Invisible ethics: the problem of anonymity on the internet

Thesis submitted to the University of Ottawa in
partial fulfillment of the requirements for the
Master of Arts (M.A.) in philosophy

Department of philosophy
Faculty of arts
University of Ottawa

Summary

This thesis aims to shed some light on the way anonymity affects behavior online, as well as the moral implications this influence supposes for internet users. We do this in three chapters, each dedicated to a specific aspect that might be problematic, and look at the technical aspects of the influence of anonymity, as well as the psychological aspects of this influence. We begin by analyzing anonymity as a tool that gives rational actors an advantage over their peers. We then gage the extent to which it constitutes a situational force, and how it coordinates with other situational forces within the online environment. Finally, through an evaluation aided by moral psychology and the literature in ethics, we use a situationist approach to identify the two main ways in which we believe anonymity affects moral behavior in the internet.

Acknowledgements

I would like, before anything else, to thank my parents Ana and Pedro for being the first readers of the final draft of this thesis. Their review, far from candid and yet unclouded by the philosophical minutiae that I often found myself lost in, allowed me to see where the rabbit holes were.

And yet, I still stumbled upon a few along the way. That is why I want to thank professors Andrew Sneddon and Vincent Bergeron for their comments, suggestions and criticisms following the presentation of my thesis project, some of which I dreaded, all of which this thesis needed in order to head the right way. Likewise, I am very thankful to them for reading my thesis and, during my thesis defense and through their final comments, for asking questions that I should have expected to hear, as well for raising points that I didn´t consider enough.

I would also like to thank the University of Ottawa, the Faculty of Arts and the Department of Philosophy for the opportunity to have worked as a teacher assistant before and during the redaction of my thesis.

The location, of course, would be little without the people. I am very grateful to all of my fellow students in the philosophy department who, by taking me away from blank pages and writing blocks through good friendship, clever wit and interesting conversation, made every day typing away a little less tiring and every day closer to the deadline less unnerving. My only regret is that I won´t get to name them all here. Still, I would specifically like to thank Nicolas Paradis, Jérôme Gosselin-Tapp, Alexis Tétreault and David Marinier for reading through early drafts, and through their advice, for helping me improve it. This thesis would not be the same without their help.

Ultimately, but most importantly, I would like to thank my supervisor, David Robichaud. For his support, dedication and most importantly, for his enthusiasm for this entire process. Our conversations always provided me with insights, ideas and changes that were fundamental to this thesis. And if at any point I happened to stop believing that this “thing” was going to work, he effortlessly convinced me otherwise with a trust that I don´t think I might have deserved at times, but that I can never thank him enough for placing in me.

Table of contents

Introduction

The 21st century has also been called the digital age, the information age or the computer age. This is because the number of aspects of our lives translated into digital media, as information coded into computers, has known a steady, one could even say exponential growth. The most common manner in which people interact through these new technologies is the internet, and, within it, the world wide web. There was a time when it served a more instrumental, fixed purpose, less about communication and more about information gathering and sharing. But what once was a lesser tool of communication, limited to desktop computers, is now made mobile by laptops, tablets and, most importantly, smartphones. Nowadays, the internet serves both as the repository of most knowledge, the medium for its transmission and the locus of a great part of human interactions. Because of this, it has become harder and harder to ignore the impact that coexisting within the environments brought by these new technologies will bring into the daily lives of people that have access to them.

Speaking in purely scientific terms, the real world and the virtual world are separate, but they are also different in the way they function. Broadly speaking, the offline world is a planet subdivided into countries with clear borders and legislation that limit our circulation. It is a world with immediate reactions to every action and consequences to every choice. And while we can customize our experience within it, our choices are often limited by our social, financial and cultural situations. The online world is an unlimited and ever-expanding aggregate of data. It is a borderless world where we can easily go wherever we aim to, and if certain places are locked to us, we can reasonably expect to find ways to access them. Within the offline world, our personal circumstances are not necessarily evident, and do not necessarily limit our freedom, thus giving extensive control over the customization of our experiences online. Given the great disparities between these two realms, it is to be expected that our balancing of the two would require some effort, and yet, over

the course of the last two decades, information technology has crept into people's lives, going from useful to necessary for some.

Indeed, one look at how communication has shifted in the last few years will yield all the evidence one needs to understand that its use has consequences. For instance, individuals have lost more and more of their private space and time away from others, with their electronics having transcended the barriers of the workplace, schoolgrounds and other social environments to bring their issues home, as the ever-growing phenomenon that is cyber-bullying seems to show.[1] This, in turn, raises a series of questions and concerns that are not only psychological, but philosophical. For instance, how does such a versatile and rapidly evolving medium affect people's behavior? From an ethical standpoint, it seems that there is a shift not only in our moral practices, but also in our conception of these moral practices whilst online.

The internet has allowed people to have access to a day to day form of coexistence that we believe is different from any other kind we have experienced before. And we believe it has effects on people that are directly tied to how strongly this environment favors anonymity, be it actual anonymity or perceived anonymity. More specifically, that it has both side-effects on people's internet demeanor and their daily, non-virtual lives. In particular, we think their ethical views can be deeply affected by their online presence, and that many of these effects can be traced back directly to anonymity performed through the online medium, like a magnetic influence over an ethical compass. The goal of this thesis is therefore to determine how anonymity online can alter our moral behavior. We will divide the problem into three chapters, all related to what we expect to be the different ways in which online anonymity affects human behavior. In the first chapter, we

[1] Wingate, V. Skye, Jessy A. Minney, and Rosanna E. Guadagno, "Sticks and stones may break your bones, but words will always hurt you: A review of cyberbullying." Social Influence 8.2-3 (2013): 87-106

will develop the more technical issue of understanding anonymity as a strategic tool that gives users objective power through the information game, and makes it harder to hold individuals morally accountable. We will begin by providing a historical analysis of the concept, moving on to describe the internet's specificities and showing how these characteristics favor problematic behavior through anonymity, by using the example of online trolls. In the second chapter, we will present the problem of anonymity as a situational force among others, how they coordinate and affect the way we interact with and perceive others while online. We will begin by presenting the literature on pre-internet anonymity, as well as its connection to Computer Mediated Communication. We will then describe the social situational forces that exist online, and explain how their shift from the real world to the virtual one has tied them to anonymity. We will then identify the internet characteristics that we believe to be significant enough to constitute situational forces that can synergize with anonymity, and finally provide some examples of problematic behavior online to expose the influence of these forces. In the third chapter, we will present anonymity's effects on self-perception and identity, and how being concealed online influences a person's preferences, beliefs and moral dispositions. We will begin by providing a detailed understanding of morality and its ties to identity through a historical review of several theories in ethics, and choose the one that better fits our requirements. We will then use this theory to make sense of the uniqueness of the internet environment, of how it affects identity and causes the problematic behaviors that occur within it. Finally, we will propose our hypothesis on anonymity's effects, and provide two main directions for elucidating this problem.

Chapter 1: Anonymity as a tool of rational actors online

1.1 Understanding anonymity

1.1.1 A matter of identifiability

In January 21st, 2008, a message was posted on YouTube. In it, with a background of rapidly moving clouds over a grey building, a computer-generated voice spoke against the church of Scientology, promising its expulsion from the internet and the inevitable dismantlement of the organization. The overtly aggressive declaration of war, motivated by discontent with the organization's actions and policies, ended with the following words: "Knowledge is free. We are Anonymous. We are Legion. We do not forgive. We do not forget. Expect us."[2] In the months that followed, the Church of Scientology was the victim of prank calls, denial-of-service attacks against their websites and other forms of self-proclaimed hacktivism.[3] Through all of it, Anonymous not only undermined the institution's message, but it also brought to light how vulnerable such organizations were to hacking. And although some culprits were tracked, most of them remained at large, nonidentified. Over the years, the group known as Anonymous would continue to perform these cyber-attacks, their victims ranging from religious organizations to governments and their institutions.[4] And every time, their grandiose, dramatic mantra rang true. They had access to an

[2] Unknown, Church0fScientology, (January, 2008). Message to scientology. Retrieved from https://www.youtube.com/watch?v=JCbKv9yiLiQ

[3] Gabriella Coleman, "Hacker, hoaxer, whistleblower, spy: The many faces of Anonymous.", Verso books (2014), Chapter 1.

[4] Ibid.

endless, free trove of knowledge on the internet. Their numbers were impossible to be determined accurately. They were judge, jury and executioner of the sentences they passed upon their victims.

But of all its components, only one truly stood out as a defining feature, the one thing that allowed them to do all these previous things. And that was the group's very name: Anonymous. It was at the same time an explanation and a justification. It was because they were anonymous that they could do what they did. Some would see it as activism by a collective of discontented, kindred souls, others as actions bordering on terror, but when boiled down to its core, it is not too different from other forms of collective action like strikes and manifestations: a group of like-minded individuals gather and share their views, before taking action to ensure that they are heard by society. What makes it novel is that the idea of having a truly anonymous organization, that is to say thousands of individuals cooperating at a moment's notice without ever seeing or knowing each other, was borderline impossible before the internet made it so. We don't mean to say by this that there have been no collectives of anonymous individuals cooperating on a large scale before anonymous, but rather that the scale and ease at which Anonymous does this is unlike anything done before it. This anarchic, perpetually shifting, and to a certain extent leaderless "legion" of individuals is a clear example of how anonymity has shifted in form with the development of information technologies. Indeed, what is most interesting is to note that the word "anonymous" that has become fundamentally entwined with this organization is different from the word we knew before it went online, and any conceptual work in regards to anonymity has to take this into account. This doesn't by any means entail that the conceptual work done before the advent of the internet is now irrelevant, but quite the contrary, that there is a need to establish what meaning is embedded into the concept.

