

The Philosophy of Online Manipulation

Edited by Fleur Jongepier
and Michael Klenk



ROUTLEDGE

Routledge
Taylor & Francis Group

NEW YORK AND LONDON

First published 2022
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2022 selection and editorial matter, Fleur Jongepier and Michael Klenk; individual chapters, the contributors

The right of Fleur Jongepier and Michael Klenk to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

The Open Access version of this book, available at www.taylorfrancis.com, has been made available under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 license.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book has been requested

ISBN: 978-1-032-03001-2 (hbk)

ISBN: 978-1-032-07114-5 (pbk)

ISBN: 978-1-003-20542-5 (ebk)

DOI: 10.4324/9781003205425

Typeset in Sabon
by Apex CoVantage, LLC

17 Manipulation and the Affective Realm of Social Media

Alexander Fischer

1 Introduction: The #StopTheSteal Manipulation

When the former republican US President Donald J. Trump used Twitter (especially in the last months of the year 2020) to vent accusations of an election being stolen by the Democrats, the goal was not to present rationally accessible proof concerning widespread voter fraud or dysfunctional voting machines. Instead, he used his favorite social media channel as a cogwheel in a broad strategy¹ to create a destructive fictitious pseudo-environment² interweaved with controversial, even outrageous accusations to stir up people's affectivity in order to spread mistrust in the democratic electoral process of the United States and get supporters and "believers" moving in opposition to what was supposedly happening. This ultimately culminated in the violent riot on Capitol Hill in Washington, D.C. on January 6, 2021.³ Furthermore, even a few days *before* the lost election in November 2020, the Trump campaign reached out to supporters via email and various social media channels with what became known as the "#StopTheSteal" campaign. This campaign was supposedly curated by the well-known, infamous strategic communication specialist Roger Stone in 2016 to be used whenever elections might not turn out victorious. Regarding the presidential election in 2020 said campaign was revitalized (Atlantic Council's DRFLab 2021; Spring 2020). As a fundraising effort (or rather scam) and, yet another tactical cogwheel, part of a thoroughly planned multifaceted manipulative strategy, it swept across inboxes, flooded Facebook and Twitter timelines spreading not only misinformation but also using destructive language and imagery to, then again, stir up people's affectivity and muster up motivation to act on these affects. Multiple agents joined in supporting the #StopTheSteal campaign, putting it in a broad frame, thus making it visible widely. On Facebook, a group with the same name ("Stop the Steal") and effort formed which was banned for the same misinformation and attempt to emotionalize quickly after its emergence. Furthermore, numerous Trump allies appeared on national TV trying to support the stolen-election-narrative that was pushed online all the more. Besides the carefully planned roll-out of the campaign, truly bizarre and dilettantish moments could be seen as well, not only showing us tools for the attempt to manipulate in a blatant way as they were handled poorly

in some cases but also how badly Trump supporters wanted this attempt to manipulate to be effective. One example: Trump's private lawyer Rudy Giuliani holding a rushed press conference erroneously in front of a landscaping business – instead of a famous hotel chain – with the name *Four Seasons Total Landscaping*. Then again, venting the same baseless claims of a stolen election while trying to suggest a (non-existent) credibility usually supported by means such as context, authority and the right timing to curate certain content. Trump's allies, in other words, tried to use an arsenal of symbols at the right time, at the right place, and by authorities that act personally worried, stating lies, formulating grave threats and painting a dark picture of the future to manufacture a public opinion and, once again, stir up people's affectivity. They provided a visually powerful blueprint of what to feel (and consequently also think) about an event no one has actually experienced for real, as “[t]he only feeling that anyone can have about an event he does not experience is the feeling aroused by his mental image of that event” (Lippmann 2008, 13).

For this extensive multi-channel effort to misinform and, amalgamated with that misinformation, *manipulate*, we can even suppose that Trump's strategists and willing allies had specific affective states in mind that they planned to stir up, such as (a) the feelings of frustration or indignation after the election night, (b) acute or persistent emotions such as anger and fear, and (c) moods of mistrust in democratic voting and the political system of democracy on the whole.⁴ Of course, it was paramount that Trump seemed to be the one able to slay the monster they build up before. To target these affective states via controversial and destructive messaging is especially effective to bundle *attention* – the main ‘online currency’ – and motivate individuals to act in a possibly destructive manner intended by a manipulator (think again of January 6, 2021). Furthermore, the systematic spreading of false information and the constant attacks on and disdain for reliable truth sources created (and still creates, as we saw yet again during the COVID-19 pandemic) a bedrock of disorientation – and thus an environment where affective states become primary tools for (self-)orientation and steering individuals.

So, the whole campaign was not at all about sound arguments and it was not based on tenable facts but instead it was about inciting all of the aforementioned affective states as a foundation to create, reassure and motivate believers. This ultimately gave ground for the suspension of Trump's mainstream social media accounts and the attempt for a second impeachment just shortly after the insurrection of the US Capitol. Once again, Trump impressively harvested the fruits of frustration, indignation, fear, anger, and mistrust in the political system mixed with faith in him in an at least partly disoriented, riven society. What we saw by Trump and his allies was a systematic attempt to *manipulate* the American people and undermine an essential democratic process by orchestrating false information and unsubstantiated accusations to muster up and connect to affective states that help motivate individuals to act in a certain direction.⁵ In this attempt to

manipulate, social media played a vital role as a digital realm of affectivity where a low threshold for publication makes it possible to anonymously reach a myriad of individuals, where their affectivity can be targeted in specifically tailored content on the grounds of big data, where our own affects are tools for orientation in the mass of stimuli, where algorithms favor controversial and destructive content, and where reviewing the substance of this content is, at times, tricky.

As social networks⁶ possess “phenomenological effects significant to the actions we take and the decisions we make . . . [and] are of no small consequence, raising questions about how and under what circumstances we are shaped by social media” (Nelson 2018, 3), my argument in this chapter will be the following: as digital realms of affectivity in which rational persuasion often plays a secondary role for many users, social networks offer potentials for manipulating users, modulating their feeling, thinking, and acting, which result from its very design. Political, economic, and also private agents try to use these potentials in *not necessarily always but often* questionable ways. Trump and his campaign can function as a crass example for a questionable use.⁷ From this starting point I, first, want to focus on the phenomenon of manipulation which is an important part of how our feeling, thinking, and acting comes about. I will offer a sketch of the phenomenon from the perspective of action theory and provide a conceptualization of what I see as the primary mechanism of manipulation which consists of rendering certain ends as pleasurable/unpleasurable, motivating us to act in a certain way without a coercing us but also without primarily using our capacities to rationally deliberate. Still, our rationality plays an important role in the context of manipulation as 1) a secondary rationalization of what and why we are feeling something and 2) the capacity to deduce reasons on the grounds of our affective states to act in a certain way.⁸ Thus, the manipulated remains, at least in a minimal sense, free to act in this manner or not. The manipulation I am going to be concerned with are instances of manipulation of affective states *suggesting* a certain direction for our decision and acting.⁹ This, in my opinion, marks the core of manipulation per se, in opposition to other types of influence like the – usually ethically esteemed – rational discourse where every decision is ideally based on rational deliberation (instead of affective states) or the – usually ethically debatable – use of coercion where there might be a lot of affectivity involved but which lacks an option to really decide freely in the end. The development of this manipulation model will be my primary focus. In a second step, I want to elaborate briefly on a assumed potentials of manipulation in social media which are based on a design of affective messaging and also on the interface’s design. My hope is to offer a fundamental perspective with an account of how specifically we are manipulated and, in more general terms, how social media supports this, thus laying the groundwork for more specific case studies regarding different phenomena like big data, interaction patterns, tailored advertising, fake reviews, and influencers as new manipulative tools are being developed constantly.

In my understanding, social media is constructive of human behavior and not with an unidirectional influence but as technology that acts on us as well as we act with it (cf. Nelson 2018, 4): “technological artifacts are not neutral intermediaries but actively coshape people’s being in the world; their perceptions and actions, experience and existence” (Verbeek 2011, 8). My focus here will primarily regard the interpersonal communication in the online realm whereas, for example, social bots will only be mentioned but not analyzed in depth.¹⁰ My proposed concept of manipulation will be applicable to the intentions of communicators who use certain interface designs and algorithmic effects to manipulate. Whereas I will claim that there *is* the potential to manipulate efficiently and effectively via social media it is neither denied that there are rational discourses and the exchange of sound arguments on social media platforms, nor am I adopting the claim that every attempt to manipulate is undermining our ability to decide freely, or that manipulation is by default morally problematic. To build this argument, I will now at look at manipulation as a certain type of influence in the landscape of influences attempting to cause an agent to feel, think, and act, differentiating it from other types of influence in order to gain a neutral account of manipulation with focus on its impact on affective states, called the *Pleasurable-Ends-Model* of manipulation (PEM) (cf. Fischer 2017, Fischer/Illies 2018, Fischer 2022).

