# Power and dignity: the ends of online behavioral advertising in the European Union

Zardiashvili, A.

# CHAPTER 3. MANIPULATION

This thesis evaluates the ability of the European Union (EU) legal framework to safeguard against consumer manipulation harms of online behavioral advertising (OBA). In order to explain how OBA leads to consumer manipulation harms, the thesis builds a coherent theory of manipulation. With this aim in mind, this chapter answers the second sub-question of the thesis:

| SQ2: what is manipulation? |
|---|

Section 3.1 describes influences on human behavior and delineates manipulation from other forms of influence. Section 3.2 defines the concept of vulnerability in the context of decision-making that can be exploited for manipulation. Section 3.3 applies the understanding of layered vulnerability to describe different levels of an influence being manipulative. Section 3.4 concludes by formulating an answer to SQ 2.

### 3.1. Influencing Human Behavior

This section identifies characteristics of manipulation that distinguish it from other forms of influence. Section 3.1.1 places manipulation in the context of influences on human behavior, section 3.1.2 defines forms of influence such as coercion and persuasion and delineates them from manipulation, and section 3.3.3 expands on the defining characteristics of manipulation as a form of influence.

#### 3.1.1. Influence

To *manipulate* something means to move it or to control it.[261] One can manipulate technical instruments, for example, a computer with a keyboard or a car with a steering wheel.[262] One can also manipulate animals—for example, snakes can be manipulated to mimic dancing ("snake charming").[263] Similarly, one can manipulate human beings—they can be moved and controlled as if they were a computer or a snake.[264] This thesis talks of manipulation as a form of influence on human behavior. Manipulation can also be understood as a form of influence that radically re-conditions the target's behavior.[265] As an illustrative analogy, behavioral scientist B.F. Skinner successfully conditioned the behavior of pigeons to play a version of ping pong.[266] This thesis differentiates the application of such strategies on human beings from ordinary interpersonal forms of manipulation and explicitly refers to it as "global manipulation".[267]

In ordinary discussions, *manipulation* as a form of influencing human behavior is morally loaded and conveys a derogatory connotation. In interpersonal relationships, manipulators are said to influence someone's behavior through a "guilt trip" – making someone feel guilty, "peer pressure" – making someone fear social disapproval, "negging" – making someone feel bad about themselves, "emotional blackmail" – making someone fear the withdrawal of affection, or

[261] Manipulate, BRITANNICA DICTIONARY, https://www.britannica.com/dictionary/manipulate (last visited Jan 24, 2023).

[262] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 12.

[263] *See Snake Charming*, WIKIPEDIA (2023), https://en.wikipedia.org/wiki/Snake_charming (last visited Jan 31, 2023).

[264] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 12.

[265] *See* Robert Noggle, *The Ethics of Manipulation*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY 1.1 (Edward N. Zalta ed., Summer 2022), https://plato.stanford.edu/archives/sum2022/entries/ethics-manipulation/ (last visited Jan 25, 2023).

[266] *See* Marina Koren, *B.F. Skinner: The Man Who Taught Pigeons to Play Ping-Pong and Rats to Pull Levers*, SMITHSONIAN MAGAZINE (Mar. 20, 2013), https://www.smithsonianmag.com/science-nature/bf-skinner-the-man-who-taught-pigeons-to-play-ping-pong-and-rats-to-pull-levers-5363946/ (last visited Jun 28, 2023).

[267] Famous fictional depictions of global manipulation include: *See e.g.,* ALDOUS HUXLEY, BRAVE NEW WORLD (1932). *See e.g.,* A CLOCKWORK ORANGE (Warner Bros., 1971). *See e.g.,* THE MATRIX (Warner Bros., 1999).

"seduction" – making something seem (sexually) appealing.[268] In philosophical discussions, there is little agreement on what binds these forms of influences together — what are the necessary and sufficient conditions for a practice to be identified as manipulation (i.e., identification question), and what makes manipulation wrong (i.e., evaluation question).[269]

Consequently, legal and policy discussions are contaminated by the variety of subjective moral standpoints one can adopt about manipulation, making it challenging to define malicious practices, identify their harms, assign responsibility, and tailor regulatory intervention.[270] This thesis aims to provide a coherent framework for understanding manipulation that can help evaluate the extent to which OBA may lead to this outcome.[271] The harms of manipulation, and therefore, the extent to which it requires regulatory intervention, are addressed separately in Chapter 5. Aiming to capture the concept of manipulation in a way that makes it useful in policy discussions, this chapter steps away from normative evaluations as much as possible and approaches the concept from a purely analytic point of view, attempting to describe it as a particular type of influence.[272]

### 3.1.2. Persuasion, Coercion, and Manipulation

As social creatures, humans depend on each other for almost everything they need, and to get those needs met, they influence each other in various ways.[273] In this sense, influence on human behavior can be understood in two dimensions: by observing what is being modified (*change*)[274] and by observing the effect of the modification on the target (*effect*).[275] Figure 3:1 illustrates the intersections of these dimensions in a quadrant (*quadrant of influence*). Firstly, in order to influence the target, an agent may change *(i)* the target's understanding of options (*perception*) or

[268] *See* Noggle, *supra* note 265.

[269] *See Id.* at 1.3.

[270] *See e.g.,* European Commission Study Dark Patterns & Manipulative Personalization, *supra* note 53 at 40.

[271] *See generally* MANIPULATION: THEORY AND PRACTICE, *supra* note 74. *See* CASS R. SUNSTEIN, THE ETHICS OF INFLUENCE (2016). *See* Robert Noggle, *Pressure, Trickery, and a Unified Account of Manipulation*, 3 AM. PHILOS. Q. 241 (2020). *See* Noggle, *supra* note 265. THE PHILOSOPHY OF ONLINE MANIPULATION, *supra* note 74. *See* Susser, Roessler, and Nissenbaum, *supra* note 38 at 17.

[272] *See also* Allen W. Wood, *Coercion, Manipulation, Exploitation*, *in* MANIPULATION: THEORY AND PRACTICE 18–21 (Christian Coons & Michael Weber eds., 2014).

[273] *See Id.* at 17. *See also* Christian Coons & Michael Weber, *Introduction: Manipulation: Investigating the Core Concept and Its Moral Status*, *in* MANIPULATION: THEORY AND PRACTICE , 1 (Christian Coons & Michael Weber eds., 2014). *See also* PLATO, REPUBLIC 59 (2008).

[274] Words formatted in *Italics* inside the parenthesis refer to how the concepts appear in Figure 3:1

[275] This view is based on dichotomy proposed by Susser, Roessler, and Nissenbaum. *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 14. In this thesis, "options" relate to "decision-space" and "perception" to "decision-making process".

*(ii)* the target's options (*options*).[276] Second, the effect of the change may be that the target of the influence has *(a)* acceptable alternative options (*choice*) or *(b)* no acceptable alternative options or no ability to exercise choice between them (*no choice*).[277] This thesis uses the model illustrated by Figure 3:1, to delineate between different forms of influences, in particular persuasion with reason (quadrant [i][a]), persuasion with incentives (quadrant [ii][a]), coercion (quadrant [ii][b]), and manipulation (quadrant [i][b]). This chapter explains each of these forms of influences and provides illustrative (but non-exhaustive list of) examples (*examples 1 –12*) that are also placed in Figure 3:1 to illustrate differences.

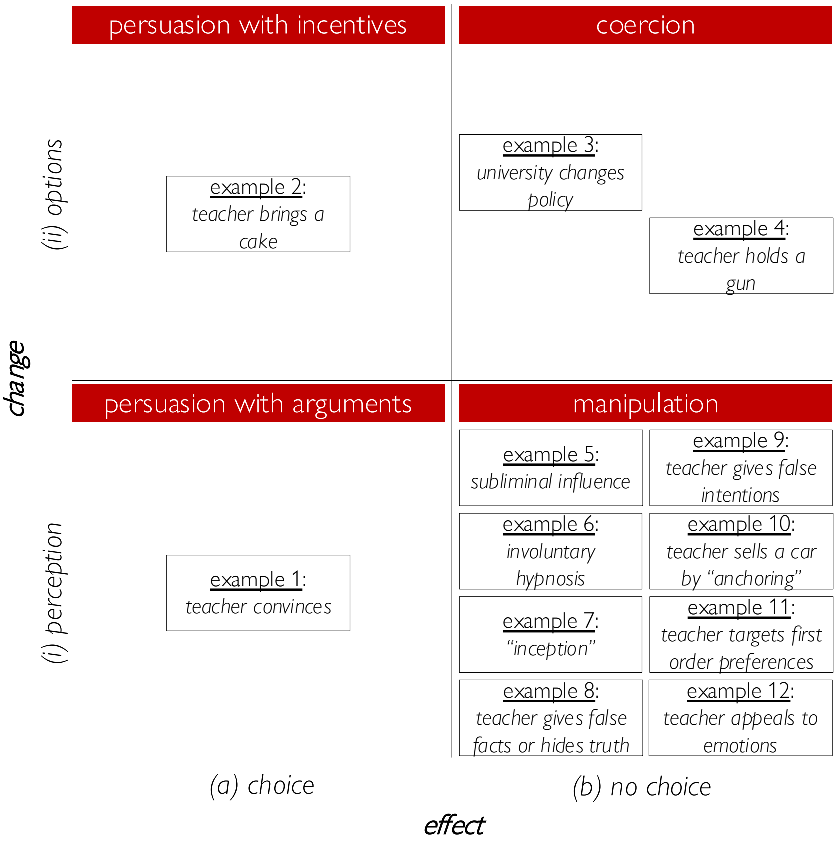

persuasion with incentives
coercion
(ii) options
example 2: teacher brings a cake
example 3: university changes policy
example 4: teacher holds a gun
change
persuasion with arguments
manipulation
(i) perception
example 1: teacher convinces
example 5: subliminal influence
example 6: involuntary hypnosis
example 7: "inception"
example 8: teacher gives false facts or hides truth
example 9: teacher gives false intentions
example 10: teacher sells a car by "anchoring"
example 11: teacher targets first order preferences
example 12: teacher appeals to emotions
(a) choice
(b) no choice
effect