In this chapter, we will try to paint a picture of anonymity. We will first define it. Then we will bring forward the key aspects of anonymity. We will also see what elements of the literature on anonymity established prior to the internet era could help us further our understanding of the problem. Then, we will see how these base problems of anonymity translate online, and both identify the problematic behaviors that rise from it and see how these often stem and benefit from it. It must be noted that in the scope of this chapter, we will not delve at length into the analysis of the situational forces that affect our behavior, or of the intricacies of the moral psychology of individuals when they engage in these behaviors (this will be the subject of the following two chapters). We will, however, evaluate these behaviors as being problematic from a social standpoint, and identify how some situations do facilitate problematic acts. Indeed, for now, our goal is to see the ways in which users transgress rules online, and how anonymity influences this by serving an instrumental purpose.

The Greek root of the word "anonymous" is "anonumos", which means "nameless",[5] and was usually employed in regards to literary authors whose name was absent from their work. The adjective was eventually broadened into a larger-encompassing noun "anonymity", used more generally to describe this state of disconnection between an individual and a written work.[6] It thus seems that, at first, anonymity was essentially linked to written communication, and more particularly to two components of it: the emitter and the message. More specifically, it answered the question of whether either of these is identifiable. Which is why, to give a good understanding of anonymity, we must determine what constitutes being identifiable, because being anonymous, at its simplest form, is being nonidentifiable. Someone being identifiable means that there are

[5] Oxford Dictionary

[6] Anne Ferry, *""Anonymity": The Literary History of a Word.",* New Literary History 33.2 (2002), 193-214.

distinct, evident links between the emitter and the message. In the literary sense we described above, identifiability for an author means that the written contents can be traced back to the one who wrote them. However, a passport or other such government-issued identification does a similar thing in a different way: it links a face with a date of birth so as to make a person identifiable. When comparing the two, it becomes evident that there is a difference in complexity when it comes to identifying the emitter, as we can do it either by looking at the message or the emitter themselves. We therefore believe that identifiability is something that should not be measured arbitrarily, but rather with a complex, context-dependent scale. Gary Marx presents a complex scale of seven dimensions of identity knowledge that distinguish the many ways in which we might be identified,[7] and we shall be using it for the time being. These criteria are complementary in allowing for identification, even if they are progressively less reliable as we go through the list. However, in specific contexts, some will be more pertinent than others. Let us look at an example to clarify:

> Miriam gets to a party and meets up with her friend Laura, who promptly warns her to look out for Sharon, whom she says is an unfathomable bore. The problem is that Miriam doesn't know Sharon. At this point, Laura telling her Sharon's social identity number or address is useless, because whilst these are often highly useful identification criteria, in the context of a party these will be utterly useless at identifying and subsequently avoiding Sharon because they will not be evidently available to Miriam. Likewise, just telling Miriam that

[7] Gary T. Marx, "*What's in a Name? Some Reflections on the Sociology of Anonymity.",* The Information Society 15.2 (1999): 99-112. The criteria are the following: 1. Legal name, 2. Locatability, 3. Pseudonyms linked to name or location, 4. Pseudonyms that are not linked to name or location, 5. Pattern knowledge, 6. Social categorization, 7. Symbols of eligibility/noneligibility

> Sharon's hair is brown is useless, given there are several other women with brown hair in the party. However, if Laura tells Miriam that Sharon is the brown-haired woman with the nose piercing, the torn jeans and the Che Guevara shirt, then Miriam can probably identify her.

This example highlights two facts about information. Firstly, the utility of certain pieces of information in identifying someone depends entirely on the observation channels available and whether this information can be taken in and processed or not. Secondly, a pertinent piece of information might only identify someone when coupled with other such pieces of information. A first name, a hair color and a description of dress mean nothing on their own, but when combined they can indicate a specific person. There are therefore several types of identifying information, and the more they are coupled with others, the more likely it is that we will more easily identify someone else. Of course, just like beauty, anonymity lies in the eye of the beholder. Even if someone were to have several identifying pieces of information regarding them in plain sight of other, they are anonymous if they find themselves surrounded by people who could not put these pieces together to make a picture (a famous football player is much more likely to be able to go unidentified if they go to a bar filled with basketball fans than if they go to one filled with fans of their team). Likewise, if a specific description is not sufficient for individuals to be identified, or if observers in that specific context can't extrapolate from that specific description, then the individual is unidentifiable. Michael Jordan could be identified through a nametag with "Michael" written on it, as well as through his initials "MJ" or the number 23 that he wore on his uniform, and yet each of these identifying pieces might be more or less effective at making him identifiable depending on the context. In summary, when either no pertinent information is available, when the pieces of information cannot reliably be assembled or when the pieces fit, but do not form a

description the observer can use, then the individual is nonidentifiable, and therefore, they can be anonymous.[8]

Imagine human beings as boxes in post-office conveyor belts. We have an exterior sheet of cardboard, our appearance, full of information. It has a color, stamps, an address number, etc. It is a container for an interior mess of pictures, drawings and texts. We are not speaking of this in the sense of whether character, identity, beliefs or other self-shaping constituents are what form our core,[9] but rather on how the interior contents are, at the simplest level, identifying information. All of our self is therefore a layer of superficial information around our core of personal information, making us who we are. And quite often, the superficial information *on* the box can give us a general idea of the information *in* the box, even if the contents are what is truly determinant in regards to our individual self. If there is a label on the box saying "7 years old", we can guess that the box might have some crayon drawings within it, maybe a picture of a dog and little to no text. But when we are anonymous, it means we have surrounded our box with an opaque plastic wrap. It can be grey, nondescript and blank just as easily as it can be colorful and full of symbols. Still, with this wrap, a boxes' exterior is concealed just as their contents are concealed. Still, in theory, even if it is hidden, the box is still the same it has always been, and its contents should be the same. If some of the contents were to spill out, we could often guess what is inside the box. Still, we can imagine the problem that could arise from all boxes looking the same despite their contents being entirely different. We didn't choose the example of a 7-year-old innocuously; children, more than any other users of the internet, are exposed to things they should never have to see precisely because of how

[8] Ferry, *"Anonymity"* (2002), p. 101

[9] We will develop on this matter in Chapter 3. For now, we will label these as mere preferences of our personality.

the internet facilitates anonymity and free circulation.[10] For now, we'll be content with deriving the following observation from the box analogy: anonymity hides the superficial information of an individual, but it does not erase this superficial information or the personal information behind it. We are still rational agents behind our opaque veil, but when anonymous, the rules in the game of communication change.

1.1.2 A matter of accountability

There is a common theme in fiction where characters conceal their identity in order to do things that otherwise would be impossible. In old myths, gods often use disguises for their ploys, and famous characters in literary works use disguises in similar ways.[11] Some go even farther, and conceal themselves entirely. H. G. Well's *Invisible Man* tells the story of a scientist coming unhinged upon making himself invisible, aiming to use this newfound power to rule Britain, and in J. R. R. Tolkien's *The Hobbit*, once the protagonist manages to slip into the den of a dragon with a ring of invisibility, he has the audacity to play word games with the beast as he loots its treasure hoard. These stories are reflections of power fantasies that people might have, where invisibility would confer us the possibility of acting in a way that we never would have, were we visible. We believe there is a clear parallel to be found here, between disguise and concealment, invisibility and anonymity. These situations endow individuals with a great amount of freedom, as to

[10] Sonia Livingstone. "*Children and the Internet",* Polity (2009)

[11] The Norse god of thunder Thor cross-dressed in order to get into the hall of an enemy giant he sought to slay, and the Greek king of the gods Zeus often transformed into animals in order to seduce unsuspecting maidens. Likewise, Ulysses hides as an old man to deceive the pretenders to his wife upon returning to Ithaca.

observers, their acts seem to be either devoid of agent or disconnected from the agent, and thus provide them with an advantage over others.

Considered the first picaresque novel, *The Life of Lazarillo de Tormes and of his Fortunes and Adversities* marked Spanish literature with its dry wit and crude humor, but most importantly, with its satirical humor that had the Catholic Church as its main victim. The book was put on the Index of Forbidden Books, and the author would have been severely punished, were it not for the fact that the audacious book was unsigned,[12] anonymity thus acting as a safeguard against repression and censorship. Indeed, authors constrained by the limitations of their context and the constraints imposed upon their liberty by others have often resorted to publishing anonymously to get their message through whilst avoiding the consequences of their words.[1314] This shows that, regardless of our intentions when using it, there is an instrumental value in anonymity. Indeed, whilst anonymous, individuals gain a net advantage over others, which expresses itself twofold: they gain increased freedom through unaccountability, and gain a strategic advantage through asymmetrical information.[15]

In the scope of this chapter, we will limit ourselves to working under the assumption that internet users are rational actors, capable of calculating their choices and making them based on their context and their individual preferences, as suggested by rational choice theory.[16] To clarify,

[12] There are however several theories regarding a coded signature inscribed within the book in code, but the spirit of its anonymous release remains significant.