2 What Is Manipulation, and How Does It Make Us Act? Conceptualizing a Type of Influence Offline and Online

The internet as a whole and social media in particular offers an environment that targets the affective side of our agency. Thus, it provides several instruments that use our peripheral routes of decision-making and also help create an impactful pseudo-environment.¹¹ Consequently, now there is talk of “online manipulation” (mostly, so far, outside of the discipline of philosophy), a term that simply aims at the exercise of manipulative influence in the digital realm of the internet (cf. Abramowitz 2017; Susser, Roessler, and Nissenbaum 2019b, 3). Here, manipulative influence is intended by human agents and then performed by means of a programmed, digital architecture that aims at various aspects of our affectivity (ranging from barely tangible qualitative reactions like us being prone to be comfortable to intense affective states like frustration, fear or anger). This architecture helps to lay the groundwork for shaping a certain action. Some of these structural intricacies are often called “dark patterns” which rely heavily on our slothfulness and avidity for convenient, functional default options and shortcuts. These effects apply to the basic design of an interface and your own profile settings (Facebook handles this masterfully): it is often used here that it is very easy, even in one click, to achieve something that is beneficial for the interface provider, whereas it just seems too hard to click through the depths of the settings to check certain privacy options or get rid of a product. Obstacles that make it hard to do something the providers do not want, are – as they

get on our nerves – effective and often found (like nagging questioning or laborious clicking). We also find social pressure via fake reviews or because of a large number of likes and shares, time pressure (in the economic context), or obscured ads all aiming at making us feel instead of thinking too much.

But social networks also serve as a potentially manipulative vehicle for private, political, economic, and other agents using a presorting algorithm logic that helps agents with an agenda to manipulate to feed specifically designed attention-seeking affective content into the realm of social media. Various algorithms support this as they are, at least partly, designed to engage us even more by showing supposedly relevant content. This content in turn has to be designed in a certain way, e.g. as controversial and destructive, in order to be considered by the algorithm. The rule of thumb here is that attention-seeking content has a better chance to reach many users. So, evoking an acute intense affective state like making someone angry counts for a lot as it helps engaging that someone to do something with the content (liking/disliking, commenting, sharing). This is at the center of my considerations. Before we dive deeper into this, let us try and understand how manipulation functions as a cogwheel in the workings of us being feeling, thinking, and acting agents.

In general, manipulation can be understood as an omnipresent form of influence of human agency which in its many costumes aims at shaping our affective states and with it our thinking and acting. It can be qualified as a form of influence on our capacities as acting agents alongside others. On a rough map of influences, rational persuasion and coercion are well explored forms whereas manipulation has been less well researched. This is not surprising as (a) its position seems to be somewhere in the messy in-between of these two, which can be seen as poles in a continuum of influence (cf. Rudinow 1978, 338; Beauchamp 1984; Coons and Weber 2014a; Fischer 2017, 53). These poles are not understood as a strict dichotomy. They are rather like brightness and darkness between which we can find many different shades. And (b) these many shades make the dissection of manipulation harder as things get opaque and more difficult to describe analytically the closer we get to our affectivity, the suggestive aspects of our communication and even our unconscious.¹²

Let us make the two mentioned poles clearer to gain a first understanding of manipulation *ex negativo*. In the context of coercion, an agent usually does not have the opportunity to choose between alternatives – or at least preferable alternatives – and thus to act freely in an extensive sense.¹³ Coercion can even be forceful so that agents have to reckon potential personal damage which often issues basic automatic behavior patterns like fight, flight, or freeze reactions as a threatening fear of consequences is evoked. In contrast, in the case of rational persuasion agents can reach a free decision on the basis of the correct and relevant information by forming good reasons without heavy pressure but with the help of their rational capacities. While coercion in extreme cases marks the absence of

free decision-making, rational persuasion is generally regarded as an ideal since it accounts for our autonomy and freedom and is not “contaminated” by the unnerving prospect of cruel consequences or factors that make free deliberation more difficult such as underhandedness or deceptive information. In general, something else is supposedly missing from rational persuasion, at least ideally conceptualized: our affectivity. The capacity to be guided by our affectivity plays an ambivalent role in the debate about the nature of free decision-making and acting, ranging from being an important part of it as our affectivity is thought of as intertwined with and a vital part of our rationality to being a grave threat to free decision-making. The truth, as it often does, might lie somewhere in between. Our affectivity seems to play an important role for a reasonable decision-making process as it shapes a meaningful perception of the world, helps us to judge, identify values, and be motivated; at the same time a decision that is worth to be called “free” also accounts to being able to prescind ourselves from what we are feeling to ultimately gain an integral decision as the foundation of our action.¹⁴

In contrast to an idealistic account of rational persuasion, manipulation is supposed to be a threat – at least in an everyday understanding. Here, underhandedness, deception, negative consequences, and selfish manipulators play a vital role. However, I want to oppose most of these characteristics as necessary conditions of manipulation. But there is one that does not seem to be possible to reason away: the “contamination” of our decision-making process by our affectivity, consisting of our feelings, emotions, and moods. In contrast, underhandedness, deception, and negative consequences can be seen as *amplifying* conditions even though they are not necessary ones. Thus, manipulation in my understanding neither accounts for a purely rational decision-making process nor does it force agents to do something specific like coercion does. It is not so much good reasons and the presentation of all the relevant information (or a gun to our head as a brutal form of coercion) that lead us to act in a certain way within the framework of manipulation but rather the curation of our affective states and the peripheral routes of decision-making.

In the literature on manipulation, we can find various suggestions on how exactly to understand manipulation and differentiate it from other forms of influence. I have already hinted to an understanding which nonetheless needs elaboration. In order to provide this elaboration, I will very briefly summarize the general discussion on manipulation and its different definitions. Following this, I will propose a new, integrative definition of the term “manipulation”, inspired by previous philosophical attempts to define it, with regard to action theory, while trying to avoid some of the problems of earlier concepts of manipulation. In the light of all this, manipulation is understood as a type of influence where a manipulator actively leads the manipulated to choose a certain end (e.g., an action or a product), but the manipulated stays at least in a minimal sense free to choose this end or not.¹⁵

Manipulation consists of intentionally modulating the affective attraction of certain ends or their realization by rendering them as pleasurable/unpleasurable, thus making some options more (or even extremely, whereas others not at all) appealing to the manipulated and consequently more likely to be chosen (cf. Fischer 2017, 2018, 2020, 2022; Fischer and Illies 2018). Our affective states are actively modulated so that the evaluation of a certain end can change, resulting in often complex affective experiences which eventually will boil down to a desire or an aversion to act accordingly to the manipulator's goal (or not).

This understanding rests on the premise that an action (in contrast to mere behavior) is a realization of a pro-attitude toward an end, selected as being more fitting than alternative ends (this is basically how Aristotle understood it). A chosen purpose then leads to an action, if we conclude that an action does not conflict with other ends that we have and if there are no limiting conditions to fulfilling that end. Choice-worthy ends for our actions are then manifold: both good and bad, objective and subjective and maybe because we just like them. To gain a better understanding of why we act it is helpful to turn to Aquinas's more general application of Aristotle's practical syllogism where he distinguishes three types of choice-worthy ends: those we (1) desire *for their own sake* as ultimate ends (such as truth), those we (2) have because they are *useful* and serve, in direct or indirect ways, other ends we have (such as healthy food that makes us healthy), or we (3) desire because they are *pleasant* (such as appreciation or, more mundane, chocolate) (for a more detailed account see Fischer and Illies 2018, 35–39).¹⁶ In the case of manipulation, our affective evaluation of an end is modulated by presenting an end as pleasurable/unpleasurable (instead of presenting something as useful or choice-worthy for its own sake). A manipulative stimulus thus is used to trigger an immediate qualitative reaction, an affective response, to a pleasurable/unpleasurable end which might create a desire or aversion to do or not do something. This desire/aversion aims at the alteration of the reality so that reality accords. (Sure, we often cannot directly act on these grounds and consequently have to cope and see where and when we can really act in accordance to our affective response.)