*Figure 3:1. Quadrant of influence with examples (by author)*[278]

Persuasion is defined as an attempt to influence targets by giving them reasons they can evaluate through conscious deliberation.[279] One can provide these reasons

[276] *See Id.* at 14-17.

[277] *See Id.* at 14-17.

[278] The figure is the author's representation of a theory of influences developed by Susser, Roesler, and Nissenbaum. *See generally Id.*

through rational argumentation (i.e., rhetoric) or through incentives.[280] In the first case, this amounts to an attempt to change the target's perception of options, and in the latter, changing the target's options. Nevertheless, persuasion is a form of influence that openly appeals to the target's capacity for conscious deliberation and leaves the target with acceptable alternative options.[281]

In example 1,[282] a university teacher wants students who have not been exposed to the *Covid-19* virus to attend class in person during a pandemic. The teacher explains why students should come to class by providing arguments. The teacher argues that in-class sessions build better rapport enable a more natural flow of interaction and that as the university facilitates hybrid education and high standards of health safety requirements, students should come to class, although they can join the class online. In this case, the teacher persuades the students with arguments – attempting to change their understanding of the options without changing available options (quadrant [i][a]). In example 2, on top of the arguments, the teacher also announces that an after-class chocolate cake will be provided for the students attending the session in class. In this case, the teacher persuades by offering a chocolate cake as an incentive – attempting to reconfigure students' options but leaving acceptable alternatives. While chocolate cake can be an attractive incentive, it cannot be regarded as irresistible (quadrant [ii][a]).[283]

Providing irresistible incentives can be considered a form of *coercion* – as it reconfigures the target's options, so that there are no acceptable alternatives (quadrant [ii][b]).[284] If persuasion equates to "making an offer", coercion can be understood as "making an offer that one cannot refuse".[285] In example 3, the teacher announces the new university policy and requests all students who have not been exposed to the virus to come to class in person, stating that all students attending the session online without medical evidence of exposure will be marked as "absent". In this case, the teacher prompts students by taking away an acceptable alternative, that

[279] Generally, persuasion can be understood in two ways: in the broader sense, the term persuasion is an umbrella term for all forms of influence, from rhetoric to violent coercion. This thesis uses the term in its narrow sense, meaning "changing someone's mind by giving reasons he or she can reflect and evaluate." *Id.* at 13–17.

[280] *Id.* at 14.

[281] *Id.*

[282] Underlined examples are placed in Figure 3:1.

[283] Rudinow formulates persuasion by providing "resistible incentives". *See* Joel Rudinow, *Manipulation*, 88 ETHICS 338, 342 (1978). *See also* Susser, Roessler, and Nissenbaum, *supra* note 38 at 15.

[284] Susser, Roessler, and Nissenbaum, *supra* note 38 at 15.

[285] Making "an offer someone cannot refuse" is a catchphrase of Vito Corleone, a fictional mafia don in the movie "The Godfather". *See* THE GODFATHER (Paramount Pictures, 1972). *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 14.

is, joining the class online without repercussions. Therefore, coercion can be defined as an overt influence that leaves the target with no acceptable alternatives.[286]

Similar to manipulation, this thesis refers to coercion in its non-moral sense. The moral and legal validity of coercion is context-dependent.[287] For instance, educational institutions are usually authorized to force students to attend class, and within the entire EU, police officers have the authority to use violence to coerce people into specific behavior, including putting them in jail. Coercion contrasts with persuasion through incentives in that it takes away acceptable alternatives. While a coerced person can physically choose an alternative, not attend the class, or run away from the police, these alternatives cannot be regarded as acceptable due to some extent to the likelihood of failure and the high severity of their consequences, such as being delisted by the education institution or being shot by police officers.[288]

Coercion and persuasion with incentives are similar in that they reconfigure the options available for the target and appeal to the target's capacity to deliberate consciously on these reconfigured options.[289] In other words, persuaders and coercers make their influence explicit and overt to encourage their targets to deliberate on their best interests. In example 4, a university teacher holds a student at gunpoint and demands that they come to class. In this case, a teacher with a gun wants the student to be consciously aware of two options: attending a class or giving away their life. While a coercer is one who modifies options in a way that a choice between alternatives is irresistible, a target of coercion is ultimately one who evaluates the worth of their time in class against the worth of their life and consciously takes a decision accordingly.[290]

[286] See *Id.*; Wood, *supra* note 272 at 17.

[287] *See* Wood, *supra* note 272 at 34.

[288] What is acceptable depends on the moral standpoint of an individual. For example, for Plato death is more acceptable than losing freedom. *See* PLATO, *supra* note 273, at 81. Likewise, stoic philosophers have long argued that death is an acceptable option and that "choice" can never be lost. *See e.g.,* EPICTETUS, DISCOURSES, FRAGMENTS, HANDBOOK 81 (2014). Such understandings of human choice are aspirational, and are not found in legal framework. *See* about "irresistible incentives" Rudinow, *supra* note 283, at 341.

[289] *See* Susser, Roessler, and Nissenbaum, *supra* note 38 at 15.

[290] *Id.* at 16. Some philosophers argue that coercion not only makes the target aware of the influence and the options, but that coercer wants the coerced to be *rational. See e.g.,* Coons and Weber, *supra* note 273, at 15. This argument can be defended only to some extent. Holding someone at gunpoint, in most instances, primarily appeals to the target's emotions rather than rationality. Purely rational analysis can allow a target to calculate the likelihood of other acceptable options (e.g., it is easy to overpower the coercer, or there is a possibility to run away). Nevertheless, a target typically makes a decision based on their fear of which the target is acutely aware. Therefore, this thesis does not argue that coercion appeals to rationality per se but conscious awareness. Similarly, this thesis argues that manipulation subverts conscious deliberation (not rationality), a self-aware decision-making process. *See also* Susser, Roessler, and Nissenbaum, *supra* note 38, at 16.

In contrast to persuasion and coercion, manipulation is a form of influence that "subverts" the target's capacity to deliberate on available options consciously (quadrant [i][b]).[291] In other words, manipulation displaces a target of influence as an agent who makes a conscious decision.[292] In manipulation, a target of influence acts like a puppet whose strings are being pulled by someone else.[293] A university teacher can be said to manipulate students when a teacher places "subliminal"[294] messages in the presentation that successfully influence students to come to class (example 5). The same would be true if a teacher induces an involuntary hypnotic state (example 6) to convince students or relies on sophisticated technology to "incept" an idea of coming to class while they are asleep (example 7).

It rarely (if ever) happens that manipulative influence completely bypasses the target's conscious deliberation process.[295] Even in subliminal influence (example 5), involuntary hypnosis (example 6), or "inception" (example 7),[296] where stimulus bypasses conscious deliberation entirely, a target maintains some agency in their decision-making process.[297] Instead, during manipulation, to steer a target towards their end, an agent inserts themselves into the target's decision-making process in a way that stays hidden from the target.[298] By hiding the influence, manipulation alters the target's perception of available options. As the target cannot consider the

[291] *See* Susser, Roessler, and Nissenbaum, *supra* note 38 at 16. In particular, this thesis agrees that the salient issue is whether or not the influence appeals to target's conscious awareness, not rationality.

[292] *Id.* at 16.

[293] *See Id.* at 17. According to authors, when the manipulative influence is found out, a target feels "played" in contrast to coercion when a target feels "used".