[13] Robert J. Griffin, "*Faces of Anonymity: Anonymous and Pseudonymous Publication, 1600-2000"*, Springer (2003):

[14]

[15] We are aware that there are also disadvantages to being unidentifiable, chief among them being the fact that we cannot rely on any advantages tied to our identity, or claim any merit for what we do when we are anonymous. Still, we believe that, within the online medium, the advantages are especially powerful and convincing.

[16] John Scott, "*Rational choice theory. Understanding contemporary society: theories of the present"*, International Encyclopedia of Social Sciences 2 (2000): 126-138.

we will not adopt rational choice theory and its evaluation of how human beings make their decisions. Rather, this approach allows us to better highlight the particularities in the interactions of users with and within the internet medium, developing on the "how" without shifting too much of the focus onto the "why" of user's actions. Given this, let us analyze the first advantage from the point of view of a rational actor. In a liberal society where all individuals have a set of rights and freedoms, but also a code of law that must be followed, information about what we do is what keeps us in line. From speed radars detecting speeders on a highway to reports on tax fraud, citizens are surrounded by measures employed by governments, companies and other citizens to keep them honest. In theory, we are all equal in that we can all keep each other accountable before the law, and make sure we all follow the rules of society. However, an individual endowed with a veil that hides them from the public eye can evade these measures as, for them to take effect, there is a due process in which reliable evidence must always be found to prove one guilty of their transgression and deserving of the sentence. But what punishment can be given to someone who can't be apprehended or worse, who can't even be connected to the misdeed? It seems, then, that what is at stake here is accountability.

Children learn about accountability at a very young age, most likely as soon as they commit their first offence and are made aware of the gravity of it by first being asked to answer for their act, and then receiving a reprimand or other form of punishment for it. Assuming the parents discipline their child by taking away a toy or depriving them of dessert, the lesson stands: acts have consequences, and individuals can and will be held accountable for not following a rule. From then on, children know that similar occurrences leave them in a situation where someone will be held accountable for the wrongful act. And every time, they have a choice to make, not only regarding whether they accept to be responsible for the act that occurred or not, but whether they let

themselves be found guilty. For instance, a child breaks a vase when they are alone at home. Regardless of whether they admit their fault or try to lie their way out of it, they are aware that they have committed an act that can be traced back to them because they were home when it happened.[17] If they were unlucky, there was a witness or other such evidence linking them to it. But maybe only they know of it, and are free to keep that information to themselves as long as they want to. This means that it is the situation in which they find themselves that determines how likely they are to succeed if they are to tell a lie regarding the accident. In other words, accountability also stems from context.

What, then, if context prevents such accessible questioning? Let us look at Halloween, and the custom that leads children to dress up and go trick or treating. After knocking on your door, children in masks ask for a reward, the absence of which will be punished with the infliction of a trick, a prank (your house will be egged, toilet paper rolls will be thrown over it, etc.). Most importantly, the children responsible will be out of your reach, because they were wearing masks. The situation is therefore one of asymmetry in available information: the children know where you live and what you look like, whereas you know nothing of them, meaning that there is a significant difference in available information. Simply put, anonymity provides a strategic advantage in the game of information, not only because it allows us to conceal our information from others, but because any information we have about others is more than what they have about us. And this leads us to the strategic advantage derived from anonymity.

[17] Talwar, Victoria, and Kang Lee, "*Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception"*, International Journal of Behavioral Development 26.5 (2002): 436-444. [17] Oxford Dictionary.

Strategy can be described as the art of planning, of using the information available to us in order to put together a series of steps and operations that would allow us to succeed in a specific enterprise. The term originates from military vocabulary[17], and history has shown that the victors in warfare are often those with the superior strategy. But this has often been more a matter of information than anything else. Take Napoleon's defeat at Waterloo, for example. The coup de grace dealt to the French Emperor's hopes of restoring his rule over Europe came when he sent his Imperial Guard over a ridge that he expected to be the weak spot of his enemy's lines, only to discover over a thousand British soldiers awaited concealed and ready to mow down his finest soldiers as soon as they came over the ridge. Strategy there simply consisted in the duke of Wellington successfully concealing information from his adversary.[18] His advantage consisted in both that Napoleon's forces didn't have the knowledge that those soldiers were waiting in ambush, and that these soldiers expected that the enemy didn't know it. This information asymmetry is the crux of the advantage that individuals gain when they are anonymous. Not only do identified individuals not know an anonymous individual's information, but by virtue of their ignorance they will always be at a disadvantage, as the anonymous individual will know at least some information about them when they know nothing of the other.

We have established the two main advantages provided by anonymity. Through concealment, an individual can avoid accountability, and therefore gain a great amount of freedom. Anonymity also gives us an advantage over identifiable users in what concerns the access to information, and lets us make more informed decisions than theirs. When anonymous, we gain the possibility of following our personal motivations and preferences. In other words, or decisions will

[18] Andrew Field, *"Waterloo: the French perspective"*, Pen and Sword (2012).

be determined by whether these preferences will restrict us from abusing the advantages that anonymity gives us, or incentivize us to do so. However, as we said earlier, anonymity might also be a matter of perception.

1.1.3 Perceived anonymity vs actual anonymity

In a study on aggressive driving behavior and anonymity, researchers determined people were quicker to blare their horns and do so for longer periods of time when they were concealed by the roofs of their car.[19] This shouldn't be a surprising fact. It isn't unusual for us to see two drivers blaring their horns and hurling expletives at each other during rush hour. Whilst driving people are not only exposed to high amounts of stress, but they are also in a relatively concealed condition. Or rather, they believe they are concealed. Let us use a thought experiment to illustrate.

> Bill is stuck in traffic. Unbeknownst to him, the roof and windows of Bill's car suddenly become one-way mirrors. People from outside can see inside as if there was no roof or windows, but he still sees himself as having an opaque roof and slightly tinted windows.

We cannot realistically expect that Bill will refrain from engaging in his usual road rage, as the change his car has undergone was imperceptible to him. He is no longer concealed, and yet, he will act exactly as he would if he thought he was concealed, until he sees evidence that his concealment has been broken. Until then, Bill is in a state of *perceived* anonymity. So far, we've seen anonymity from a practical point of view, as a property that individuals and messages either

[19] Ellison, P. A., John M. Govern, Herbert L. Petri, & Michael H. Figler, "*Anonymity and aggressive driving behavior: A field study",* Journal of Social Behavior and Personality 10 (1995): 265.

have or don't have. Either the right pieces of information come together in the right context to make someone identifiable, or these same pieces are hidden and the context is muddy enough to conceal their identity. Our true interest, however, is on how individuals act when they *think* they are anonymous, regardless of whether they actually are or not. In other words, how people act when they believe they have the freedom and the advantages provided by anonymity. For now, we will not delve into the psychological implications this may have too deeply, and we will instead focus on the issue of how the benefits anonymity provides affect the choices of rational actors.

We've defined anonymity as non-identifiability. Much like invisibility, anonymity frees individuals from accountability and provides them with a net advantage over users who have are not as thoroughly concealed. We have also determined that it is reasonable to expect that individuals act this way because of the belief that they are anonymous, regardless of whether they are truly non-identifiable. This raises the question of whether they will abuse these advantages or not once they believe themselves to be anonymous. Whilst invisibility is still not technologically achievable, anonymity has become commonplace, more specifically within the internet environment. We will now turn to the internet environment and describe it. We hope to show how it not only makes anonymity very accessible, but also facilitates certain pervasive behaviors in rational actors.

1.2 The internet environment: a descriptive framework

1.2.1 History and characteristics of the internet medium

One can try to set a before and after for the internet, be it the year the first computer was created, or when the first bit-based message was sent. In the light of this study, we would place it in 1989, the year the world wide web was invented, an invention accredited to computer scientist Tim Berners-Lee.[20] We say this because this was the beginning of the internet's rapid domination of nearly all communication worldwide, as it was through the world wide web that the average individual managed to both expand the borders of their personal world and engage with the rest of the world. In the following years, many companies would develop web browsers to grant their customers access to this web, and through it, to an exponentially bigger world, filled with a rapidly increasing number of websites of all kinds. Blogs and forums, followed by social media; video and image sharing boards, followed by streaming platforms; personal messaging systems, videogames and countless other applications that come and go with the trends of the season and the whims of the userbase. This wide variety of sites to pick from, and the ease with which these can be made (most require only a minimum of coding knowledge, and there are plenty of services available for those who want to bypass the coding work) makes the internet a perpetually expanding virtual space. It is a new social environment, or rather, a galaxy of social environments. And through its many forms of Computer Mediated Communication (CMC), we exist online and engage with others. This coexistence is far from being a time-framed, limited thing. The generations born from the nineties and onwards were born into a digital age, and have become deeply entangled with technology in their daily lives, thanks in no small part to the spread of personal computers through the popularization of smartphones (there are now over 3.8 billion smartphone users in the planet).[21]

[20] Berners-Lee, Tim. "Answers for young people." World Wide Web Con (2011).