In order to achieve this, underhandedness, deception, negative consequences and even careless, selfish manipulators are not necessarily needed but can function as amplifying conditions. This may seem like an atypical definition, as it partly leads away from our everyday understanding of the phenomenon, which usually degrades “manipulation” to a fighting word (even though the term was long used neutrally but, in a neo-Marxist tradition, manipulation was tinted as unethical), thus often blocking a clear view of what is happening in detail as we are so convinced it is something devilish.¹⁷ So let us take a quick look at the steps that lead me to this definition and thereby summarize existing accounts (for a detailed discussion, see Fischer 2017, 26–78).

3 The Different “Schools” of How to Understand Manipulation

The first perspective on manipulation emphasizes the character of manipulation as an intentionally underhand influence that unfolds beyond our consciousness and is therefore almost not at all controllable for the manipulated (cf. Baron 2003; van Dijk 1998; Goodin 1980; Ware 1981; Noggle 2018). Daniel Susser et al. also identify the essential feature of online manipulation “as the use of information technology to covertly influence another person’s decision-making, by targeting and exploiting decision-making vulnerabilities” (Susser, Roessler, and Nissenbaum 2019a, 6; for an argument against this see: Klenk 2021). This applies to many dark patterns where something is, for example, secretly placed in our shopping cart, where tricky questions lead us to answers that we did not intend, where additional costs remain hidden or where advertising is disguised so that we click on it because we simply think it is something else than advertising. However, highlighting “covertly” does not help much to differentiate manipulation. Almost everything that happens in a hidden manner, trickily or secretly might then be understood as manipulative – such as lies, cheating (e.g., in a game), or even magic tricks. While there are many examples in which underhandedness is part of an attempt to manipulate and might be considered an amplifying condition, it is not a necessary or sufficient feature of manipulation. This can be seen not only in the personal context (e.g., when relatives blatantly induce guilt) but also in the affective realm of social media: it is widely known that advertising tries to grab us by our affectivity and that this is done in a specifically personalized manner or that right-wing trolls want to stir up an acute emotion like anger (often as a part of a long-term strategy). It still works.

Closely related to this notion of manipulation as a hidden and secretive tactic to influence is the understanding of manipulation as an encroachment of an individual’s perception of reality, in other words: as a form of deception or trickery (cf. Scanlon 1998, 298; Noggle 1996, 44; Cave 2007). Very often the terms “deception” and “manipulation” are even used synonymously.¹⁸ For the specific online context, Susser, Roessler, and Nissenbaum (2019b) refer to the weaknesses of agents in the decision-making process which are exploited, for example, the difficulty of being able to directly falsify every piece of deceptive information. This association with deception explains the often-made (and not false) affiliation of “fake news” with manipulation.¹⁹ But also dark patterns can be associated with deception, for example, involuntarily sharing more information than wanted without knowing it (ironically called “privacy Zuckering” – aiming at Facebook’s CEO Zuckerberg). However, these attempts to influence an individual in a certain direction are not necessarily manipulative as they are simply obscure paths and/or false information, which represent a form of deceptive rationality but do not always aim at influencing our affective states (but *can* be used for that as #StopTheSteal illustrates). This characterization of

manipulation also seems to be over-inclusive since it, for example, includes any form of marketing as manipulation that goes beyond the presentation of factual statements about the product offered.²⁰ Emphasizing deception as inherently manipulative also narrows our vision of manipulation down to the presentation of incorrect information and false reasons. This, again, can be seen as an amplifying condition of manipulation by specifically using false information to affect and to arrange the perception of reality of an agent. That helps us to understand how editorially curated and preselected content via algorithms can be part of an elaborate manipulation scheme. Nevertheless, deception does not have to be understood either as a necessary or as a sufficient condition of manipulation because it is also quite possible to use actual facts to manipulate. An adequate concept of manipulation should consider both cases. Those where underhandedness and deception play a vital role and those where they do not.

Another characterization of manipulation associates it with the manipulator's pursuit of egoistic purposes which yield negative consequences for the manipulated. Here, two components are addressed: first, that manipulation is tied to corrupt, selfish characters, who often carelessly use manipulation as a means for a clearly selfish end; second, that they pursue ends that are useful and pleasurable to *them* and harm the manipulated (Green and Pawlak 1983, 35–37). Against the background of egoism Marcia Baron describes manipulation as a condemnable form of harmful and selfish behavior, even as a “vice” involving “arrogance” (Baron 2003, 37, 49). However, this also seems to distort the perspective on manipulation (just think of Shakespeare's Iago). Ultimately, there is no doubt that there are many cases where manipulation involves careless, harmful and/or selfish intentions on the part of a manipulator and where it consequently does damage to the manipulated. However, at the same time it is neither reasonable to claim that any kind of manipulation is careless and harmful, nor that it serves only negative purposes of a selfish manipulator. In general, the same mechanism that leads people to bad actions can also lead them to good actions (Fischer and Illies 2018, 31). Manipulation is also often very thoroughly planned (and not careless at all) and can even be regardful. In such cases manipulation might even do something good for the manipulated as they are nudged, for example, to a healthier or more environmentally friendly life (cf. Noggle 2018). It is also far reaching to say that anyone who manipulates is careless, has a corrupt character, or that the benefit of the manipulator always constitutes the direction of the manipulation. We just have to think of romantic relationships in which one person is selflessly concerned about the welfare of the other and yet does not try to convince rationally (e.g., because they know the other one will stubbornly reject a good argument or else). In consequence, negative consequences and careless, selfish manipulators should not be considered as necessary or sufficient conditions of manipulation. It is, in fact, interesting why we seem to cling to a negative understanding of manipulation (more on this in Fischer 2022).

The most important characterization of manipulation, which is often implicitly contained in the previous ones (because the manipulated do not notice being manipulated, are misinformed, or have fewer abilities, etc.), conceptualizes manipulation as a form of influence that at least partially bypasses our rationality and possibly even undermines it completely (cf. e.g., Wood 2014; Gorin 2014b). Robert Noggle suggests that manipulation is an act where the manipulator controls someone by using their “psychological levers” (Noggle 1996, 44). He suggests that manipulation leads the manipulated “astray from certain paths toward certain *ideals*” (Noggle 1996, 44) by the already mentioned deception (changing a belief), changing situations or conditioning (changing desires) or inducing emotional states like guilt (changing feelings). Noggle seems to assume that the ideal way of decision-making lies in the rational deliberation on the basis of good reasons. Susser et al. assume with regard to online manipulation that ideal decision-making processes are prevented by the use of weaknesses. Noggle’s view of the circumvention of rationality points in an important direction: as assumed previously, manipulation does, on the one hand, not primarily use rationality and good reasons or, on the other hand, coercion; it at least circumvents our rationality to a certain extent and strengthens the role of our affectivity in decision-making. The view that manipulation disconnects the links between good reasons and our decisions is still very popular (Fischer 2004; Wood 2014; see also Barnhill 2014; Gorin 2014a). This is because of a threat to autonomy that manipulation supposedly entails. Ethical concerns usually take over by that point.²¹ But, in the case of manipulation, even though it uses the biological and bounded rational side of our being, there still seems to be room for a rational and free decision, not always following the path of a modulated affectivity that motivates us to act in a certain way. Free agency remains robust but at least might be challenged. This is probably one of the reasons why manipulation is a particularly interesting type of influence in liberal societies as we are still able to act on the basis of *our own* affectivity and can usually decide for or against its suggested direction – even if this is not always easy. Manipulation can make it difficult to act in a way we would rationally choose, it can lead us in a certain direction (although hardly generate completely new feelings, emotions or moods) while flying beyond our rational radar, but it does not establish a one-way street of decision-making (a manipulative influence can be very weak when, e.g., default options use our slothfulness²²). Otherwise, we should speak of “coercion” and not “manipulation”. Consequently, at least partially circumventing the rationality of an agent is a necessary condition for manipulation. We are finite, boundedly rational beings with a talent for rational deliberation and a colorful affectivity – both constantly interacting with each other. To overstress our rational capacities and dodging our complex affectivity seems to be one of the standard moves or even the “life-style” of Western society (Gellner 1992, 136) – something Martha Nussbaum once

called (borrowing a term by Frans de Waal) “anthropodenial” (Nussbaum 2008; see also Fischer 2017, 91–103).²³

4 An Integrative Understanding of Manipulation: The *Pleasurable-Ends-Model*

Even if it is not reasonable to simply define manipulation as the vitiation of human rationality (and our freedom and autonomy) and as underhand, deceptive, or a harmful means to achieve selfish ends as a manipulator (and all of these features together), all of the mentioned approaches point to items worth discussing. What they lack, though, is a detailed description of the mechanism as a foundation of manipulation that allows us to include or not include the mentioned (and at times amplifying) features. Generally speaking, manipulation introduces an influence into the development of our thoughts, decisions, and acting by modulating our affectivity. It is especially the evaluation that is strongly suggested by our affectivity that seems to be the target of a manipulation (and where all kinds of measures are used for it); in other words, the affective significance of an end is tried to be changed. In order to focus on the description of manipulation and its mechanism (without simultaneously involving an ethical assessment), I suggest understanding the phenomenon as follows.