[294] "Subliminal stimuli" refers to visual, audible or any other sensory stimuli below the threshold for conscious perception. While effects of subliminal stimuli on human behavior is disputed, some studies find that such stimuli can affect decision-making processes. *See* S. J. Brooks et al., *Exposure to Subliminal Arousing Stimuli Induces Robust Activation in the Amygdala, Hippocampus, Anterior Cingulate, Insular Cortex and Primary Visual Cortex: A Systematic Meta-Analysis of fMRI Studies*, 59 NEUROIMAGE 2962 (2012). ("[D]ata suggest that despite stimulus presentation being presented outside of conscious awareness, the brain remains able to respond to such stimuli, mainly in sub-cortical regions associated with bodily arousal, implicit memory, conflict monitoring and detection of unpredictability. Activation in these brain regions, using subliminal paradigms, provides robust evidence that specific arousal systems in the brain can be activated outside of conscious awareness.")

[295] *See* SUNSTEIN, *supra* note 271 at 82. *See also* Shaun B. Spencer, *The Problem of Online Manipulation*, 3 UNIV. ILL. L. REV. 960, 990 (2020). *See* Susser, Roessler, and Nissenbaum, *supra* note 38 at 17.

[296] *See* INCEPTION (Warner Bros., 2010). Thanks to Agata Szczepańska for suggesting this example.

[297] *See* SUNSTEIN, *supra* note 271 at 82. *See* Coons and Weber, *supra* note 273 at 6.

[298] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 17. This account of manipulation has been criticized because of ambiguity about exactly what remains hidden. *See* SUNSTEIN, *supra* note 271 at 102–107. Sunstein argues that the issue is that the influence does not "sufficiently" appeal to conscious deliberation. Sunstein focuses on "manipulative practices" as attempts of manipulation not on "manipulation" that is a successful outcome of manipulative practices addressed by Susser, Roessler, and Nissenbaum.

influence as part of evaluating the options, it takes away from their ability to exercise choice.[299]

### 3.1.3. Manipulation: Hidden, Successful, Intentional Influence

In summary, manipulation can be understood as a *hidden influence* on human behavior. The manipulator hides something important from the target.[300] While some forms of manipulation, for example, subliminal influence (example 5), may hide the manipulative stimulus itself, other forms may make the stimulus visible but hide the manipulator's role or intentions (example 8, example 9). In example 8, the university teacher refers to the new university policy to announce that all students attending the session online without medical evidence of exposure to the virus will be marked as "absent". However, unlike example 3, university policy has not been changed, yet the teacher hides that students are not obligated to attend class. This scenario is an example of *deception*, a specific type of manipulation. Students are aware of the stimulus: the teacher wants them to be in class. However, they are unaware that the teacher is giving them false information by referring to a non-existing policy and that students are, in fact, not required to attend the class in person. Therefore, the teacher's role and the mechanism of influence remain hidden from the students.

As soon as a target of influence becomes aware of a covert influence, influence becomes part of their decision-making.[301] For example, once students become aware that the teacher shared false information and that university policy allows them to voluntarily choose between in-class and online attendance, they face a different set of acceptable options. In example 9, when students confront the teacher, the teacher admits that they did not have the authority to demand in-class attendance but tells them that they resorted to announcing it anyway because coming to class is in the student's best interests, and this was the only way they could convince them. This time, the teacher does not disclose that the scarcity of students in the class makes it difficult for the teacher to concentrate. Therefore, the teacher hides some part of the intention or a reason for the influence, again classifying the influence in example 9 as deception and, thus, manipulation.

Manipulation is not limited to deception, including hiding stimulus (example 5), falsifying facts (example 8), or hiding intentions (example 9); it can also be a

[299] Even in retrospective analysis, when assessing how a person made a decision, it can be complicated to understand if the manipulator influenced them (and to what extent). The question arises: is it ever possible to be sure that the target would make a different decision without manipulative influence? Such certainty is not required for defining manipulation. On the contrary, uncertainty about forming a decision can be seen as an element of manipulation. People who "feel manipulated" can not fully understand why they acted the way they did or if they did so in their or someone else's best interests. *See also* Susser, Roessler, and Nissenbaum, *supra* note 38, at 17.

[300] *See* SUNSTEIN, *supra* note 271, at 102.

[301] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 20.

form of pressure, in which manipulator is hiding psychological mechanisms by which a stimulus steers the target's behavior (section 3.2).[302] That being said, manipulation is always *intentional*.[303] Exaggerated portrayals of manipulators include depicting them scheming with an evil smile or laughter.[304] Nevertheless, manipulation does not always involve conscious deliberation to hide some aspect of influence and exploit vulnerability.

In essence, manipulation can occur when a manipulator elects to influence a target and neglects to deliberate the means through which influence is achieved.[305] In other words, for influence to be classified as manipulation, the *intention* to influence must always be present, but the hiddenness of influence can be caused by negligence.[306] For example, negging involves an attempt to influence another person's behavior by making that person feel bad about themselves or the situation. In intimate, friend, and family relationships, people do not always deliberately mean to make others feel bad but might do so anyway and somewhat unconsciously to compel them to do something. Such dual nature of manipulation as deception and hidden pressure comes back throughout later sections of the thesis (e.g., section 3.2).

Manipulation is also a "success concept" – it reflects that the stimulus hiddenly and successfully influenced a target towards an outcome.[307] Manipulation itself is blind to the methods and strategies; instead, it suggests that intentional influence has taken place in a way that remained hidden from the target of this influence.[308] There are no degrees in manipulation: it has either occurred or not. In contrast, a practice can be *manipulative* if it is an attempt to manipulate, whether or not such an attempt is successful, and, thus, leads to manipulation.[309]

For example, in the previously mentioned example 9, if students had recognized that their teacher did not disclose their true intentions, there would be no case of manipulation: students would deliberate on the true intentions of the teacher (*i.e.,* that he could not concentrate without students attending class in person) and decide whether they wanted to join in person or not. Nevertheless, the teacher's

[302] *See* Spencer, *supra* note 295, at 990.

[303] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 26.

[304] *See e.g.,* Heath Ledger's depiction of "Jocker" in Christopher Nolan's "The Dark Knight". *See* THE DARK KNIGHT (Warner Bros., 2008).

[305] The account of manipulation as "careless influence" was first developed by Klenk. *See* Michael Klenk, *(Online) Manipulation: Sometimes Hidden, Always Careless*, 80 REV. SOC. ECON. 85, 13 (2022). Klenk argues that an action is manipulative if "a) M[anipulator] aims for S[ubject] to do think, or feel b through some method m and b) M disregards whether m reveals eventually existing reasons for S to do, think or feel b to S"). *See also* Noggle, *supra* note 265. Klenk explicitly states to disagree with the view that manipulation is hidden.

[306] *See Id.*

[307] *See* Note 298.

[308] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 27. *See* Wood, *supra* note 272, at 11.

[309] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 27.

actions would still be manipulative because an aspect of the influence intentionally hidden can still be regarded as manipulative, as he attempted to manipulate the students but failed to do so.

Therefore, this thesis defines manipulation as:

> *Agent's successful and intentional attempt to influence the target's behavior where an essential aspect of an influence remains hidden from the target and an agent is aware that the method of influence is likely to exploit the target's decision-making vulnerabilities*.[310]

In summary, when the target acts towards the agent's desired outcome, the agent has manipulated the target if:

1) the agent intended to influence a target towards an outcome;
2) an essential aspect of an influence remains hidden from the target and
3) the agent is aware that the method of influence is likely to exploit the target's decision-making vulnerabilities.

This thesis aims to define manipulation in a helpful way for policymakers and enforcers. As policy may entail preventing manipulation from occurring, it is essential to evaluate not only the situations that can be evaluated as successful manipulation but also manipulative practices that may remain unsuccessful. With this in mind, Section 3.2 further elaborates on methods of manipulation, and Section 3.3 formulates the way to measure the "manipulativeness" of an agent's attempts to influence a target. While essential aspects of influence are context-dependent, section 4.1.1 elaborates on some of those aspects in the context of advertising.

### 3.2. Methods: Exploitation of Vulnerability

There are various ways one can conceptualize how a manipulator manages to exert hidden influence on their target. This thesis adopts the framing closely aligned with Susser, Roesler, and Nissenbaum's view that the various means of manipulation, such as deception or pressure, can be summarized as methods that exploit human decision-making *vulnerabilities.*[311] This section defines and describes

[310] Manipulation as *hidden influence* is one of at least three ways manipulation can be understood. Other two ways include manipulation as *trickery* (deception) and manipulation as *pressure*. *See* Noggle, *supra* note 265.

[311] This thesis also refers to "exploitation" in non-moralized sense to describe that a vulnerability is used by an agent of influence towards agent's pre-determined end. *See* Wood, *supra* note 272, at 43. Deception essentially exploits vulnerability of "unavailability of perfect information" in human decision-making context, and pressure, generally, exploits the need for social approval. Susser, Roessler, and Nissenbaum combine all means of manipulation under the umbrella of "cognitive,

what these vulnerabilities are and how their exploitation leads to manipulation. Section 3.2.1 addresses cognitive biases, and section 3.2.2 other vulnerabilities that manipulators exploit.