[21] Newzoo. "Number of smartphone users worldwide from 2016 to 2021 (in billions)." Chart. September 11, 2018. Statista. Accessed October 11, 2019. https://www.statista.com/statistics/330695/number-of-smartphone-usersworldwide/

The smartphone is an encyclopedia one moment, a streamed soap opera the next; it is a tool so versatile that it has become, to all intents and purposes, necessary in the daily life of many. This means that the gateway to the internet is open to us twenty-four hours a day, seven days of the week, at any given place. There is, therefore, a pressing need to better understand the internet world. And to understand what makes this environment unique, we must first understand how it works.

Let us begin by talking about information online. We can say that it is easily accessible, overwhelmingly abundant, relatively credible and ephemeral, but persistent. Let us elaborate. Users not only have easy access to computers, but most search engines and websites are designed to be easily navigable by newcomers. Average users can surf the net with a minimal amount of technical knowledge, and those with more know-how can refine their use to fit their more specific needs and wants.[22] With more than half of the world's population having internet usage,[23] the absence of significant technical hurdles for a basic usage of the internet only makes it more likely to expand into our daily lives. This is demonstrated by the absurdly vast amount of information available online, with an estimated 2.5 quintillion bytes of data being generated each day.[23] This growth is seemingly exponential, and yet most of it is carefully mapped out to be accessible to the average user. Think, for example, of the internet's most used encyclopedia, Wikipedia. Though it is curated and monitored by certain users and the Wikimedia Foundation, it is very accessible to the general public and actually expects individuals to expand it with whatever knowledge they can bring. This is, in a sense, what makes information relatively credible: the plentifulness of sources

[22] Deborah Fallows, "*Search engine users: Internet searchers are confident, satisfied and trusting--but they are also unaware and naïve"*, Pew Internet & American Life Project (2005). [23] Internet World Stats: https://www.internetworldstats.com/stats.htm

[23] Bernard Marr, "*How much data do we create every day? The mind-blowing stats everyone should read"*, Forbes, May 21st (2018).

available for individuals to cross-reference makes it possible to double check a piece of information, or at least expect that it has already been curated (of course, this doesn't mean that people usually go through the trouble of critically evaluating information found online).[24] The internet has consequently become the answer to most of our ordinary daily questions, and, while there is a wide amount of disinformation online, most daily tidbits of information can be expected to be relatively reliable. For instance, the average user can reasonably expect that a trick to catch fruit flies or unclog a drain will probably work if others have validated it. Finally, let us look at the persistence and ephemerality of information online. The idea of something being both persistent and ephemeral might seem a bit counterintuitive, so we'll elucidate with an example. Maria posts a picture of herself on Facebook. After a couple of hours, Maria decides she doesn't like the picture as much as she thought, and decides to take down the picture, effectively removing her post from the platform. Two things have happened here. On the one hand, Maria shared her picture to a limited amount of people, for a limited amount of time. It was, to all intents and purposes, ephemeral. On the other hand, in the space of a few hours, anyone might have taken a screenshot or otherwise saved a copy of Maria's picture. If so, then there is a possibility that this picture will now never leave the internet's repository of information, be it as data stored in the computer memory of a friend of Maria, or lost in the endless flow of information. It would then become a persistent piece of information. To summarize the situation, information online is ephemeral, because it can be put up and taken down at a moment's notice, and given it has no physical support, no trace of it might remain. But it can also be endlessly copied, shared and stored by other users or even by computer algorithms, be it as an individual piece or as metadata linked to a person. The

[24] Miriam J. Metzger, *"Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research"*, Journal of the American Society for Information Science and Technology 58.13 (2007): 2078-2091.

issue of what is done with this information has raised many debates, be it regarding government surveillance[25] or on the evident growth in targeted online marketing in the last few years.[26] These problems are some among the many raised by the way data exists online, and who has a right to what once something becomes virtual. For now, we limit ourselves to saying that once information makes it online, it is very easy for it to be absorbed into the large heap of internet memorabilia.[27]

Having detailed the way information works online, let us now move to user interaction. There are several key aspects of online interaction identified by the literature. Regarding user-touser interaction, we can safely say that people online can connect with each other in a consistently easy way. On the one hand, reaching others literally becomes easier with every new generation of hardware and software. The difference between old switchboard-and-operator phones and our current smartphones is obviously inmense, but even comparing messaging applications from the early 2000's to those that we have now shows a progress in making them more accessible, intuitive and efficient. In other words, the number of steps necessary for us to successfully communicate with one another is reduced with every new technological breakthrough. Which also means it is easier for people to work together online, as communication can help coordination. Email and other such messaging systems have effectively replaced fax and postal services for a lot of day to day activities. Working from one's home, videoconferencing and simultaneously editing documents with one's peers are only some of the many functionalities of current technology that give us great

[25] Susan Landau, "*Making sense from Snowden: What's significant in the NSA surveillance revelations"*, IEEE Security & Privacy 11.4 (2013): 54-63.

[26] It has become evident that companies make use of the information we leave behind as we surf the net to better tailor advertisements, taking full advantage of the fact most users believe it to be ephemeral. See the proceedings of the Grand Committee on Data, Ethics and Democracy held in May 2019 for more details.

[27] There is also a great amount of data that is lost amidst the ever-changing programming structures and technological advances of the internet. For many researchers interested in keeping track of the medium's history, this supposes a significant problem. For more information, see Daniel J. Cohen, "*Digital history: A guide to gathering, preserving, and presenting the past on the web"*, University of Pennsylvania (2005).

freedom to choose the ideal way to coordinate with others. And as we said earlier, a great part of our online activity relies on others and their input, as all of the information online has been generated by a user in one way or another.

On the other hand, being online removes physical and social barriers. Indeed, CMC makes interactions between people who already know each other but are separated by long distances more frequent, and interactions between complete strangers commonplace. Two people that passed by each other in the street or rode on the same bus together and never even dreamed of starting a conversation in that context may inadvertently begin talking in a forum about a tv show that they both like. When we are online, through the creation of profiles and accounts, we often curate our appearance and interests to match the websites we find ourselves in. A fifty-six-year-old, divorced father of three may prioritize certain elements of his self-presentation depending on whether he is visiting the official fan website of his football team, an online dating site or a discussion forum on tv shows from the eighties. Website communities establish guidelines and codes of conduct, and usually set the bar for how much information must be shared by users at the get-go, but users ultimately end up having the choice of how much information is actually shared. Which raises the question: why do people conceal their information when online?

1.2.2 Intended use versus emerging use of anonymity

One would expect your average internet user to be someone simply seeing it as a tool. They can use it to talk with a friend, play a videogame, share a picture or look up a review for a restaurant. These are simple tasks, with usually simple goals, that are a far cry away from necessitating an anonymized state. There are, however, clear merits to both anonymity and identifiability online.

Alongside his scale of identifiability, Gary Marx provides a good summary of both sides of the arguments for and against anonymity online. Anonymity allows for the free, unburdened circulation of information. This is a main tenet of the internet, one its users often go to great lengths to protect. It also allows for the protection of one's identity and personal information, be this in regards to privacy or overall security. Not only that, but with anonymity individuals can exist online without the fear of prejudice. Finally, it favors certain kinds social interactions and facilitates the formation of bonds between people who would otherwise be impeded by social and cultural barriers.[28] All of these merits can be derived from the central idea of freedom, so we can state that the main merit of online anonymity is the freedom it provides. This is consistent with what we have seen previously.

On the contrary, identifiability can be perceived as limiting freedom, but it guarantees the accountability of users, facilitates certain logistical aspects of online coexistence and also provides certain safeguards for personal information.[29] Therefore, the merit of identifiability lies in that it allows us to keep tabs on others whilst online. Indeed, identifiability serves more as a dissuasive force, a way to deter individuals from acting in irresponsible ways. It seems, then, that not much is different between real-life anonymity and virtual anonymity in what concerns the basic problem they bring forward, namely the large amount of freedom and the advantage that they give to people. It would stand to reason that those using anonymity online would either highly value the aforementioned advantages of anonymity or would not value the benefits of identifiability highly enough to consider identifying themselves. The intended use of online anonymity, then, can be boiled down to three main purposes:

[28] Marx, "*What´s in a Name"* (1999), 103
[29] Ibid. 105

1. *As a safeguard that preserves our privacy and information.*
2. *As a social lubricant that allows us to interact with others without prejudice.*
3. *As a key to free, unconstrained circulation and expression within the internet space.* [30]

Anonymity would therefore, in theory, be a careful, calculated choice, done in order to maximize the benefits we reap from our online experiences. People seem to value their privacy and the protection of their personal data, sometimes to the point of going the extra mile for its protection;[31] anonymity provides people with a lot of social "wiggle room", with examples of this going from women who want to be able to frequent gaming circles without being discriminated[32] to forums for personal self-disclosure regarding deep emotional troubles people might be reluctant to share with those close to them;[33] and while many enjoy making loud, highly identifiable statements on Twitter, it seems a significant portion of Twitter users prefer to do so anonymously, or limit themselves to "lurk" and simply observe.[34] However, while these practices are relatively coherent with the intended use of anonymity, much of the actual use of anonymity online seems to differ significantly from the intended use in two main ways. Firstly, a great many people exist online whilst being relatively unidentified for no particular reason. While it is evident that many people exist online with rather transparent and maybe even careless amounts of personal information