The *how* of manipulation includes three steps. (1) An attempt is made to actively change the affective attraction of certain ends or their realization, in the sense that the realization of the respective end is more pleasant or unpleasant than the felt status quo. This is usually done by depicting a change that reaches us by our affectivity through effectively contrasting what is and what could be (cf. Ben Ze’ev 2001, 15) – for example, thieves stealing an election and the dark future after that. The prospect of a pleasant or unpleasant change in the status quo then makes (2) an option more attractive (or even extremely attractive) for the manipulated, whereas others not at all and thus (3) more or less likely that this option is chosen (Fischer and Illies 2018, 27; Fischer 2017, chapter 1). Attractive is everything that is connected to a sense of well-being (or vice versa in regard to unattractiveness). Well-being is founded in our interests, dispositions, and artificial and natural needs. As these things want to be satisfied, they can be used manipulatively. The evaluation of a certain end changes, resulting in often (but not always) complex affective responses which eventually boil down to a desire or an aversion regarding a certain end. This motivates the manipulated to act in a manner according to the manipulator’s goal (Fischer 2022). Thus, manipulation can be seen as a form of influence where a manipulator leads the manipulated person to choose an end (an action, a product, etc.) but where the manipulated remains free at least in a minimal sense to decide whether she or he adapts this end or not. This is where the necessary and sufficient prerequisites for manipulation lie.²⁴

By shedding more light on the mechanism an integrative and neutral understanding is achieved that leaves space for the various aspects discussed earlier and thus also upholds a connection to the everyday usage of the term “manipulation”, while refining it. The *Pleasurable-Ends-Model* of manipulation can, but does not have to, include deception, underhandedness, negative consequences and selfish characters as possibly amplifying conditions of a manipulation since it focusses on the mechanism, which can be supported with certain more or less effective and applaudable means. According to this definition, it is necessary to push rationality aside (at least to a certain extent and as a primary mode of judging) as well as modulating our affectivity to change the evaluation of an end along the lines described previously. Just bypassing rational capacities is not sufficient, as this can also be done in ways other than manipulation (e.g., through underhandedness or deception). The active modulation of our affectivity, more precisely: our feelings, emotions, and moods with regard to the attraction of an end, is necessary (cf. for the properties of our various affects see Ben Ze’ev 2018, 112–137).

These three spheres of our affectivity make clear why there are many different suitable ways and instruments of manipulating as they have to be addressed specifically. Feelings are qualitative inner triggers that seem to be primarily responsible for making us act in a short-termed manner and can be used manipulatively by triggering impulses. Emotions, acute or persistent, are more complicated as they make us feel, think, evaluate, believe, and ultimately decide and motivated to act; they consist of cognitive and affective states simultaneously.²⁵ They can be used manipulatively for acute purposes or in broader, long-termed schemes. Moods tint certain things in life in a long-term manner, they often “belong” to us, can sometimes become firm dispositions, and establish certain manipulation-relevant triggers that can reach us more effectively (you can seldom catch the melancholic with outright fun).

With this characterization, we gave manipulation a place on the map of influences by determining its mechanism which enables us to distinguish it from other forms such as rational persuasion and coercion. If and how manipulation can count as a morally legitimate type of influence is so far only hinted at: if selfish manipulators, deception, underhandedness, and negative consequences are involved, manipulation tends to become morally problematic. However, there is more differentiation needed to dissect the difficult question about the morality of manipulation (for more see, e.g., Fischer 2017, chapter 3, 2018; Noggle 2018; Wood 2014).

But back to social media: due to the hinted at many shades in the evolution of an affective state/thought/decision/act, we are faced with a conceptual problem that becomes particularly clear in the online world, where technology stands between users and an interest pursuing beneficiary who uses certain technology, e.g., Trump and his campaign. It is already just not always clear how exactly an affective state/thought/decision/act comes about in the analogue world. The online world maybe makes this even

messier. Here, when trying to get a better grasp of online manipulation, we can make a difference between (a) built-in structures of an interface that are leaning towards coercion and some that are outright manipulative and (b) interpersonal communication and the message design online.²⁶ This kind of online interpersonal communication via digital surfaces seems to make, in comparison to offline attempts, a pointed version of manipulative communication possible.²⁷ Sure, social media is used in various ways, e.g., as a major news source, for social interaction and self-presentation, but exactly the mixture of these features makes it so interesting for manipulation. Not just because people create a vital part of their construct of reality there (when using it as a news channel and collecting basin for perspectives) but also because individuals can be reached easily and are generally interested in interacting in the realm of social media (from clicking the “like” button to sharing, commenting, and posting). All the more, relatively few agents can reach large, interacting groups creating certain dynamics. But how then does social media support to actively change an attraction of an end or its realization? I will make a few remarks about this toward the end of this chapter.

5 Social Media as an Affective Realm Providing an Environment and Tools for Efficient and Effective Manipulation

Now that we have conceptualized manipulation, it is important to outline the characteristics of social media that render an efficient and effective manipulation possible. Let us come back to the example from the beginning for a moment. #StopTheSteal by Trump and his allies counted on misinformation. But it did more: it used affectively loaded, destructive language and imagery prone to set peoples’ affectivity on fire with a certain narrative rendered salient, counting on controversy to stir up affective responses like frustration, indignation, fear, anger, mistrust, and a belief in Trump; it provided a memorable phrase for this to simplify a very complex issue (as, e.g., hashtags on Twitter often do) and used video clips and pictures reminiscing a dark and possibly violent future (if the election stays “stolen”) whereas a “heroic leader” like Trump could slay the threatening monster – of course with the help of the recipients. By means of all this, #StopTheSteal literally tweaked algorithms and created widespread attention. So, there are two aspects to be differentiated: first, there is the designed controversial messaging trying to effectively carry content – how is it trying to influence (by using controversy, hope, . . .)? In what way does it primarily aim at our affectivity? Is it trying to depict an end as pleasurable/unpleasurable (in a blatant or subtle sense²⁸)? Which means does it use do that? Does it stand alone or is it framed in a bigger context? And second, there is the interface that serves as a vehicle for that affective messaging and is designed for this purpose – is the platform aiming at affectively engaging its users, making them prone to

affective messaging? How does the affective message benefit from certain interaction possibilities, platform rules and algorithms?

Regarding #StopTheSteal, social media served as an ideal surface for a campaign like this that was a multichannel purpose using a multitude of communication devices where heavily affective messaging with controversial and destructive dark and heroic imagery, was pushed and constantly repeated to call for action. It aimed at user's affectivity, presented an end state as pleasurable (Trump staying president)/unpleasurable (the election staying "stolen"). The design of social media platforms like Facebook is a useful vehicle for such affective messaging as the interface itself aims at affectively engaging us. Involvement is a key concern for social media that thus try to provide a convenient, surprising, fun and informative platform with its algorithmic news feeds, videos, pictures, written messages provided by various agents calling us to react by consuming, commenting, liking, sharing. Publishers want the attention of users, which is created by constant repetition and flashy messages. Reaching the user's affectivity guarantees the biggest success: more followers, more likes and more shares – all of which broaden the distribution of a message (which is of course supported by programmed social bots which create even more likes and shares). Rational arguments and the confirmability arguments need are drowned out, as primarily affective content goes viral more effectively. Thus, it is only logical to design messages this way for successfully being recognized and achieving a goal. This fundamental construction of social media creates a manipulative potential as the affective realm can ultimately function as an affectivity catalyst, modifying the attraction of certain ends and thus making it more or less likely for them to be chosen. Let us look at the different bricks that provide the walls of this realm and connect them to how they help manipulative messages that render an end as pleasurable/unpleasurable to be successful and widely recognized as well as pointing out specific interface features that rely on, invite, and reward affective interaction, thus making it a useful tool for efficient and effective manipulation.