3.2.1. Cognitive Biases

One way to understand the conscious deliberation process through which humans make decisions is by the interplay of a person's beliefs, preferences, and emotions that precede their actions.[312] Ideally, a decider would hold *beliefs* that truthfully reflect circumstances; they would form *preferences* that accurately reflect these beliefs and experience *emotions* that help them gauge their proximity to their preferences.[313] As people have many beliefs, desires, and emotions, conscious deliberation is a process through which one makes up one's mind or adapts beliefs, prioritizes desires, and interprets emotions.[314] *Rationality* – a state of being governed by reason – is one form of conscious deliberation that allows a decision-maker to advance toward their self-interest by always choosing the best available option.[315]

Rationality has often been an aspirational state for human beings.[316] Historically, rationality has also been ascribed to humans as their descriptive characteristic: some economic theories and legal frameworks are constructed around a view of human beings as rational beings.[317] Nevertheless, almost a century of cognitive, behavioral, and social psychology studies reveal that human beings rarely,

emotional, or other decision-making vulnerabilities". *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 26.

[312] *See* Robert Noggle, *Manipulative Actions: A Conceptual and Moral Analysis*, 33 AM. PHIL. Q. 43, 4 (1996).

[313] *See Id.* For a simplified example to illustrate interplay of beliefs, preferences, and emotions: imagine that one believes their purpose in life is to create value for the community through their work (belief #1); they may desire to get to work constantly on time (preference #1). They feel guilty (emotion #1) when they are late. They believe that a bicycle is a faster means of transportation than being on foot (belief #2), and they feel excited (emotion #2) when considering buying a bicycle (preference #2).

[314] *See Id.* at 44–47. *See generally* Susser, Roessler, and Nissenbaum, *supra* note 38.

[315] R. Jay Wallace, *Practical Reason*, THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., Spring 2020 Edition), https://plato.stanford.edu/archives/spr2020/entries/practical-reason/ (last visited Feb 2, 2023).

[316] *Id.* at 6.

[317] In law and economics, human beings can be imagined as economic agents who are consistently rational, and optimize for their self-interest (often referred to as "homo economicus" or "economic man"). Such views were promoted by early economic theorists, such as John Stuart Mill and Adam Smith. *See e.g.,* JOHN STUART MILL, ESSAYS ON SOME UNSETTLED QUESTIONS OF POLITICAL ECONOMY (2011). *See e.g.,* ADAM SMITH, THE WEALTH OF NATIONS (Robert B. Reich ed., 2000). The EU legal framework sometimes resembles such a view of humans as rational agents. For example, when referring to "average consumer" consumer protection legislation considers a consumer that is "reasonably well informed, observant, and circumspect". *See* European Commission Study Dark Patterns & Manipulative Personalization, *supra* note 53, at 90.

if ever, behave entirely rationally.[318] These studies further conclude that most everyday human decision-making does not even happen consciously and deliberately.[319] Instead, they suggest that for evolutionary purposes, the human brain developed mechanisms that they call *heuristics* and *automated behavior patterns* – to shortcut the decision-making process, reduce complexity, and save energy in the face of repetitive and unimportant tasks.[320]

Cognitive psychologists refer to the conscious decision-making process as *System 2* and describe it as a *slow,* reflective, effortful, controlled way of thinking that requires time, energy, and attention (hereafter, *slow* thinking).[321] In contrast, they explain, humans make most of their decisions using the thinking paradigm they call *System 1,* which is *fast,* non-reflective, automatic, simple, and requires much less time, energy, and attention (hereafter, *fast* thinking).[322] Studies reveal that humans only mobilize slow thinking when fast thinking cannot handle the task at hand.[323] Even then, System 1 continues to generate cues that a person receives in the form of impressions, intuitions, and feelings that they consider during their slow thinking process.[324] Therefore, in many situations, these fast-thinking shortcuts are prone to errors in the decision-making process called *cognitive biases* that may lead to sub-optimal decisions.[325]

The "anchoring effect" is one such cognitive bias that distorts a person's estimates by causing them to rely on a pre-existing piece of information, such as a number (an anchor) when making a decision.[326] For example, presenting original/discounted prices can influence a viewer's price perceptions. The "availability heuristic" influences a person to provide further weight to a specific scenario or an occurrence already available in a person's memory compared to other scenarios of objectively similar weight.[327] Through the "framing effect," people

[318] Three influential works analyzing the decision-making shortcuts are: *See* DANIEL KAHNEMAN, THINKING FAST AND SLOW (2011). *See* ROBERT B. AUTHOR CIALDINI, INFLUENCE: THE PSYCHOLOGY OF PERSUASION (Revised edition.; First Collins business essentials edition. ed. 2007); *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: THE FINAL EDITION (Updated edition. ed. 2021).

[319] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 21.

[320] *See* for "heuristics" Amos Tversky & Daniel Kahneman, Judgment under Uncertainty: Heuristics and Biases, 185 SCIENCE 1124 (1974). *See* for "automated behavior patterns" CIALDINI, *supra* note 318.

[321] *See* KAHNEMAN, *supra* note 318, at 21. Thaler and Sunstein refer to System 1 as the "Automatic System" and "Gut", and to System 2 as the "Reflective System" and "Conscious Though". *See* THALER AND SUNSTEIN, *supra* note 318, at 19.

[322] *See* KAHNEMAN, *supra* note 318, at 25.

[323] *Id.* at 24. Spencer, *supra* note 295, at 964.

[324] *See* KAHNEMAN, *supra* note 318, at 24.

[325] *Id.* at 25.

[326] *Id.* at 119. *See* Spencer, *supra* note 295 at 964.

[327] *See* Tversky and Kahneman, *supra* note 320, at 1128. *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 22.

draw different conclusions and sometimes make contrasting decisions based on identical information framed differently.[328] Also, because of the "social proof principle," people view specific behavior as correct if they see others performing it.[329] These cognitive biases can be triggered accidentally, but they are also susceptible to being exploited by an intentional external influence. In example 10 (*Figure 3:1*), a university teacher is selling his car and negotiating the price. They attempt to get the best deal by initially suggesting an inflated price and lowering it during negotiations. The initial offer acted as an anchor, and the target thought they got a good deal, even though they still paid a higher price than the car's actual market value.

As cognitive biases are susceptible to exploitation by others, this thesis refers to them as decision-making vulnerabilities. In example 10, a target of influence is consciously deliberating, but the process is skewed, as the university teacher activates the fast-thinking brain of the buyer, introducing an influence in their conscious thinking that the target is unaware of. Alternatively, manipulators could exploit cognitive biases to bypass the conscious deliberation process altogether.[330] Subliminal stimulus (example 5) or involuntary hypnosis (example 6) would be examples of such manipulation (*Figure 3:1*). Therefore, manipulation can take the form of hidden pressure that is targeted (or otherwise is likely) to exploit the target's decision-making vulnerabilities. Being hidden is what makes such pressure a form of manipulation. If all essential aspects of influence are overt – that is, the target is aware that the influence is likely to exploit their vulnerability – an influence can be classified as *coercive*. Manipulation and coercion can be regarded as two forms of exploitation.

### 3.2.2. Beliefs, Desires, Emotions, and Nudges

Cognitive biases are not the only aspects of decision-making susceptible to exploitation. Human beliefs, desires, and emotions are also vulnerable to outside influences.[331] For example, when deciding, people can never fully cover all available information, as data that can be considered in any given situation is infinite.[332] Others may exploit this lack of perfect information to encourage their targets to hold false beliefs. Such influence on the target's beliefs is called *deception*. Example 8 and example 9, described in section 3.1, where the university teacher provides students with false information about university policy and then

[328] *See* Amos Tversky & Daniel Kahneman, *Rational Choice and the Framing of Decisions*, 59 J. BUS. 251, 257 (1986). *See also* Susser, Roessler, and Nissenbaum, *supra* note 38, at 22.

[329] *See* CIALDINI, *supra* note 318 at xiii.

[330] *See* Noggle, *supra* note 265.

[331] *See* Noggle, *supra* note 312, at 44. One of the earliest accounts for such a view is Plato's *tripartite mind*: of reason, desire and passion. *See* PLATO, *supra* note 273, at 143–152.