[30] Deborah G. Johnson, "*Ethics online"*, Communications of the ACM 40.1 (1997): 60-65.
[31] Kim B. Sheehan, "*Toward a typology of Internet users and online privacy concerns"*, The Information Society 18.1 (2002): 21-32.
[32] Gray, Kishonna, and Wanju Huang, "*More than addiction: Examining the role of anonymity, endless narrative, and socialization in prolonged gaming and instant messaging practices",* Journal of Comparative Research in Anthropology & Sociology 6.1 (2015).
[33] Sabina Misoch, *"Stranger on the internet: Online self-disclosure and the role of visual anonymity."*, Computers in Human Behavior 48 (2015): 535-541.
[34] Peddinti, Sai Teja, Keith W. Ross, and Justin Cappos, *"On the internet, nobody knows you're a dog: A Twitter case study of anonymity in social networks."*, Proceedings of the second ACM conference on Online social networks (2014).

exposed,[35] it seems that it is also the case that people exist online anonymously without a very clear reason for concealing their identity.[36] We share in this intuition after collecting data on the profiles of YouTube commenters on the ten most watched videos of the site's trending page for a month on a daily basis. We found that only a few could be considered highly identifiable by our standards.[37] It seems reasonable to assume, given the rather innocuous nature of most comments, that a portion of users browse websites anonymously even though they don't technically *need* to.

When browsing websites, we rarely identify ourselves unless prompted to do so, by creating an account for example. Even when we create this account, most websites don't demand more than an email and a password, thus making it very easy for individuals to proceed with information that is practically irrelevant to other users; the name of a stranger will usually mean nothing to us, it is empty information. Again, anonymity is in the eye of the beholder, and most people online are not looking to actively identify other users while they browse because they have other objectives in mind. On top of that, fabricating information or providing false information in our profile is not only easy, but intuitive. Most individuals seem to value their privacy, and so are naturally skeptical of sharing their personal information when no functional reason is provided as a justification.[38] Let us think, for example, of a social media website like Facebook versus a videosharing platform like YouTube. Facebook is a platform for sharing life experiences with

[35] Gross, Ralph, and Alessandro Acquisti, *"Information revelation and privacy in online social networks.",* Proceedings of the 2005 ACM workshop on Privacy in the electronic society (2005).
[36] Mislove, Alan, Bimal Viswatah, Krishna P. Gummadi & Peter Druschel, *"You are who you know: inferring user profiles in online social networks."*, Proceedings of the third ACM international conference on Web search and data mining (2010).
[37] We compiled this on a daily basis between June 11, 2018 and July 21, 2018. We classified users as highly identifiable, partially identifiable and nonidentifiable. We found that 60% of the 1521 comments posted were unidentifiable by our standards (the profile picture and username provided no direct link to the person's personal information).
[38] Sheehan, *"Toward"* (2002), 25-27

acquaintances and keeping up with the lives of others, so it fundamentally requires that one give their personal information. When we enter Facebook, we expect others to have shared their information, and are to a great extent invited to share because of the website's dynamic. Meanwhile, within a video-streaming platform like YouTube, there is little to no incentive to share our information with others. The main purpose of the platform is entertainment, and the vast majority of the content, be it the videos themselves or the comments of other users, is produced by people we do not know and will most likely never meet in person. Not only do we not have the expectation that others will share their personal information without a good reason, but we have no pertinent reason to do it ourselves unless we are personally motivated to do so. It must be added to this that, whilst online, our interactions with the internet environment leave indelible prints of metadata (search histories, cookies, IP number tags, etc.). Users online are therefore not only being watched from all angles, but all they do might be recorded, which leads to growing concern and desire among the average user to protect their data.[39] As a consequence, when people enter these websites, they often do so in a state of low identifiability or even anonymity not because they rationally chose this as the optimal option, but because the alternative of identifying themselves is not intuitive to them. In other words, people are often anonymous online not because they want to be anonymous for the advantages it gives them, but because they see no reason to share their information. In a sense, they are in a state of coincidental anonymity.

Secondly, and most importantly, many people use anonymity because of all its advantages, but not according to its three main purposes. Rather, they engage in what we will consider to be problematic acts. We will develop a more detailed profile of what we consider to be ethical acts

[39] Feigenbaum, Joan, and Bryan Ford, *"Seeking anonymity in an internet panopticon."*, arXiv preprint arXiv:1312.5307 (2013).

online in the next section of this chapter, but for now, suffice to say that a problematic act is fundamentally one that is mostly focused on our own preferences, even if these come to disregard or harm others. In this sense, there are three different, emerging uses of anonymity:

1. *As a safeguard against retaliation for our acts and words.*
2. *As a social barrier that allows us to interact with others based exclusively on our personal choice.*
3. *As a key to free, unconstrained circulation and expression within the internet space.*

As is evident upon comparison, reasons 1 and 2 have changed into twisted corollaries of their original, whereas reason 3 remains the same. I will now develop on why this is. The first use follows from our earlier thoughts on invisibility and accountability. Being truly anonymous and undetectable online is extremely hard from a technical point of view,[40] but as we have said, what is truly important is the belief of anonymity. And if individuals believe that they are truly untraceable, and more specifically, that those who would have a problem with their actions cannot trace them, then they are wholly incentivized to continue acting in that way.

The second use essentially means that being online often entails presenting a set of information of our choosing. This means we are inevitably putting on a mask, be it one that we hope to describe us as accurately as possible or as favorably as possible. The degree to which our online-self differs from our offline-self is something we determine through our actions, and so we are also given free rein on how much our acts and words differ from those that would be expected of us on a day-to-day basis. Even if this chapter will not delve into the psychological implications of this, we will indicate that, for now, we are working under the assumption that individuals have a certain degree of coherence between their online and offline identities, and so when they engage

[40] Kang, Ruogu, Stephanie Brown, and Sara Kiesler. *"Why do people seek anonymity on the internet?: informing policy and design."*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2013).

in cyber-behavior that is incompatible with their offline preferences, then it most likely involved a rational choice to do so.

The third reason, namely the free expression and circulation of users, remains the same because it is fundamentally a double-edged sword. This evident problem is very much the same in the real world: free speech entails that the extremist has the same right to voice their opinion as the moderate, and that the abuser has the same right to be heard as the abused. The debates on what constitutes hate speech or what should be the legal consequences for certain speech-acts only elicit how complex this issue is, and how difficult it is to try and preserve free speech while protecting citizens from abuse, and to avoid limiting free speech to the point of undermining liberty.[41] Of course, being online adds the extra layer of complexity that comes with the endless possibilities of virtual existence.

In a sense, the three reasons complement each other, and make each other possible. Without the guarantee of privacy, there is no freedom in our circulation and expression. Without freedom, there is no possibility of having a truly personal choice in how we present ourselves online. Without a personal choice in what we share, there is no privacy to our information. These necessary qualities of the internet also enable their antinomies, and impeding one necessarily impedes the rest. In other words, if we want to keep anonymity for its good uses, then we have to accept the possibility of its bad uses.

"On the internet, nobody knows you're a dog." The famous catchphrase, taken from a comic strip from 1993 (picturing a dog sitting in front of a computer talking to another dog beside it), came to embody the sentiment of existing online. Users exist as sets of data, and all they create, exchange

[41] Cass Sunstein, *"Democracy and the problem of free speech.",* Publishing Research Quarterly 11.4 (1995): 58-72.

and find is data made by them or others. People seemingly always have a choice regarding what information they are going to share, or not share, with others, but the amount of control they have over this information is deceptive. Anonymity serves similar purposes online as it does offline. However, we have indicated two things that are novel about anonymity online: people can be anonymous whilst being relatively unaware of what it entails, and people can abuse the fact that the main tenets of online anonymity are grounded on socially acceptable reasons. We will now focus on the way users can act unethically, as rational actors embedded in an online structure.

