelled as immoral, judged within the system by not complying with it. This, the thesis argues, leads to a situation where actors defending the ideal act in such a way that they undermine their own legitimacy—as they do not abide by their own moral rules. All of this is manifesting as strategic moral communication, through discursive structures morally grouping people into categories of us and them, deciding on values and their normative valence—all captured in wide representations of the rational and irrational on the underlying sole basis that the moral conform with one's normative system while the immoral do not. While normative frameworks, such as critical discourse analysis, are presented as emancipatory tools, the thesis' analysis argues that they instead build the type of hegemonic systems of power that they themselves criticise. The thesis' methodology of moral discourse analysis is suggested as a framework that can be used to shine light on these, and other, normative frameworks. It is argued to form an emancipatory tool for scholars and other actors who fail to recognise the normative character of the frameworks that they approach as common sense.

Conclusion

This thesis is formed as a response to pre-existing moralised presentations of strategic communication. From those descriptions of strategic communication as constituting the immoral, such a normative position has been challenged in this thesis by moving towards an empirical understanding of strategic communication's relation to morals. It has done so in three parts, responding to the three research questions guiding this exploration of an empirical perspective.

Part 1 started by rebuilding a theoretical stance that moved from a normative towards an empirical point of departure. It first unpacked Habermas' normative theory of discourse ethics to describe the philosophical assumptions building his ideal. Finding that he rests the theory on a transdisciplinary stance—drawing upon empirical findings from psychology up until the 1980s—the subsequent chapter showed how this stance has been challenged in the 40 years of research published in moral psychology since. This second chapter pointed to three overarching trends in contemporary moral psychology; first emphasising the need to understand social structures dividing people into us and them that challenges Habermas' universal ideal, then pointing to the various value-systems building different context's sense of morality, which would frame Habermas' discourse ethics as a WEIRD morality, and finally the field's emphasis that morals are not only a matter of rational conclusions, but show how we use communication as a tool to rationalise our moral intuitions. Finally, the last chapter in Part 1 used these insights as a stance to propose an empirical theory of strategic communication's relation to morals. The framework was embedded in a new theory of strategic communication, suggesting that the field move from focusing on organisations' use of communication to achieve a pre-set goal to a philosophical view of strategic communication as a type of communication used by various human actors with the intent to, or indeed succeed in, achieving influence. This, the theoretical framework argued, expands to a new moral theory of communication in which strategic communication is not considered immoral, but recognises from the new stance in moral psychology that moral communication is strategic.

Part 2 built on this new theoretical stance to propose a methodological alternative to critical discourse analysis, presented as moral discourse analysis. This new methodology departs from a new set of principles, researching moral phenomenologies through an inductive and immanent analysis. The goal is not to judge or moralise, but to describe morals through their expression as strategic moral communication. After showing how these general principles translate to a particular sampling, coding, and analysis in the methodology's design, the second chapter showed how the methodology formed the thesis' case-study of the fake news discourse. Through these two chapters, Part 2 thus responded to the second research question by suggesting a new methodology to study strategic moral communication.

Part 3 finally addressed the thesis' third research question, illustrating how the methodological framework could be used in a study to investigate and describe strategic moral communication in real discourses. It presented the analysis from the case-study of the fake news discourse, moving from a representation of the normative system advocated to an immanent review of to what extent these norms are followed in their own conduct. Through this review, the third part illustrated how strategic moral communication manifests in-between those two forms, forming a system setting out to regulate actors by binding good conduct to a moral system, but also showing how the actors themselves move from their own ideal when rationalising moral conclusions in the moral discourse on fake news.

Immanent reflections

As Bloor (1991) points out, any theory of the social should rest on an internal validity. That is, recognising that the theories we propose are explaining human behaviour, one should see that the theory is present in one's own theorising. This would also be relevant for this thesis, presenting below a final critical reflection on the impact of the proposed frameworks upon the very development of the proposition.

First, I should inform the reader that I am Swedish, from a cultural context in which Habermas' discourse ethics serves as the moral axiom. Swedes often take pride in our freedom of speech and press as national treasures, with a near religious belief in consensus as the model for all forms of decision-making. Contrary to the impression I might have given when embarking on a critical review of Habermas' framework, I believe that his ideal model of discourse ethics is one that we should aspire to realise and that it is one worth defending. In that sense, this might also have affected the theoretical proposition, as it not only explains what is happening but can easily be paired with a suggestion for a constructive path forward for those who would like to bridge this empirical understanding with a strategic aim of moving towards Habermas' ideal (continuous discussion in the later discussion on empirical implication, see p. 194).

A core in the thesis' suggestion is that we use communication strategically to rationalise an intuition. This thesis can also be seen as developed from such a process, as initially presented (see p. ix), developing because of my experience of the fake news discourse. During the emergence of the discourse, I was engaged in projects intended to help organisations understand the fake news phenomena and find ways to counter them. It took time, but eventually I recognised that the initial presentation of the values we were protecting were not aligned with our presentation of these actors. It was also easy to propose suggestions on how to deal with the problem, however it was much harder to find ways to align such counter-measures with the ideals we set out to protect. For example, we argued that some ought to be excluded from the debate, while at the same time advocating for an inclusive conversation where all had a chance to be heard. Another example would be the way we urged platforms to regulate the conversation to ensure that unethical communication was excluded, while criticising them for influencing the conversation through algorithmic solutions. The very framework presented is thus a product of this initial intuition that something was wrong—an experienced dissonance—seeking a theoretical framework beyond Habermas' discourse ethics which would capture the observable discursive patterns in the fake news discourse.

The inductive search for harmony has manifested in many forms throughout this process, perhaps most fundamentally in the way I have used Habermas' discourse ethics. While many suggested I rely on his theory, I quickly realised that it did not help me understand the discursive patterns in the fake news discourse. At my halfway seminar, I still only mentioned Habermas *once* in the 100-page manuscript. Instead, I moved from readings on 'morals' to new seeds from theories about 'empathy' and how to take the perspective of 'the other' as a constructive way for my project to not echo the commonly published critique of the fake news phenomena, but rather seek an explanation for it. Stumbling upon the 'theory of mind', and continuously finding myself drawn to moral psychologists' research, I slowly formed a cluster of research that helped me understand why seemingly irrational and absurd texts gained influence, and why the initial approach of merely informing the public in the fake news discourse might not be the most strategic way forward. However, it was not until finding patterns among the claims resurfacing in the literature on moral psychology that I realised how they were in opposition to the key principles in Habermas' discourse ethics. In this light, I saw the purpose of using his theory as a resource, an anti-thesis that helped crystallise the key points of my initially diffuse gut-feeling, now supported by studies from moral psychology. Another reason for why I was drawn to fields such as philosophy and psychology's theories on ethics and morals was that there was little academic knowledge about unethical public communications' expression in the new political and communication landscape. The discourse on fake news was also of a particular character in the beginning, initially foremost formed through think tanks and government reports (thus here studied in the thesis). By following the discourse, I gained a longitudinal insight into the concepts used and the key events influencing the rapid changes of themes and attitudes characterising the debate. For me, it was always apparent that these reports were strategic—unveiled in the choice of words and how they were defined, forming cues for the underlying message of what was communicated as 'good', 'constructive', 'real', and 'true' as opposed to 'malicious', 'false', 'fake', 'harmful', 'inauthentic' etc. All these examples show how the very process of arriving at the theory and the methodology as proposed in this thesis is resting on the same inductive principle promoted.

Some would argue that this non-linear process is perfectly okay for a scholarly thesis, others would surely disagree, and a few might recognise that they do the same thing but do not write out the messy (irrational) process. And while I recognise my biased, subjective account, I have sought ways to address this weakness. First, even if a critic would see the thesis as a post-constructed thesis, as based on rationalised intuitions, I would argue that the thesis gains weight by drawing upon extensive empirical research from moral psychology that is based on a completely different methodology. In that way, the conclusions drawn in this thesis and the theory and methodology proposed is built on the shoulders of entire fields of accumulated common sense from extensive empirical research. Secondly, I would argue that the inductive approach has helped me in presenting a thesis based on at least a new intuition, rather than on a pre-existing one. I have not been bound to a particular theory and such theory's pre-determined stance, including the axioms that one then is set to ascribe to. Instead of using a pre-set frame from which I then seek to fit my findings, I have continuously sought to test different approaches, finding them ill-functioning and rejecting them, again finding a good point of departure while later recognising its flaws and moving on or supplementing the very same in a search for a greater fit—and allowing myself repeatedly to be wrong. In many ways, I have aimed to live up to Habermas' ideal, while recognising that realistically I am bound to fail in living up to his notion of rationally derived knowledge. Furthermore, I would argue that thesis itself is not to be seen as a rationally derived conclusion, but a post-rationalised theory based on a combination of readings from the cognitive sciences and the immanent criticism emerging when taking such as stance to look at Habermas' theoretical ideal of discourse ethics as empirically manifested through paradoxes in the fake news debate.

On a final note, I concur with Bloor (1991) that 'immersed as we are in society we cannot grasp it as a whole in our reflective consciousness except by using a simplified picture, an image, or what may be called an "ideology"' (p. 46). In many ways, this thesis makes a suggestion for another theory which reduces complex realities into neat categories. And while presenting an argument for why we need to start the conversation about how our own phenomenological boundaries restrict us in our theorising, the thesis has been limited to our current vocabulary. In future studies, I hope that we will be able to break out of such rational structures of theorising. While this thesis is still bound to a language of labels such as 'ethical' or 'unethical', 'good' or 'bad', 'right' or 'wrong', 'real' or 'fake'—this very representation of binary oppositions, where the one is normatively phrased as the moral and immoral, may restrict us in scholarly language to a not only politicised but simplistic understanding of the lifeworld.

Implications for public discourse

Finally, if one were to recognise this empirical perspective and use it as a new stance in practical and scholarly discourses on strategic communication's relation to morals, one can see that it could have various implications. The section below expands on a discussion on what such implications might be when reviewed from a practically professional perspective in relation to the fake news discourse and the scholarly discourse on strategic communication in the social sciences respectively.

First, just as Habermas notes that Hegel is right in his critique against Kant and other moral theorists as 'they focus on questions of *justification*, leaving the question of *application* unanswered' (Habermas, 1999, p. 206), the thesis' proposition can be seen to have empirical implications. For, how should we approach public discourse if not resting on Habermas' ideal? I would argue instead of aiming to achieve consensus as an agreed direction that rests on a common recognition of valid claims that drive this conclusion, we could also search for a consensus on an agreed direction while accepting that different people might have different reasons for arriving at such a conclusion. One can also imagine that we would evaluate and engage in discourse in a different way if we do not assume that actors engage in discourse with the intent to live up to discourse ethics' ideal. For what would happen if we assume that people present rationalised arguments that supports their moral intuitions? This in itself may help us gain a better understanding of *why* people are proposing some things rather than others.

The very recognition that actors in the fake news discourse break with their own normative ideals is also grounds for immanent critique, by illustrating how actors in power have to adjust their behaviour to abide by their own communicative norms. A simple example is to review the discourses created through the discussion about fake news and its conceptual siblings. This would mean critically reviewing who is allowed access to speak their opinion, and to evaluate the risk of excluding actors from the conversation as this might drive the very polarization criticised. Deplatforming would drive people to form their own discursive communities that share the common denominator of being branded as 'the other'. By opening up to conversations with adversaries and actors appearing in-between one's allies and enemies, a true discourse ethics can take its first steps towards inclusiveness in the discussions and testing of validity claims through communication rather than pre-determining the result of such conversations by all stemming from the same sense of truth, truthfulness, and rightness. Acknowledging diverging stances in the fake news discourse is not enough, as there are actors set apart by being framed as part of 'the other'. What we need is ideographical presentations of these actors on their own terms, critics and their claims need to be included if the ideal is to be achieved.

Moving from the thesis' implications on the empirical to the scholarly discourse, firstly, this thesis has argued the necessity of recognising that strategic communication is an empirically neutral label, as an empirical perspective cannot hold a moral valence. This shift, the thesis argues, also allows scholars to study strategic communication in a new way—not having to justify its moral legitimacy but simply pay interest to the communicative structures that are formed and sustained by strategic communication. To rid strategic communication of the immoral label means departing from a new set of theories and methodologies—here making a first suggestion by moving from Habermas' discourse ethics to strategic moral communication, from critical discourse analysis to moral discourse analysis—where the change in perspective may allow us to form a transdisciplinary common interest in the way communication, when strategic, links humans together *or* work to drive them apart. From this stance, strategic communication could move from its dichotomous presentation in the social sciences—as either a theoretical field discussing a professional practice or alternatively as an immoral label—to form a common phenomenon of interest. It could serve as a common point of discussion for scholars in psychology, sociology, political science, and beyond, as the phenomenon forming social constructions.

In a broader scheme, the thesis has also sought to illustrate the value of understanding strategic communication in the social sciences from this empirical stance. It has implicitly argued for strategic communication's value beyond its own disciplinary borders, seeking an understanding of why it should matter in a broader sense within the studies of humans and their societies. For this, however, strategic communication research needs to claim the position of holding knowledge beyond being a field researching the professional practice of actors using communication strategically to advance an organisational goal (see Chapter 3, starting on p. 57). Instead, the field needs to recognise the potential if re-framing such interests as holding knowledge in *what makes communication strategic*. From years of studies, we hold insights (together with affiliated fields such as rhetoric, propaganda studies, public relations, and advertising research to name a few) into what types of communication hold influential qualities. Our times' moral scepticism towards such knowledge has, however, held us back in making it part of acceptable frameworks in the wider social sciences. When approached as a single-minded pursuit of interests in a Machiavellian style, we neglect seeing that it could just as well be approached as a key to understand people's world-view and how they were formed through actions of strategic communication with other people. The field's common sense, i.e., that one should stick to simplicity, unexpectedness, concreteness, credibility, emotions, and use stories to be communicatively influential is today knowledge conveyed as tips by and for strategic communicators. These practitioners hold insights about what works when one seeks to influence other people, and scholars in strategic communication have an opportunity to conceptualise and idealise these insights into theoretical principles, frameworks, models, and concept that should be of interest to the wider social sciences.

Appendix

The following appendix provides an elaborated presentation of the theoretical and empirical material used in the thesis. The first sections present the literature reviews, in which it describes what material has grounded the theoretical presentation of Habermas' discourse ethics, and the 40 years of research in moral psychology since his engagement with the field. After presenting these literature reviews, the appendix continues with a presentation of the empirical material. While introducing the sampling procedure and material selected in Part 2—which is used for the analysis in Part 3—the below section presents a list of the 100 most frequent words and a list of definitions for the key concepts used throughout the thesis in relation to the fake news discourse. This empirical presentation thus supplements the previously presented methodological application of moral discourse analysis in the case-study, and provides additional transparency in relation to the analysis in Part 3. As to the theoretical and empirical references used in the thesis, these are all presented after the appendix in the thesis' list of reference (see p. 221).

Literature reviews

The thesis is supported by a set of literature reviews in which the literature has been found using searches for academic journal articles primarily through the database *Web of Science* (WoS)[112] or *Google Scholar*[113] and searching for books through the platform *LIBRIS*[114].

HABERMAS' DISCOURSE ETHICS

There is no single publication where Habermas presents the theory of discourse ethics. As Regh (1994) describes, the 'various presentations of discourse ethics are scattered throughout a variety of books and articles' where 'these presentations are often occasional in character and sketchy in details' (p. xiii). In his own review, Regh settles on a presentation of Habermas' books *Moral Consciousness and Communicative Action* (1999) and *Justification and Application* (1994) to account for Habermas' discourse ethics. Others would surely argue for another selection as proper. Even Habermas seems to consider text central at different times, as both books above-mentioned were added a new essay when translated to their English editions.[115] The thesis started taking the same route as Regh—tracing the development of Habermas' discourse ethics in his works—continuing with a more systematic review based on a selection of texts compiled by Habermas to represent his discourse ethics. Both tiers have influenced the thesis' presentation of discourse ethics in Chapter 1, as the first tier provided a context that contributed to the understanding of the theory, while the second tier ended up as the primary corpus of text to represent Habermas' discourse ethics. The following sections expands upon this process and elaborates on the material found to be central to describe discourse ethics.

---

[112] A scientific database previously known as *Web of Knowledge* which contains over 34,600 peer-reviewed journals published worldwide in over 250 sciences in 171 million records (including articles, review articles, book chapters, conference proceedings etc.), accessible online through www.webofscience.com.

[113] The latter as WoS captures peer-reviewed academic pieces only from a selection of journals, while Google Scholar captures a wider scope of academic publications in journals as well as e.g., reports and other types of publications not necessarily written by academics.

[114] A joint catalogue for over 7 million titles held at over 500 Swedish academic and research libraries, which also provides digital or physical access to such recorded books.

[115] Adding the essay Morality and Ethical Life: Does Hegel's Critique of Kant Apply to Discourse Ethics? (originally published in German in 1986 as Moralität und Sittlichkeit: Treffen Hegels Einwände gegen Kant auch auf die Diskursethik zu?) to the 1990 English edition Moral Consciousness and Communicative Action as otherwise a direct translation of the 1983 original German Moralbewußtsein und kommunikatives Handel. Likewise, the essay Morality, Society, and Ethics: An Interview with Torben Hviid Nielsen (originally published in German in 1990 as Interview mit T. Hviid Nielsen) was added to the 1993 English edition Justification

The first tier began in line with Regh's strategy and with the same limitations,[xxii] sampling texts with the aim to capture Habermas' discourse ethics by reviewing a compiled bibliography of his works,[116] to trace the emerging contours of his discourse ethics. This went beyond Reghs' starting point, rather arguing that we can find antecedents in Habermas' historic account of the bourgeois public sphere in *The Structural Transformation of the Public Sphere* (Habermas, 1989), clarifying the communicative dimension in this theory in the two volumes on *The Theory of Communicative Action* (1984, 1987), formed as a moral theory of communication in *Moral Consciousness and Communicative Action* (1999), after which he addressed key aspects such as the difference in moral and ethics in *Between Facts and Norms* (2009) and adding additional clarifications in *Justification and Application: Remarks on Discourse Ethics* (1994). This is not, however, to be confused with how Habermas' thoughts develop in relation to his theoretical contribution in a wider sense, where his serious engagement on the matter of discourse ethics emerges in what Dews (2017) describes as the third phase in Habermas' level of confidence on matters of power and communicative rationality. This, Dews argues, started in the late 1980s, or as this thesis finds, coincides with the emergence of Habermas' discourse ethics.

The second tier of literature was found upon reading one of Habermas' more recent publications—the book *Philosophical Investigations* (2018b). The book was written by Habermas on the occasion of his 80th birthday

---

and Application: Remarks on Discourse Ethics of the 1991 original German Erläuterungen zur Diskursethik (as detailed in the bibliography by Deflem, 1994).