[332] *See* Belfast Buddhist, *Alan Watts - Choice*, YOUTUBE (2016), https://www.youtube.com/watch?v=wyUJ5l3hyTo (last visited Feb 3, 2023).

about their intentions, are the paradigm examples of direct deception.[333] Deception is always manipulation as the falsehood of the proposition is always hidden, undermining the target's ability to understand their options.[334]

Manipulators can also influence people's desires.[335] Any given individual has a myriad of interrelated desires. A person may want to fill up their water bottle because they are thirsty, continue to work at the desk to meet their desired writing goal, and want to be outside enjoying the rare sunlight, all at the same time. Ideally, people would order these desires into preferences to maximize their self-interest.[336] A person may fill up water, return to the desk immediately, and decide to go outside to enjoy some sunlight only after and if they meet their writing goal. Such orders of desires which sort out preferences about preferences are called *second-order preferences*. This ordering is rarely fully conscious and often fluid, and others can exploit this fluidity. In example 11, a university teacher is aware of their colleague's fascination with sunlight and suggests that they join them for coffee outside with the hidden intention that the colleague misses their writing goal.[337] Sexual seduction is another form of manipulative influence on a target's desires.[338]

Human emotions also play an essential role in the decisions people make.[339] Ideally, people get excited when they are about to satisfy their preferences and get

[333] Noggle refers to "direct deception" as "any assertion of a proposition that the asserter does not believe, with the intention of causing someone to believe that the proposition is true." Noggle, *supra* note 312, at 44. Deception is largely discussed mean of manipulation. For more in depth analysis *See* Noggle, *supra* note 265, at 2.2.

[334] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 21.

[335] This thesis refers to the desire to include all of the motives for acting, including visceral factors such as hunger, thirst, and sex drive, to other relatively more consciously formed preferences (e.g., attaining a law degree). Some works combine these in the word "motive". *See* Eric M. Cave, *What's Wrong with Motive Manipulation?*, 10 ETHICAL THEORY MORAL PRAC. 129, 130 (2007). Some framew them into "preferences" *See* Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: The Problem of Market Manipulation*, 76 N.Y.U. L. REV. 630, 733–743 (1999).

[336] *See* Hanson and Kysar, *supra* note 335, at 672.

[337] It is essential here that the teacher's intentions are hidden, and the colleague is not aware that a teacher wants them to fail in their writing goal. In a way, the teacher may think they have been persuaded; they may even feel coerced – if they feel that the temptation was too much for them to handle; but as the intention is hidden, the influence is manipulation. In contrast, Noggle sees Christ's temptation by Satan as a form of manipulation. This thesis does not agree with this view. *See* Noggle, *supra* note 265, at 43. If we assume that Christ knows that Satan intends to make him break the fast by tempting the visceral factor of hunger, Satan provides an irresistible incentive and coerces Christ. In case Christ could resist but chose not to, then he has been persuaded. *See* the "martini example" in Susser, Roessler, and Nissenbaum, *supra* note 38, at 18–19.

[338] For example, Cave refers to "unsavory" seduction, when a person uses cognitive biases, such as anchoring to arouse another person's sexual interest referring to Neil Straus's *The Game* and the culture of "pick-up" artists. This thesis agrees that some seduction is manipulative, but does not evaluate such manipulation is "unsavory" or not. This requires normative evaluation, that this chapter refrains from. *See* Eric M. Cave, *Unsavory Seduction and Manipulation*, *in* MANIPULATION: THEORY AND PRACTICE , 176–177 (Christian Coons & Michael Weber eds., 2014).

[339] *See* Noggle, *supra* note 312, at 44.

depressed when they think satisfying these preferences is impossible.[340] In a way, emotions help humans to scan through life's complexity to determine what to focus on.[341] For example, when a colleague follows a university teacher outside to catch some sunlight, they may feel regret, which reminds them of the second-order preference for their writing goal. However, emotions are also vulnerable to outside influence. In example 12, the colleague regrets leaving their desk and is about to return inside, but the university teacher starts to sulk about their personal life. In this case, the teacher appeals to the colleague's sympathy and tries to get them to consider this in their deliberation process. Guilt trips, peer pressure, and emotional blackmail similarly play on people's emotions to influence their behavior.[342]

Finally, human beings are also influenced by the context in which they make decisions (e.g., their physical environment).[343] For example, when people decide what to buy in the cafeteria, the arrangement of options (e.g., some are at eye level, some more challenging to reach), also called "choice architecture", influences them to select the closest options.[344] The aspects of the choice architecture that influence people's behavior are called "nudges".[345] By definition, nudges alter people's behavior "without forbidding options or *significantly* changing their economic incentives".[346] Such nudges can be in the environment accidentally, but they can also be designed intentionally to influence human behavior.[347] Many intentionally designed nudges influence to appeal to conscious deliberation (e.g., graphic health warnings on cigarette packages nudge people to consider the health effects of smoking). However, some nudges bypass conscious deliberation and influence people in a hidden way (e.g., many public bathrooms in the Netherlands introduced "fly in the urinal" that men unconsciously target and eventually minimize the spill outside the urinal). Manipulators can also nudge people by changing their decision-making contexts.[348]

In summary, humans trying to reach a decision are susceptible to manipulation in various ways: cognitive biases, beliefs, desires, emotions, and decision-making

[340] *Id.* at 46.

[341] *Id.*

[342] *See* Noggle, *supra* note 265 at 4.2.

[343] This is largely due to cognitive biases discussed in the Section 3.2.1. *See generally* THALER AND SUNSTEIN, *supra* note 318.

[344] *Id.* at 1–4. Susser, Roessler, and Nissenbaum, *supra* note 38, at 23.

[345] *See* THALER AND SUNSTEIN, *supra* note 318, at 6.

[346] *Id.*

[347] Much has been said about an overlap between nudging and manipulation. *See e.g.,* Susser, Roessler, and Nissenbaum, *supra* note 38, at 23. *See also* Robert Noggle, *Manipulation, Salience, and Nudges*, 32 BIOETHICS 164 (2018). *See also* Thomas RV Nys & Bart Engelen, *Judging Nudging: Answering the Manipulation Objection*, 65 POLIT. STUD. 199 (2017).

[348] Within the theory of manipulation adopted in this thesis, some nudges are not manipulative, but many are. Some manipulative nudges may lead to harm. *See generally* Nys and Engelen, *supra* note 347.

contexts create the vulnerabilities that manipulators can exploit to sway their targets towards their predetermined ends. Nevertheless, evaluating whether decision-making vulnerabilities have been exploited and, therefore, if manipulation has occurred is particularly challenging. This is also the case in commercial practices, such as OBA. Section 3.3. proposes how commercial practices can be investigated to evaluate whether they lead to manipulation.

### 3.3. Measuring Manipulativeness

In this thesis, manipulation is defined as a successful and intentional attempt to influence someone's behavior that remains hidden from the target's conscious awareness during the influence (Section 3.1). This thesis has described the exploitation of decision-making vulnerabilities, including cognitive biases, beliefs, desires, and emotions, as the means through which an agent manipulates a target (Section 3.2). Evaluating whether a particular practice manipulated a target requires concluding that it was "manipulative" and successfully affected the outcome. An influence can be considered manipulative if (1) an agent intended to direct a specific target toward a particular outcome (i.e., influence is targeted) and if (2) an agent intended or disregarded that an aspect of the influence remained hidden from the target (i.e., influence is hidden).[349] This thesis supports the view of manipulative influence, in which an agent may overlook the hidden aspect because they are negligent towards the means through which they influence the target.[350]

However, in contrast to "manipulation", "manipulativeness" is not a binary concept; instead, it can be best imagined as a spectrum – some attempts and practices are more manipulative than others.[351] Such a degree of manipulativeness depends on the *likelihood* that targeted and hidden influence will exploit the target's decision-making vulnerabilities. When taken to a commercial context, this thesis defines manipulative practices as attempts of the business to influence a consumer towards a targeted outcome (e.g., business profit) while willing to keep some aspect of the influence hidden in a way that can exploit their decision-making vulnerabilities. Therefore, manipulative practices have three elements:

[349] Such distinction between "manipulation" and "manipulativeness" is in line with the argumentation of Susser, Roesler and Nissenbaum. *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 26–29. One deviation from their theory may be that in this thesis intentionality of manipulation does not assume *deliberateness* of hiding the influence. Such hiddenness can be due to influencer's *negligence* about the means of influence.

[350] *See generally* Klenk, *supra* note 305.

[351] People often say that a person or a strategy is "very manipulative". One of the main attributes of manipulation is how far a manipulator is willing to take their influence. Let's imagine most striking examples of manipulation to illustrate this point: In the movie *Truman Show*, entire world is designed for the target (in this case protagonist) so that he has a false belief about his situation. In *Inception*, state of the art technology is precisely developed and used to hide the influence. In these cases manipulators go in great lengths to hide the influence.

1) they are targeted;
2) they are hidden, and
3) they are likely to exploit the target's decision-making vulnerabilities.[352]

While elements (1) and (2) are essential for a practice to be considered manipulative, element (3) provides a way to measure the degree to which the practice is manipulative (Figure 3:3). In order to illustrate how the likelihood of exploitation is differential, section 3.3.1 elaborates on the distinction between labeled and layered conceptions of vulnerability; Section 3.3.2 describes different sources that layers of vulnerabilities may stem from. Section 3.3.3 illustrates how different layers of the concept of vulnerability can be used to measure degrees of manipulativeness.