1. Social media interfaces are designed in a manner that tries to make users stay. Nir Eyal offers the thesis that it is a discomfort (feeling bored, lonely, confused, fearful, lost, or indecisive) that brings us online to find (often very short-termed) relief in interactions that distract us, even make us feel good and thus offer relief (Eyal and Hoover 2013). So, we are using social media (at least partly) to fulfill affective needs. If our own content is recognized and actually evaluated positively we like to come back. Recognition feels good (at least often). The short-high that comes with it seems to be close to what we are feeling when we shop. The interface design is in a simple way focused on basic conditioning, as rewards bring us back now and again. But, usually rewards do not carry us too far; keeping someone at it in regard to a certain direction and, on this foundation, the development of habits and at last manufacturing an

inner cognitive and affective connection need more than simple reward systems. Thus, e.g., Facebook offers far more than quick rewards: a news feed which is pleasurable in itself by being convenient, informative but all the more affectively stimulating by being sometimes surprising, the positive amazement or the outrage conveyed through affective messaging and the attention through receiving likes and shares. This is the multifaceted foundation social media is built on, creating lust and a routine to use it. As the interface already aims at engaging us primarily on an affective level, users are, at worst, ultimately made to be manipulable as we are kept in ‘affective mode’ when using social media; thus rationality is potentially put on the back burner.

2. Being an informative platform plays an important role in modifying an end in the context of manipulation. As soon as we dive into the world of social media, it becomes clear how much a significant part of our reality consists of using the shimmering bluish screens of our computers, smartphones, and tablets. Here, it is the representation of controversial messages suggesting drastic (often depicted as destructive) changes of personal relevance that grasp our attention and stimulate our affectivity but not so much a rational discourse. On the message side controversial and destructive content like #StopTheSteal becomes effective in creating a pseudo-environment that users turn to and use as a basis for their feeling (including the evaluation of a certain attraction of ends), thinking (i.e., their beliefs), and ultimately acting. To communicate affectively, this is hardly surprising, pictures, videos, and rather short messages are often more effective than long texts (Döveling 2015; Sachs-Hombach 2003). The evolution of advertising in the twentieth century gives proof to this tendency: words are less and less important (if, at all, they are important in the form of slogans), and the focus on pictorial messages is of growing importance. The increasing significance of Instagram and TikTok seems to show exactly this. The convenient presentation due to algorithms helps to manifest a pseudo-environment as it connects well to our slothfulness whereas the multitude of stimuli supports affective heuristics to sort through all the content. Difficult, potentially blinding, but intense states of affectivity (that can stand out), like fear and anger, on the grounds of the formula “excitement instead of information” are thus guiding principles for designing social media content in order to reach high visibility and, for example, unravel users.²⁹ This bears not only the danger of a rational discourse being drowned out in certain contexts and regarding certain topics but also to provide a distorted affectively loaded perspective onto reality. Shortcutting the ways to create a construct of reality opens up potential for misinformation. This also helps create a foundation on which manipulation is advantaged by the possibility of a presentation of certain ends as pleasurable/unpleasurable so that an effort would be needed for users to differentiate or distance themselves from the

attempt to modulate our affectivity by this presentation and, in broader strategies, a whole pseudo-environment.

3. With regard to the interface design, where algorithms conveniently provide you with related posts, groups, and sites “you might like”, a strong temptation might occur to design your social media space in the form of an echo chamber in which existing convictions and affective states are reinforced within a relatively hermetic system. So, this is not only done by algorithms themselves but also supported by conveniently being able to subscribe in one click without checking certain agendas before. This invitation to create echo chambers can support social and political polarizations and the normalization of problematic opinions, since outside influences are hardly able to penetrate this bubble. Again, to be worthy of entering a user’s echo chamber there is a need to be visible with which affective messaging helps. Also, it helps to drown out the need for rational checking as convenience and social pressure might just make a user give in. The phenomenon of being presented, searching for, and interpreting information according to one’s own expectations is also known under the term “confirmation bias” (Pohl 2004, 93). In addition, claims might turn into felt truth (something Trump liked to legitimize constantly while talking about the “stolen election”) if they are often and constantly repeated, ultimately manifesting an end as pleasurable/unpleasurable and supporting the realization of an end. So, sharing and liking of posts, which thus receive greater distribution and attention through algorithms, helps produce this phenomenon of felt (not known) “truth” solely through the widespread attention and an accompanying principle of repetition (Heath 2015, 191) rendering certain narratives and their images salient, highlighting certain ends as especially pleasurable/unpleasurable.³⁰
4. The design of social media platforms with its masses of stimuli also invites using shortcuts to evaluate something. In our modern societies we find a steady high frequency of information stimuli. The media has always played an important role in structuring these stimuli and the environment they stem from for us: “For the real environment is altogether too big, too complex, and too fleeting for direct acquaintance. We are not equipped to deal with so much subtlety, so much variety, so many permutations and combinations” (Lippmann 2008, 16). Since much more content is being produced than we can look at, affective orientation is a useful method. Again, algorithms support this form of orientation by showing posts and related ones with which users interact the most. Also, we receive more and more sometimes crude suggestions that possibly cement an individual’s echo chamber and the pseudo-environment that comes with it. Whatever most effectively appeals to our affectivity, thus successfully creating attention, can go viral. After all, what counts in the realm of social media is evaluation: “like” (on Facebook there is more options like “love”, “haha”, “wow”, “sad”,

“angry”); there is no “needs fact-checking” button. Like this, we are constantly engaged affectively and less so rationally, giving way for effective affective messaging.

5. Such affective messaging design seldom causes a problem in regard to publishing rules in social media. Anyone can say or show something about almost anything at any time, often without having to give up anonymity. Trump and his allies did not have to go through *any* process of certification to forgo their manipulation, they just published and the only obstacles they had to face (very late in the process) were deleted groups on Facebook or tweets marked with a little flag telling us such and such claim “is debated” (in the end this campaign seems to be one of the main reasons why Trump and some allies got banned on social media platforms for years – a historic intervention). Other than that, they could count on reaching millions of recipients within seconds, distributing their attempt to manipulate even more with the help of affective-design-rewarding algorithms.³¹ With the interface design disregarding identities, a manipulative attempt becomes intransparent whereas at the same time the efficient distribution of affective messaging provides users with a suggestion of something being true and heart-felt. Selfish manipulators are hidden, motives unclear, and self-regulation by authors as well as recipients seems in some case to be massively weakened (Trump is a perfect example for this). On a sidenote: the lack of efficient social and legal control supports the virtue of “temperance” to crumble in digital communication (Vallor 2010).³²

Let us sum up: by using the logics of seeking attention, social media tries to engage its users by means of its interactive design and strongly algorithmically selected affective content. Attention is created by flashy affective messaging depicting controversial, often destructive and drastic changes, trying to touch users; the visibility of these messages is often supported artificially by social bots and/or trolls, thus blowing up specific topics. The content itself mainly contains pictures, videos, mostly (very) short texts, if any, rich with simple messages and/or symbolism, so that the interpretation of this specific content is left to a mixture of confiding in friends who shared it, confiding in friends of the friends who are engaged in groups, making up the foundation of an echo chamber, and an affective evaluation of what is shown. Data analyses helps tailoring (economic but also political) ads groupwise and even individually, making it possible to target specific feelings, emotions, and moods.³³ Here lies a big potential for online manipulation; as well as with convenient dark patterns using our slothfulness, fake reviews and other means creating social pressure, and influencers providing a parasocial interaction where ends are rendered pleasurable/unpleasurable efficiently and effectively.³⁴

Christopher Wylie, the whistleblower in the 2018 *Cambridge-Analytica*-scandal, once said: “We exploited Facebook to harvest millions of people’s

profiles and built models to exploit what we knew about them and target their inner demons” (Cadwalladr and Graham-Harrison 2018). What Wylie stated to the journalists of the *Guardian* describes the core of Trump’s #StopTheSteal campaign: to target inner demons and to stir up possibly problematic affective states to make certain individuals become a part in what could be said was a tried – and luckily unsuccessful – *coup d’etat*. What Trump and his campaign did is clearly not ethically applaudable. They tried to create a pseudo-environment that had nothing to do with the facts; they targeted affective states in a multi-channel effort that are especially motivating while possibly blinding us for being able to see the complexity of things: frustration, indignation, fear, anger, mistrust, and an affection for Trump. They well knew the affectively engaging attention-focused design of social media and instrumentalized it for a harmful purpose on the grounds of selfish ends with deceptive (whereas not so much underhanded) and, most of all, heavily affective controversial content, suggesting that a *status quo* would result in an unpleasant reality, thus making democracy-undermining actions a pleasurable end. They were influenced in a way that can serve as a destructive example of an illegitimate (online) manipulation.