1.3 Challenges to social norms online

1.3.1 Computer Mediated Misbehavior

The common understanding one might have of social norms is that they are rules by which we can guide ourselves to be more socially-minded and respectful of others. Day to day examples that could come to mind would be helping a mother carry her baby stroller up some stairs, or not littering. Of course, these social norms are often informed by codes of ethics. So far, in the scope of this thesis, we have not aimed at giving a strong, closed definition of ethics and what it consists of, and we prefer that the definition be kept relatively open and general for now, choosing instead to use the optic of social norms to judge the problematic aspects of online behavior.[42] After all, in the scope of this chapter we are not interested in delving too deep into moral considerations, but focus on judging the acts in the daily lives of average internet users. These will not have extensive knowledge of theories in ethics, but will most likely have a set of guidelines and teachings they follow, such as "don't litter" and "help those in need", social and moral intuitions that are based

[42] We will provide a more detailed account of morality and ethics once we reach the sections of this thesis concerned with moral psychology, in chapters 2 and 3.

on ethical standards given to them by the societal context surrounding them. We will therefore need to define problematic actions a little more carefully, as they constitute the core of this chapter's interest. Problematic actions can either be acts that are actively problematic because of their implications for other people (littering) or because they ignore a problematic situation (not helping the mother in need). On the one hand, we can consider acts that are selfish and selfcentered, made with complete disregard of others and even disrespecting their basic rights, freedoms and dignity to be problematic. When an individual performs such an act, we will consider it an actively problematic act. An example of this would be a husband who physically and verbally abuses his wife. On the other hand, there are acts that are problematic in their disregard in the face of what is socially acceptable. These might be considered less grave than actively problematic acts, but they are far from being innocuous. When an individual ignores the expectations of others when performing an act, we will call it a passively problematic act. An example of this would be the silent complicity of a neighbor who sees and hears the wife being abused and doesn't report it to the authorities.[43]

The internet has its own share of social problems, be they problems that existed in the real world long before the internet and that were transported into the online environment, still essentially the same but with new "twists", or entirely new forms of problematic behavior. As we stated earlier in the chapter, supposing individuals aren't held back by their personal convictions, they usually refrain from committing problematic acts in fear of repercussions, unless they have a way to avoid them. We considered anonymity as one of these ways, but there are other online

[43] We want to avoid embracing a specific theory of ethics in the scope of this chapter. However, it is evident that our description of "actively problematic" and "passively problematic" acts is essentially a reiteration of Kant's perfect and imperfect duties, which remains a relatively intuitive moral description and distinction of problematic acts.

variables in play that must be considered if we are to properly gauge the weight of anonymity in this question. The main competing variable we've identified so far is Computer Mediated Communication. Indeed, when we engage with others, there is a clear and evident distance between us and them. This is manifested by the physical distance separating our two computers, but also by the temporal distance between our interactions and the emotional distance caused by the absence of visual cues.[44] And just as children doing phone-call pranks rely on the distance between them and their victim to be able to do what they do, those committing offenses online rely on the fact that no-one will bother finding them. This is, in a sense, a feeling of "unreachability", where individuals believe that they can get away with their actions because they are far from the consequences of the latter (again, this seems like a matter purely of perception, as one can easily see how these behaviors can be traced to the offender).[45] This constitutes a problem of accountability, one that is similar and yet distinct from the one that is generated by anonymity, and that we should take into consideration going forward, in order to avoid conflating the effects of mediated communication with those of anonymity.

Take, for example, the phenomenon of cyberbullying. While there is no consensus on the proper definition of what constitutes cyberbullying online, we can gather from the many definitions provided that cyberbullying consists in computer-mediated verbal abuse, which can be constant and repeated.[46] There is a consensus, however, regarding the fact that this new form of bullying has grown to complement traditional bullying to worrying degrees, as people bring home abuse that was once limited to the schoolyard or the workplace. This is confirmed by plenty of statistical

[44] Joseph B. Walther, *"Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction."*, Communication research 23.1 (1996): 3-43.

[45] We will develop on the more psychological implications of this in chapter 2.

[46] Wingate, V. Skye, Jessy A. Minney, and Rosanna E. Guadagno, *"Sticks and stones may break your bones, but words will always hurt you: A review of cyberbullying."* Social Influence 8.2-3 (2013): 87-106, 90

evidence that goes from surveys on the self-image of students to data on teenage suicide rates.[47] Cyberbullying is, therefore, a perfect example of problematic, unethical behavior online that would benefit from anonymity and even be incentivized by it. And yet, it doesn't necessarily require anonymity. What it does require, however, is for the communication to be mediated through the cyberspace, as this allows for the harassment to be done at any given time, and become indelibly engraved online. Notable examples are students posting offensive remarks regarding their peers on social media, spreading privately sent pictures or messages to the web or even actual hacking and other sorts of technical meddling with their accounts (this is not to say that all cyberbullying needs to have a public side to it, but rather that the kind that does is usually the one that is performed by identifiable bullies). It is often obvious who has committed the offense, sometimes to the point of authorship being explicitly declared, so anonymity in its practical sense is lost.[48] However, one could argue that the feeling of anonymity is still embedded in cyberbullying. Indeed, Wingate, Minney and Guadagno believe cyberbullying is fueled by the way the internet environment facilitates a feeling of distance from other users, in part through anonymity. The chasm created by the online borders would render most observers indifferent to the abuse, as opposed to the complicity students share when they ignore obvious cases of bullying in the schoolyard. To quote:

> "If the individuals engaging in online interaction feel anonymous, they are not thinking about the repercussions of their behaviors online, and are not feeling any uncomfortable tension regarding their Internet expression, then they will most likely continue that normative group behavior. This would suggest that

[47] Hinduja, Sameer, and Justin W. Patchin, *"Bullying, cyberbullying, and suicide."*, Archives of suicide research 14.3 (2010): 206-221.
[48] Agatston, Patricia, Robin Kowalski, and Susan Limber. *"Youth views on cyberbullying."*, Cyberbullying prevention and response, Routledge (2012): 68-82. [50] Wingate et al., "Sticks and stones", 92-93

> the bullying in cyberspace is perpetuated by the environment itself, and its unique characteristics exacerbate the prevalence of cyberbullying."[50]

We can here begin to see the differences and similarities between anonymity and CMC as variables in the behavior of users. While CMC doesn't necessarily imply a lack of information, it does generate a sense of distance and impunity in perpetrators, much like anonymity does. In a sense, anonymity online is always dependent on mediated communication to give users the sense of being truly separated from our own personal information, and in some cases of cyberbullying, anonymity serves not as a tool for the bully, but as a tinted screen for the passive observers to distance themselves from the problematic act.

A completely different, but altogether pertinent variable is the varying semantic value of information online. As we've addressed earlier, when online, everything is both effectively ephemeral and potentially persistent. This in turn not only means that interactions between individuals are necessarily different, but that the nature of these bits of information is itself unique. The way users communicate online is in perpetual shift, with memetic references continually being recycled and repurposed to borderline cryptic degrees. To grasp how this can become problematic, one can simply look at the tone that humor has taken in certain online circles and subcultures, where parody videos depicting the 9/11 attacks, jokes regarding suicide and unabashedly racist lingo circulate freely, as if these things were unattached from their historical significance or worse, derived their value from making light of this significance. This is indicative of the way internet subcultures have developed in parallel to the offline world, and this distance might complicate the task of producing informed moral evaluations of the behavior of online actors. It should not be forgotten that, more often than not, these are done from behind an anonymous façade, but this

dissonance in meaning and understanding has a significance all its own.[49] This constant attribution and reattribution of sense to online lingo means that it is hard to determine whether apparently problematic declarations or acts have any actual ill intent behind them. For instance, given how marginal openly racist movements are in democratic societies nowadays, when the most famous Youtuber in the world is consistently labeled as a racist by media outlets despite being adored by his racially-diverse, worldwide legion of fans,[50] it is safe to assume that something was lost in the translation between virtual and physical world. Of course, this semantic disconnection between words and their real-world weight also indicates a certain degree of emotional disconnection from the real world.

We have seen that problematic behaviors online do not necessarily stem from anonymity, as other variables, such as CMC, seem to be in play. Having taken them into consideration, we could say that problematic online activity is usually the result of a complex convergence of factors. And yet, the feeling of anonymity seems to seep through the medium. We will now, through the character of the internet troll, show how online anonymity has a significant, unique role in shaping the online environment.

1.3.2 Trolling as a case study of problematic practices

Whilst the origin of the word's use in the online dialects is often attributed to the Norse mythological creature with which it shares a name, the more likely origin is derived from the act of "trolling", fishing by trailing a baited hook from a moving boat.[53] An online troll engages with

[49] Whitney, Phillips *"This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture."*, Mit Press (2015).
[50] Fägersten, Kristy Beers, *"The role of swearing in creating an online persona: The case of YouTuber PewDiePie.",* Discourse, context & media 18 (2017): 1-10. [53] Merriam Webster Dictionary

other users by deceptively luring them into dialogue in which the end goal is to upset, confuse and generally make a fool out of their victim. More often than not, this is done by using provocative declarations, purposefully botched arguments or nonsensical claims. Trolls have a particular taste for sensitive topics, like political opinions or religious matters, and usually prowl discussion forums and other such chat-based websites where users most often lose their cool and let things become personal whilst exchanging messages. In this final section of the chapter, we hope to show that trolls are one of the most telling signs of the problem of anonymity online.

Whitney Phillips provides an in-depth analysis of the troll, from how the practice of trolling emerged in the early days of the internet to how it became a widespread subcultural phenomenon, showing how the practice was deeply tied to the internet's structure. Trolls, she says, fashion themselves as trickster figures, who expose the inconsistencies in the arguments and ideas of others, and just generally thrive in sowing disorder and confusion.[51] This sense of superiority above the formless mass of internet users makes trolls usually behave in a rather egotistical manner, trying to outperform each other and go ever farther in the ludicrousness of their "feats".