### 3.3.1. Vulnerability

In ordinary language, vulnerability means exposure to attack or damage.[353] It is typically ascribed to a subject and describes relative exposure toward a particular outcome (subject *X* is vulnerable to outcome *Y*). For example, computer systems (subject) are vulnerable to cybersecurity breaches (outcome).[354] Human beings are vulnerable to being physically or emotionally wounded (in Latin, "vulnus" means "wound").[355] Human beings are vulnerable to various types of harm, making human vulnerability a complicated concept to untangle.[356] This is more so in legal theory, which borrows terminology and conceptual frameworks of vulnerability from

[352] *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 27.

[353] *See Definition of Vulnerable*, MERRIAM-WEBSTER (2023), https://www.merriam-webster.com/dictionary/vulnerable (last visited Feb 6, 2023). Other definitions refer to exposure to "harm". Harm has a specific meaning in legal discourse and is discussed in detail in Chapter 5 in the context of manipulation via OBA. To avoid confusion, this thesis refers to vulnerability as exposure to attack.

[354] The terminology of "exploiting vulnerabilities" is widely used in cybersecurity, where it refers to hackers using vulnerabilities in computer systems to breach the security and access stored data. *See e.g., Cybersecurity Vulnerabilities: Types, Examples, and More*, GREAT LEARNING BLOG: FREE RESOURCES WHAT MATTERS TO SHAPE YOUR CAREER! (2023), https://www.mygreatlearning.com/blog/cybersecurity-vulnerabilities/ (last visited Feb 7, 2023).

[355] *Definition of Vulnerable*, *supra* note 353. *See* VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY, 4–5 (Catriona Mackenzie, Wendy Rogers, & Sandy Dodds eds., 2014). Also note, that in contrast to how it is often used in academic literature, human vulnerability in this thesis does not mean human fragility. This thesis endorses the view of humans being vulnerable like plants, not fragile like jewels: vulnerability that exposes plants (and humans) to injury is also the source of their growth. *See* Will Buckingham, *Vulnerability and Flourishing—Martha Nussbaum*, HIGHBROW (Oct. 26, 2017), https://gohighbrow.com/vulnerability-and-flourishing-martha-nussbaum/ (last visited Feb 7, 2023). In a way, it can be argued that vulnerability is "antifragility" *See* for the concept of antifragility NASSIM NICHOLAS TALEB, ANTIFRAGILE: THINGS THAT GAIN FROM DISORDER (2013).

[356] *See* Gianclaudio Malgieri & Jedrzej Niklas, *Vulnerable Data Subjects*, 37 COMPUTER L. SECURITY REV., 3–5 (2020).

various external disciplines, such as political philosophy, gender studies, and bioethics.[357]

These disciplines conceptualize vulnerability to address a broad range of problems.[358] For example, bioethics considers the concept of vulnerability for protecting human research participants.[359] In comparison, political theorists view it as a human condition ("la condition humaine") that triggers state responsibility and places it at the roots of political organization.[360] Such multiplicity of meanings and functions makes an overarching definition of vulnerability elusive.[361] This thesis scopes the use of the concept solely in a decision-making context, with a particular emphasis on commercial relationships.[362]

Even with such a scope, the quest for defining vulnerability may lead to stereotyping sub-populations or excluding the vulnerable from rigid taxonomies that cannot fully grasp the complexity of vulnerability in real-life.[363] Nevertheless, this thesis recognizes the need to formulate a coherent way of thinking about vulnerability in a decision-making context to support legal discussions about the likelihood of manipulation and risks. Historically, legal discussions have adopted a "labeled" understanding of vulnerability that labels particular sub-populations (e.g.,

[357] *Id.* at 3.

[358] *Id.* at 3–5. *See also* VULNERABILITY, *supra* note 355, at 4–5.

[359] Vulnerability is a foundational concept in bioethics. In particular, it arose as the need to give express consent for participation in human research. *See* Wendy Rogers, *Vulnerability and Bioethics*, *in* VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY (Catriona Mackenzie, Wendy Rogers, & Susan Dodds eds., 2013). *See also* Florencia Luna, *Identifying and Evaluating Layers of Vulnerability – A Way Forward*, 19 DEV. WORLD BIOETHICS 86 (2019).

[360] *See* ROBERT E. GOODIN, PROTECTING THE VULNERABLE: A RE-ANALYSIS OF OUR SOCIAL RESPONSIBILITIES (1986). *See also* MARTHA C. NUSSBAUM, FRONTIERS OF JUSTICE: DISABILITY, NATIONALITY, SPECIES MEMBERSHIP (2007). *See also* Martha Fineman, *The Vulnerable Subject and the Responsive State*, 60 EMORY L. J. 251 (2010). Chapter 5 returns to vulnerability in context of political theory.

[361] Luna particularly argues against developing taxonomies of vulnerability and develops theory of vulnerability as *layers* and not *labels*. Her theory steps away from stereotyping vulnerable groups by "labeling", and, maintains conceptual flexibility to cover variety of forms of vulnerability. *See generally* Luna, *supra* note 359. This thesis adopts Luna's point of view of vulnerability as layered. However, while this thesis agrees that it is impossible to categorize reality in particular in such complex contexts as human behavior, it finds it necessary to create a taxonomy that resembles the real-life complexity of vulnerability at least more accurately than taxonomies in current legal instruments. This is particularly the case in the context of the view of vulnerability in the EU consumer protection law. *See* Joanna Strycharz & Bram Duivenvoorde, *The Exploitation of Vulnerability Through Personalised Marketing Communication: Are Consumers Protected?*, 10 INTERNET POLICY REV. (2021).

[362] Thesis analyzes the extent to which human beings are vulnerable to manipulation, and the extent to which decision-making is vulnerable to exploitation (such exploitation can come from coercion or manipulation).

[363] *See* Luna, *supra* note 359, at 90.

minors, persons with mental disabilities) as “vulnerable groups”.[364] Studies from other disciplines have criticized such a model and argued that membership in a group can be understood only as one of several “layers” of an individual’s vulnerability to manipulation.[365] These layers rarely, if ever, apply in isolation to any given individual, but they interplay with each other to form a complex figure of a person’s vulnerability.[366]

Therefore, while entirely capturing and precisely measuring such complexity may be impossible, without outlining better contours of vulnerability to manipulation, legal instruments may fall strikingly short of meeting their aims and leave vulnerable individuals unprotected. This is important in the EU legal framework for OBA, where vulnerability is a key concept. For example, vulnerability plays a definitive role in regulating manipulative practices in the discussions on the Artificial Intelligence Act (AIA). In the proposal for AIA, the European Commission endorsed vulnerability as a labeled concept in Article 5, and also introduced vulnerability due to “an imbalance of power, knowledge, economic or social circumsntaces” in Article 7 that resembles the layered vulnerability approach.[367] Therefore, to support the legal discussions in better capturing human vulnerability, this thesis builds upon neighboring disciplines and endorses the view of vulnerability as a layered concept.[368] Section 3.3.1 explains how different layers interplay to create a spectrum of vulnerability.

### 3.3.2. Levels of Vulnerability

This section differentiates between three sources of vulnerability: (1) *intrinsic vulnerabilities* stem from the target of the influence; (2) *situational vulnerabilities* stem from the circumstances, and (3) *relational vulnerabilities* stem from the asymmetries in the relationship between a target and the agent of the influence. Such delineation of sources is intended to capture, rather than to limit, various types of vulnerability. In specific contexts, the line between sources of vulnerability may be blurred. For example, relational factors can be considered situational, and situational factors as intrinsic. Therefore, this thesis merely refers to sources to explicate the potentiality of different layers and suggests a way to measure vulnerability to

[364] *See e.g.,* Unfair Commercial Practices Directive, *supra* note 42; *See also* AI Act Proposal referring to technologies that “exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities”. *See* AI Act Proposal, *supra* note 52 at 13.

[365] *See* Luna, *supra* note 359, at 90.

[366] *See Id.*

[367] The European Parliament has suggested updating the model to include other layers (e.g., socio-economic factors) *See* Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021/0106 (COD) Draft 20-06-2023, art. 5, 7. [hereinafter AI Act Mandates].

[368] *See Definition of Vulnerable*, *supra* note 353.

manipulation on the spectrum by adding them up. The following paragraphs expand on these three sources of vulnerability.

Firstly, the target of an influence can be intrinsically vulnerable to manipulation. In a decision-making context, on a fundamental level, all human beings are *inherently* vulnerable to manipulation.[369] The source of inherent, intrinsic vulnerability in human decision-making is the human embodiment and their social nature.[370] In particular, human thinking is shaped by the physiological and psychological needs that arise within their bodies, which also predisposes humans to cognitive biases (see more detail in section 3.2.1).[371] Humans need other human beings to meet many of their needs, and as they can never fully process all information available in a given situation, they need to rely on emotions and assumptions about other humans (trust others) and the world around them (have beliefs). Therefore, in this way, vulnerability is a constant condition in decision-making for all human beings.[372] One can think of it as a baseline to be a mentally and physically healthy adult capable of forming decisions independently in a situational vacuum.[373]

On top of the baseline or inherent, intrinsic vulnerability, each individual has other intrinsic traits that can make them particularly vulnerable to erroneous decision-making.[374] Generally, such differential understanding of intrinsic

[369] *See* MARIJN SAX, BETWEEN EMPOWERMENT AND MANIPULATION: THE ETHICS AND REGULATION OF FOR-PROFIT HEALTH APPS 76 (2021).