These basic structures certainly do not seduce everyone to move affectively guided in the social media world, to be manipulated, or use these structures to manipulate. However, social media has a strong potential for this in the sense of an almost optimally designed affective realm with an efficient interface using our human condition (especially by patterns), para-social interaction and our need to be connected, as well as the possibility to retrieve data about users via tracking their behavior online which can make an attempt to manipulate even more efficient and effective.

Notes

I want to thank Klaus Sachs-Hombach (University of Tübingen), Damian Cox (Bond University), Christian Illies (University of Bamberg), Fabian Geier and Sebastian Krebs (CODE University Berlin) for comments on parts of this chapter which I presented at various occasions. Also, I want to thank the organizers and participants of the workshop preceding this volume for their suggestions.

1. Trump – who was not the only but certainly the most prominent one – first began tweeting allegations of fraud in April 2020. Since then, he made these allegations occasionally here and there, soon tweeting more often and regularly about it, ultimately leading to the systematic attempt we saw in the last weeks before, then during (establishing the hashtag #StopTheSteal and videos of suggested election fraud going viral), and after the election. Trump and his allies sure did not just tweet. They used every outlet possible to weave in the fraud allegations into the public discourse and the heads of Americans. What we witnessed here was a classic build-up of fear, frustration, and anger by using a fictitious scenario to create images in the heads of US citizens that should ultimately make them act in certain manner: to fight a “rigged” election, possibly making it possible for Trump and his allies to stage a coup and try to subvert democracy. Social media played a vital role in this attempt which means that

there comes a whole lot of responsibility with it when designing and curating this online realm.

2. This is a term I am borrowing from Walter Lippmann, meaning fictions, images in our heads, that shape our perception of the reality of the “world out there”: “We shall assume that what each man does is based not on direct and certain knowledge, but on pictures made by himself or given to him” (2008, 25).
3. Lippmann wrote, that “it is clear enough that under certain conditions men respond as powerfully to fictions as they do to realities” (Lippmann 2008, 14). This has become painfully clear once again when looking at how events unfolded over the course of Trump’s presidency, especially in regard to the presidential election of 2020, but also during post-election times where the republican party continued to keep the narrative of a stolen election alive, leading to severe changes in the appearance of the party and its political direction which in part seems to approach a fascist posture.
4. In my understanding we can divide our affectivity as a whole into these three categories which are all related to one another but not quite the same. *Feelings* are qualitative bodily impulses that can be very basic just like pain. Pain is clearly a feeling but not yet an emotion. *Emotions* are more complex because they contain not only feelings but also other components of a cognitive, evaluative, and motivational nature. They are intentionally related to an object in the environment and usually acute (like anger) or extensive and persistent (like a very complex emotion like love). Love, on the other hand, despite its persistence is not a *mood* like melancholia (which can also be constant), because it is more specifically related to an object, that is, the loved one, while a mood has a generalized scope and colors our lives in many areas (cf. Ben Ze’ev 2001).
5. This stands as a warning for every other nation grappling with right-wing populists and others with a disrespect for reason, truth, and democracy as the latter rests in large parts on exactly this: reason and the will to adhere to a factual basis and solid measures of how to count something as evidently true or false.
6. I am going to use “social media” and “social network” synonymously.
7. Also briefly consider this positive example: in the beginning of the COVID-19 pandemic Vietnam’s Health Ministry produced an upbeat song (called “Ghen Cô Vy”), inspired by a precursory dance challenge on TikTok, teaching the necessary measures like handwashing, how to correctly wear a face mask, and so on. The video went viral all over social media and reached millions of people with its slightly kitschy animation which includes demonstrations of handwashing, warnings about face masks and public gatherings, and a gloved hand flicking away an angry-looking green coronavirus particle. This song (there are many others in the Asian music world) as fun and upbeat it is (while also being informative), grabbed lots of individuals by their affectivity manipulatively nudging them to act in the safe way necessary.
8. I developed a detailed account in my book *Manipulation. Zur Theorie und Ethik einer Form der Beeinflussung* (published in late 2017). Some thoughts from it are published in English; see, for example, Fischer and Illies (2018) and Fischer (2022).
9. I take it that affective states always help to shape our thinking and thus consequently our decision-making process.
10. At least as long as they are not trying to imitate interpersonal communication (which they, as of yet, often don’t do very well). If that is the case, many of my theses should be transferable.
11. See again endnote 2.
12. Literature often offers even more insightful accounts, for example, on specific emotions than philosophical analysis or empirical psychological studies.

Our affectivity does not seem to be measurable to the full extent in a scientific sense or to be broken down into clear propositions. We can conceptualize the outlines of it and describe its main workings in cold, technical terms, but by this we will not grasp them to the full extent. Literature masterfully fills this gap and warmly plasticizes our affectivity and shows itself as a vehicle of non-propositional knowledge, an essential foundation for our practical and phenomenological perception of the world that can function as a completion of the propositional kind of knowledge we can find in science and most areas of philosophy (cf. Fischer 2018). See with a slightly different focus than here Olivia Sudjic's novel *Sympathy* (2017).

13. Of course, you can, for example, act against a law (which counts as a form of coercion) but to understand that as freedom is, at least in judicial systems where fairness is fundamental, understanding freedom in a distorted I-do-everything-as-I-think-is-best-way. Sure, coercion also leaves an option to just abide by the law but not to decide out of unencumbered alternatives.
14. Even though our affectivity seems to have a bad reputation, we can find accounts that attest our affectivity a share of an integrated rationality. Aristotle famously thought that our affects are rational when they show themselves at the right time, for the right amount, and the right reason, for example, grief when a loved one died. Robert C. Solomon, as an example for a modern-day emotion theorist, also conceptualizes our affectivity as a vital part of rationality as it provides meaningful judgments, discerns value, is trainable, functions as an engine of our actions, is strategic, and creates meaning (Solomon 2001). In consequence, this also means that our affectivity does not in every case and necessarily undermine our autonomy.
15. I will look at active, intentional attempts of manipulation which does not mean that manipulation does not happen unconsciously or by just careless individuals. Nonetheless, many attempts of manipulation are thoroughly planned and it is not enough to assume that every manipulating agent is just careless. It is something else that seems to be at the core of the phenomenon: trying to use affectivity and our peripheral routes of decision-making to bring about a certain action.
16. These three types of ends are, of course, in many cases mixed with one another. Often, too, an agent is not fully aware of her (unacknowledged) ends.
17. Suggestions to not speak of "manipulation" anymore open up the question if it is possible to meaningfully speak about the phenomenon without the term "manipulation". Listeners just would not know anymore what we are talking about. There seems to be no everyday language term that marks a positive emotional influence anyhow. "Emotional influence" itself may be a neutral candidate which is less pejoratively connotated than "manipulation" (a connotation that can be closely linked to neo-Marxist thinking and the grim times of national socialism in Germany and capitalism worldwide). But the negative connotation seems to have more to do with a critique of challenges to our rationality than with the actual literal sense of the word "manipulation" (which is nonetheless understandable against the background of Nazi propaganda and the rise of capitalism). Since we are all finite, boundedly rational beings, it seems that we must admit that manipulation can count as a normal mode of communication. I am voting for keeping the term "manipulation" instead of erasing it because it marks an influence that goes beyond our rational radar and can still be connected to the everyday use of it, even if I try to carve out its characteristics, the when and how a bit more (cf. Fischer 2022).
18. In 2020, Sven Feurer of the Bern University of Applied Sciences (Switzerland) and I conducted an empirical study with a representative sample (in regard to

gender, age, and education) of 1000 German consumers to research the perception of marketing as manipulative. We asked the sample what they perceive as manipulation and ethically problematic with regard to new marketing measures that heavily rely on the internet and social media platforms like influencer marketing, fake reviews, targeted ads, and so on. The study confirmed that manipulation is not necessarily associated with deception but stronger with an attempt to affectively involve consumers to buy a product. Whereas deception was seen as very morally problematic, manipulation was perceived with a general skepticism but not necessarily seen as ethically problematic in *every* case. Targeted ads were seen as much less problematic than sentiment analysis or fake reviews, for example. For more, see Feurer and Fischer (2022).