[116] Guided by the bibliography by Corchia (2021), limited, however, in my readings to publications or translations in English.

where he looks back at a lifetime's worth of publications. Here, Habermas outlines his key approaches to communicative reason with one essay for each of the 'five main areas on which my philosophical work has focused' (p. vii). In each essay, he refers to a set of previously published texts that together 'could take the place of unwritten monographs' (p. vii). The essay itself would then provide 'an introduction to each of the five thematic volumes' (p. vii). One of these chapters is dedicated to discourse ethics (pp. 100–121), where Habermas highlights 'five texts in which I developed discourse ethics' (p. 104). However, as he points out, 'publishing a series of introductions without the texts to which they refer is, of course, an imposition on the reader, who will if necessary have to track down these texts using the references' (p. viii). This is what I have done. The essay has served as the sampling strategy, using the referenced literature as the key publications on discourse ethics—selected by the author himself.

This essay and the texts which it refers to has served as the basis for the literature review. From it, additional texts were identified immanently in a second tier, seeking understanding of key concepts by going back to the texts referenced in the essays. Among other references widely used in Habermas' essay on discourse ethics (2018b, pp. 100–121), he often references books such as *Legitimation Crisis* (2007) and *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (2009). This is used to frame his theory in a wider sense, connecting 'moral theory with the discourse theory of law and the democratic constitutional state' (2018b, p. 104).[xxiii] This also pointed to the thesis' delimitation to focus on the *communicative* dimension of his moral theory—just as Habermas separated the presentation of discourse ethics from political philosophy, the theory of communicative action, the foundations of sociology in the theory of language, and the formal-pragmatic conception of language and rationality (2018b, p. vii).[117]

---

[117] Discourse ethics is described as a theory of morality discussing moral validity claims as one type of normative validity claim that helps us also understand other types, such as legal validity and political legitimacy (Habermas, 2018b, p. 104). Similarly, Habermas continues to describe Apel's critique by offering 'an account of the relationship between the discourse theory of truth, morality and law, while at the same time providing an introduction to the normative principles of political theory' (p. 121).

As the essay on discourse ethics was compiled in 2009, a search was also made for later publications by Habermas on morals and ethics. This amounted to a review of 18 titles. The first literature review focused on finding academic journal articles published by Habermas in English since 2009, using the scientific database WoS (see p. 198). Out of the 41 journal articles recorded as published by Habermas, 20 were published in English. Upon review of these 20 publications, four were found to explicitly mention ethics or morals, and one was indirectly connected to the topic of the thesis by reflecting on the new challenging communication landscape. In addition to journal articles, seven additional publications were identified by reviewing a compilation of Habermas' publications (Gregersen, 2022a, 2022b, 2022c, 2023a, 2023b, 2023c) published up until the end of 2022. For a full list of these publications, see below in Table 1. This table contains the final set of 18 reviewed primary literature as: 1) six texts referenced in the essay from 2009, 2) five articles published since, and 3) seven other publications.

Table 1. Habermas' publications on discourse ethics

List texts by Habermas, sampled to capture his presentation of discourse ethics.

Sampling	Title
Essays	Discourse ethics (Habermas, 2018b, pp. 100–121)
	Discourse Ethics: Notes on a Program of Philosophical Justification (Habermas, 1999, pp. 43–115)
	Morality and Ethical Life: Does Hegel's Critique of Kant Apply to Discourse Ethics? (Habermas, 1999, pp. 19–216)
	Remarks on Discourse Ethics (Habermas, 1994, pp. 19–112)
	Morality, Society, and Ethics: An Interview With Torben Hviid Nielsen (Habermas, 1994, pp. 147–176)
	A Genealogical Analysis of the Cognitive Content of Morality' (Habermas, 1998, pp. 3–48)
Articles	Once Again: On the Relationship Between Morality and Ethical Life (Habermas, 2021)
	From Formal Semantics to Transcendental Pragmatics: Karl-Otto Apel's Original Insight (Habermas, 2020b)
	The Concept of Human Dignity and the Realistic Utopia of Human Rights (Habermas, 2010a)
	The 'Good Life'—A 'Detestable Phrase': The Significance of the Young Rawls's Religious Ethics for his Political Theory (Habermas, 2010b)
	Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere (Habermas, 2022b)

Chapters etc.	Foreword (Habermas, 2022a)
	Commentary on, Christina Lafront, Democracy Without Shortcuts (Habermas, 2020a)
	For a Democratic Polarisation: How to Pull the Ground From Under Right-Wing Populism (Habermas, 2016)
	Interview With Jürgen Habermas (Habermas, 2018a)
	Moral Universalism at a Time of Political Regression: A Conversation With Jürgen Habermas About the Present and His Life's Work (Czingon, Diefenbach, & Kempf, 2020)
	Critique and Communication: Philosophy's Missions. A Conversation With Jürgen (Foessel, 2018)
	For God's Sake, Spare Us Governing Philosophers! (Hermoso, 2018)

At one point, I considered adding more references to Habermas' critics as a way to add to the contextual nature of his theory (context as criticism rather than understanding, see p. 87). A balance was made by adding some critics in the presentation of discourse ethics in Chapter 1, but foremost focusing on Habermas' own acknowledgement of some of these critical voices. As outlined in his interview with Torben Hviid Nielsen—one of the texts he himself describe as central to the presentation of discourse ethics—the text is described as a '*tour d'horizon* of objections that had been raised against discourse ethics' (Habermas, 2018b, p. 111). This is also present in his essay *Remarks on Discourse Ethics* where he situates the proposition in relation to competing theories, adding in later publications that it is a matter of 'clarification of various details rather than in an improvement of the construction as a whole' (2018b). For example, he addresses the accusations of a 'Eurocentric bias of the deontological concept of justice', against the concepts he uses, and 'the still widespread "postmodern" misconception that moral universalism is a prescription for a kind of equal treatment that imposes uniformity and inclusion through assimilation', responding that discourse ethics enables 'inclusions of others *in their otherness* that is sensitive to differences' (p. 111). As such, Habermas' critiques have been integrated into the presentation of discourse ethics in this thesis, however as the aim of the chapter is to make sense of what *his* normative proposition is, the way in which these critical remarks are presented is rather to clarify and seek sharper contours that can help describe his proposition.

A META-META REVIEW OF MORAL PSYCHOLOGY RESEARCH

Chapter 2 presents the results from a literature review of the past 40 years of research on moral reasoning. The goal is to capture the most common contemporary discussions, thus not a complete review of all studies, nor elaborating on the criticism and ongoing discussions in relation to these theories. I recognise the limits of my ability to assess the quality and determine the key strands of research in a field that is not my own. Because of that, the literature review is a review of previously published summaries and reviews by scholars with better insights into these discussions than I have. A final sample of five books and 19 review articles were selected to that end, presented below.

To first acquaint myself with the field, I started by reading what is referred to as 'popular books', accessible summaries of the field by its researchers. Some books focus on rational reasoning in general, such as the psychologist and economist Kahneman's *Thinking Fast and Slow* (2011), while the majority were focused on moral thinking and behaviour in particular, with a list of readings spanning from the social psychologist Haidt's *The Righteous Mind* (2012), the neuroscientist Greene's *Moral Tribes* (2013), the psychologist Bloom's *Against Empathy* (2016), and the biological anthropologist Henrich's *The WEIRDest people in the World* (2010). I have approached these readings as providing a general understanding of the shift towards modern moral psychology, using these and the work they reference as a first step in this meta-literature review.

After gaining a broad understanding of the discussions emerging on moral reasoning, I continued to search for scholarly reviews in the form of 'review articles'. [118] A search on 2 May 2023 on the WoS (see p. 198) revealed 141 review articles published in English within psychology that either mentioned *ethics* or *morals* as well as *reasoning* or *judgment* in the title, abstract, or among the key words.[119] When reviewing this list of publications, I found that there were, however, many missing the intended mark—such as papers presenting 'a review of telework research' and 'the folk psychology of souls'. After excluding themes such as studies on moral development in children and discussions about psychological diagnosis, a total of 19 publications were selected as a sample to ground the literature review. Particular weight was given to comprehensive reviews, and therefore foremost with direct references to one particular review in the running text—a review article rested on a review of '1,278 relevant research articles published from 1940 through 2017. These were subjected to expert content analysis and standardized bibliometric analysis to classify research questions and relate these to (trends in) empirical approaches that characterize research on morality' (Ellemers et al., 2019, p. 332).

---

[118] Described by WoS as 'a renewed study of material previously studied' which includes 'review articles and surveys of previously published literature' while these publications will usually 'not present any new information on the subject'.

[119] Search string expanded as *ethic** or *moral** to include variations such as *ethics, ethical, morality, moralism* etc. and similarly searching for *reasoning* or *judgement.*

And, as always, from these sampled texts, additional publications were found as referenced in the above-mentioned publications.

Empirical data

The thesis proposes a new methodology of moral discourse analysis, introduced in Chapter 4, describing the application in relation to the thesis' own case-study of the fake news discourse in Chapter 5, and presenting the results of such analysis in Part 3 (Chapter 6–9). The following section introduce a table presenting the 100 most frequent words appearing in the sample, and a list of 100 concepts that illustrate some examples of the concepts used in the analysis to represent the moral discourse on fake news (see Chapter 8).

FREQUENT WORDS

As part of the methodology, the thesis has in part relied on a review of the quantitatively most frequently appearing words in the sampled discourse. The section below thus shows a list of the 100 most frequent words appearing in this sample.

Table 2 The 100 most frequent words
List of the 100 most frequent words after the exclusion of a list of stop words.

No	Word	Times
1	media	32418
2	news	30551
3	information	22879
4	content	19210
5	Facebook	19079
6	political	14616
7	social	14506
8	disinformation	14046
9	online	13935
10	Russian	13792
11	public	12764
12	data	12114
13	people	11436
14	Russia	10460
15	twitter	9724
16	fake	8984
17	election	8062
18	government	8031
19	states	7847
20	report	7662
21	policy	7611
22	accounts	7474
23	research	7296
24	digital	7180
25	internet	7080
26	state	6582
27	transparency	6495
28	users	6113
29	community	6093
30	influence	5987
31	security	5950
32	right	5883
33	European	5829
34	fact	5617
35	literacy	5615
36	platforms	5576
37	groups	5418
38	national	5383
39	law	5292

40	world	5183
41	rights	5126
42	foreign	4980
43	international	4947
44	support	4942
45	policies	4904
46	elections	4801
47	society	4573
48	united	4568
49	center	4487
50	false	4445
51	service	4420
52	companies	4368
53	countries	4352
54	human	4280
55	campaign	4279
56	operations	4274
57	like	4237
58	group	4201
59	Europe	4149
60	standards	4101
61	network	4007
62	post	3953
63	propaganda	3935
64	access	3854
65	technology	3804
66	campaigns	3799
67	global	3783
68	ads	3704
69	Ukraine	3612
70	actors	3573
71	trust	3491
72	account	3450
73	services	3412
74	user	3397
75	misinformation	3395
76	platform	3389
77	communication	3359
78	analysis	3339
79	YouTube	3333
80	Kremlin	3252
81	activities	3213
82	freedom	3211
83	democratic	3191
84	company	3190
85	advertising	3183
86	activity	3177
87	truth	3153
88	party	3149
89	action	3144
90	measures	3126
91	pages	3116
92	tools	3116
93	cyber	3101
94	committee	3101
95	Google	3099
96	China	3086
97	journalism	3071
98	speech	3036
99	privacy	2968
100	issues	2949

Stop words include: 0 000 09 1 10 11 12 13 14 15 16 17 18 19 2 20 2014 2015 2016 2017 2018 2019 2020 2021 21 22 24 25 28 3 30 4 5 6 7 8 9 a about above across act after again against all also am an and any are aren't aren’t around article as at available based be because been before being below between both but by can can't cannot case com could couldn't did didn't different do does doesn't doing don't down during e each efforts en eu even example far fb few first for from further had hadn't has hasn't have haven't having he he'd he'll he's help her here here's hers herself high him himself his how how's however http https i i'd i'll i'm i've if important in including into is isn't it it's its itself july june key let's local made make many march may me more most mustn't my myself need new no nor not now number of off often on once one only or order org other ought our ours ourselves out over own page part pro project projects provide related role s said same say says see set shall shan't she she'd she'll she's should shouldn't so some such take than that that's the their theirs them themselves then there there's these they they'd they'll they're they've this those through time times to too two u uk under until up upon us use used using very was wasn't way we we'd we'll we're we've well were weren't what what's when when's where where's which while who who's whom whose why why's will with within won't work would wouldn't www you you'd you'll you're you've your yours yourself yourselves

CONCEPTS AND DEFINITIONS

Of particular interest in the methodology of moral discourse analysis are the concepts appearing in the moral discourse. The methodology is presented in Chapter 4 and 5, using these concepts throughout Part 3 (particularly in Chapter 8) to illustrate how the discourse-specific notion of moral and immoral communication manifests through phenomena conceptualised and defined. The table below lists 100 examples of concepts used for this analysis.

Table 3 100 concepts

A selection of 100 concepts identified in the fake news discourse.

Concept	Example of coded definition
Algorithm	An algorithm is a fixed series of steps that a computer performs in order to solve a problem or complete a task. For instance, social media platforms use algorithms to compile the content that users see. These algorithms in particular are designed to show users material that they will be interested in, based on each user's history of engagement on that platform. For example, algorithms can often filter content so that users primarily see types of content with which they have previously engaged. Users tend to engage with content that provokes an emotional reactions like fear and anger. As such, it's argued that algorithms designed to take advantage of users' emotions create an environment wherein disinformation created to play into deep-seated fears and cultural identities will flourish. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Algorithmic bias	…algorithmic bias toward sensational content, driven by commercial preference for user engagement but easily exploited by those with intention to inflame and distort. (Atlantic Council, 2019a, p. 19)
Alt-right	The term 'alt-right' was coined by Richard Spencer in 2008 to describe right-wing political views at odds with the conservative establishment. (Data&Society, 2017, p. 9)
Alternative facts	…alternative facts imply the existence of objective facts that are present somewhere but contextualized to drive the story. In other words, alternative facts match the existing data and the facts are cast in a context that suits political powers. (Heinrich Böll Foundation, 2020, p. 11)

Alternative media	…the porous border between social media and hyper-partisan media outlets creates an 'alternative media ecosystem' that enables online misinformation to be amplified on television, on the radio, or in newspapers (Centre for the Study of Democratic Institutions, 2018, p. 17)
Artificial intelligence (AI)	Artificial intelligence (AI) describes computer programs that are 'trained' to solve problems that would normally be difficult for a computer to solve. These programs 'learn' from data parsed through them, adapting methods and responses in a way that will maximize accuracy. As disinformation grows in its scope and sophistication, some look to AI as a way to effectively detect and moderate concerning content. AI also contributes to the problem, automating the processes that enable the creation of more persuasive manipulations of visual imagery, and enabling disinformation campaigns that can be targeted and personalized much more efficiently. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Astroturfing	…a few trolls can, by posting a number of comments, give the impression of a majority opinion even when it is not at all the case—it is enough to have a paralyzing effect on others. This technique consisting of giving an appearance of popularity is called 'astroturfing,' a reference to a brand of artificial turf (AstroTurf). (French CAPS & IRSAM, 2018, p. 87)
Bandwagon effect	A cognitive effect where beliefs increase in strength because they are shared by others. (UK Government Communication Services, 2019, p. 47)
Black hat SEO	Black hat SEO (search engine optimization) describes aggressive and illicit strategies used to artificially increase a website's position within a search engine's results, for example changing the content of a website after it has been ranked. These practices generally violate the given search engine's terms of service as they drive traffic to a website at the expense of the user's experience. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Bot	Bots are "algorithmically driven computer programs designed to do specific tasks online" (Woolley and Howard 2016, 4885). McKelvy and Dubois (2017) propose four types of political bots: dampeners suppress messages, amplifiers make messages appear more popular than they are, transparency bots share information relevant to informed citizenship, and servant bots are used by government and organizations to answer questions or provide other services. (Centre for the Study of Democratic Institutions, 2018, p. 17)
Botnet	Hacker bots are sometimes used to spread infested code, or malware, to hijack internet-connected devices. A botnet is a group of such hijacked devices that can be deployed from

	a remote location without the knowledge of the device’s owner to provide computing resources that can be used for a variety of malicious purposes, such as distributing phishing email and orchestrating distributed denial of service (DDoS) attacks. (Swedish Civil Contingency Agency, 2018, p. 59)
Bullshit	…a catch-all term that includes half-truths, lies, and misrepresentations; but most importantly for bullshit, it involves a disregard for truth altogether, for the purpose of crafting a narrative. (NATO StratCom COE & King’s Centre for Strategic Communications, 2018, p. 18)
Cheerleading	…fake social media accounts are used to cheerlead for state talking points, making them seem like organic, popular grassroots points of view. (Stanford Internet Observatory Cyber Policy Center & Institution, 2020, p. 19)
Clickbait	When a publisher posts a link with a headline that encourages people to click to see more, without telling them much information about what they will see (Facebook, 2014b)
Cloaking	These bad actors disguise the true destination of an ad or post, or the real content of the destination page, in order to bypass Facebook’s review processes. … when a Facebook reviewer clicks a link to check whether it’s consistent with our policies, they are taken to a different web page than when someone using the Facebook app clicks that same link. (Facebook, 2017a)
Clone site	A website that copies and displays Instagram or Facebook profiles, posts and other information without the users’ knowledge or consent (Facebook, 2020c)
Computer propaganda	A multidiscipline effort led and coordinated by Info ops function to analyse an adversary’s information activities, its source content, intended audience, media selection, and effectiveness. (NATO StratCom COE, 2017, p. 67)
Confirmation bias	We tend to favor information that confirms our preexisting assumptions, supports our positions and does not offend our sensibilities. (French CAPS & IRSAM, 2018, p. 31)
Conspiracy	Conspiracy theories amplify a natural tendency we harbor, namely to believe that every effect is caused by intentional action, especially those effects that benefit certain people. For this reason, conspiracy theories are inevitable and feed mostly on crises and violent events. Fortunately, not all are dangerous. Some are harmless. But others can have destabilizing effects, even if they are only shared by a very small percentage of the population, as long as that small minority is ready to take violent action. (French CAPS & IRSAM, 2018, p. 35)
Content moderation	Technology today isn't so advanced that we can solely rely on it to enforce our policies. For instance, context can be important when determining whether certain content, like satire, is violative. As such, our team of trained moderators

	helps to review and remove content. In some cases, this team proactively removes evolving or trending violative content, such as dangerous challenges or harmful misinformation. Another way we moderate content is based on reports we receive from our users. (TikTok, 2019b)
Coordinated inauthentic behaviour	…online behavior intended to deceive audiences about its authors or motives, conducted by a number of users in coordination with one another. CIB is banned on Facebook, and was the reason used for removing a number of Russian and Iranian disinformation assets. (Atlantic Council, 2019b, p. 48)
Critical thinking	For one of them, integrating critical thinking meant focusing on principles of journalism—skepticism, using informed judgment, asking questions, and assessing information-collecting processes. (RAND, 2019, p. 20)
Cyber bullying	The use of the internet enabled forms of communication to bully a person, typically by sending messages of an intimidating or threatening nature. (UK Law Commission, 2018, p. v)
Cyborg	There are also cyborg accounts87, which are jointly operated by people and software. (Council of Europe, 2017, p. 38)
Dark ads	Targeted advertisement based on an individual user's psychographic profile, 'dark' insofar as they are only visible to targeted users. (UK Government Communication Services, 2019, p. 47)
Data mining	Data mining is the process of monitoring large volumes of data by combining tools from statistics and artificial intelligence to recognize useful patterns. Through collecting information about an individual's activity, disinformation agents have a mechanism by which they can target users on the basis of their posts, likes, and browsing history. A common fear among researchers is that, as psychological profiles fed by data mining become more sophisticated, users could be targeted based on how susceptible they are to believing certain false narratives. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Debunking	A study found that debunking was ineffective when the consumer could generate competing explanations supporting the misinformation. Furthermore, a debunking message that simply identified misinformation as incorrect without offering corrective information was similarly ineffective. Even when a detailed debunking message included corrective information, the debunking did not always reduce participants' belief in the misinformation. (PARK Advisors, 2019, p. 11)
Deep fake	Videos altered by artificial intelligence tools to either misrepresent an event that occurred or manufacture an event that never occurred. (Brookings Institution, 2019)