[370] *See* Introduction: What Is Vulnerability, and Why Does It Matter for Moral Theory?, *in* VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY, 7 (Catriona Mackenzie et al. eds., 2013).

[371] *Id.*

[372] Vulnerability theorists also distinguish *dispositional* v. *occurrent* vulnerabilities that relate to the potential vulnerabilities and the fact that vulnerabilities are actualized. *See Id.* Imagine the example of the solubility of a sugar lump. Sugar has solubility – if it is placed in water, it will dissolve. However, it is not dissolved unless it is exposed to water. Similarly, there are potential and actual vulnerabilities in the decision-making context. All human beings are universally vulnerable to manipulation. This potential exists even when a person is alone in a dark room (without a phone or an internet connection). However, manipulation can only happen if a social interaction occurs. This thesis only discusses the actualized vulnerability, that is, when the target is being influenced toward a particular outcome.

[373] Humans rarely make decisions in a vacuum, fully autonomously. Relationships with others often provide catalysis for human decision-making. Some philosophers call this phenomenon "relational autonomy" and understand a human being as a "relational self". *See* JONATHAN HERRING, LAW AND THE RELATIONAL SELF (1 ed. 2019). Due to this relational nature some scholars see autonomy and vulnerability as *entwined*. *See* Joel Anderson, *Autonomy and Vulnerability Entwined*, *in* VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY 134 (Catorina Mackenzie ed., 2013). *See also* SAX, *supra* note 369, at 78. This thesis agrees with such an understanding human decision-making. It suggests that vulnerability is very tissue of autonomy. That is, autonomy in humans is contingent on vulnerability. An eye sees only when it is exposed; the "Self" expresses autonomy in society or in relation to others.

[374] *See* Mackenzie et al., *supra* note 349, at 7.

vulnerability is based on different degrees of resiliency and coping ability.[375] In a decision-making context, one such intrinsic vulnerability may come from belonging to a particular age group, such as minors or the elderly. A mental disability, such as obsessive-compulsive disorder (OCD), or chronic physical illness, such as diabetes, can be considered an intrinsic personal vulnerability to manipulation. Some intrinsic traits, such as introversion or personality type, may not make a person vulnerable to manipulation *per se* but can be triggered as a vulnerability by a particular stimulus or other circumstantial or relational factors. For situations when personality traits become actualized as vulnerabilities, this thesis considers them as personal intrinsic vulnerabilities.[376]

Secondly, some vulnerabilities are *situational* in that they stem from the particularities of the circumstances that are extrinsic to individuals.[377] Such situational vulnerabilities can be short-lived, intermittent, or long-term and typically involve environmental, social, economic, political, and personal circumstances.[378] Environmental circumstances, such as a pandemic or an earthquake, may significantly affect individuals' decision-making processes.[379] For some, it may become challenging to deliberate due to political circumstances, such as riots or armed conflicts. Social factors can also have a significant effect: for example, depending on the community, having a particular race, sexual orientation, or gender may be a reason for a person's oppression, making them vulnerable to a wide range of interferences with their decision-making.[380]

Systematic racism, sexism, or homophobia may blur the line between situational and intrinsic vulnerabilities: while systematic racism is not an intrinsic state to the human condition, for some, it can feel like an "inescapable" feature of their life experience.[381] Such distinction between intrinsic and situational is even more blurred when vulnerabilities stem from personal circumstances. For example, becoming a parent, going through a divorce, losing a loved one, or losing a job are personal circumstances in which people are more susceptible to influences in their decision-making.[382] This thesis identifies such vulnerabilities as situational because their source is the situation, not an intrinsic trait.[383] The precise delineation between intrinsic and situational vulnerabilities is not necessary for the coherence of a

[375] *Id.*

[376] *See* Strycharz and Duijvenvoorde, *supra* note 361, at 5.

[377] *See* SAX, *supra* note 369, at 77.

[378] *See* Mackenzie et al., *supra* note 349, at 7.

[379] *See* SAX, *supra* note 369, at 78.

[380] *See Id.*

[381] *See Id.* at 77.

[382] *See Id.* at 77.

[383] This thesis also recognizes the need for future research in this area for better delineating between different sources, and forms of vulnerability in decision-making contexts. Developing coherent framework is essential for protection of vulnerable.

layered vulnerability framework. Instead, such conceptualization intends to capture both sources of vulnerabilities and regard them as layers that may compound and exacerbate the overall vulnerability of a target to be manipulated.

Thirdly, a person can also be vulnerable to manipulation due to the particularities of their relationship with the agent of influence.[384] Humans are vulnerable in all relationships because humans need to trust other human beings.[385] As Keymolen puts it: "We, as human beings, cannot exactly predict the thoughts and actions of others; they are—to a certain extent—black boxes to us and, consequently, constitute a source of insecurity."[386] In other words, uncertainty about others' potential actions constitutes a form of vulnerability in a person's decision-making capabilities.[387] Humans are particularly vulnerable when in hierarchical relationships or relationships with information or power asymmetries.[388] For example, in a decision-making context, vulnerability occurs in teacher-to-student, employer-employee, business-to-consumer, caretaker-patient, or similar relationships where one party has the authority or the other way sets the rules of the interaction.[389] Therefore, relationships can act as a layer of vulnerability.

In summary, illustrating different sources of vulnerability reveals various layers that can compound one another and create varying levels of vulnerability that can be imagined on a spectrum instead of a monolithic label applied to specific groups (Figure 3:2). Every human being can be regarded as having at least a baseline level of vulnerability to manipulation (*ordinary vulnerability*). A personal trait, situational circumstance, or relational asymmetry can provide a second layer and deem a person more than ordinarily vulnerable (*vulnerable*). Vulnerabilities can compound: a personal trait, a situational circumstance, or the nature of a relationship which can act as an additional layer and create a state of *heightened vulnerability.* The compound effect of vulnerability can be exaggerated due to further compounding; people who have four or more layers of vulnerability can be regarded as presenting *extreme vulnerability* to manipulation.

[384] Mackenzie, Rogers and Dodds identify pathogenic vulnerabilities, which are "a subset of situational vulnerabilities that are particularly ethically troubling." *See* Mackenzie et al., *supra* note 349 at 9. They discuss examples when a person is assigned a caretaker because of their vulnerability, and the caretaker abuses their role to exploit vulnerability. Relational vulnerability is not the same as pathogenic. Instead, it focuses on the authority and hierarchical relationships that may cause pathogenic consequences.

[385] *See also* SAX, *supra* note 369, at 80.

[386] *See generally* Esther Keymolen, *Trust In the Networked Era: When Phones Become Hotel Keys*, 22 TECHNÉ RES. PHIL. & TECH. 7 (2018).

[387] *See Id.*

[388] *See* Rogers, *supra* note 359, at 68–69.

[389] Margaret Urban Walker, *Moral Vulnerability and the Task of Reparations*, *in* VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY (2013).

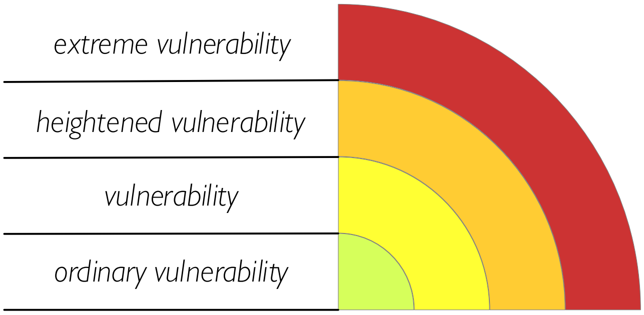


*Figure 3:2. Levels of vulnerability (by author)*[390]

Lastly, these levels of vulnerability can be used to evaluate how manipulative (and coercive) the practice is, which can be linked to the likelihood of an influence to exploit the vulnerability.[391] Section 3.3.3 uses the levels of vulnerability in Figure 3:2 to develop a framework that can evaluate the levels of manipulative practices and other forms of influence on a spectrum (Figure 3:3). Section 3.3.3 explains these levels by providing illustrative situations (*s1 – s7*) that exemplify the differences in levels and forms of influences. These situations are also placed in Figure 3:3.

### 3.3.3. The Spectrum of Influences

The likelihood of exploitation may depend on the specificity with which the influence is tailored to the target's vulnerabilities.[392] In order for an influence to be considered manipulative under the definition of this thesis, the influence does not have to be intentionally targeted to these vulnerabilities. Instead, manipulative influence involves a deliberate attempt to influence a person, coupled with the agent's expected awareness that the influence can exploit the target's vulnerabilities.[393] Therefore, how manipulative the practice is can depend on the target's level of vulnerability.