19. This association is not necessarily wrong due to the usually strong emotionalization of fake news which plays a vital role besides the objectively wrong content. Manipulation, in the end, always also influences our thinking even when it primarily tries to modulate our affective states.
20. The empirical study Sven Feurer and I conducted shows that most people perceive marketing strategies usually as manipulative as they are aiming at our affectivity instead of presenting a product in a non-affective kind of way (Feurer and Fischer 2022).
21. Moral concerns arise, since an unencumbered rationality is seen as a necessary condition for autonomy. Manipulation, however, does not “*sufficiently* engage or appeal to [agents’] capacities for reflective and deliberative choice” (Sunstein 2016, 443; my highlighting) or even “*perverts* the way that [a] person reaches decisions, forms preferences or adapts goals” (Raz 1986, 377, my highlighting). Not being able to dive deeper into this particular discussion, I want to put forward that a manipulatively induced behavior does not automatically yield a degradation of an agent to an object as it is claimed in a Kantian tradition (e.g., Wood 2014). See for the link of our affectivity and our autonomy again endnote 14.
22. This is almost not tangible but still works with an affective state: that of not wanting to invest anything but instead staying comfortable. This points us to a certain problem in conceptualizing manipulation: some attempts at influencing an agent are, due to the use of many different factors, not 100% or 0% manipulative, but of a more or less manipulative quality and certainly tangible to almost not tangible. So there could be rational devices like arguments involved in manipulatively influencing an agent while certain forms of contextualization (e.g., an incident like the US election 2020), framing (e.g., as a “fraudulent election” of “corrupt individuals”), or presentation (e.g., by supposedly trustworthy authorities, with effective images etc.) add an affective and maybe manipulative character to what tries to reach an agent by argument cursorily (even though it might be objectively wrong). So, it is not always easy to distinguish how we influence each other between the mentioned poles in the analogue but also digital realm as it is not easy to determine what exactly makes an agent think, act, or decide in this or that direction. After all, this is not only based on a more or less rational *and* affective basis but always shaped by concrete situations, specific contexts, habits, and individual character traits. Burrhus F. Skinner’s operant condition might perhaps be the simplest example of manipulation through its use of rewards that are pleasurable and motivate us to do something again and punishments that make us avoid repeating certain behavior. But there are many more ways of using our affective states to influence us.
23. It becomes clear that all three ways of characterizing manipulation tend to blur the boundaries between descriptively and normatively defining said phenomenon by usually seeing manipulation as a negative type of influence. This

coincides with the everyday use of the word as an encumbered term which usually is intended to highlight that an outrageous type of influence has been used. Regarding the case of manipulation it makes sense to separate the question of how it works from the question of whether it is ethical or unethical (cf. Wood 2014, 19; Coons and Weber 2014a, 6–8). First, this can be explained by the fact that the term was once used neutrally and received its negative connotation only in the course of the twentieth century – which may not come as a surprise if one considers, for example, the horrors of Nazi propaganda (cf. Fischer 2017). However, this may at the same time stimulate us to look carefully and try to understand what exactly happens in the context of manipulation instead of leaving it blurry by immediately rejecting it as something evil. Stripped from this tinted looking glass, it becomes clearer that manipulation is constantly present in our social life and might even be qualified as a rather normal mode of communication between individuals that is not just malicious (although it can be) because, after all, nobody communicates purely rationally throughout. Additionally, a normatively loaded definition from the outset threatens to block a differentiated ethical debate because it supposedly seems clear from the get-go that manipulation is devilish. If we turn to the history of rationality we can quickly learn how rationality became the sun that supposedly helped grow the bulk of the grass of our humanity and that, especially by the discipline of philosophy, became an even dazzling light that might perhaps have blinded us for being at peace and with trust in regard to our affectivity. For more thoughts on our attitude regarding manipulation, see Fischer 2022.

24. Defining manipulation this way finds precursors in the concepts of Baron (2014, 109), that manipulation plays upon emotions, uses pressure to acquiescence (which is not yet coercion) or weaknesses of character, as well as in Noggle's (1996) and Barnhill's (2014) examples using guilt or Marcuse's observation that manipulation works via systematically inducing libidinal needs (Marcuse 1969, 31).
25. The relationship of our cognitive and affective states in the case of emotions is often complicated. Just think of jealousy where a strong feeling component contaminates our thoughts drastically, even creating tunnel vision, while we feel bad, evaluate harshly, and are motivated to act in an often destructive manner. It sure would be interesting to fan out a phenomenology of other difficult affects like anger, indignation, fear, and so on. But this is not the place for that.
26. Recommender systems, for example, can merge a) and b); for more detail, see Klenk, in this volume.
27. This is applicable especially for cases of, e.g., strategic political communication and marketing, whereas this is not necessarily true for every case of interpersonal communication where body language, facial expression, pitch of voice, and various other factors of nonverbal communication can intensify an attempt to manipulate more effectively than online.
28. In Sven Feurer's and my empirical study, we found that consumers themselves believe that manipulation in marketing has become much more subtle over the last two decades.
29. Sure, excitement can also be gained by other affective states than a feeling of indignation. I have also mentioned the acute emotions fear and anger or the persistent mood of mistrust in regard to the #StopTheSteal campaign. However, we can also be, for example, humorously affected or enthused in a positive manner and be guided by this through the affective online realm. See again endnote 7.
30. On December 10, 2020, Trump tweeted: "78% of the people feel (know!) the Election was RIGGED". With this tweet we get a small affidavit of means as it shows the (often) undifferentiated and in consequence dangerous equalization of feeling and knowing something is true. For Trump, it is convenient to stylize

feelings to equal truth as his only goal is to stir up the affectivity of his recipients without providing proper evidence that can actually be verified or falsified and thus *known* in a rational sense.

31. The suspension of Trump's social media accounts happened extremely late. After the U.S. Capitol was insurrected, Twitter and Facebook were able to argue that violence is actively incited (which breaches their rules – not an actual (and maybe needed) law, which regulates what is allowed to be done online). Before this specific, huge outbreak of violence both social networks held back, pointing to Trump's status as the president of the United States and thus a person of interest, when, of course, inciting violence played a big role for all of the years of Trump's presidency just on a different scale.
32. This often is because of a crooked understanding of freedom of expression. Trump tried to depict himself as a victim of censorship after his accounts were closed down. But objectively considered he was not at all a victim. He had a press room in the White House where he could address the nation and answer questions. Consequently, he was neither censored nor was his freedom of expression destroyed.
33. Because of its growing importance for the future the phenomenon called "microtargeting" should be kept in mind. Here, psychographic profiles are built to let content creators decide which advertisement or campaign design can make the biggest impression to which group of individuals – something that is heavily used by campaign strategists to efficiently grab voters by their affectivity. Interestingly enough the empirical study Sven Feurer and I conducted with regard to e-commerce marketing strategies and their manipulateness showed that targeted ads are perceived as manipulative in general but not too morally problematic (in comparison to influencer advertising, sentiment analysis or fake reviews). Of course, this evaluation concerns product advertisement and not politically used targeted advertising where one can expect a different answer (cf. Feurer and Fischer 2022).
34. Cf. Kramer, Guillory, and Hancock 2014 for an empirical work that has investigated affective influences on Facebook via the modification of hundreds of thousands of news feeds, without the knowledge of the users. The authors come to the (admittedly quite general) conclusion: "Online messages influence our experience of emotions, which may affect a variety of offline behaviors" (Kramer, Guillory, and Hancock 2014, 8788). Kramer himself is a data analyst at Facebook. This gives the work an interesting and a bit dazzling component: it is not only that users were experimented with without their consent. But Facebook also has an interest to show its advertising customers that manipulation works on the platform, which is why a scientific output can help to substantiate this claim (at the same time, users must be told that they are not easily manipulated on the platform). The cautious wording cited might result from ultimately not excessively strong experimental effects observed by the authors. However, they rightly point out that against the background of the size of social networks even small effects are based on a large number of people. It is also interesting to know that the paper had to take a lot of criticism. For an overview of this criticism see Grohol 2018.