Whitney also argues that trolls depend on the internet's social structures much like a parasite depends on its host. Without the influx of content and users, trolls would not have victims to pursue. And while trolling and cyberbullying are often conflated, it must be noted that there is a fundamental distinction between the two: indifference. Indeed, while cyberbullying is a highly targeted and specific form of abuse, trolling is aimless and scattered, indifferent in who its victim is; while cyberbullying is often highly personal, trolling is rather impersonal.[55] Trolls do not seek to torment one specific individual because of personal gripes they have with them in real life, but

[51] Phillips, *"This is why"* (2015): 9-10, 23-26, 33-35
[55] Ibid. 27-32

rather because doing so would result in something they would deem entertaining. Phillips, along with most of the literature on trolling, has come to argue that traditional trolls are fundamentally reliant on anonymity and its emerging uses to exist online. Indeed, the possibility to hide one's true identity means that there is little to no way to hold users responsible for or consistent with their acts, which in turn means that we can't reliably know whether or not other users are being honest when they exchange with us. This a perfect situation for trolls, as by their very nature, trolls do not aim to be consistent with their position, because they do not believe in it in the first place. Their position is often simply an exaggerated or overblown reflection of their victim's views. A troll aiming to rile up an alt-right enthusiast will use liberal rhetoric, and vice-versa.[52]

On top of that, mediated communication means there are several degrees of separation between us and other users, and anonymity means we can't see who is who. Drawing a clear line of division between other users is practically impossible when we are online, especially if most other users are nonidentifiable to us. The opacity of the internet allows for users to easily meld themselves into a crowd and separating trolls from the rest in a reliable manner is practically impossible. This makes such environments into the ideal hunting ground for trolls. One could even say trolling is only possible because online communication is mediated. The careful planning that goes into a troll's game requires time and distance, and the lies that will be fed to the victim can be made much more believable. On top of that, users might be using the same language, but it doesn't mean the words have the same meaning to all those involved. Trolls take the semantic dissonance to the extreme, as they communicate using a channel and a language all their own, that

[52] It might also be interesting to consider Harry G. Frankfurt´s take on "bullshit" as a defending a position in which we do not really believe. See Harry G. Frankfurt, "*On bullshit*.", Princeton University Press (2009).

functions in a parasitical manner to any other semantic channel.[53] What we mean by this is that in an online conversation led in English, words have different meanings for the troll and the victim, and thus there is an asymmetry in understanding between them as only the trolls are "in" on the trick. For example, trolls often engage with the goal of getting a single word with memetic significance or any other such reference to their subculture to seamlessly fit into the debate. Getting their victim to say the word themselves is often akin to a checkmate in their rhetorical game (as Phillips shows with the example how a troll got reality TV host Oprah Winfrey to say "over ninethousand penises").[54]

What we have hoped to show through this case study is not that trolls are the villains of the online world, or that they bear more responsibility for their problematic acts than other users. It would be a stretch to consider that some misguided individuals that con other users for their own entertainment to be responsible for the online environment's oversaturation of problematic practices. But they embody what happens when one uses anonymity to its full extent, and how it allows for problematic behaviors to develop and grow unchecked. On the one hand, trolls instill a sense of doubt, a sort of wariness and skepticism on internet users, especially if they have already been victims of trolling. Like a city known for its con artists will have a tourist doubting every person that comes to them with an elaborate plight or request, the internet is a comfortable nest for trolling. Users become trapped in a dilemma, where every exchange they have online could be a trap where they invest their time and emotional investment on a fake person, a ruse set up by an individual completely unengaged with the issue at hand who only wants to ridicule them. On the other hand, they themselves become entangled in this self-justifying crusade against the general

[53] Phillips, *"This is why"* (2015): 61-69
[54] Ibid.

public, seeing in their online activities a righteous and yet unpretentious humbling of other online users. And the best part is that, according to trolls themselves, it is all play, a hobby that can be put on hold and stopped as soon as one steps out of the virtual world.[55] Anonymity, in this sense, not only gives them freedom from constraints and accountability. It gives them freedom from their daily life, and distance from whatever standard of morality they have in it.

Conclusion: Rings of power

Over the course of this chapter, we tackled the concept of anonymity, seeing that it is essentially a matter of nonidentifiability in the right context. We aimed to show that, from a purely rational point of view, anonymity has a series of merits; nonetheless, these can be repurposed to our favor and become clear advantages over our peers. In theory, by giving us an increased freedom and an upper hand in the game of information, they leave us with a rationally valid option of ignoring ethics, our choice then becoming a personal matter of preference. We also noted that, just as actual anonymity, the feeling of anonymity ought to produce the same results in people as actual anonymity. We then transitioned to the online environment, and found that the dynamics of anonymity remain essentially the same apart from two things, namely that the anonymity of certain users is not a calculated, but rather coincidental one, and that other variables run parallel with anonymity online. We finished by analyzing the internet troll, and using this unique internet denizen to show that the nature of online environments allows for the development of problematic behavior through its widespread and facilitated access to anonymity. However, over the course of

[55] Ibid. 21-23

the chapter, it became harder and harder to ignore that there is something about problematic acts covered by anonymity that goes beyond rational choice. Can problematic behavior online simply be reduced to the rational choices of rational actors unconcerned by ethics?

Earlier in this chapter we made a reference to *The Hobbit* of J. R. R. Tolkien, and of its magical ring that turned its wearer invisible. However, we did not mention that this ring not only concealed its user, but also whispered in the ear of its user, slowly corrupting their heart. We did not choose the example randomly; indeed, it seems that something quite similar happens when individuals are provided with anonymity. Even if anonymity is only one among several factors that can lead individuals to behave in problematic ways online, it is undeniably the one variable that endows users with a newfound power that is most certain to affect their choices more than just as a shift in the advantages calculated by a rational actor. Namely, when anonymous, people's actions are not only changed because they are out of the range of consequences, but because being anonymous in turn changes their perception of themselves and everything surrounding them.

Chapter 2: Lost in the bit crowd: situational forces and socialvirtual normativity

In the *Republic* of Plato, Glaucon shares the story of the King Gyges of Lydia, and of his mystical ring that made its wearer invisible. The tale went that his ancestor, a humble shepherd, quickly changed his ways upon finding the ring by chance, using its power to seduce the queen of Lydia, murder the king and take the throne for himself. Glaucon then wondered whether a just man and an unjust man, if provided with such a ring, would act differently. Glaucon finds it unlikely, his reasoning being that, when given such overwhelming power, it would be too much to expect for even the most noble person to not use it in their own advantage:

> "No man would keep his hands off what was not his own when he could safely take what he liked out of the market, or go into houses and lie with any one at his pleasure, or kill or release from prison whom he would, and in all respects be like a God among men."[56]

As Glaucon suggests, we may wonder why would an invisible individual be bound by the laws of those who can't see him, much less apprehend him? Likewise, we've seen that anonymity provides us with clear advantages over others, and that there is often little to no reason beyond our own dispositions and preferences to refrain from fully employing these advantages in a variety of online situations.

[56] Asscher, Sue, and Widger David. "Plato-The Republic." Project Gutenberg. https://www. gutenberg. org/files/1497/1497-h/1497-h. htm. Retrieved on 2 (2008): 2016.

Plato's answer in the dialogue is articulated in Socrates' moral evaluation of the man who would use such a ring for evil as a lowly man falling to the temptations of his lower instincts, saying that a man with the rational capacity of holding himself back would know that not abusing the ring's power is the only key to being truly happy.[57] The problem with this answer is that it relies on the preferences of the user, thus not straying too far from rational choice theory in how close-knit it is around the individual and their personal choices. And if the user is being perfectly rational in their calculation of benefit, then problematic choices might often be seen as insignificant or even unjustified in the eyes of their preferences. This then becomes a matter of how much the moral preferences of someone continue to influence their decisions when they are provided with anonymity.

In the previous chapter we avoided the question of morality, choosing instead to view actions as purely rational choices based on preferences and uninfluenced by other forces. We did not do this because we believe this is the way things are, but because we wished to give a description of anonymity and of the internet structures that modulate it without losing ourselves within ethical minutiae. In a sense, our main goal was not to ask the "why" of people's behavior online, but rather the "how" of this behavior, to develop an understanding of the ways in which anonymity changes the conditions of the social game online. We looked at the objective ways in which it provided an advantage, and how rational actors could take advantage of these to perform problematic acts. However, if we are to answer the "why", the previous chapter did provide some insights.

[57] Ibid.

If humans were purely rational actors, acting like unemotional and amoral robots, then their behavior while anonymous would become a matter of pure cost-benefit calculation and personal preference not unlike what we were describing before. If so, then the shepherd gets to decide what he wants to do with his newfound power, and what he decides is a merely reflection of his own personal preferences, which can be calculated or estimated and then used to predict behavior. Most importantly, this entails that these decisions can all be traced back to a conscious process of decision-making. If this was the case, problematic acts committed by an anonymous person would essentially be problems in a vacuum, case by case studies of the rational choice of one individual determining whether they will abuse or not abuse their power, as it depends entirely on their preference. However, in the same way that it seems odd to think that a good family man can become a tyrant after acquiring power solely and immediately out of a rational calculation of personal benefit, it would be odd to consider that the plethora of problematic behaviors online stemming from anonymity is the product of a preference emitted by a rational calculation. There are two reasons for this. Firstly, there is the barrier preventing the shepherd from becoming a selfish tyrant, namely, the man's preexisting set of moral beliefs or preferences. Glaucon believes this would systematically fail because these moral norms only hold as dissuasive forces, and that the absence of punishment removes the need to be moral.[58] Plato believes that if the individual was truly rational and listened to their higher functions instead of the baser ones, their moral beliefs would succeed. Evaluating problematic behavior online through this lens would mean developing a full picture of the moral composition of individuals and how it can endure the rational calculation of benefit through the rational evaluation of the moral value of their acts.