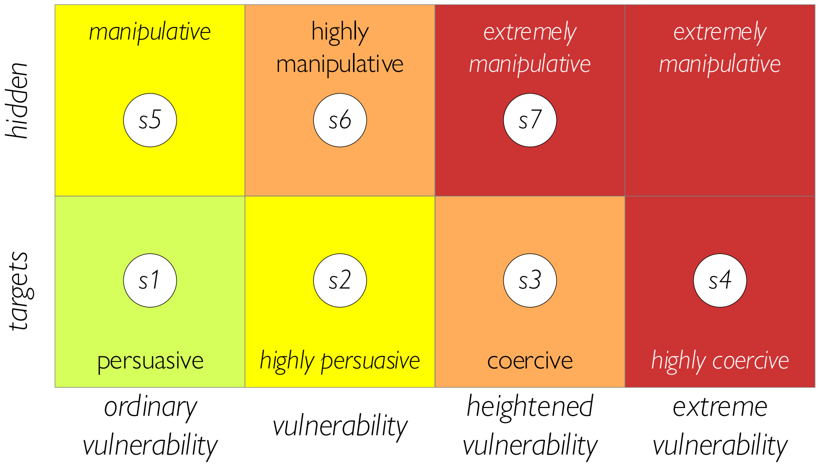
[390] The figure is a representation by the author of the "layered" concept of vulnerability developed in sections 3.3.1 and 3.3.2, where adding one layer of vulnerability increases the level of vulnerability by one on the spectrum.

[391] *See* Susser, Roessler, and Nissenbaum, *supra* note 38 at 27.

[392] Susser, Roessler & Nissenbaum argue: "Indeed, targeting is best understood as an exacerbating condition: the more closely targeted a strategy is to the specific vulnerabilities of a particular manipulee, the more effective one can expect that strategy to be." *Id.*

[393] Klenk defends this point of view. *See generally* Klenk, *supra* note 305. Klenk argues against Susser, Roessler & Nissenbaum's manipulation as a "hiddenness" view. However, this thesis finds that carelessness and hiddenness conditions are not self-excluding; instead, the fact that influence itself can be overt, but the vulnerability exploited due to the manipulator's negligence of other person's exposure supports the hiddenness argument. Critics may argue that such failure can be in any social situation or advertisement. However, when an agent of influence takes steps to turn up the notch of specificity with which they are to target a person's characteristics to influence them – they also increase the likelihood that they will exploit the vulnerabilities.

Generally, targeting vulnerabilities can also be employed as a method for overt forms of influence. Vito Corleone, Mafia don from the movie The Godfather, increases the likelihood of his *coercive* attempts being effective by placing the head of his target's favorite horse into his bed.[394] Mr. Keating, the English teacher from the movie Dead Poets Society, also increases the likelihood of his *persuasive* attempts being effective by showing his students the picture of the dead alumni to encourage them to live extraordinary lives.[395] As long as the target can become conscious of their own vulnerability, an influence that is likely to exploit this vulnerability cannot be classified as manipulative. Figure 3:3 illustrates how the specificity of targeting, hiddenness, and the likelihood of exploitation of vulnerability interact with forms of influence.



| | | | | |
|---|---|---|---|---|
| hidden | manipulative<br>s5 | highly manipulative<br>s6 | extremely manipulative<br>s7 | extremely manipulative |
| targets | s1<br>persuasive | s2<br>highly persuasive | s3<br>coercive | s4<br>highly coercive |
| | ordinary vulnerability | vulnerability | heightened vulnerability | extreme vulnerability |

*Figure 3:3. Spectrum of influences with situations (by author)*[396]

This section illustrates differences between forms of influence and their levels based on levels of vulnerability in seven situations (reflected in Figure 3:3 as *s1*-*s7*) in which agent *y* aims to get target *x* to drink.[397] Suppose a <u>situation 1</u> (*s1)*, where *y* asks *x* if they are up for having a drink. This situation reveals a *persuasive* attempt – it is clear that *y* wants *x* to have a drink, where *x* is ordinarily vulnerable. Suppose <u>situation 2</u> (*s2*), where *x* has disclosed to *y* that martini is their favorite drink, and *y* asks *x* if they are up for having martinis. This situation reveals another but more persuasive attempt that appeals to the personal preference of *x*. Personal preference

[394] THE GODFATHER (Paramount Pictures, 1972).

[395] DEAD POETS SOCIETY (Touchstone Pictures, 1990).

[396] The figure was developed by the author to illustrate differences in forms and levels of influence based on the theory of vulnerability and manipulation developed in Chapter 3 of this thesis.

[397] The situations described in this section are adaptations of the "martini example" introduced by Susser, Roessler, and Nisenbaum. *See* Susser, Roessler, and Nissenbaum, *supra* note 38, at 18.

for martinis acts as a layer of vulnerability for *x,* who has the second-order preference of not having a drink that day. This situation reveals an influence to be *very persuasive,* as *x* is more than ordinarily vulnerable. Suppose situation 3 (*s3*), where *x* has also disclosed that they want to stop drinking, and they cannot be anywhere near vodka martinis; *y* makes a fine glass of shaken vodka martini and puts it in front of *x*. This situation reveals an influence that is as *coercive* as it is overtly tailored to the target, which is *highly vulnerable* to such an influence.[398] Suppose situation 4 (*s4*), where *x* discloses to *y* that it is the end of a particularly stressful workday and he could use a drink if he was not trying to quit – in response to which *y* parades a fine glass of shaken vodka martini in front of them. This situation reveals a *highly coercive* attempt that overtly tries to influence a target that is extremely vulnerable.

An influence is overt in all of these four situations (s1 - s4). *x* knows that *y* is aware of the extent of their vulnerability and is conscious of *y*'s objectives and the nature of the influence. Let us suppose situation 5 (*s5*), where *x* does not explicitly ask *y* to join him for a drink; instead, *y* appears to *x,* sipping on the drink in front of him. This situation reveals a *manipulative* attempt in which *y*'s intentions are hidden, and x, as the target of the influence, is ordinarily vulnerable to such hidden influence. Suppose situation 6 (*s6*), where *y* learned from *z* that *x* is a lover of martinis, and they are holding not any other drink but a martini when they approach *x*. In this situation, *y's* influence is hidden, and tailored to *x,* who is more than ordinarily vulnerable. Such influence is *highly manipulative*. Lastly, suppose situation 7 (*s7*) where *y* knows that *x* associates drinking with jazz and turns the music on while starting to drink a martini in front of *x*. This situation reveals an attempt at the hidden influence that is tailored to target those extremely vulnerable to the tailored influence and, therefore, is an *extremely manipulative* practice.[399]

In sum, a layered understanding of vulnerability provides a way to understand the different levels at which the influence can be manipulative as well as coercive. In the end, this thesis understands the actions of the agent to be manipulative to the extent to which they can exploit the target's decision-making vulnerabilities, given that the agent intends to influence a target towards a particular outcome, and some aspect of the influence remains hidden. All influence that overtly exploits human vulnerabilities can be conceptualized as coercion.

### 3.4. Conclusion: Manipulation

Thus, this section offers an answer to the second sub-question of this thesis:

[398] Some Georgians will regard this as coercive but morally justified.

[399] Again, the concept of manipulation in this thesis is morally neutral. In some situations, a person may find it morally justified to manipulate their friend to have a drink, and, indeed, this can be a form of entertainment. Similarly seduction in romantic relationships can be highly manipulative, and a form of play.

SQ2: what is manipulation?

Manipulation is an agent's successful and intentional attempt to influence a target towards an outcome, where an essential aspect of the influence remains hidden from the target's awareness, and the agent is aware that the method of influence can exploit the target's decision-making vulnerabilities. Manipulation is a successful influence – the target has behaved like the agent wanted. Manipulation can happen through deception or pressure. Manipulation contrasts with other forms of influence, such as persuasion and coercion, in that it is *hidden.* It contrasts with persuasion in that it takes away the target's ability to exercise choice. It contrasts with coercion in that it maintains the illusion of choice – as individuals believe they are exercising choice.

The success of an agent's attempts to manipulate depends on the likelihood of the actions exploiting human decision-making vulnerabilities. An action of an agent can be said to be manipulative if (1) an agent intends to influence a target towards a particular outcome, (2) an essential aspect of the influence remains hidden, and (3) the method of influence is likely to exploit the target's vulnerability. The likelihood of exploitation can be mapped out on the different levels of the target's susceptibility to manipulative influence, that can be evaluated based on target's inherent, situational, or relational vulnerabilities. An action can be "manipulative" if it is tailored to a target who is ordinarily vulnerable to being manipulated by this action. Additional layers of vulnerability can make a target "more than ordinarily vulnerable" or with "heightened vulnerability", and tailoring action to such a target can be considered "highly manipulative" or "extremely manipulative", depending on the level of vulnerability.