Deepweb and darkweb	The Deepweb refers to any parts of the World Wide Web that cannot be found using conventional search engines like Google. This could be because the content is restricted by the website creators. The Darkweb refers to the small portion of the Deepweb that can only be accessed through the use of specific software, such as the TOR browser. It has both legitimate and illegitimate uses, and is commonly used for facilitating the distribution of controlled drugs and indecent photographs of people aged under 18 years. (UK Law Commission, 2018, p. vi)
Digital literacy	The ability to use digital technology, communication tools or networks to locate, evaluate, use and create information. It also refers to the ability to understand and use information in multiple formats from a wide range of sources when presented via computers, or to a person's ability to perform tasks effectively in a digital environment. Digital literacy includes the ability to read and interpret media, reproduce data and images through digital manipulation, and evaluate and apply new knowledge gained from digital environments. (UNESCO, 2011, p. 183)
Disinformation	Dissemination of false information with the deliberate intent to deceive or mislead. (NATO StratCom COE, 2017, p. 67)
Dormant account	A dormant account is a social media account that has not posted or engaged with other accounts for an extended period of time. In the context of disinformation, this description is used for accounts that may be humanor bot-operated, which remain inactive until they are 'programmed' or instructed to perform another task. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Doxxing	Searching for and publishing private or identifying information about a particular individual on the web, typically with malicious intent. (UK Law Commission, 2018, p. vi)
Echo chambers	An ideological environment in which ideas and opinions are amplified and reinforced by their repetition, creating a mainstreaming effect of likemindedness. (NATO StratCom COE, 2017, p. 67)
Election interference (election manipulation, election integrity)	…an attempt to influence the outcome of an election through covert, undeclared, or illegal means, or by disinformation. Election manipulation can be domestic or foreign in origin. It is distinguished from electioneering by the use of covert, undeclared, or illegal means, and by the use of disinformation. (Atlantic Council, 2019b, p. 48)
Established media	With the exception of Yahoo! News (39%), almost all the popular online brands are digital versions of established media outlets. Online alternative news brands including Bastille Post and Stand News maintain their popularity by targeting niche audience tastes, such as soft news and politics, respectively. (Reuters Institute, 2018, p. 128)

Evidence	Facts, data, or other information supporting a belief, argument, or view. This can be empirical (based on observation or experiment) or episodic (based on occasional or unsystematic observation). (RAND, 2018b, p. 9)
Fact	Objective information that can be proven and verified and is consistent with reality (RAND, 2018b, p. 9)
Fact-checking	…the process of determining the truthfulness and accuracy of official, published information such as politicians' statements and news reports. (Institute for Strategic Dialogue, 2020, p. 6)
Fake account	By "fake account," we refer either to an account that is managed by someone pretending to be someone else or to accounts that are not managed by people, but are automated (bots). The fake accounts on social media are "the foot soldiers in this form of warfare." They work to amplify the message, introduce hashtags and intimidate or block other users. (French CAPS & IRSAM, 2018, p. 83)
Fake followers	Fake followers are anonymous or imposter social media accounts created to portray false impressions of popularity about another account. Social media users can pay for fake followers as well as fake likes, views, and shares to give the appearance of a larger audience. For example, one Englishbased service offers YouTube users a million "high-quality" views and 50,000 likes for $3,150. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Fake news	False, often sensational, information disseminated under the guise of news reporting. (UK Law Commission, 2018, p. vi)
Fake platform	Identity of a web platform is disguised to promote fabricated content. (UK Government Communication Services, 2019, p. 48)
False amplifiers	Coordinated activity by inauthentic accounts with the intent of manipulating political discussion (e.g., by discouraging specific parties from participating in discussion, or amplifying sensationalistic voices over others). (NATO StratCom COE, 2017, p. 68)
Falsehood	Information that is inaccurate, inconsistent with reality, or is based on fabrication or fallacy. (RAND, 2018b, p. 10)
Filter bubbles (information bubble, information cocoon)	A phenomenon whereby the ideological perspectives of internet users are reinforced as a result of the selective algorithmic tailoring of search engine results to individual users (as reflected in recorded data such as search history, click data, and location). (NATO StratCom COE, 2017, p. 68)
Firehose of falsehoods	Even before the 2016 election, observers in Europe and the United States alleged that Russian efforts produced a 'firehose of falsehood,' defined by 'high numbers of channels and messages and a shameless willingness to disseminate partial truths or outright fictions.' (Knight Foundation, 2018, p. 11)

Flagging	User flags give individuals the ability to log complaints of inappropriate content with content moderators. (UN, 2015, p. 12)
Flooding	Flooding operates with a large but limited number of more or less spuriously substantiated narratives, pushed in multiple channels and amplified by botnets, in order to overload the target system's capacity to differentiate credible from incredible. (Swedish Civil Contingency Agency, 2018, p. 75)
Foreign interference	…an attempt to adversely affect, or undermine confidence in, any political, governmental, or democratic process, or prevent the exercise of human or democratic rights, through coercion, corruption, or the use of covert, malicious, or deceptive means, acting from abroad. (Atlantic Council, 2019b, p. 48)
Forgery	Product or content is wholly or partly fabricated to falsely ascribe the identity of the source. (UK Government Communication Services, 2019, p. 48)
Government interference	…conducted on behalf of a government entity or by a foreign actor. (Facebook, 2020b)
Hacking	Use of illegitimate means to unlawfully gain access to, or otherwise disturb the function of, a platform. (UK Government Communication Services, 2019, p. 48)
Hashtag poisoning	Coordinating content on social media to dominate hashtags (reducing the quality of the communications environment) (DEMOS, 2019, p. 19)
Hate speech	We define hate speech as a direct attack on people based on what we call protected characteristics — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We protect against attacks on the basis of age when age is paired with another protected characteristic, and also provide certain protections for immigration status. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation. (Facebook, n.d.)
Hijacking	Unlawful seizure of a computer or an account. (UK Government Communication Services, 2019, p. 48)
Hybrid warfare	…a widespread but confusing notion, which in reality refers to war waged across the full spectrum—from conventional means to information and cyber means, from clandestine operations to nuclear intimidation. (French CAPS & IRSAM, 2018, p. 19)
Hyper-partisan sites	Benkler et al. describe hyper-partisan sites as 'combining decontextualized truths, repeated falsehoods, and leaps of logic to create fundamentally misleading view of the world.' This is not fake news per se, but propaganda rooted in an ideologically-driven worldview. (Data&Society, 2017, p. 21)

Impersonation	Another common - but not universal - feature of information operations is a disguised messenger - that is, the identity of the person who is transmitting the information or engaging in the information activity is concealed, either through anonymity or through impersonation. (DEMOS, 2019, p. 22)
Inauthentic behaviour	…behavior on social media that is conducted for a reason other than the apparent one. Examples of inauthentic behavior include selling likes, following other accounts for pay, and setting up anonymous accounts to covertly promote other organizations. (Atlantic Council, 2019b, p. 48)
Influence operation (influence campaign, information influence)	…the coordinated efforts of a foreign power comprised of several influence activities and/or influence operations where each activity (or operation) has one or several ends of their own intended to help achieve the ends of the influence campaign as a whole. (Swedish Civil Contingency Agency, 2018, pp. 14–15)
Information literacy	Refers to the ability to recognize when information is needed and to locate, evaluate, effectively use and communicate information in its various formats. (UNESCO, 2011, p. 186)
Information operations	A non-kinetic, coordinated attempt to inauthentically manipulate an information environment in a systemic/strategic way, using means which are coordinated, covert and inauthentic in order to achieve political or social objectives. (DEMOS, 2019, p. 12)
Information warfare	Warfare that integrates electronic warfare, cyberwarfare, and psychological operations (PSYOPS) into a single fighting organisation. (NATO StratCom COE, 2017, p. 70)
Laundering	The process of passing of disinformation as legitimate information by gradually distorting it and obscuring its true origin. (UK Government Communication Services, 2019, p. 48)
Leaking	Disseminating unlawfully obtained information. (UK Government Communication Services, 2019, p. 49)
Like farm	…an organization that uses a substantial number of social media accounts to sell likes, reposts, and follows to real users who want to make themselves look more important. Like farms are typically commercial, but the return is low: a like farm exposed in 2018 by DFRLab sold ten thousand likes on Instagram. Like farms can use automated accounts, human users, or any combination of the two. (Atlantic Council, 2019b, p. 48)
Like-bating	When a post explicitly asks News Feed readers to like, comment or share the post in order to get additional distribution beyond what the post would normally receive. (Facebook, 2014a)
Mainstream media	Media disseminated via the largest distribution channels, which are therefore representative of what

	the majority of media consumers are likely to encounter. The term also denotes media that generally reflect the prevailing currents of thought, influence or activity. (UNESCO, 2011, p. 187)
Malinformation	…genuine information that is shared to cause harm. This include private or revealing information spread to harm a person or a reputation. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Maling rhetoric	Lingual ruses aimed at undermining reasonable and legitimate debate and silencing opinions. (UK Government Communication Services, 2019, p. 49)
Manipulated media	Manipulations can be made through simple technology like Photoshop or through sophisticated tools that use artificial intelligence or "deep learning" techniques to create videos that distort reality – usually called "deepfakes." (Facebook, 2020d)
Manufactured amplification	Manufactured Amplification occurs when the reach or spread of information is boosted through artificial means. This includes human and automated manipulation of search engine results and trending lists, and the promotion of certain links or hashtags on social media. There are online price lists for different types of amplification, including prices for generating fake votes and signatures in online polls and petitions, and the cost of downranking specific content from search engine results. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Media and information literacy (MIL)	MIL stands for media and information literacy, and refers to the essential competencies (knowledge, skills and attitude) that allow citizens to engage with media and other information providers effectively and develop critical thinking and life-long learning skills for socializing and becoming active citizens. (UNESCO, 2011, p. 187)
Media literacy	Understanding and using mass media in either an assertive or non-assertive way, including an informed and critical understanding of media, the techniques they employ and their effects. Also the ability to read, analyze, evaluate and produce communication in a variety of media forms (e.g. television, print, radio, computers etc.). Another understanding of the term is the ability to decode, analyze, evaluate and produce communication in a variety of forms. (UNESCO, 2011, p. 188)
Meme	The formal definition of the term meme, coined by biologist Richard Dawkins in 1976, is an idea or behavior that spreads person to person throughout a culture by propagating rapidly, and changing over time. The term is now used most frequently to describe captioned photos or GIFs that spread online, and the most effective are humorous or critical of society. They are increasingly being used as powerful vehicles of disinformation. (Harvard

	Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Microtargeting	Internet platforms enable advertisers to send selective messages to narrow groups of voters based on demographics, interests, geography, and other political variables. In fact, they will also allow targeting of Moreover, despite all the criticism surrounding it, microtargeting allows campaigns and interest groups to efficiently deliver messages to the people they want to reach, as marketers have done for years with direct mail, phone calls, and later—email. Finally, the Internet has proven particularly beneficial for small-donor fundraising, as candidates and groups have used it to raise significant amounts of money from large numbers of donors, thereby 'democratizing' political finance. But online political advertising has also proven fertile ground for all of the democracy- related pathologies attributed to digital communication technologies. Just as microtargeting can assist in mobilization and fundraising efforts, it enables campaigns to send 'custom audiences' derived from lists of individual email addresses provided by campaigns, and 'lookalike audiences', which the platform derives from the immense amount of data it has to match the group provided by a campaign. (Kofi Annan Commission on Elections and Democracy in the Digital Age, 2020)
Misinformation	False or misleading information that is spread, unintentionally, by error or mistake. (RAND, 2018b, p. 10)
News media	The section of the mass media that focuses on presenting current news to the public. It includes print media (e.g. newspapers and magazines), broadcast media (radio and television), and increasingly, Internet-based media (e.g. World Wide Web pages and blogs). (UNESCO, 2011, p. 188)
Organic content	…content generated by human users. (Atlantic Council, 2019b, p. 49)
Organic traffic	…social media traffic that appears to be primarily generated by spontaneous users, rather than a coordinated campaign. (Given the nature of online conversations, even organic traffic is likely to have some automated component—for example, by advertising bots picking up a popular hashtag to increase users' awareness of the product being promoted.) (Atlantic Council, 2019b, p. 49)
Phishing	These fraudsters impersonate legitimate companies or people, sending emails and links that attempt to direct you to false websites, or infect your computer with malware. (LinkedIn, 2019)
Polarization	Polarization contributes both to increasing disagreement regarding facts and analytical interpretations of facts and data and to the blurring of the line between opinion and fact by creating opposing sides, each with its own narrative, worldview, and facts. The groups on each side can become insular in their thinking and communication,

	creating a closed environment in which false information proliferates. (RAND, 2018b, p. xiv—xv)
Populism	Since the 1890s, populism has been used to refer to any sort of movement, on the left or the right, that is anti-elite or anti-status and that promotes the interests of the "forgotten common man" over the interests of big banks, big industry, and other technocrats. (RAND, 2018b, p. 44)
Post-truth	adjective relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief (Ash Center for Democratic Governance and Innovation & Harvard Kennedy School, 2018, p. 5)
Potemkin village	A smoke-screen of institutions and/or platforms established to deceive audiences. (UK Government Communication Services, 2019, p. 50)
Propaganda	Propaganda is true or false information spread to persuade an audience, but often has a political connotation and is often connected to information produced by governments. It is worth noting that the lines between advertising, publicity, and propaganda are often unclear. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Resilience	A third category comprises media and information literacy practices, as education of the public at large is another way of building resilience among citizens, end-users and voters. The purpose is to bolster prevention and to reduce the appeal of disinformation and conspiracy theories. (European Commission, 2018, pp. 16–17)
Satire	Satire is writing that uses literary devices such as ridicule and irony to criticize elements of society. Satire can become misinformation if audiences misinterpret it as fact. There is a known trend of disinformation agents labelling content as satire to prevent it from being flagged by fact-checkers. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Scraping	Scraping is the process of extracting data from a website without the use of an API. It is often used by researchers and computational journalists to monitor mis- and disinformation on different social platforms and forums. Typically, scraping violates a website's terms of service (i.e., the rules that users agree to in order to use a platform). However, researchers and journalists often justify scraping because of the lack of any other option when trying to investigate and study the impact of algorithms. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Shilling	To give credibility to a person or a message without disclosing intentions or relationships. (UK Government Communication Services, 2019, p. 50)
Sock puppet	A sock puppet is an online account that uses a false identity designed specifically to deceive. Sock puppets are used on

	social platforms to inflate another account's follower numbers and to spread or amplify false information to a mass audience. The term is considered by some to be synonymous with the term "bot". (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
Spamming	Sending unsolicited and unwanted e-mails in bulk for advertising purposes. The proliferation of such material, which now accounts for some 85% of all e-mails sent, has become a serious nuisance to business users. (NATO StratCom COE, 2017, p. 71)
Synthetic media	…the catch-all term for AI-manipulated media and data. This includes deepfakes but could also include text generation, speech, etc. (University of Washington Center for an Informed Public, 2020)
Troll	Today, the term "troll" is most often used to refer to any person harassing or insulting others online. However, it has also been used to describe human-controlled accounts performing bot-like activities.
Trollfarm	A troll farm is a group of individuals engaging in trolling or bot-like promotion of narratives in a coordinated fashion. One prominent troll farm was the Russia-based Internet Research Agency that spread inflammatory content online in an attempt to interfere in the U.S. presidential election. (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2018)
User generated content	Also known as consumer-generated media (CGM) and usercreated content, UGC refers to various kinds of publicly-available media content that can be produced by the users of digital media. Those consuming the content therefore also produce content. (UNESCO, 2011, p. 191)
Verified account	Essentially, an account that has been vetted to ensure that it represents whom it says it does receives a blue check mark next to the account name. (RAND, 2018a, p. 39)
Yellow journalism	Most Americans are familiar with "yellow journalism," sensational coverage that sold newspapers at the expense of factual accuracy at the turn of the 20th century. Yellow journalism strongly contributed to the start of the Spanish-American War and (arguably) the U.S. entry into World War I. (Knight Foundation, 2018, p. 9)

Index

Alternative facts, 140
Argumentation, 27
Author, 89
Binding/bonding effect, 25, 26
Categorical imperative, 26
Colonisation, 42
Communication, 59
    Modes of, 35
Communicative
    action, 25
    competences, 25
Context, 87
Corpus, 86
Crisis, 40
Critical discourse analysis, 76
Discourse, 28
    (D), 27
    analysis, 76
    Critical discourse analysis, 76
    Fake news discourse, 96
    Moral discourse, 84
    SKAD, 81
Domains of reality, 34
Echo chambers, 121
Ethics, 30
Fake news, 118
    discourse, 96
Filter-bubbles, 121
Heuristics, 44
Ideal types, 92
Ideology, 42
Immanent analysis, 83
Inductive, 82
Instrumental act, 41
Intuitions
    Moral intuitions, 52
    Social intuitions model, 52
Language, 59
Lifeworld, 23
Modes of communication, 35
Moral, 30
    discourse, 84
    foundations theory, 49
    intuitions, 52
Othering, 29
Ought and Is, 31
Perspective taking, 27
Post-truth, 157
Principle of
    discourse (D), 27
    universalisation (U), 26
Propaganda, 4
Rational, 89
    reconstruction, 75
Rightness, 35
SKAD, 81
Social intuitions model, 52
Speech act, 25
    Communicative action, 25
    Instrumental act, 41
    Strategic act, 41
Strategic act, 41
System, 41
Text, 85
Truth, 35
Truthfulness, 35
Understanding, 24
Universalisation (U), 26
Validity claim, 34
    Rightness, 35
    Truth, 35
    Truthfulness, 35
WEIRD societies, 50
Zeitgeist, 33

References

Adorno, T. W., & Horkheimer, M. (1997). *Dialectic of Enlightenment* (J. Cumming, Trans.). London: Verso Books. [*Dialektik der Aufklarung*, original work published in 1944]

Agamben, G. (1998). *Homo sacer: Sovereign power and bare life* (D. Heller-Roazen, Trans.). United States: Stanford University Press. [*Homo sacer: Il potere sovrano e la nuda vida*, original work published 1995]

Arnett, R. (1987). The status of communication ethics scholarship in speech communication journals from 1915 to 1985. *Central States Speech Journal, 38*(1), 44–61. doi:10.1080/10510978709368229

Arnett, R. C., Arneson, P., & Bell, L. M. (2006). Communcation ethics: The dialogical turn. *Review of Communication, 6*(1–2), 62–92. doi:10.1080/15358590600763334

Ash Center for Democratic Governance and Innovation, & Harvard Kennedy School. (2018). *Something has cracked: Post-truth politics and Richard Rorty's postmodernist burgeois liberalism.* Retrieved from https://ash.harvard.edu/files/ash/files/post-truth-politics-rorty.pdf

Atlantic Council. (2018). *Democratic Defense Against Disinformation.* Retrieved from https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/resources/docs/Atlantic%20Council_DemocraticDefenseDisinformation.pdf

Atlantic Council. (2019a). *Democratic Defense Against Disinformation 2.0.* Retrieved from https://www.atlanticcouncil.org/wp-content/uploads/2019/06/Democratic_Defense_Against_Disinformation_2.0.pdf

Atlantic Council. (2019b). *Disinformation in democracies: Strengthening digital resilience in Latin America.* Retrieved from https://www.atlanticcouncil.org/wp-content/uploads/2019/09/Disinformation-in-Democracies.pdf

Austin, J. L. (1962). *How to do things with words.* Great Britain: Oxford University Press.

Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2013). Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping, 34*(12), 3168–3181. doi:10.1002/hbm.22133

Bagozzi, R. P., Verbeke, W. J. M. I., Dietvorst, R. C., Belschak, F. D., Berg, W. E. v. d., & Rietdijk, W. J. R. (2013). Theory of mind and empathic explanations of Machiavellianism: A neuroscience perspective. *Journal of Management, 39*(7), 1760–1798. doi:10.1177/0149206312471393

Barton, E., & Eggly, S. (2009). Ethical or unethical persuasion? The rhetoric of offers to participate in clinical trials. *Written Communication, 26*(3), 295–319. doi:10.1177/0741088309336936

Batson, C. D., Thompson, E. R., & Chen, H. (2002). Moral hypocrisy: Addressing some alternatives. *Journal of Personality and Social Psychology, 83*(2), 330–339. doi:10.1037/0022-3514.83.2.330

Baym, G. (2005). The Daily Show: Discursive integration and the reinvention of political journalism. *Political Communication, 22*(3), 259–276. doi:10.1080/10584600591006492

Benabou, R., & Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics, 107*(3), 921–958. doi:10.2307/2118369

Benjamin, W. (1999). *The arcades project* (H. Eiland & K. McLaughlin, Trans.). London, England: Belknap Press of Harvard University Press. [*Das Passagen-Werk*, original work published 1982]

Bentele, G., & Nothhaft, H. (2010). Strategic communication and the public sphere from a European perspective. *International Journal of Strategic Communication, 4*(2), 93–116. doi:10.1080/15531181003701954

Bentele, G., & Nothhaft, H. (2015). Strategic communication and the public sphere from a European perspective. In D. Holtzhausen & A. Zerfaß (Eds.), *The Routledge handbook of strategic communication* (pp. 53–73). Routledge.

Berger, P. L., & Luckmann, T. (2011). *The social construction of reality: A treatise in the sociology of knowledge*. New York: Open Road Integrated Media.

Bernays, E. L. (2005). *Propaganda*. Brooklyn, New York: IG Publishing.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin, 88*(1), 1–45. doi:10.1037/0033-2909.88.1.1

Bloom, P. (2010). How do morals change? *Nature, 454*, 490. doi:10.1038/464490a

Bloom, P. (2016). *Against empathy: The case for rational compassion*. London: Bodley Head.

Bloor, D. (1991). *Knowledge and social imagery* (2 ed.). London: Univeristy of Chicago Press.

Bowen, S. A. (2018). Strategic communication, ethics of. In R. L. Heath & W. Johansen (Eds.), *The international encyclopedia of strategic communication* (pp. 1–11). doi:10.1002/9781119010722.iesc0074

Braddock, R. (1958). An extension of the "Lasswell formula". *Journal of Communication, 8*(2), 88–93. doi:10.1111/j.1460-2466.1958.tb01138.x

Brookings Institution. (2019, 10 January). How campaigns can protect themselves from deepfakes, disinformation, and social media manipulation. Retrieved from https://www.brookings.edu/articles/how-campaigns-can-protect-themselves-from-deepfakes-disinformation-and-social-media-manipulation/

Bruns, A. (2007, 18 February). Habermas and/against the Internet. Retrieved from https://snurb.info/node/621

Buhmann, A., Paßmann, J., & Fieseler, C. (2020). Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse. *Journal of Business Ethics, 163*, 265–280. doi:10.1007/s10551-019-04226-4

Canadian House of Commons. (2018). *Democracy under threat: Risks and solutions in teh era of disinformation and data monopoly*. Retrieved from https://publications.gc.ca/collections/collection_2018/parl/xc73-1/XC73-1-1-421-17-eng.pdf

Carey, J. W. (2009). *Communication as culture: Essays on media and society*. New York: Routledge.

Cassaniti, J. L., & Hickman, J. R. (2014). New directions in the anthropology of morality. *Anthropological Theory, 14*(3), 251–262. doi:10.1177/1463499614534371

Cecil, K., Marshall, J., & Bloom, P. (2020). Children's reasoning about in-group and out-group obligations. *The Yale Undergraduate Research Journal, 1*(1), 1–9.

Center for Business and Human Rights. (2019). *Tackling domestic disinformation*. Retrieved from https://media.business-humanrights.org/media/documents/files/documents/NYU20Domestic20Disinformation_Digital20Version.pdf

Center for European Policy Analysis. (2016). *Winning the information war*. Retrieved from https://cepa.org/comprehensive-reports/winning-the-information-war/

Centre for Countering Digital Hate. (2020). *The anti-vaxx industry*. Retrieved from https://counterhate.com/wp-content/uploads/2022/05/200112-The-Anti-Vaxx-Industry.pdf

Centre for the Study of Democratic Institutions. (2018). *Digital threats to democratic elections: How foreign actors use digital techniques to undermine democracy*. Retrieved from https://democracy2017.sites.olt.ubc.ca/files/2018/01/DigitalThreats_Report-FINAL.pdf

Chatham house. (2016). *Agents of the Russian world proxy groups in teh contested neighbourhood*. Retrieved from https://www.chathamhouse.org/sites/default/files/publications/research/2016-04-14-agents-russian-world-lutsevych.pdf

Chen, C. (2017). Science mapping: A systematic review of the literature. *Journal of Data and Information Science, 2*(2), 1–40. doi:10.1515/jdis-2017-0006

Cheney, G., May, S., & Munshi, D. (Eds.). (2011). *The handbook of communication ethics*: Routledge.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christensen, L. T., & Svensson, E. (2017). The nature of strategic communication: A rejoinder to Nothhaft. *International Journal of Strategic Communication, 11*(3), 180–183. doi:10.1080/1553118X.2017.1318883

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science, 22*(3), 306–313. doi:10.1177/0956797610397667

Cikara, M., Bruneau, E. G., Bavel, J. J. v., & Saxe, R. R. (2014). Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of Experimental Social Psychology, 55*, 110–125. doi:10.1016/j.jesp.2014.06.007

Cooke, M. (2019). Transcendence in Postmetaphysical Thinking: Habermas's God. *European Journal for Philosophy of Religion, 11*(4), 21–44. doi:10.24204/EJPR.V11I4.2685

Cooren, F. o. (2000). Toward another ideal speech situation: A critique of Habermas' reinterpretation of speech act theory. *Quarterly Journal of Speech, 86*(3), 295–317. doi:10.1080/00335630009384298

Corchia. (2021). *A bibliography. 1. Works of Jürgen Habermas (1952–2021).*

Council of Europe. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Retrieved from https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Craig, R. T. (1999). Communication theory as a field. *Communication Theory, 9*(2), 119–161. doi:10.1111/j.1468-2885.1999.tb00355.x

Cukier, W., Ngwenyama, O., Bauer, R., & Middleton, C. (2009). A critical analysis of media discourse on information technology: Preliminary results of a proposed method for critical discourse analysis. *Information Systems Journal, 19*, 175–196. doi:10.1111/j.1365-2575.2008.00296.x

Czingon, C., Diefenbach, A., & Kempf, V. (2020). Moral universalism at a time of political regression: A conversation with Jürgen Habermas about the present and his life's work. *Theory, Culture & Society, 37*(7–8), 11–36. doi:10.1177/0263276420961146

Dahlgren, P. (2005). The internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication, 22*(2), 147–162. doi:10.1080/10584600590933160

Danish Ministry of Defence. (2018, 7 September). Strengthened safeguards against foreign influence on Danish elections and democracy. Retrieved from https://www.fmn.dk/en/news/english/strengthened-safeguards-against-foreign-influence-on-danish-elections-and-democracy/

Das, V. (2012). Ordinary ethics. In D. Fassin (Ed.), *A companion to moral anthropology* (pp. 133–149). doi:10.1002/9781118290620.ch8

Data&Society. (2017). *Media manipulation and disinformation online*. Retrieved from https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf

Data&Society. (2018). *The oxygen of amplification: Better practices for reporting on extremism, antagonism, and manipulation online*. Retrieved from https://datasociety.net/wp-content/uploads/2018/05/1_PART_1_Oxygen_of_Amplification_DS.pdf

Dawkins, R. (1999). *The selfish gene*. New york, NY: Oxford University Press.

Deetz, S. (1994). Future of the discipline: The challenges, the research, and the social contribution. *Annals of the International Comunication Association, 17*(1), 565–600. doi:10.1080/23808985.1994.11678904

Deflem, M. (1994). Habermas, modernity and law. *Philosophy & Social Criticism, 20*(4), 151–166. doi:10.1177/019145379402000407

DEMOS. (2019). *Warring Songs: Information operations in the digital age*. Retrieved from https://demos.co.uk/wp-content/uploads/2019/10/Warring-Songs-final-1.pdf

Dews, P. (2017, 10 November). Review: Postmetaphysical thinking II. *Notre Dame Philosophical Reviews.* Retrieved from https://ndpr.nd.edu/reviews/postmetaphysical-thinking-ii/

Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences, 151*, 1–6. doi:10.1016/j.paid.2019.109547

Eberhardt, N. (2014). Piaget and Durkheim: Competing paradigms in the anthropology of morality. *Anthropological Theory, 14*(3), 301–316. doi:10.1177/1463499614534552

Ellemers, N., Toorn, J. v. d., Paunov, Y., & Leeuwen, T. v. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review, 23*(4), 332–366. doi:10.1177/1088868318811759

EU Committee of Ministers. (2019, 13 February). Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes. Retrieved from https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

European Commission. (2016). *Recommendations and declarations of the Committee of Ministers of the Council of Europe in the field of media and information society*. Retrieved from https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680645b44

European Commission. (2018). *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*. Retrieved from https://coinform.eu/wp-content/uploads/2019/02/EU-High-Level-Group-on-Disinformation-A-multi-dimensionalapproachtodisinformation.pdf

European Commission. (2018a). *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*. Retrieved from https://coinform.eu/wp-content/uploads/2019/02/EU-High-Level-Group-on-Disinformation-A-multi-dimensionalapproachtodisinformation.pdf

European Commission. (2018b). *EU Code of Practice on Disinformation*. Retrieved from https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

European Data Protection Supervisor. (2018). *Opinion and online manipulation and personal data*. Retrieved from https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf

European Parliament. (2019). *Disinformation and propaganda: Impact on the functioning of the rule of law in the EU and its member states*. Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf

EUvsDisinformation. (2018). *The strategy and tactics of the pro-Kremlin disinformation campaign*. Retrieved from https://euvsdisinfo.eu/the-strategy-and-tactics-of-the-pro-kremlin-disinformation-campaign/

Facebook. (2014a, 10 April). Cleaning up news feed spam. Retrieved from https://about.fb.com/news/2014/04/news-feed-fyi-cleaning-up-news-feed-spam/

Facebook. (2014b, 25 August). News feed FIY: Clickbaiting. Retrieved from https://about.fb.com/news/2014/08/news-feed-fyi-click-baiting/

Facebook. (2016, 15 December). Addressing hoaxes and fake news. Retrieved from https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/

Facebook. (2017a, 9 August). Addressing cloaking so people see more authentic posts. Retrieved from https://about.fb.com/news/2017/08/news-feed-fyi-addressing-cloaking-so-people-see-more-authentic-posts/

Facebook. (2017b, 6 September). An update on information operations on Facebook. Retrieved from https://about.fb.com/news/2017/09/information-operations-update/

Facebook. (2017c, 22 November). Continuing transparency on russian activity. Retrieved from https://about.fb.com/news/2017/11/continuing-transparency-on-russian-activity/

Facebook. (2018a, 25 January). Ariadne Vromen: Is social media good or bad for democracy? Retrieved from https://about.fb.com/news/2018/01/vromen-democracy/

Facebook. (2018b, 3 April). Authenticity matters: The IRA has no place on Facebook. Retrieved from https://about.fb.com/news/2018/04/authenticity-matters/

Facebook. (2018c, 22 January). Cass R. Sunstein: Is social media good or bad for democracy? Retrieved from https://about.fb.com/news/2018/01/sunstein-democracy/

Facebook. (2018d, 19 January). Helping ensure news on Facebook is from trusted sources. Retrieved from https://about.fb.com/news/2018/01/trusted-sources/#:~:text=We%20decided%20that%20having%20the,whether%20they%20trust%20that%20source.

Facebook. (2019a, 12 June). Automation plus expert journalism: How full fact is fighting misinformation. Retrieved from https://about.fb.com/news/2019/06/inside-feed-full-fact-interview/

Facebook. (2019b, 21 October). How we respond to inauthentic behaviour on our platforms: Policy update. Retrieved from https://about.fb.com/news/2019/10/inauthentic-behavior-policy-update/

Facebook. (2019c, 24 Jnauary). Understanding Facebook's business model. Retrieved from https://about.fb.com/news/2019/01/understanding-facebooks-business-model/

Facebook. (2019d). Understanding social media and conflict. Retrieved from https://about.fb.com/news/2019/06/social-media-and-conflict/

Facebook. (2019e, 28 August). Updates to ads about social issues, elections or politics in the US. Retrieved from https://about.fb.com/news/2019/08/updates-to-ads-about-social-issues-elections-or-politics-in-the-us/

Facebook. (2020a, 22 October). A privacy-protected way to refer cases to the oversight board. Retrieved from https://about.fb.com/news/2020/10/privacy-protected-way-to-refer-cases-to-oversight-board/

Facebook. (2020b, 5 May). April 2020 coordinated inauthentic behaviour report. Retrieved from https://about.fb.com/news/2020/05/april-cib-report/#:~:text=Summary%20of%20April%202020%20Findings,their%20respective%20countries%20(CIB).

Facebook. (2020c, 19 November). Combating clone sites. Retrieved from https://about.fb.com/news/2020/11/combating-clone-sites/

Facebook. (2020d, 6 January). Enforcing against manipulated media. Retrieved from https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/

Facebook. (2020e, 27 May). Investment to fight polarization. Retrieved from https://about.fb.com/news/2020/05/investments-to-fight-polarization/

Facebook. (2020f, 20 October). Taking action against fake engagement and ad scams. Retrieved from https://about.fb.com/news/2020/10/taking-action-against-fake-engagement-and-ad-scams/

Facebook. (2020g, 9 October). The next web conversation on preparing for the US elections and balancing regulation and innovation. Retrieved from https://about.fb.com/news/2020/10/the-next-web-preparing-for-elections-regulation-innovation/

Facebook. (n.d.). Hatespeech. Retrieved from https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/#

Facebook. (n.d.a). Bullying and harrassment. Retrieved from https://transparency.fb.com/en-gb/policies/community-standards/bullying-harassment/

Facebook. (n.d.b). Hatespeech. Retrieved from https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/#

Facebook. (n.d.c). Spam. Retrieved from https://transparency.fb.com/en-gb/policies/community-standards/spam/

Farkas, J., & Schou, J. (2020). *Post-truth, fake news and democracy*: Routledge.

Feldner, S. B., & Meisenbach, R. J. (2007). SaveDisney.com and activist challenges: A Habermasian perspective on corporate legitimacy. *International Journal of Strategic Communication, 1*(4), 207–226. doi:10.1080/15531180701623569

Finlayson, J. G. (2014). Hegel, Adorno and the origins of immanent criticism. *British Journal for the History of Philosophy, 22*(6), 1142–1166. doi:10.1080/09608788.2014.993918

Fjällhed, A. (2020). Managing disinformation through public diplomacy. In P. Surowiec & I. Manor (Eds.), *Public Diplomacy and the Politics of Uncertainty* (pp. 227–253). Palgrave Macmillian.

Fjällhed, A., & Pamment, J. (2023). Disinformation. In E. Gilboa (Ed.), *A research agenda for public diplomacy* (pp. 173–186). Edward Elgar Publishing.

Fjällhed, A., Pamment, J., & Bay, S. (2021). A Swedish perspective on foreign election interference. In J. D. Ohlin & D. B. Hollis (Eds.), *Defending democracies: Combatting foreign election interference in a digital age* (pp. 139–162). doi:10.1093/oso/9780197556979.003.0007

Fjällhed, A., Sandre, A., & Lüfkens, M. (2024). New trends in digital diplomacy: The rise of TikTok and the geopolitics of algorithmic governance. In C. Bjola & I. Manor (Eds.), *The Oxford handbook of public diplomacy* (pp. 173–186).

Flowerdew, J., & Richardson, J. R. (2018). *The Routledge handbook of critical discourse studies*. London: Routledge.

Foessel, M. (2018). Critique and communication: Philosophy's missions. A conversation with Jürgen Habermas. In P. E. Gordon, E. Hammer, & A. Honneth (Eds.), *The Routledge companion to the Frankfurt School.*

Fornäs, J. (2013). The dialectics of communicative and immanent critique. *TripleC, 11*(2), 504–514. doi:10.31269/triplec.v11i2.504

Foucault, M. (1980). *Power/Knowledge*. New York: Pantheon.

Foucault, M. (1981). The order of discourse. In R. J. C. Young (Ed.), *Untying the text: A post-structuralist reader* (pp. 8–78). Routledge Kegan & Paul.

Foucault, M. (1990). *The history of sexuality. Volume 1: An introduction* (R. Hurley, Trans.). New York: Vintage Books. [*La volenté de savoir* original work published 1976]

Foucault, M. (1994a). *The order of things: An archeology of human sciences* (A. Sheridan, Trans.). New York: Vintage Books. [*Les mots et les choses: Une archéologie des sciences humaines*, original work published 1966]

Foucault, M. (1994b). Two lectures. In M. Kelly (Ed.), *Critique and power: Recasting the Foucault/Habermas Debate* (pp. 17–46). Cambridge, Massachusetts: The MIT Press.