6 References

- Abramowitz, Michael J. 2017. "Opinion: Stop the Manipulation of Democracy Online." *The New York Times*, November 12. Accessed August 20, 2021. www.nytimes.com/2017/12/11/opinion/fake-news-russia-kenya.html.

- Atlantic Council's DRFLab. 2021. "#StopTheSteal: Timeline of Social Media and Extremist Activities Leading to 1/6 Insurrection." Accessed August 20, 2021. www.justsecurity.org/74622/stopthesteal-timeline-of-social-media-and-extremist-activities-leading-to-1-6-insurrection/.
- Barnhill, Anne. 2014. "What is Manipulation?" In Coons and Weber 2014, 51–72.
- Baron, Marcia. 2003. "Manipulativeness." *Proceedings and Addresses of the American Philosophical Association* 77 (2): 37. doi:10.2307/3219740.
- Baron, Marcia. 2014. "The Mens Rea and Moral Status of Manipulation." In Coons and Weber 2014, 98–109.
- Beauchamp, Tom L. 1984. "Manipulative Advertising." *Business and Professional Ethics Journal* 3 (3/4): 1–30.
- Ben Ze'ev, Aaron. 2001. *The Subtlety of Emotions*. Cambridge, MA: Bradford Books.
- Ben Ze'ev, Aaron. 2018. "The Thing Called Emotion: A Subtle Perspective." In *Philosophy of Emotion*, edited by Aaron Ben Ze'ev and Angelika Krebs, 112–37. London: Routledge.
- Cadwalladr, Carole, and Emma Graham-Harrison. 2018. "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach." *The Guardian*, March 17. Accessed August 20, 2021. www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.
- Cave, Eric M. 2007. "What's Wrong with Motive Manipulation?" *Ethical Theory and Moral Practice* 10 (2): 129–44.
- Coons, Christian, and Michael Weber. 2014a. "Manipulation: Investigating the Core Concept and its Moral Status." In Coons and Weber 2014, 1–16.
- Coons, Christian, and Michael Weber, eds. 2014b. *Manipulation: Theory and Practice*. Oxford: Oxford University Press.
- Döveling, Katrin. 2015. "Emotion Regulation in Bereavement: Searching and Finding Emotional Support in Social Net Platforms." *New Review of Hypermedia and Multimedia* 21 (1–2): 106–22.
- Eyal, Nir, and Ryan Hoover. 2013. *Hooked: How to Build Habit-Forming Products*. London: Penguin Books.
- Feurer, Sven, and Alexander Fischer. 2022. *Exploring the Ethical Limits of Manipulation in Marketing: A Discussion Based on Consumer Perceptions*. under review.
- Fischer, Alexander. 2017. *Manipulation: Zur Theorie und Ethik einer Form der Beeinflussung*. Berlin: Suhrkamp.
- Fischer, Alexander. 2018. "Die Moral durch die Geschicht': Erzählen als vernünftige Integration von Verstand und Affekten im moralischen Handeln." *Internationale Zeitschrift für Philosophie und Psychosomatik* 10 (1): 1–15.
- Fischer, Alexander. 2020. "Im Schraubstock der Angst: Manipulation und unsere Disposition zur Ängstlichkeit." *Hermeneutische Blätter* 26 (1): 20–37.
- Fischer, Alexander. 2022. "Then Again, What is Manipulation? A Broader View of a Much-Maligned Concept." *Philosophical Explorations*. doi: 10.1080/13869795.2022.2042586
- Fischer, Alexander, and Christian Illies. 2018. "Modulated Feelings: The Pleasurable-Ends-Model of Manipulation." *Philosophical Inquiries* 1 (2): 25–44. Accessed August 06, 2020.
- Fischer, John M. 2004. "Responsibility and Manipulation." *Journal of Ethics* 8 (2): 145–77.

- Gellner, Ernest. 1992. *Reason and Culture: The Historic Role of Rationality and Rationalism*. Oxford: Oxford University Press.
- Goodin, Robert E. 1980. *Manipulatory Politics*. New Haven, CT: Yale University Press.
- Gorin, Moti. 2014a. "Do Manipulators Always Threaten Rationality?" *American Philosophical Quarterly* 51 (1): 51–61. Accessed June 04, 2019.
- Gorin, Moti. 2014b. "Towards a Theory of Interpersonal Manipulation." In Coons and Weber 2014, 73–97.
- Green, Ronald K., and Edward J. Pawlak. 1983. "Ethics and Manipulation in Organizations." *Social Service Review* 57 (1): 35–43.
- Grohol, John M. 2018. "How Facebook's Squishy Ethics Got Them into Trouble." *PsychCentral*, July 8. <https://psychcentral.com/blog/how-facebooks-squishy-ethics-got-them-into-trouble>.
- Heath, Joseph. 2015. *Enlightenment 2.0: Restoring Sanity to Our Politics, Our Economy, and Our Lives*. Toronto: Harper.
- Klenk, Michael. 2022 "Manipulation, Injustice, and Technology." In *The Philosophy of Online Manipulation*, edited by Jongepier, F. and Klenk, M., 108–132. New York, NY: Routledge.
- Klenk, Michael. 2021. "Manipulation (Online): Sometimes Hidden, Always Careless." *Review of Social Economy*. 80: 1, 85–105. doi:10.1080/00346764.2021.1894350.
- Kramer, A. D. I., J. E. Guillory, and Jeff Hancock. 2014. "Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 111: 8788–90.
- Lippmann, Walter. 2008. *Public Opinion*. Miami: BN Publishing.
- Marcuse, Herbert. 1969. "Aggressivität in der gegenwärtigen Industriegesellschaft." In *Aggression und Anpassung in der Industriegesellschaft*, edited by Herbert Marcuse, 7–29. Frankfurt a.M.: Suhrkamp.
- Nelson, Lisa S. 2018. *Social Media and Morality: Losing Our Self Control*. Cambridge: Cambridge University Press.
- Noggle, Robert. 1996. "Manipulative Actions: A Conceptual and Moral Analysis." *American Philosophical Quarterly* 33 (1): 43–55.
- Noggle, Robert. 2018. "The Ethics of Manipulation." In *Stanford Encyclopedia of Philosophy: Summer 2018*, edited by Edward N. Zalta.
- Nussbaum, Martha. 2008. "Compassion: Human and Animal." <http://idsk.edu.in/wp-content/uploads/2015/07/SL-4.pd>. IDSK Special Lecture 4.
- Pohl, Rüdiger. 2004. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. New York, NY: Psychology Press.
- Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.
- Rudinow, Joel. 1978. "Manipulation." *Ethics* 88 (4): 338–47. doi:10.1086/292086.
- Sachs-Hombach, Klaus. 2003. *Das Bild als kommunikatives Medium: Elemente einer allgemeinen Bildwissenschaft*. Köln: Halem.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Solomon, Robert C. 2001. *True To Our Feelings: What Our Emotions Are Really Telling us*. Oxford: Oxford University Press.
- Spring, Marianne. 2020. "'Stop the Steal': The Deep Roots of Trump's 'Voter Fraud' Strategy." *BBC News*, November 23. www.bbc.com/news/blogs-trending-5500995.

- Sunstein, Cass R. 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.
- Susser, Daniel, Beate Roessler, and Helen Nissenbaum. 2019a. "Online Manipulation: Hidden Influences in A Digital World." *Georgetown Law Technology Review* 4 (1): 1–45. Accessed February 27, 2020.
- Susser, Daniel, Beate Roessler, and Helen Nissenbaum. 2019b. "Technology, Autonomy, and Manipulation." *Internet Policy Review* 8 (2): 1–22. doi:10.14763/2019.2.1410.
- Vallor, Shannon. 2010. "Social Networking Technology and the Virtues." *Ethics and Information Technology* 12: 157–70.
- van Dijk, Teun A. 1998. *Ideology: A Multidisciplinary Approach*. London: SAGE Publications.
- Verbeek, Peter-Paul. 2011. *Moralising Technology: Understanding and Designing the Morality of Things*. Chicago, IL: University of Chicago.
- Ware, Alan. 1981. "The Concept of Manipulation: Its Relation to Democracy and Power." *British Journal of Political Science* 11 (2): 163–81.
- Wood, Allen W. 2014. "Coercion, Manipulation, Exploitation." In Coons and Weber 2014, 17–50.