Foucault, M. (1995). *Discipline and punishment: The birth of the prison* (A. Sheridan, Trans.). Vintage Books. [*Surveiller et punir: Naissance de la prison*, original work published 1975]

Foucault, M. (1999). What is an author? In J. D. Fabion (Ed.), *Aesthetics, method, and epistemology* (pp. 205–222). The New Press. [original work published 1977]

Foucault, M. (2010). *The archeology of knowledge: And the discourse on language* (A. M. S. Smith, Trans.). New York: Vintage Books. [*L'Archéologie du Savoir*, original work published 1969]

Foucault, M. (2013). *Madness and civilization: A history of insanity in the age of reason* (R. Howard, Trans.). New York: Vintage Books. [*Folie et* *déraison: Histoire de la folie à l'âge classique*, original work published 1961]

Foucault, M. (2019). *Ethics: Subjectivity and truth* (R. Hurley, Trans.). New York: New Press. [*Dits and ecrits, 1954–1988*, original work published 1994]

Frankental, P. (2001). Corporate social responsibility—a PR invention? *Corporate Communications: An International Journal, 6*(1), 18–23. doi:10.1108/13563280110381170

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text, 25/26*(56-80). doi:10.2307/466240

French CAPS & IRSAM. (2018). *Information manipulation: A challange for our democracies*. Retrieved from https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf

French CAPS & IRSAM. (2018). *Information manipulation: A challange for our democracies*. Retrieved from https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf

French, W., & Allbright, D. (1998). Resolving a moral conflict through discourse. *Journal of Business Ethics, 17*(177–194). doi:10.1023/A:1005792208264

Gaus, D. (2019). Rational reconstruction. In A. Allen & E. Mendieta (Eds.), *The Cambridge Habermas lexicon* (pp. 369–378).

Gazzaniga, M. S. (1990). *The social brain*: Basic Books.

Gee, J. P., & Handford, M. (2013). *The Routledge handbook of discourse analysis*: Routledge.

Gellen, S., & Lowe, R. (2021). (Re) constructing social hierarchies: A critical discourse analysis of an international charity’s visual appeals. *Critical Discourse Studies, 18*(2), 280–300. doi:10.1080/17405904.2019.1708762

Georgakopoulou, A., & Fina, A. D. (2020). *The Cambridge handbook of discourse studies*: Cambridge University Press.

German Marshall Fond. (2018, 19 April). Ten legislative proposals to defend america against foreign influence operations. Retrieved from https://www.gmfus.org/news/ten-legislative-proposals-defend-america-against-foreign-influence-operations

Gert, B. (1988). *A new justification of the moral rule*. Oxford, UK: Oxford University Press.

Global Health Security Network. (2020). *The COVID-19 pandemic vs post-truth*. Retrieved from https://www.ghsn.org/Policy-Reports/

GLOBSEC Policy Institute. (2016). *Russia's information war in central Europe: New trends and counter-measures*. Retrieved from https://www.globsec.org/sites/default/files/2016-08/russias_information_war_in_central_europe.pdf

Gode, H. E. (2019). Employee ideation on internal social media: addressing uncertainty through dialogue strategies. *International Journal of Strategic Communication, 13*(5), 385–403. doi:10.1080/1553118X.2019.1654481

Google. (2017a, 6 July). Make it easier for publishers to share fact check content. Retrieved from https://blog.google/outreach-initiatives/google-news-initiative/making-it-easier-publishers-share-fact-check-content/

Google. (2017b, 30 October). Security and disinformation in the U.S. 2016 election. Retrieved from https://blog.google/outreach-initiatives/public-policy/security-and-disinformation-us-2016-election/

Google. (2017c, 30 October). Security and disinformation in the U.S. 2016 election: Our work going forward.

Google. (2017d, 30 October 2017). Security and disinformation in the U.S. 2016 election: Steps against phishing and hacking.

Google. (2017e, 30 October). Security and disinformation in the U.S. 2016 election: What we found.

Google. (2018, 23 August). An update on state-sponsored activity. Retrieved from https://blog.google/technology/safety-security/update-state-sponsored-activity/

Google. (2020, 22 June). Bringing fact check information to Google images. Retrieved from https://blog.google/products/search/bringing-fact-check-information-google-images/

Government of Ireland. (2019). *Interdepartmental group on security of Ireland's electoral process and disinformation*. Retrieved from https://assets.gov.ie/39188/8c7b6bc1d0d046be915963abfe427e90.pdf

Graber, S. M. (2017). War of perception: A Habermasian discourse analysis of human shield newspaper reporting during the 2014 Gaza War. *Critical Studies in Media Communication, 34*(3), 293–307. doi:10.1080/15295036.2017.1294759

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (pp. 55–130). Academic press.

Graham, P. (2018). Ethics in critical discourse analysis. *Critical Discourse Studies, 15*(2), 186–203. doi:10.1080/17405904.2017.1421243

Graphica. (2019). *The IRA CopyPasta Campaign*. Retrieved from https://www.congress.gov/116/meeting/house/110106/documents/HHRG-116-JU00-20191022-SD003.pdf

Graphica. (2020). *From Russia with blogs: GRU operators leveraged blogs, social media accounts and private messaging to reach audiences across Europe*.

Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin Press.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgement work. *Trends in Cognitive Science, 6*(12), 157–523. doi:10.1016/S1364-6613(02)02011-9

Gregersen, T. (2022a). *Books and Articles on Jürgen Habermas, Part 3: 1992–2000*. Retrieved from https://www.habermasforum.dk/secondary-literature/articles-and-books-on-habermas-1992-2000?sort=p.sort_order&order=ASC&limit=7

Gregersen, T. (2022b). *Books and Articles on Jürgen Habermas, Part 4: 2001–2009*. Retrieved from https://www.habermasforum.dk/secondary-literature/articles-and-books-on-habermas-2001-2009?sort=p.sort_order&order=ASC&limit=7

Gregersen, T. (2022c). *Books and Articles on Jürgen Habermas, Part 5: 2010–2018*. Retrieved from https://www.habermasforum.dk/secondary-literature/articles-and-books-on-habermas-2010-2018?sort=p.sort_order&order=ASC&limit=7

Gregersen, T. (2023a). *Books and Articles on Jürgen Habermas, Part 1: 1961–1981*. Retrieved from https://www.habermasforum.dk/secondary-literature/articles-and-books-on-habermas-1961-1981?sort=p.sort_order&order=ASC&limit=7

Gregersen, T. (2023b). *Books and Articles on Jürgen Habermas, Part 2: 1982–1991*. Retrieved from https://www.habermasforum.dk/secondary-literature/articles-and-books-on-habermas-1982-1991?sort=p.sort_order&order=ASC&limit=7

Gregersen, T. (2023c). *Books and Articles on Jürgen Habermas, Part 6: 2019–2023*. Retrieved from https://www.habermasforum.dk/secondary-literature/articles-and-books-on-habermas-2019-2022?sort=p.sort_order&order=ASC&limit=7

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. doi:0.1177/1525822X05279903

Habermas, J. (1970). Towards a theory of communicative competence. *Inquiry, 13*(1–4), 360–375. doi:10.1080/00201747008601597

Habermas, J. (1979). *Communication and the evolution of society* (T. McCarthy, Trans.). Boston: Polity Press. [*Sprachpragmatik und philosophie and Zur rekonstruktion des historischen materialismus*, original work published 1976]

Habermas, J. (1982). The entwinement of myth and enlightenment: Re-reading dialectic of enlightenment. *New German Critique, 26*, 13–30. doi:10.2307/488023

Habermas, J. (1984). *The theory of communicative action: Vol 1: Reason and the rationalization of society* (T. McCarthy, Trans.). Polity Press. [*Theorie des kommunikativen ahndelns. Band 1: Handlungsrationalität und gesellschaftliche rationalisierung*, original work published 1981]

Habermas, J. (1986). The new obscurity: The crisis of the welfare state and the exhaustion of utopian energies. *Philosophy & Social Criticism, 11*(2), 1–18. doi:10.1177/019145378601100201

Habermas, J. (1987). *The theory of communicative action: Vol 2: Lifeworld and system: A critique of functionalist reason* (T. McCarthy, Trans.). Beacon Press. [*Theorie des kommunikativen ahndelns. Band 2: Zur kritik der funktionalistischen vernunft*, original work published 1981]

Habermas, J. (1988). *On the logic of the social sciences* (S. W. Nicholsen & J. A. Stark, Trans.). Polity Press. [*Zur logik der sozialwissenschaften*, original work published 1970]

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of Burgeois society* (T. Burger, Trans.). Polity Press. [*Strukturwandel der Öffenlicheit*, original work published 1962]

Habermas, J. (1989a). *The new conservatism: Cultural criticism and the historians' debate* (S. W. Nicholsen, Trans.). Polity Press. [original work published 1985]

Habermas, J. (1994). *Justification and application: Remarks on discourse ethics* (C. P. Cronin, Trans.). MIT Press. [*Erläuterungen zur diskursethik*, original work published 1991]

Habermas, J. (1997). *A Berlin republic: Writings on Germany* (S. Rendall, Trans.). University of Nebraska Press. [*Die normalität einer Berliner republik*, original work published 1995]

Habermas, J. (1998). *The inclusion of the Other: Studies in political theory* (C. Cronin, Trans.). Polity Press. [*Die einbziehung des anderen: Studien zur politischen theorie,* original work published 1996]

Habermas, J. (1999). *Moral consciousness and communicative action* (C. Lenhardt & S. W. Nicholsen, Trans.). Cambridge, Massachusetts: MIT Press. [*Moralbewußtsein und kommunikatives handeln*, original work published 1983]

Habermas, J. (2002). *Religion and rationality: Essays on reason, God, and modernity* (E. Mendieta Ed.). Polity Press. [Original work published 1981–1997]

Habermas, J. (2003a). *On the pragmatics of social interaction: Preliminary studies in the theory of communicative action* (B. Fultner, Trans.). Polity Press.

[*Vorstudien und ergänzungen zur theorie des kommunikativen handelns,* original work published 2003]

Habermas, J. (2003b). *The future of human nature*: Polity Press. [Original work published 2001]

Habermas, J. (2006). Religious tolerance—The Peacemaker for cultural rights. In L. Thomassen (Ed.), *The Derrida-Habermas Reader* (pp. 195–207). Edinburgh: Edinburgh University Press.

Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory, 16*(4), 411–426. doi:10.1111/j.1468-2885.2006.00280.x

Habermas, J. (2007). *Legitimation crisis* (T. McCarthy, Trans.). Polity Press. [*Legitimationsprobleme im spätkapitalismus,* original work published 1973]

Habermas, J. (2007b). *The philosophical discourse of modernity: Twelve lectures* (F. Lawrence, Trans.). Polity Press. [*Der philosophische diskurs der moderne: Zwölf vorlesungen*, original work published 1985]

Habermas, J. (2008). *Between naturalism and religion* (C. Cronin, Trans.). Polity Press. [*Zwischen naturalismus und religion*, original work published 2005]

Habermas, J. (2009). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Regh, Trans.). Polity Press. [*Faktizität und geltung: Beiträge zur diskurstheorie des rechts und des demokratischen rechtsstaats*, original work published 1992]

Habermas, J. (2010a). The concept of human dignity and the realistic utopia of human rights. *Metaphilosophy, 41*(4), 464–480. doi:10.1111/j.1467-9973.2010.01648.x

Habermas, J. (2010b). The 'Good Life'—A 'detestable phrase': The significance of the young Rawls's religious ethics for his political theory. *European Journal of Philosophy, 18*(3), 443–454. doi:10.1111/j.1468-0378.2010.00424.x

Habermas, J. (2016). For a democratic polarisation: How to pull the ground from under right-wing populism. Retrieved from https://www.socialeurope.eu/democratic-polarisation-pull-ground-right-wing-populism

Habermas, J. (2017). *Postmetaphysical thinking II: Essays and replies* (C. Cronin, Trans.). Polity Press. [*Nachmetaphysisches denken II: Aufsätze und repliken*, original work published 2012]

Habermas, J. (2018a). Interview with Jürgen Habermas. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. Warren (Eds.), *The Oxford handbook of deliberative democracy* (pp. 871–882). doi:10.1093/oxfordhb/9780198747369.013.60

Habermas, J. (2018b). *Philosophical introductions: Five approaches to communicative reason* (C. Cronin, Trans.). Polity Press. [*Philosophische texte: Studienausgabe in fünf bänden*, original work published 2009]

Habermas, J. (2020a). Commentary on, Christina Lafront, Democracy withouth shortcuts. *Journal of Deliberative Democracy, 16*(2), 10–14. doi:10.16997/jdd.397

Habermas, J. (2020b). From formal semantics to transcendental pragmatics: Karl-Otto Apel's original insight. *Philosophy & Social Criticism, 46*(6), 627–650. doi:10.1177/0191453720930837

Habermas, J. (2021). Once again: On the relationship between morality and ethical life. *European Journal of Philosophy, 29*(3), 543–551. doi:10.1111/ejop.12716

Habermas, J. (2022a). Foreword. In E. Prattico (Ed.), *Habermas and the crisis of democracy: Interviews with leading thinkers* (pp. xiii–xix). Routledge.

Habermas, J. (2022b). Reflections and hypotheses on a further structural transformation of the political public sphere. *Theory, Culture & Society, 39*(4), 145–117. doi:10.1177/02632764221112341

Habermas, J., & Ratzinger, J. (2006). *The dialectics of secularization: On reason and religion*: Ignatius Press. [*Dialektik der säkulariserung: Uber vernunft und religion,* original work published 2005]

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. doi:10.1037/0033-295X.108.4.814

Haidt, J. (2003). The emotional dog does learn new tricks: A reply to Pizarro and Bloom. *Psychological Review, 110*(1), 197–198. doi:10.1037/0033-295X.110.1.197

Haidt, J. (2004). The emotional dog gets mistaken for a possum. *Review of General Psychology, 8*(4), 283–290. doi:10.1037/1089-2680.8.4.283

Haidt, J. (2008). The emotional dog and its rational tail: A social intuitionist approach to moral judgement. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 1024–1052).

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*: Pantheon Books.

Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education, 42*(3), 281–297. doi:10.1080/03057240.2013.817327

Haidt, J., Björklund, F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason*. University of Virginia. Retrieved from https://polpsy.ca/wp-content/uploads/2019/05/haidt.bjorklund.pdf

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*, 98–116. doi:10.1007/s11211-007-0034-z

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*(4), 55–66. doi:10.1162/0011526042365555

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*(4), 613–628. doi:10.1037/0022-3514.65.4.613

Hallahan, K., Holtzhausen, D. R., Ruler, B. v., Verčič, D., & Sriramesh, K. (2007). Defining strategic communication. *International Journal of Strategic Communication, 1*(1), 3–35. doi:10.1080/15531180701285244

Hammersley, M. (2020). Ethics and the study of disinformation. In A. D. Fina & A. Georgakopoulou (Eds.), *The Cambridge Handbook of Discourse Studies* (pp. 465–486). Cambridge University Press.

Hartman, R. S. (1967). *The structure of value: Foundations of scientific axiology.* Carondale: Southern Illinois University Press.

Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy. (2018). *Information Disorder: The Essential Glossary.* Retrieved from https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf

Heath, C., & Heath, D. (2008). *Made to stick: Why some ideas survive and others die*: Random House UK.

Heath, J. (2019). Validity claim. In A. Allen & E. Mendieta (Eds.), *The Cambridge Habermas Lexicon* (pp. 462–463). Cambridge University Press.

Hegel. (1979). *The phenomenology of spirit* (A. V. Miller, Trans.). Oxford University Press. [*Phänomenologie des geistes*, original work published 1807]

Heinrich Böll Foundation. (2020). *Misinformation, disinformation, malinformation: Causes, trends, and their influence on democracy.* Retrieved from https://www.boell.de/en/2020/08/25/misinformation-disinformation-malinformation

Hendley, S. (2000). *From communicative action to the face of the other: Levinas and Habermas on language, obligation, and community*: Lexington Books.

Hendley, S. (2004). Speech and sensibility: Levinas and Habermas on the constitution of the moral point of view. *Contintental Philosophy Review, 37*, 153–173. doi:0.1007/s11007-005-3926-3

Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*: Farrar, Straus and Giroux.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioural and Brain Sciences, 33*(2–3), 62–83.

Hermoso, B. (2018, 25 May). Jürgen Habermas: 'For God's sake, spare us governing philosophers!'. Retrieved from https://english.elpais.com/elpais/2018/05/07/inenglish/1525683618_145760.html

Herzog, B. (2018). Invisibilization and silencing as an ethical and sociollgical challenge. *Social Epistemology, 32*(1), 13–23. doi:10.1080/02691728.2017.1383529

Hickman, L. A. (2007). *Ragmatism as post-postmodernism: Lessons from John Dewey* (Vol. 21). Fordham University Press.

Hogan, B., & Marcelle, L. (2022). Real interests and incoherent desires. *The Journal of Speculative Philosophy, 36*(1), 51–68. doi:10.5325/jspecphil.36.1.0051

Holtzhausen, D., & Zerfaß, A. (2014). Strategic communication: Opportunities and challenges of the research area. In D. Holtzhausen & A. Zerfaß (Eds.), *The Routledge handbook of strategic communciation* (pp. 3–17). Routledge.

Holtzhausen, D. R. (2010). Communication in the public sphere: The political context of strategic communication. *International Journal of Strategic Communication, 4*(2), 75–77. doi:10.1080/15531181003730037

Holtzhausen, D. R., & Hallahan, K. (2007). Strategic directions for new journal. *International Journal of Strategic Communication, 1*(1), 1–2. doi:10.1080/15531180701285087

Holtzhausen, D. R., & Zerfaß, A. (2013). Strategic communication: Pillars and perspectives on an alternate paradigm. In Sriramesh, A. Zerfaß, & Kim (Eds.), *Current trends and emerging topics in public relations and communication management* (pp. 283–302). New York: Routledge.

Honneth, A., & Joas, H. (Eds.). (1991). *Communicative action: Essays on Jürgen Habermas' the theory of communicative action*. Cambridge, MA: MIT Press.

Hook, D. (2001). Discourse, knowledge, materiality, history: Foucault and discourse analysis. *Theory & Psychology, 11*(4), 521–547. doi:10.1177/0959354301114006

Huang, Y.-H. C., Ao, S., Lu, Y., Ip, C., & Kao, L. (2017). How trust and dialogue shape political participation in mainland China. *International Journal of Strategic Communication, 11*(5), 395–414. doi:10.1080/1553118X.2017.1368020

Husserl, E. (1970). *The crisis of European science and transcendental phenomenology: An introduction to phenomenological philosophy* (D. Carr, Trans.). Northwestern University Press. [*Die Krisis der europäishen Wissenschaften und die transzendentale phänomenologie: Eine einleitung in die ph'nomenologische philosophie*, original work published 1954]

Ikonen, P., Luoma-aho, V., & Bowen, S. A. (2017). Transparency for sponsored content: Analysing codes of ethics in public relations, marketing, advertising and journalism. *International Journal of Strategic Communication, 11*(2), 165–178. doi:10.1080/1553118X.2016.1252917

Indian National Cyber Research & Innovation Centre. (2020). *Fake news and disinformation: How to spot and investigate?: A guide for law enforcement agencies*. Retrieved from

Institut Montaigne, & Institute for Strategic Dialogue. (2020). *Information manipulation around COVID-19: France under attack*. Retrieved from https://www.institutmontaigne.org/ressources/pdfs/publications/information-manipulations-around-covid-19-france-under-attack-policy-paper_0.pdf

Institute for Strategic Dialogue. (2020). *Anatomy of a disinformation empire: Investigating NaturalNews*. Retrieved from https://www.isdglobal.org/wp-content/uploads/2021/10/20211013-ISDG-NaturalNews-Briefing.pdf

Institute of Modern Russia, & The Interpreter. (2014). *The menace of unreality: How the Kremlin weaponizes information, culture and money*. Retrieved from https://imrussia.org/media/pdf/Research/Michael_Weiss_and_Peter_Pomerantsev__The_Menace_of_Unreality.pdf

International Telecommunication Union. (2020). *Balancing act: Countering digital disinformation while respecting freedom of expression*. Retrieved from https://www.broadbandcommission.org/wp-content/uploads/dlm_uploads/2023/01/WGFoEDisinfo_Report2020-compress.pdf

ISD. (n.d.). Make Germany Great again.

ISD. (2019). *Extract from ISD's submitted response to the UK government online harms white paper*.

ISD, & LSE. (2018). *Smearing Sweden: International Influence Campaigns in the 2018 Swedish Election*.

Iser, M. (2019). Axel Honneth (1949–). In Allen & E. Mendieta (Eds.), *The Cambridge Habermas Lexicon* (pp. 570–572). Cambridge: Cambridge University Press.

Johansen, T. S. (2012). The narrated organization: Implications of a narrative corporate identity vocabulary for strategic self-storying. *International Journal of Strategic Communication, 6*(3), 232–245. doi:10.1080/1553118X.2012.664222

Johansson, S., Nothhaft, H., & Fjällhed, A. (2023). Digital corporate communication and hostile hijacking of organizational crises. In V.

Luoma-aho & M. Badham (Eds.), *Handbook on digital corporate commuication* (pp. 208–211). Edward Elgar Publishing.

Johnson, J. (1991). Habermas on strategic and communicative action. *Political Theory, 19*(2), 181–201. doi:10.1177/0090591791019002003

Joint Research Centre. (2018). *The digital transformation of news media and the rise of disinformation and fake news*. Retrieved from https://joint-research-centre.ec.europa.eu/system/files/2018-04/jrc111529.pdf

Jones, P. E. (2007). Why there is no such thing as 'critical discourse analysis'. *Language & Communication, 27*(4), 337–368. doi:10.1016/j.langcom.2006.08.001

Joseph, S. M. (2002). *The trivium: The liberal arts of logics, grammar, and rhetoric*: Paul Dry Books.

Kagan, S. (2018). *Normative ethics*: Routledge.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissable harm*. Oxford: Oxford University Press.

Kang, D. S., & Davenport, L. (2009). Strategic communication in multicultural markets: Korean and American cultural differences toward leisure. *International Journal of Strategic Communication, 3*(3), 183–198. doi:10.1080/15531180902984208

Kant, I. (2012). *Groundwork of the metaphysics of morals* (M. Gregor & J. Timmermann, Trans.). Cambridge University Press.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies, 76*(4), 1359–1395. doi:10.111l/j.l467-937X.2009.00559.x

Kartika, N., Ottavianib, M., & Squintanic, F. (2006). Credulity, lies, and costly talk. *Journal of Economic Theory, 134*(1), 93–116. doi:10.1016/j.jet.2006.04.003

Keller, R. (2011). The sociology of knowledge approach to discourse (SKAD). *Human Studies, 34*(1), 43. doi:10.1007/s10746-011-9175-z

Keller, R. (2016). Has critique run out of steam?—On discourse research as critical inquiry. *Qualitative Inquiry, 23*(1), 58–68. doi:10.1177/1077800416657103

Keller, R., Hornidge, A.-K., & Schünemann, W. J. (2018). *The sociology of knowledge approach to discourse: Investigating the politics of knowledge and meaning-making*: Routledge.

Kelm, O., Dohle, M., & Bernhard, U. (2017). Social media activities of political communication practitioners: The impact of strategic orientation and in-group orientation. *International Journal of Strategic Communication, 11*(4), 306–323. doi:10.1080/1553118X.2017.1323756

Knight Foundation. (2018). *Disinformation, 'fake news' and influence campaigns on Twitter*. Retrieved from https://s3.amazonaws.com/kf-site-legacy-media/feature_assets/www/misinfo/kf-disinformation-report.0cdbb232.pdf

Koehler, K. (2014). Dialogue and relationship building in online financial communications. *International Journal of Strategic Communication, 8*(3), 177–195. doi:10.1080/1553118X.2014.905477

Kofi Annan Commission on Elections and Democracy in the Digital Age. (2020). *Protecting electoral integrity in the digital age*.

Kohlberg, L. (1958). *The development of modes of thinking and choices in years 10 to 16.* (PhD). University of Chicago,

Kohlberg, L. (1964). Development of moral character and moral ideology. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research* (Vol. 1, pp. 383–431). New York, NY: Russel Sage Foundation.

Kohlberg, L. (1977). Moral development: A review of theory. *Theory Into Practice, 16*(2), 53–59. doi:10.1080/00405847709542675

Koren, R. (2009). Can Perleman's NR be viewed as an ethics of discourse? *Argumentation, 23*(3), 421–431. doi:10.1007/s10503-009-9149-8

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*. doi:10.1037/0033-2909.108.3.480

Kyratzis, A., & Cook-Gumperz, J. (2015). Child discourse. In H. E. H. a. D. S. D. Tannen (Ed.), *The handbook of discourse analysis*. doi:10.1002/9781118584194.ch32

Laaksonen, S.-M. (2016). Casting roles to stakeholders: A narrative analysis of reputational storytelling in the digital public sphere. *International Journal of Strategic Communication, 10*(4), 238–254. doi:10.1080/1553118X.2016.1159564

Laidlaw, J. (2020). For an anthropology of ethics and freedom. *Journal of the Royal Anthropological Institute, 8*(2), 311–332. doi:10.1111/1467-9655.00110

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (Vol. 37, pp. 136–139). New York: Harper and Row.

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry, 30*(2), 225–248. doi:10.1086/421123

Leeuwen, T. v. (2018). Moral evaluation in critical discourse analysis. *Critical Discourse Studies, 15*(2), 367–381. doi:10.1080/17405904.2018.1427120

LinkedIn. (2019, 25 November). Phishing emails. Retrieved from https://www.linkedin.com/help/linkedin/answer/5342

Loader, B. D., & Mercea, D. (2011). Networking democracy? Social media innovations and participatory politics. *Information, Communication & Society, 14*(6), 757–769. doi:10.1080/1369118X.2011.592648

Lock, I., Wonneberger, A., Verhoeven, P., & Hellsten, I. (2020). Back to the roots? The applications of communication science theories in strategic communication research. *International Journal of Strategic Communication, 14*(1), 1–24. doi:10.1080/1553118X.2019.1666398

Luhmann, N. (1996). The sociology of the moral and ethics. *International Sociology, 11*(1), 27–36. doi:10.1177/026858096011001003

Lévinas, E. (1991). *Totality and infinity: An essay on exteriority* (A. Lingis, Trans.). Kluwer Academic Publishers. [*Totalité et infini: Essai sur l'exteriorité*, original work published 1961]

Macdonald-Laurier Institute. (2019). *Stemming the virus: Understanding and responding to the threat of Russian disinformation*. Retrieved from https://macdonaldlaurier.ca/files/pdf/20181211_MLI_Russian_Disinformation%20PAPER_FWeb.pdf

Mannheim. (1997). *Diagnosis of our time: Wartime essays of a sociologist* (Vol. 3). London: Routledge.

Marxen, C. (2019). Discourse theory of law in times of populism. *European Law Journal, 25*, 480–486. doi:10.1111/eulj.12343

Mattingly, C. (2012). Two virtue ethics and the anthropology of morality. *Anthropological Theory, 12*(2), 161–184. doi:10.1177/1463499612455284

McCarthy, T. (1987). *The critical theory of Jürgen Habermas*: MIT Press.

Men, L. R., Yang, A., Song, B., & Kiousis, S. (2018). Examining the impact of public engagement and presidental leadership communiction on social media in China: Implications for government–public relationship cultivation. *International Journal of Strategic Communication, 12*(3), 252–268. doi:10.1080/1553118X.2018.1445090

Mendieta, E., & Vanantwerpen, J. (Eds.). (2011). *The power of religion in the public sphere*: Columbia University Press.

Mouffe. (2005). *The democratic paradox*. London: Verso.

Mouffe, C. (1999). Deliberative democracy or agonistic pluralism? *Social Research: An International Quarterly, 88*(3), 745–758.

Mouffe, C. (2005a). *The democratic paradox*. London: Verso.

Mouffe, c. (2005). *The return of the political*. London: Verso.

Mozilla. (2014, 19 June). Mozilla announces a partnership with The New York Times, The Washington Post and Knight Foundation for a new online community platform. Retrieved from https://blog.mozilla.org/en/mozilla/mozilla-announces-a-partnership-with-the-new-york-times-the-washington-post-and-knight-foundation-for-a-new-online-community-platform/

Mumby, D. K. (1997). Modernism, postmodernism, and communication studies: A rereading of an ongoing debate. *Communication Theory, 7*(1), 1–28. doi:10.1111/j.1468-2885.1997.tb00140.x

Myndigheten för Psykologiskt försvar. (2019). *Bli inte lurad: En handbok om att känna igen och hantera falsk och vilseledande information*. Retrieved from https://www.bliintelurad.se/assets/uploads/2022/07/HANDBOK_MPF.pdf

Myndigheten för Samhällsskydd och Beredskap. (2018). *Att möta informationspåverkan: Handbok för kommunikatörer*. Retrieved from https://www.msb.se/RibData/Filer/pdf/28692.pdf

Nagel, J., & Waldmann, M. R. (2013). Deconfounding distance effects in judgments of moral obligation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 237. doi:10.1037/a0028641

Natani, Y. (2002). *Axel Honneth and the Immanent Critique of Society.* (Doctor of Philosophy (Sociology)). The University of Wisconsin,

NATO StratCom COE. (2017). *Digital hydra: Security implications of false information online*. Retrieved from https://stratcomcoe.org/pdfjs/?file=/publications/download/WEB-Digital-Hydra-FULL-16NOV.pdf?zoom=page-fit

NATO StratCom COE. (2019). *Hybrid threats: A strategic communications perspective*. Retrieved from https://stratcomcoe.org/publications/hybrid-threats-a-strategic-communications-perspective/79

NATO StratCom COE, & King's Centre for Strategic Communications. (2018). *Fake news: A roadmap*. Retrieved from https://stratcomcoe.org/pdfjs/?file=/publications/download/fake_news_book_final_full_version.pdf?zoom=page-fit

New Knowledge. (2018). *SSCI research summary: An assessment of the Internet Research Agency's U.S.-directed activities in 2015–2017 based on platform-provided data*. Retrieved from https://www.scribd.com/document/544375202/2018-12-01-New-Knowledge-SSCI-Research-Summary

NGO Internews Ukraine, & Ukraine World. (2018). *Taming the hydra: How to resist Kremlin's information agression*.

Nietzsche, F. (2006). *On the geneology of morality: A polemic* (C. Dieth, Trans.). Cambridge University Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. doi:10.1037/0033-295X.84.3.231

Nordic Council of Ministers. (2018). *Fighting fake news: The Nordic way*. Retrieved from http://www.diva-portal.org/smash/get/diva2:1200348/FULLTEXT01.pdf

Nothhaft, H. (2016). A framework for strategic communication research: A call for synthesis and consilience. *International Journal of Strategic Communication, 10*(2), 69–86. doi:10.1080/1553118X.2015.1124277

Nothhaft, H. (2017). Disagreement about the therapy, not the diagnosis: A reply to the rejoinders. *International Journal of Strategic Communication, 11*(3), 189–193. doi:10.1080/1553118X.2017.1318884

Nothhaft, H., Werder, K. P., Verčič, D., & Zerfaß, A. (2018a). Editor's introduction. *International Journal of Strategic Communication, 12*(4), 329–332. doi:10.1080/1553118X.2018.1493484

Nothhaft, H., Werder, K. P., Verčič, D., & Zerfaß, A. (2018b). Strategic communication: Reflections on an elusive concept. *International Journal of Strategic Communication, 12*(4), 352–366. doi:10.1080/1553118X.2018.1492412

Nothhaft, H., Werder, K. P., Verčič, D., & Zerfaß, A. (2020). *Future directions of strategic communication*: Routledge.

Open Society Institute. (2018). *Common sense wanted: Resilience to 'post-truth' and its predictors in teh new media index 2018.*

Organization for Security and Co-operation in Europe. (2011). *Joint declaration on freedom of expression and the internet.* Retrieved from https://www.osce.org/files/f/documents/e/9/78309.pdf

Oxford Internet Institute. (2018). *The IRA, social media and political polarization in the United States, 2012–2018.* Retrieved from https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf

PARK Advisors. (2019). *Weapons of mass distraction: Foreign state-sponsored disinformation in the digital age.*

Paul, C. (2011). *Strategic communication: Origins, concepts, and current debates: origins, concepts, and current debates*: Praeger.

Pearche, W. B. (1989). *Communication and the human condition*: Southern Illinois University Press.

Peel, C., Simensen, E., Larsson, I., Love, J. S., & Djärv, U. (2020). *A lexicon of medival nordic law*: Open Book Publishers.

Piaget, J. (1932). *The moral judgement of the child.* New York: Harcourt, Brace & World.

Pinterest. (2019, 28 August). Bringing authoritative vaccine results to Pinterest search. Retrieved from https://newsroom-archive.pinterest.com/en-gb/bringing-authoritative-vaccine-results-to-pinterest-search

Plato. (360 BC). *Sophist.*

Power, M. K. (1995). Habermas and the counterfactual imagination. *Cardozo Law Review, 17*, 1005–1026.

RAND. (2018a). *Countering Russian social media influence.*

RAND. (2018b). *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Retrieved from https://www.rand.org/pubs/research_reports/RR2314.html

RAND. (2018a). *Modern political warfare*. Retrieved from https://www.rand.org/pubs/research_reports/RR1772.html

RAND. (2019). *Countering truth decay: Exploring media literacy education as a tool for mitigating truth decay*. Retrieved from

Regh, W. (1994). *Insight and solidarity: The discourse ethics of Jürgen Habermas*. Berkeley, CA: University California Press.

Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Reuters Institute. (2017). *Digital News Report 2017*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

Reuters Institute. (2018). *Digital News Report 2018*. Retrieved from https://www.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf

Reuters Institute. (2020). *Digital News Report 2020*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. Cambridge, mass: MIT Press.

Rheingold, H. (2007, 5 November). Habermas blows off question about the internet and the public sphere. Retrieved from http://www.smartmobs.com/2007/11/05/habermas-blows-off-question-about-the-internet-and-the-public-sphere/

Rich, M. H., & Craig, R. T. (2012). Habermas and Bateson in a world gone M.A.D.: Metacommunication, paradox, and the inverted speech situation. *Communication Theory, 22*, 383–402. doi:10.1111/j.1468-2885.2012.01412.x

Roderick, I. (2018). Multimodal critical discourse analysis as ethical praxis. *Critical Discourse Studies, 15*(2), 154–168. doi:10.1080/17405904.2017.1418401

RSIS. (2018). *Fake news: National security in the post-truth era*. Retrieved from https://www.rsis.edu.sg/wp-content/uploads/2018/01/PR180313_Fake-News_WEB.pdf

Sandhu, S. (2017). Heading for mars while we haven't been on the moon: A reply to Nothhaft. *International Journal of Strategic Communication, 11*(3), 184–188. doi:10.1080/1553118X.2017.1319837

Sceptical science. (2012). *The debunking handbook*. Retrieved from https://skepticalscience.com/docs/Debunking_Handbook_2011.pdf

Schiffrin, D., Tannen, D., & Hamilton, H. E. (Eds.). (2005). *The handbook of discourse analysis* (Vol. 1). John Wiley & Sons.

Schopenhauer, A. (1831). *The art of being right: 38 ways to win an argument [Alternative title 'The art of controversy' or 'Eristic dialectic: The art of winning an argument'.* [*Eristische dialektik: Die kunst, recht zu behalten*, original work published 1831]

Schwarz, A. (2008). Covariation-based causal attributions during organizational crises: Suggestions for extending situational crisis communication theory (SCCT). *International Journal of Strategic Communication, 2*(1), 31–53. doi:10.1080/15531180701816601

Searl, J. R. (2006). Social ontology: Some basic principles. *Anthropological Theory, 6*(1), 12–29. doi:10.1177/1463499606061731

Self, C. C. (2010). Hegel, Habermas, and community: The public in the new media era. *International Journal of Strategic Communication, 4*(2), 78–92. doi:10.1080/15531181003704651

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shepherd, G. J. (1993). Building a discipline of communication. *Journal of Communication, 43*(3), 83–91. doi:10.1111/j.1460-2466.1993.tb01279.x

Shweder, R. A., & Menon, U. (2014). Old questions for the new anthropology of morality: A commentary. *Anthropological Theory, 14*(3), 356–370. doi:10.1177/1463499614534555

Silverman. (2017, 31 December). I helped popularize the term 'Fake news' and now I cringe every time I hear it. Retrieved from https://www.buzzfeednews.com/article/craigsilverman/i-helped-popularize-the-term-fake-news-and-now-i-cringe

Simon, H. A. (1955). A behavioural model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118. doi:10.2307/1884852

Singer, P. (1972). Famine, affluence, and morality. *Philosophy & Public Affairs*, 229–243.

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics, 9*(3), 331–352. doi:10.1007/s10892-005-3508-y

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science, 50*(9), 799–813. doi:10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Special Counsel Robert S. Mueller. (2019a). *Report on the investigation into Russian interference in the 2016 Presidential election. Volume I of II.* Retrieved from https://storage.courtlistener.com/recap/gov.uscourts.dcd.205521/gov.uscourts.dcd.205521.122.1.pdf

Special Counsel Robert S. Mueller. (2019b). *Report on the investigation into Russian interference in the 2016 Presidential election. Volume II of II.* Retrieved from https://storage.courtlistener.com/recap/gov.uscourts.dcd.205521/gov.uscourts.dcd.205521.122.2.pdf

Stadsrådets Kansli. (2019). *Att möta informationspåverkan: Handbok för kommunikatörer*. Retrieved from https://julkaisut.valtioneuvosto.fi/handle/10024/161586

Stanford Internet Observatory Cyber Policy Center, & Institution, H. (2020). *Telling China's story: The Chinese communist party's campaign to shape global narratives*. Retrieved from https://purl.stanford.edu/pf306sw8941

Swedish Civil Contingency Agency. (2018). *Countering information influence activities: The state of the art*. Retrieved from https://www.msb.se/RibData/Filer/pdf/28697.pdf

Swedish Civil Contingency Agency. (2018a). *Countering information influence activities: A handbook for communicators*. Retrieved from https://www.msb.se/ribdata/filer/pdf/28698.pdf

Swedish Defence University, CATS, & Hybrid CoE. (2018). *Addressing hybrid threats*. Retrieved from https://www.hybridcoe.fi/wp-content/uploads/2020/07/Treverton-AddressingHybridThreats.pdf

Tannen, D., Hamiltin, H. E., & Schiffrin, D. (Eds.). (2015). *The handbook of discourse analysis* (Vol. 2). John Wiley & Sons.

Taylor, P. (1963). The ethnocentric fallacy. *Monist, 47*(4), 563–584.

Taylor, P. M. (2003). *Munitions of the mind: A history of propaganda from the ancient world to the present era*. Manchester: Manchester University Press.

Thomassen, L. (2019). J. L. Austin (1911–1960). In Allen & E. Mendieta (Eds.), *The Cambridge Habermas Lexicon* (pp. 488–489). Cambridge University Press.

Tiberius, V. (2015). *Moral psychology: A contemporary introduction*: Routledge.

TikTok. (2019a). Effective in-app reporting for a healthy community. Retrieved from https://newsroom.tiktok.com/en-us/effective-in-app-reporting-for-a-healthy-community

TikTok. (2019b). Our H2 2019 transparency report. Retrieved from https://newsroom.tiktok.com/en-us/our-h-2-2019-transparency-report

TikTok. (2019c). Statement on TikTok's content moderation and data security practices. Retrieved from https://newsroom.tiktok.com/en-us/statement-on-tiktoks-content-moderation-and-data-security-practices

TikTok. (2020). Community guidelines. Retrieved from https://www.tiktok.com/community-guidelines/en/

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Rewiev of Psychology, 64*, 231–255. doi:10.1146/annurev-psych-113011-143812

Twitter. (2018, 25 September). Creating new policies together. Retrieved from https://blog.twitter.com/en_us/topics/company/2018/Creating-new-policies-together

Twitter. (2019). *Strenghtening our approach to deliberate attemtps to mislead voters.*

Twitter. (2020, 9 September). Amplifying #SuicidePrevention resources on Twitter. Retrieved from https://blog.twitter.com/en_us/topics/company/2020/amplifying-suicideprevention-resources-on-twitter

Twitter. (2020a, 9 December). Crossroad for the open internet. Retrieved from https://blog.twitter.com/en_us/topics/company/2020/crossroads-for-the-open-internet

Twitter. (2020b, 9 December). Platform manipulation and spam policy. Retrieved from https://help.twitter.com/en/rules-and-policies/platform-manipulation

UK Government Communication Services. (2019). *RESIST: Counter-disinformation toolkit.* Retrieved from https://gcs.civilservice.gov.uk/wp-content/uploads/2019/03/RESIST_Toolkit.pdf

UK Government Communication Services. (2021). *RESIST 2: Counter-disinformation toolkit.* Retrieved from https://gcs.civilservice.gov.uk/wp-content/uploads/2021/11/RESIST-2-counter-disinformation-toolkit.pdf

UK House of Parliament. (2017). *Online information and fake news.* Retrieved from https://researchbriefings.files.parliament.uk/documents/POST-PN-0559/POST-PN-0559.pdf

UK Law Commission. (2018). *Abusive and offensive online communications: A scoping report.* Retrieved from https://s3-eu-west-2.amazonaws.com/cloud-platform-e218f50a4812967ba1215eaecede923f/uploads/sites/30/2018/10/6_5039_LC_Online_Comms_Report_FINAL_291018_WEB.pdf

UN. (2015). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* Retrieved from https://digitallibrary.un.org/record/798709?ln=en

UNESCO. (2011). *Media and information literacy: Curriculum for teachers.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000192971

UNESCO. (2018). *Handbook for journalism education and training: Journalism, 'fake news' and disinformation.* Retrieved from https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf

Unger, P. (1996). *Living high and letting die: Our illusion of innocence*: Oxford University Press.

University of Washington Center for an Informed Public. (2020). *Deepfakes in the 2020 elections and beyond: Lessons from the 2020 workshop series*. Retrieved from https://jevinwest.org/papers/Prochaska2020CIP_Deepfake_Report.pdf

US Committee on Foreign Relations. (2018). *Putin's assymmetric assault on democracy in Russia and Europe: Implications for U.S. national security*. Retrieved from https://www.govinfo.gov/content/pkg/CPRT-115SPRT28110/pdf/CPRT-115SPRT28110.pdf

US Intelligence Community Assessement. (2017). *Assessing Russian Activities and Intentions in Recent US Elections*. Retrieved from https://www.intelligence.senate.gov/sites/default/files/publications/ICA_2017_01.pdf

Vaara, E., Sorsa, V., & Pälli, P. (2010). On the force potential of strategy texts: a critical discourse analysis of a strategic plan and its power effects in a city organization. *Organization, 17*(6), 685–702. doi:10.1177/1350508410367326

Valtioneuvoston Kanslia. (2019). *Informaatiovaikuttamiseen vastaaminen: Opas viestijöille*. Retrieved from https://julkaisut.valtioneuvosto.fi/handle/10024/161512

Van Ruler, B. (2018). Communication theory: An underrated pillar on which strategic communication rests. *International Journal of Strategic Communication, 12*(4), 367–381.

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology, 11*, 59–63. doi:10.1016/j.copsyc.2016.06.007

Vetlesen, A. J. (1997). Worlds apart? Habermas and Levinas. *Philosophy & Social Criticism, 23*(1), 1–20.

Vorster, N. (2019). Land and identity in South Africa: An immanent moral critique of dominant discourses in the debate on expropriation without compensation. *Theological Studies, 75*(4). doi:10.4102/hts.v75i4.5290

Weaver, W. (1949). The mathematics of communication. *Scientific American, 181*(1), 11–15.

Weber, M. (1963). Objectivity in social sciences and social policy (E. A. Shils & H. A. Finch, Trans.). In M. Natanson (Ed.), *Philosophy of the social sciences* (pp. 355–418). New York: Random house. (Reprinted from: Methodology in the social sciences).

Weber, M. (1968). *Economy and society: An outline of interpretive sociology*. New York: Bedminster Press.

Weber, P., & Grauer, Y. (2019). The effectiveness of social media storytelling in strategic innovation communication: Narrative form matters.

International Journal of Strategic Communication, 13(2), 152–166. doi:10.1080/1553118X.2019.1589475

Werder, K. P., Nothhaft, H., Verčič, D., & Zerfaß, A. (2018). Strategic communication as an emerging interdisciplinary paradigm. *International Journal of Strategic Communication, 12*(4), 333–351. doi:10.1080/1553118X.2018.1494181

Wilfred Martens Center for European Studies. (2018). *Political subversion in the age of social media*. Retrieved from https://www.martenscentre.eu/wp-content/uploads/2020/06/ces_policybrief_political-subversion-v4.pdf

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans. 4 ed.). Whiley-Blackwell.

Wittgenstein, L. (1974). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuinness, Trans.). London: Routledge. [*Annalen der Naturphilosophie*, original work published 1921]

Young, R. J. C. (Ed.) (1982). *Untying the text: A post-structural anthology*: Routledge. [original work published 1981]

YouTube. (2006, 8 October). How flagging works. Retrieved from https://blog.youtube/news-and-events/how-flagging-works/

YouTube. (2009, 13 August). 'Just-in-time' learning about flagged videos. Retrieved from https://blog.youtube/news-and-events/just-in-time-learning-about-flagged/

YouTube. (2012, 24 May). A new human rights channel on YouTube. Retrieved from https://blog.youtube/news-and-events/new-human-rights-channel-on-youtube/

YouTube. (2019, 19 February). Making our strikes system clear and consistent. Retrieved from https://blog.youtube/news-and-events/making-our-strikes-system-clear-and/

YouTube. (2020, 28 April). Expanding fact checks on YouTube to the United States. Retrieved from https://blog.youtube/news-and-events/expanding-fact-checks-on-youtube-to-united-states/

Zerfaß, A., Verčič, D., Nothhaft, H., & Werder, K. P. (2018). Strategic communication: Defining the field and its contribution to research and practice. *International Journal of Strategic Communication, 12*(4), 487–505. doi:10.1080/1553118X.2018.1493485

Zerfaß, A., Werder, K. P., & Holtzhausen, D. R. (2012). Strategic directions revisited. *International Journal of Strategic Communication, 6*(3), 197–198. doi:10.1080/1553118X.2012.691333

Zhao, H. (2017). The impacts of contextual factors on social media crises: Implications for crisis communication strategy selection. *International Journal of Strategic Communication, 11*(1), 42–60. doi:10.1080/1553118X.2016.1244061

Notes

---

[i] While Habermas is clear about both the theory of ideal speech situations and discourse ethics being ideal theories, from the beginning he described that it 'can only be analysed with regard to a consensus achieved in unrestrained and universal discourse. … independent of the empirical structures of the social system to which we belong' as 'we are quite unable to realize the ideal speech situation; we can only anticipate it' (Habermas, 1970, p. 372). In discourse ethics, he moves from presenting this as an ideal form of communication, to argue that 'we must show that these rules of discourse are not mere *conventions*; rather, they are inescapable presuppositions.' (1999, p. 89; c.f., 1994, p. 148) thus in discourse ethics providing the justification for such a statement. Nonetheless, the theory has been a subject of harsh criticism on this very point (see i.e., Honneth & Joas, 1991). Mouffe (1999, p. 751) for example writes how 'far from being merely empirical, or epistemological, the obstacles to the realization of the ideal speech situation are ontological' as ideal speech situations would be a '*conceptual* impossibility' (2005, p. 33). In an attempt to 'clarify a persistent misunderstanding about the concept' (Habermas, 2018a, p. 871), Habermas gives the following answer in an interview published in 2018 on the question about ideal speech acts…

> Aside from the fact that I have not used this misleading expression since my 1972 essay on 'Theories of Truth,' one must take into account the context in which the concept is introduced. At the time I used the expression to refer to the set of pragmatic presuppositions that we must (!) assume as a matter of fact whenever we engage in argumentation about the validity of propositions. As participants in discourse, we 'know' that we are not arguing 'seriously' if, in such an exchange of reasons, coercion or manipulation is at work, some of those involved are excluded, or relevant opinions and positions are suppressed. We must *presuppose* that, in the given situation, only the unforced force of the better argument comes into play. This 'know how'—our 'knowing how to participate in a rational discourse'—has a regulating influence on the actual behavior of participants in argumentation even if they are aware that they can only approximately fulfil these pragmatic presuppositions. In view of this counterfactual status, one can say perhaps that *the idealizing content* of the pragmatic presuppositions of discourse plays the role of a regulative idea for the participants. (Habermas, 2018a, p. 871)

From his first observation of its historic antecedents in his first book on the coffee houses, later through his presentation of ideal speech theory, and finally in his theory of discourse ethics in the late 20th century, Habermas moral theory of communication is presented as an ideal idea rather than an observed reality. It is an ideal—and perhaps utopian—presentation of a reality that is theorised as a set of communicative axioms in his normative theory which has since been read as an imperative and prescriptive guide used by those agreeing with such ideology (Regh, 1994).

[ii] The goal to come 'to an understanding [*Verständigung*] is to bring about an agreement [*Einverständnis*] that terminates in the intersubjective mutuality of reciprocal understanding, shared knowledge, mutual trust, and accord with one another' (Habermas, 1979, p. 24). To Habermas, understanding can be interpreted as in a minimal sense an agreement on what is being said, in a maximal sense that 'there exist an accord concerning the Tightness of a text in relation to a mutually recognized normative background' or what is meant, or indeed that the 'two participants in communication can come to an understanding about something in the world, and they can make their intentions understandable to one another' (p. 24). For Habermas, this is the default purpose of any form of communication, as 'for example, conflict, competition, strategic actions in general—are derivatives of action oriented towards reaching understanding [*verständigungsorientiert*]' (p. 23) among people.

[iii] This shared interpretation has also been explored in phenomenology as a shared 'universal horizon' (Husserl, 1970) or as Habermas (1999) would describe it, 'a storehouse of unquestioned givens from which those participating in communication draw agreed-upon patterns of interpretation for use in their interpretive efforts.' (Habermas, 1999, p. 135) Just as Searl, Habermas argues that we can understand these lifeworlds by looking at their use of language, as its structures 'are determined by linguistically produced intersubjectivity' (2007, p. 14).

> …if you share a common language and are already involved in conversations in that common language, you already have a social contract. The standard account that presupposes language and then tries to explain society has things back to front. You cannot begin to understand what is special about human society, how it differs from primate societies and other animal societies, unless you first understand some special features of human language. Language is the presupposition of the existence of other social institutions in a way that they are not the presupposition of language. … Institutions such as money, property, government and marriage cannot exist without language, but language can exist without them. (Searl, 2006, p. 14)

[iv] Continuing to describe that (U) means that where '*all* affected can accept the consequences and the side effects its *general* observance can be anticipated to have for the satisfaction of *everyone's* interest (and these consequences are preferred to

---

those of known alternative for possibilities for regulation)' or that 'for a norm to be valid, the consequences and side effects of its general observance for the satisfaction of each person's particular interest must be acceptable to all' (Habermas, 1999, p. 40, original emphasis). Earlier publications similarly present that *communication ethics* would require 'not only generality of norms but a discursively attained consensus about the *generalizability* of the normatively prescribed interests' (2007, p. 23, original emphasis). That is, in 'their underlying idea, which is designed to take into account the impersonal or general character of valid universal commands. The moral principle is so conceived as to exclude as invalid any norm that could not meet with the qualified assent of all who are or might be affected by it. This bridging principle, which makes consensus possible, ensures the only those norms are accepted as valid that express a *general will*' (1999, p. 63, original emphasis)

[v] The two principles are described further when having a clearer idea of the distinction between ethics and morals as a differentiation between the principle's application where 'for the justification of moral norms, the discourse principle takes the form of a universalization principle' and 'functions as a rule of argumentation' where 'for the application of moral norms to particular cases, the universalization principle is replaced by a principle of appropriateness' (Habermas, 2009, p. 109). From this point, axiomatic maxims such as Kant's categorical imperative can either be judged from an ethical or moral point of view—'whether a maxim is good for me … [or] whether I can will that maxim should be followed by everyone else as a general law' (1994). The problem with Kant is that it does not account for the multiverse of moral egos who may hold different views on what ought to be considered a general law. For that, Habermas argues, we need to engage in moral deliberations with others.

[vi] In this, Habermas' (1999) duality emerges, in his belief in rationality appearing in the frock of *moral argumentation* where 'ultimately, there is only one criterion by which beliefs can be judged valid, and that is that they are based on agreements reached by argumentation' (Habermas, 1999, p. 14). Such a practice with two tasks, as 'analytically testing the consistency of the value premises (or the preference system taken as a basis); and empirically testing the realizability of goals selected from value perspectives' (pp. 105–106).

[vii] Habermas writes that 'I use the evolutionary argument that there is an increasing convergence between the diversity of historical concepts of justice and the discursive procedure of judging moral-practical conflicts. The idea of justice, which first assumes the form of concrete conceptions in the particularistic contexts of tribal societies and early civilizations, progressively loses its substantive content with increasing social complexity, until finally the propositional content of 'justice' withdraws into the procedural form of impartial judgement. The scenario of the disintegration of the communal forms of ethos and the increasing pluralism of forms of life and worldviews explains why the substantive ideas of justice, in terms

---

of which the worthiness of moral principles and of norms to be recognized is measured, finally dissolve into the idea of discursive procedural justice. The semantic content of justice converges with a procedural notion of impartiality that is operationalized in the form of an agreement achieved in rational discourse' (2018b, p. 119). Later on, he continues that, 'I use the evolutionary argument that there is an increasing convergence between the diversity of historical concepts of justice and the discursive procedure of judging moral-practical conflicts. The idea of justice, which first assumes the form of concrete conceptions in the particularistic contexts of tribal societies and early civilizations, progressively loses its substantive content with increasing social complexity, until finally the propositional content of "justice" withdraws into the procedural form of *impartial judgement*' (p. 119). In other writings, Habermas continues that 'This is why Kohlberg speaks of a transition to the *postconventional* stage of moral consciousness. At this stage, moral judgment becomes dissociated from the local conventions and historical coloration of a particular form of life. It can no longer appeal to the naive validity of the context of the lifeworld. Moral answers retain only the rationally motivating force of insights. Along with the naive self-certainty of their lifeworld background they lose the thrust and efficacy of empirical motives for action. To become effective in practice, every universalist morality has to make up for this loss of concrete ethical substance, which is initially accepted because of the cognitive advantages attending it. Universalist moralities are dependent on forms of life that are rationalized in that they make possible the prudent application of universal moral insights and support motivations for translating insights into moral action. Only those forms of life that meet universalist moralities halfway in this sense fulfill the conditions necessary to reverse the abstractive achievements of decontextualization and demotivation' (1999, p. 109).

[viii] As a sociologist, Habermas can be seen to align with an idea that the study of ethics 'would first have to be empirical and not philosophical' (Luhmann, 1996, p. 27) and so opened up the perspective 'to think of philosophical problems as *human* problems in that they arise in the lives of actual people and not only in the hallowed hall of philosophy' (Das, 2012, p. 133). From this perspective, he aligns with *applied ethics*, moving from normative frameworks to discuss moral behaviour (c.f., Kagan, 2018) as he formulates his discourse ethics as a toolbox to enable moral discourse. This is particularly clear in his second book on discourse ethics called *Justification and Application: Remarks on discourse ethics* (1994), but is also described to have developed from the beginning 'in the arena of intersection of practical philosophy and social theory' to deal with 'the specific problem of rational justification of decisions' (2018b, p. 101).

[ix] In some writings describing how this merges Webers' idea of three separate cultures in Western rationalism as 'cognitive, aesthetic-expressive, and moral-practical' culture (Habermas, 1999, p. 107), whereas his discourse ethics shows how in everyday life these theoretical categories are merged in one lifeworld. In

---

other passages, Weber rather draws on Plato's description of an objective realm, a realm of individual internal experiences, and a realm of abstract types such as words or numbers, and indeed corresponding to Habermas' reference to Aristotle's suggestions on argumentation as either 'those at the logical level of products, those at the dialectical level of procedures, and those at the rhetorical level of processes.' Where *procedure* arguments refer to logical and semantic rules that determine the validity of one's claims and have no ethical content, while *procedural* arguments correspond to his idea of discourse as 'processes of reaching understanding that are ordered in such a way that proponents and opponents, having assumed a hypothetical attitude and being relieved of the pressures of actions and experience, can test validity claims that have become problematic' and the final *process* argumentative speech would be aligned with his former proposition for ideal speech situations wherein 'participants in argumentation cannot avoid the presupposition that, owing to certain characteristics that require formal description, the structure of their communication rules out all external and internal coercion other than the force of the better argument and thereby also neutralizes all motives other than that of the cooperative search for truth.' Indeed the process argument, 'in light of its goal of reaching a rationally motivated agreement, must satisfy improbable conditions' (Habermas, 1999, pp. 87–89).

[x] This means that in the evaluation of a claim, we can use these three values to come to an understanding of whether it is accepted or not accepted, an embedded dual opportunity based on 'the yes or no position taken by the hearer whereby he accepts or rejects the claim to validity that has been raised by the speaker' based on three factors—either (1) as objective *truth* refers to 'the claim that the text in question is true', (2) as normative *rightness* 'that the speech act is right in terms of a given normative context', and (3) as subjective *truthfulness* 'that the speaker's manifest intentions are meant as they are expressed' (Habermas, 1982, p. 19, 1999, p. 137). He even describes how myths, as linguistic constitution, inevitably connects to a real reality and that 'language, for example, as the medium of representation mythical narrative, is not so far removed from reality that the conventional sign is completely divorced from its semantic content and for its referent; speech and world view remain in some way interwoven with the order of the world. Mythical traditions cannot be revised without endangering the order of things and the identity of the tribe that is embedded in them. Categories of validity such as 'true' and 'false', 'good' and 'evil', are still linked to empirical concepts such as exchange, causality, health, substance, and wealth.' In a later reformulation of Habermas' notion of rationality, he continues that 'instrumental and strategic action raise claims to truth or effectiveness. Normative action raises a claim to rightness. Dramaturgical action raises a claim to sincerity or authenticity. Parties to communicative action can—whether implicitly or explicitly—raise validity claims of each sort' (Johnson, 1991, p. 184). Habermas' stance in moral cognitivism means that he approaches any valid claims as presenting a good reason

---

for something, adding that 'ultimately, there is only one criterion by which beliefs can be judged valid, and that is that they are based on agreements reached by argumentation' (1999, p. 14). At the same time, the nature of moral statements means that they cannot be determined on an objective basis, but their validity is constructed through communication…

> …moral validity claims lack the ontological connotation of reference to a world of objects *existing independently of our descriptions of them.* … Whereas we find and encounter the objective world of which we can have knowledge, it is we who first give rise to the social world of interpersonal relationships, which we can judge from the moral point of view. *Moral judgements and practical discourses are themselves part of that ongoing constructive activity.* (Habermas, 2018b, p. 117)

Valid claims to truth and rightness is settled discursively by presenting reasons, while truthfulness can only be validated through behaviour (Habermas, 1999, p. 59)

[xi] Where he responds to their thesis that 'reason itself destroys the humanity which it had made possible in the first place', he continues in the same text that this is substantiated 'with the argument that the process of enlightenment is from the very beginning dependent on an impulse of self-preservation which mutilates reason because it can only make use of it in the form of purposive-rational domination of nature and instinct, i.e., in the form of instrumental reason' (Habermas, 1982, p. 17), continuing that the book 'offers hardly any prospect of escape from of instrumental rationality' (1982, p. 18) and instead himself proposing a solution by returning to the ideals of the Enlightenment. This is a continuous point of return in Habermas works, stretching from *The Philosophical Discourse of Modernity* (2007b) to the very first presentation of discourse ethics in *Moral Consciousness and Communicative Action* (1999) by responding to Alasdair MacIntyre's comments on the Enlightenment which echoes Horkheimer's position. Here, Habermas position discourse ethics in relation to this position by stating that 'since Kant this conclusion has been opposed by cognitivist moral philosophies that maintain in one sense or another practical questions amit of truth' (1999, p. 43).

[xii] Habermas elaborates on the relation between Western morals and its historic religious morality in i.e., 1998 (p. 7). Again, we see the deep ties between linguistics and ethics in his view, not least in his observation of 'the linguistification of the sacred' (Habermas, 1987, p. 77).

> …it makes no sense to oppose one sort of reason, secular, against religious reasons on the assumption that religious reasons are coming out of a worldview which is inherently irrational. … At a general cognitive level, there is only one and the same human reason. However, if it comes to lumping together Kantianism and utilitarianism, Hegelianism and so on with religious doctrines, then I would say

> there are differences in kind between reasons. One way to put it is that 'secular' reasons can be expressed in a 'public' or generally shared, language. (from interview with Habermas published in Mendieta & Vanantwerpen, 2011, p. 61)

Religious perspectives run as a continuous interest in Habermas' works, ranging from books such as *Religion and Rationality* (2002), *The Future of Human Nature* (2003b), *The Dialectics of Secularization* (2006), and *Between Naturalism and Religion* (2008). Again, the link is made to the introductory description that 'modernity is characterised by a rejection of the substantive rationality typical of religious and metaphysical worldviews and by a belief in procedural rationality and its ability to give credence to our views in the three areas of objective knowledge, moral-practical insight, and aesthetic judgements' (Habermas, 1999, pp. 3–4).

[xiii] Also distinguished by Weber (1968, pp. 24–25), where *instrumentally rational* (*zweckrational*) would refer to acts motivated by conditions while *value-rational* (*wertrational*) would be based on one's common sense. Or the first as 'calculation of material interests' or 'purposive conduct' and the latter as 'ideal motives enjoyed by religion or magic'. Similar categorizations of different types of communication can for example be found in Buber's distinction between *genuine dialogue* (seeking subjective understanding), *technical dialogue* (seeking objective understanding), and *monologue disguised as dialogue* (under the illusion of having a conversation, while the participants only focus on their own stance).

[xiv] In the 21st century, the field of *strategic communication* emerged as new a trend in practice, soon resulting in a new field studying how language could be used to construct meanings. As a result of new information and communication technology blurring previous boundaries between communication activities (Hallahan et al., 2007, p. 10), organisations found themselves 'seeking integration as well as enhanced effectiveness through synergy, enhanced efficiencies, and reduced redundancies' (p. 5). Soon enough, such integration was mirrored in the emergence of a new scholarly enterprise—introduced in a seminal article in the first issue of the *International Journal of Strategic Communication* (Hallahan et al., 2007)—and with a growing call for scholarly integration 'extending ideas and issues grounded in various traditional communications disciplines' (p. 7). While strategic communication 'evolved from professions practiced as crafts', 'for a science to develop from immature to mature status requires a community of scholars working within an organized structure of assumptions, theories, and research traditions' (Holtzhausen & Hallahan, 2007, p. 1). Some 10 years later, the field still had not found such structure of assumptions, theories, or traditions that captured the scholarly interest. Still in search for the field's *essence*, Zerfaß, Werder, and Holtzhausen (Zerfaß, Werder, & Holtzhausen) renewed this call, urging the community to 'help us better define this field of study' (p. 198). Nothhaft (2016) was the first to take up this call to action, pointing to the problem with the development of the field as *fractured* and urging methods to *converge*. At one point he described how 'researchers in strategic communication should engage

---

in a debate about the very core of the field, i.e., clarify what exactly constitutes our object of research, and what exactly makes our research *scientific* (or at least scholarly)' (p. 75). Two rejoinders followed (Christensen & Svensson, 2017; Sandhu, 2017) where Nothhaft (2017) later points out that there seems to be an agreement about the *diagnosis* of the field and the issue with e.g., fragmentation, while there are different proposed *therapies* or *cures*. As Christensen and Svensson (2017) proposed, the journal celebrated its 10th year by engaging in a more in-depth discussion on the basis of the field, first through an ICA pre-conference, later through the special issue on the *Future Directions of Strategic Communication* and book (Nothhaft, Werder, Verčič, & Zerfaß, 2020) with the same name. In the editors' introduction to this special issue, Nothhaft, Werder, Verčič and Zerfaß (2018a) re-stated the conclusion that 'there is still no universal understanding of the pillars on which strategic communication rests' (p. 329), including articles finding that 'strategic communication scholarship can be described as lacking in its attempt to achieve high levels of disciplinary integration' (Werder, Nothhaft, Verčič, & Zerfaß, 2018, p. 346), that 'refinements to the way strategic communication has been defined may strengthen the consistency of purpose for research and theory building among scholars' (p. 347), while other papers described how 'the best is yet to come for strategic communication' (Nothhaft, Werder, Verčič, & Zerfaß, 2018b, p. 353).

[xv] This means that we can also consider strategic moral communication manifesting in societal structures. Conceptualised in Habermas' *zeitgeist* (see p. 33), the concept would signal that it is a time-and-place-specific moral discourse, introduced by Hegel (1979) as literally translating to the spirit (*Geist*) of one time (*Zeit*). Following a canon of other concepts, it is captured in other concepts such as Gadamer's *historical consciousness*, Benjamin's (1999) *zeitkerne* as 'a nucleus of time lying hidden within the knower and the known alike', Foucault's *episteme* (see p. 81), and perhaps even in Tugendhat's understanding of *morality* as a 'system of norms that exist in a society as a product of social pressures' (quoted in Habermas, 1994, p. 43). Habermas similarly speaks of zeitgeist as appearing not 'within an objective context of events but with the symbolic context of a spirit that expresses itself in them' (1988, p. 17), or as 'culture which generates structures of rationality or cultural value spheres' (1999, p. 17).

[xvi] Habermas, however, emphasises that rational reconstruction ought to be a methodology used by philosophers, solving the problems from which he derives his criticism towards these scholars' research.

[xvii] The two were scheduled to meet in a debate, Habermas writing an essay as a reply to Foucault's comments on his theory (described here in the text) in 1984 titled *Taking Aim at the Heart of the Present*. The debate did not, however take place as Foucault at that time had passed away, and the essay was revised by Habermas in consideration of this fact. While the two are presented as incompatible, some have indeed combined them in studies, as for example using a

---

'Habermasian discourse analysis' (Graber, 2017) or in other studies phrased as aiming to 'systematically apply Habermas' validity claims to empirical research' by developing a new form of critical discourse analysis (Cukier, Ngwenyama, Bauer, & Middleton, 2009, p. 175).

[xviii] In many studies, it might be interesting to look at discursive patterns over time, thus clustering texts not only in relation to the issuing author, but also in relation to the time it was published. In some studies, comparing the discursive patterns across types of material, it might be relevant to cluster all policy documents as one type while analysing all interviews as a separate type of material. Others might take an interest in an even smaller unit of interest, analysing the use of pronouns across the coding of quotes. As hopefully conveyed, moral discourse analysis would suggest that the first three levels are of interest in most studies, however neither arguing that these are necessary nor exclusive points of interest in every study.

[xix] In this first text, further texts were referenced in footnotes. In later sampled text, citations were found by e.g., following hyperlinks or by reviewing lists of recommended readings. While neither the first nor the second referenced text fitted the sampling criteria, the third reference was sampled as the second text in the corpus, the *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* (UN, 2015). This process continued throughout the document, then repeated with the new texts in the corpus, in turn identifying new texts and thus expanding the corpus backwards in time.

[xx] Through Google and Google scholar searchers, newer texts were identified by searching for pdfs where the title of a sampled text occurred. For example, finding a list of 600 citations to UNESCO's handbook in Google scholar, where many results point to scholarly publications. Through this process, later texts were identified such as the study *Disinformation and propaganda—impact on the functioning of the rule of law in the EU and its Member States* requested by the European Parliament's Committee on Civil Liberties, Justice and Home Affairs (European Parliament, 2019) that references both the handbook by UNESCO and the report by the UN presented above.

[xxi] Habermas touching upon this in i.e., 1994 (p. 160) and 1998 (p. 5). The very idea of *moral sociology* as a field is, however, nonsensical to some, arguing that morals are embedded in all sociological studies—dating back to Durkheim's 'view of society as the ultimate source of moral authority' (Eberhardt, 2014, p. 302) or as a 'system of rules for conduct' (Shweder & Menon, 2014, p. 50) where he recognised morality as the very basis for social life (Laidlaw, 2020). This, scholars argue, would mean that by 'identifying the social order within the moral order' it 'effectively stripped the latter of its analytical power' and thus all sociological studies 'in effect study morality' (Laidlaw, 2020). Recently, however, there has been a surge in publications on morality in sociology, reflecting the need for this new focus—as emphasised by Hickman (2007), Mattingly (2012), Cassaniti and

---

Hickman (2014), and Shweder and Menon (2014). The field is now focused on morals explicitly, concerned with…

> …how moral questions are posed and addressed or, symmetrically, how nonmoral questions are rephrased as moral. It explores the moral categories via which we apprehend the world and identifies the moral communities that we construe, examines the moral signification of action and the moral labor of agents, analyzes moral issues and moral debates at an individual or collective level (Das, 2012, p. 4)

[xxii] The limitations for the presentation of discourse ethics follows those previously outlined by Regh (1994), that 'there are a number of things I will not do here' including a) not presenting discourse ethics beyond Habermas' own proposition 'at most I can refer to such alternative views in passing or for the purpose of bringing out a systematic point in Habermas' approach', b) not a historical account 'although at certain points it will be necessary to note certain shifts in his position', and c) not even providing 'a full account of discourse ethics in all its systematic ramification' as this 'reach across the boundaries of moral theory into philosophical theories of language and agency, as well as into psychological theories of moral development' but 'can only touch on these ramifications, noting the various points where discourse ethics, as a theory of practical reasoning, opens on other discussions' (p. xv).

[xxiii] Elaborated, that together these two books 'serves the purpose of classifying practical discourse' as dealing both with 'the differentiation between the use of practical reason, then address the relationship between normative validity, understood in epistemic terms, and the non-epistemic concept of truth and, finally, connect moral theory with the discourse theory of law and the democratic constitutional state' (Habermas, 2018b, p. 104).

Publikationer från Medie- och kommunikationsvetenskap Lunds universitet

Beställning och aktuella priser på: http://www.bokshop.lu.se/
Böckerna levereras mot faktura.

Lund Studies in Media and Communication (ISSN 1104-4330)

---

1 Lars Palm, Övertalningsstrategier. *Att välja budskap efter utgångsläge* ISBN 9179662560 (1994)

2 Thorbjörn Broddason, *Television in Time. Research Images and Empirical Findings* ISBN 9179663931 (1996)

3 Inger Lindholm, *Målgruppen och budskapet. En modell för målgruppsanpassning av trafiksäkerhetsbudskap till unga manliga trafikanter* ISBN 917966430X (1997)

4 Åsa Thelander, *En resa till naturen på reklamens villkor* 216 sidor ISBN 91-7267-125-4 (ak. avh. 2002)

5 Ulrika Sjöberg,Screen *Rites: A study of Swedish young people's use and meaning-making of screen-based media in everyday life* 314 sidor ISBN 91-7267- 128-9 (ak. avh. 2002)

6 Charlotte Simonsson, *Den kommunikativa utmaningen: En studie av kommunikationen mellan chef och medarbetare i en modern organisation* 272 sidor ISBN 91-7267-131-9 (ak. avh. 2002)

7 Mats Heide, *Intranät – en ny arena för kommunikation och lärande* 244 sidor ISBN 91-7267-130-0 (ak. avh. 2002)

8 Helena Sandberg, *Medier & fetma: En analys av vikt* 297 sidor ISBN 91-7267-170-X (ak. avh. 2004)

9 Michael Karlsson, *Nätjournalistik – En explorativ fallstudie av digitala mediers karaktärsdrag på fyra svenska nyhetssajter* 240 sidor ISBN: 91-7267-212-9 (ak. avh. 2006)

10 Jakob Svensson, *Kommunikation Medborgarskap och Deltagardemokrati – En studie av medborgarutskotten i Helsingborg* 302 sidor ISBN 978-91-628-7512-1 (ak. avh. 2008)

11 Malin Nilsson, *Att förklara människan – Diskurser i populärvetenskapliga tv-program* 252 sidor ISBN 91-7267-296-X (ak. avh. 2009)

12 Michael Krona, *Från orsak till verkan – Berättarstrategier i Sveriges Televisions inrikespolitiska nyhetsförmedling* 1978-2005 305 sidor ISBN 978-91-628-7876-4 (ak. avh. 2009)

13 Sara Hamquist, *Dagspress, sport och doping – Mediaskandaler i ett samtida Sverige* 250 sidor ISBN 978-91-628-7979-2 (ak. avh. 2009)

14 Inger Larsson, *Att bygga broar över kulturgränser – Om svenskars kommunikation med icke-svenskar* 231 sidor ISBN 91-7267-312-5 (ak. avh. 2010)

15 Marja Åkerström, *Den kosmetiska demokratin – En studie av den politiska diskursiva praktiken i Sjöbo och Ystad* 261 sidor ISBN 91-7267-324-9 (ak. avh. 2010)

16 Veselinka Möllerström, *Malmös omvandling från arbetarstad till kunskapsstad – En diskursanalytisk studie av Malmös förnyelse* 243 sidor ISBN 978-91-7473-171-2 (ak. avh. 2011)

17 Tina Askanius, *Radical Online Video – YouTube, video activism and social movement media practices* 260 sidor ISBN 978-91-7473-393-8 (ak. avh. 2012)

18 Susanna Magnusson, *Att säkerställa att vi är välkomna: om organisationers strategiska kommunikation och förtroendeskapande arbete i det mångkulturella samhället.* 204 pages. ISBN 978-91-7623-207-1 (2014)

19 Jacob Stenberg, *The Communicative State: political public relations & the rise of the innovation hype in Sweden*. 266 pages. ISBN 978-91-7623-737-3 (2016)

20 Joanna Doona, *Political comedy engagement: Genre work, political identity and cultural citizenship*. 256 pages. ISBN 978-91-7623-895-0 (2016)

21 Nils Holmberg, *Effects of online advertising on children’s visual attention and task performance during free and goaldirected internet use -A media psychology approach to children’s website interaction and advert distraction* 212 pages. ISBN 978-91-7753-052-7 (2016)

22 Carolina Martínez, *Targeting children online: Young internet users and producers in the commercial media environment* 268 pages ISBN 978-91-7753-012-1 (2017)

23 Zaki Habibi, *Creative voices of the city – Articulating media, space and cultural identities by creative collectives in Southeast Asia* 222 pages ISBN 978-91-7895-666-1 (2020)

24 Rickard Andersson, *Strategic communication at the organizational frontline: Towards a better understanding of employees as communicators*. 101 pages. ISBN 978-91-7895-523-7 (2020)

25 Hui Zhao, *Politicizing crisis communication via social media: A contextual understanding of organizational crises in China.* 107 pages. ISBN 978-91-7895-464-3 (2020)

26 Maria Rosén *Social mättnad, relationell hunger: Strategisk kommunikation i förändring när matindustrin utmanas av folklig misstro*. 216 pages ISBN 978-91-7895-761-3 (2021)

27 Michael Rübsamen, *Om kändisjournalistik.* 231 pages ISBN 978-91-8039-285-3 (2022)

28 Alicia Fjällhed, *Strategic moral communication: A metatheoretical and methodological response to the normative perspective on strategic communication.* 260 pages ISBN 978-91-8039-908-1 (2024)

Strategic moral communication

Strategic communication has long suffered from a reputation of immorality, often being conflated with derogatory terms such as sophistry and propaganda. Scholars in the social sciences have been trained to accept this moralizing idea of strategic communication as a normative point of departure, through influential theories such as Habermas' discourse ethics and methodological approaches such as critical discourse analysis. This thesis challenges such assumptions. Shifting from a normative to an empirical stance, new perspectives on strategic dimensions in moral communication are presented. The thesis introduces 'strategic moral communication' as a new theory and 'moral discourse analysis' as a new methodology to explore such strategic dimensions in moral discourses. This approach is used to shine new light on the present moral zeitgeist, in an analysis of the discourse around fake news which the thesis argues reflect how our time defines moral and immoral public communication. The immorality of strategic communication may appear commonsensical, but this thesis argues that it is only common sense in our contemporary moral context. While many have assumed that strategic communication is immoral, this thesis rather finds that moral communication is strategic.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

Department of Strategic Communication
Faculty of Social Sciences
Lund University
ISBN 978-91-8039-908-1
ISSN 1104-4330