[58] Ibid.

However, there is a second element that must be considered, and that is the fact that an invisible shepherd and visible shepherd, who should have the same preferences, might act differently when conditions are changed despite these preferences. In other words, given the shepherd does not make his decision in a vacuum, there are forces at play involved in his decision beyond the calculation of personal benefit and the moral evaluation of the action. If we are to truly understand the conflict of personal preferences and gauge the balance of power between the selfish preference and the moral compass of the shepherd, we will first need to understand the other variables at play that can tip the scales. In this second chapter, we will analyze anonymity as a variable supported by and coordinating with several others in affecting our behavior. Firstly, we will make a review of the debate in the literature within philosophy and psychology regarding problematic behavior to better develop an understanding of CMC as a situational force that can coordinate with anonymity. We will then analyze the situational forces of a more social nature, such as crowds and groups, and how these translate online. Thirdly, we will list the forces that seem unique to the online environment, how they complement anonymity online and how they can bleed into the real world, finishing the chapter with some concrete examples.

2.1 Deindividuation and other monsters: the psychological effects of anonymity

2.1.1 Rational morality meets the situation

Most western philosophical theories from the Enlightenment era leading up to the twentieth century fostered a conception of man that was highly dependent on rationality. This was a logical thing, as scientific method and discovery paved the way for an ideal centered around the human as a rational animal, an increasingly more independent creature capable of making its own choices based

entirely on the exemplar faculty of reasoning, not unlike a finely-tuned machine. Rational choices led to proper decisions, much unlike irrational or emotionally-driven ones, and any contesting this usually found their views belittled to poetry without substance.[59] While there was interest in the irrational inconsistencies that could be detected in general human behavior through the lens of psychoanalysis,[60] and despite how incomplete medical knowledge was regarding the human brain, the mechanics of its cognition and the problems that could afflict it,[61] the consensus in an era striving for continuous scientific progress was to value rationality above all things, and to hold educated human beings as the highest standard. After all, our choices as individuals were both the exercise of our liberty and the fruit of rational thought and decision, and so we were not only completely responsible for our actions, but we were the main driving force behind them.[62] Logic could always be found within every action, and every decision could be explained by interest or preference, even within the major ethical theories of the time. For Immanuel Kant, a moral act was necessarily the product of careful consideration on the part of an individual. It was the following of a moral rule apt to resolve the situation at hand, but a rule unconcerned with consequences.[67] Likewise, Bentham and Mill's utilitarian theories all conceived of human beings as distinct actors,

[59] Bertrand Russell's account of the history of western philosophy can be read as a history of rationalism's rise to power and consequent dominance of the philosophical field for over three centuries. Occasional rogues that favored a more emotional approach, such as Rousseau and Nietzsche, can be found scattered throughout it. See Bertrand Russell, *"History of western philosophy: Collectors edition."*, Routledge (2013).

[60] While the works of Sigmund Freud and Carl Jung are often disputed and disavowed by psychologists given their rather philosophical, pen and paper nature, it is undeniable that it was a catalyst that sparked interest in the inner workings of the human mind and, ultimately, might have led to an interest in developing a more scientific approach to psychology.

[61] Neurobiology's great breakthroughs as an independent science happened in the second half of the twentieth century. For a fascinating description of this moment, see the introduction of Eric Kandel, *"In Search of Memory: The Emergence of a New Science of Mind"*, WW Norton & Company (2007).

[62] It is interesting to note how visions as distant in time and subject as John Locke's political philosophy and Sartre's existentialism present individual decisions as a rational choice for which we must be held fully responsible.

[67] Kant's theory, as presented in his "Critique of Practical Reason", should be understood as the face of the monolith of moral rationalism we will refer to later in this chapter.

capable of calculating the consequences of their acts and making their decisions with a mindfulness of whether they benefitted the majority or not, and consequently determining them to be moral or not.[63] Morality was the rationally expected outcome of the careful consideration of reasonable beings, and so a failure to be moral could only happen when one was either acting against reason, therefore being irrational and maybe even insane,[69] or willingly ignoring their moral dictates after evaluating them rationally.[64]

However, over the course of the twentieth century, this conception of the rational individual met its greatest challenge in historical developments that showed that many philosophers had either highly overestimated the human capacity to be rational regardless of our context, or underestimated how easily individuals bend when under the influence of situations. The most terrifying signs of this were the ethnic cleansings, mass murders and genocides that marked the century, many of which occurred within the same societies that had birthed the ethical theories meant to promote and ensure social behavior. The possibility of these tragedies being the result of the choices of a limited number of "bad apples" seemed implausible given how widespread they were, and a clear weakness in the explanatory power of these theories demanded a shift in perspective. Of course, philosophers, anthropologists and sociologists alike rose to the challenge, and soon found that they struggled to understand how people so easily strayed from what should be morally evident results of rational calculation without taking into consideration the possibility of certain exterior variables.

---

[63] While putting Mill and Bentham in the same bag as Kant might be oversimplifying the complexity of their respective theories, it remains a fact that the common thread in their theory is very much that valid moral decisions cannot be based on emotion or intuition and require a rational evaluation to truly be moral. [69] Interestingly, the intuition that people with psychological pathologies cannot be held responsible for their actions in the same way a rational person can was a very early intuition that most moral philosophers incorporated into their theories. For more details, see Shoshana Felman, *"Madness and philosophy or literature's reason."*, Yale French Studies 52 (1975): 206-228.

[64] After all, in rational choice theory, choosing the moral option is merely the result of a calculation of preference leading us to subjectively prefer the moral option.

Let us take Hannah Arendt as an example. The German political theorist proposed that the individuals engaged in what she described as acts of evil were not fully aware of their significance. These acts were reduced in importance and strategically played down by interested parties to reduce the possibility of moral outcry and to normalize them. Things that would have been problematic within the previous standard of day to day life were thus rendered commonplace, and so for example functionaries of the German government during the Third Reich continued their lives as bureaucrats while engaging in unethical and often cruel acts, what Arendt calls the "banality of evil".[65] One of the reasons behind the banalization of these acts was the fact that they were done by government officials, military men or individuals otherwise involved in a communal effort, and not lone actors.[66] The weight of these acts would thus be distributed among the many, and become more acceptable to the average individual. Parallel to this, the fact that many of these horrors occurred in totalitarian regimes raised the question of the importance of figures of authority in both the minimizing of the importance of these unethical acts and in facilitating the distribution of responsibility among those actively engaged and the complicity of the general public, unengaged, unaware or indifferent to these crimes. Arendt's famous example of Adolf Eichmann is the embodiment of the phenomenon of an individual entirely disaffected from their acts, limiting their decisions to simply following orders despite their position of authority. We choose Arendt specifically because, even if her aim was not to develop a theory on behavior, her work brings out a significant aspect of human behavior in many ways, a truth that thinkers had to contend with for the rest of the century: human decisions are not invulnerable to circumstance. On the contrary, they

[65] Hannah Arendt, "*Eichmann in Jerusalem*.", Penguin (2006).
[66] Ibid. 18

are often affected and impeded by forces outside of the individual's control, like habit and indifference or the existence of a plausible excuse.

These were, however, no more than philosophical intuitions, and they went against a longstanding tradition that was skeptical of the power of these forces in the face of the monolith of moral rationalism, and many philosophical theories informed by these theories would continue to work under the metaethical assumption that moral justification is necessarily rational justification.[67] However, the emotional, irrational side of morality was no longer ignored: other fields, chief among them that of psychology, were invested in verifying these intuitions.

2.1.2 Psychology and the unraveling of the situation

Solomon Asch's conformity study showed how easily people conformed with positions they knew to be wrong in order to avoid going against the majority,[68] and during Stanley Milgram's infamous experiments on submission to authority, despite clear hesitations, most test subjects didn't stop inflicting pain on others when pressed to continue the experiment by the experimenter, a scientifically empowered figure of authority.[69] Findings like these showed that ethics needed to be reevaluated through the lens of psychology, and showed that context mattered greatly when it came

[67] As we stated in an earlier comment, the most obvious cases of this are Kantian Deontology, and contemporary proponents of the tradition such as Alan Gerwith. However, we believe that most ethical theories placing rationality above all things in ethics are part of the monolith in one way or another. This would include most moral theories with a prescriptive nature, such as deontology, but also consequentialist-utilitarian theories or contractualist theories. For more details, see the introduction of Alasdair McIntyre, "*After virtue*.", A&C Black (1981/2013).

[68] Solomon E. Asch, *"Studies of independence and conformity: I. A minority of one against a unanimous majority.",* Psychological monographs: General and applied 70.9 (1956): 28-31, 36-50.

[69] Stanley Milgram, *"Behavioral study of obedience."*, The Journal of abnormal and social psychology 67.4 (1963): 371-378.

to people acting on their moral preferences, especially when they were put in a context with other people.

A clear example of this is Phillip Zimbardo's Stanford Prison Experiment, in which 24 students were divided into two groups, guards and prisoners, and proceeded to become so entranced in their roles that the experiment had to be brought to an end to preserve their health.[70] Even if the moral validity of the methodology of the experiment is often questioned, it still holds relevance to us in two ways. Firstly, through the experiment, Zimbardo coined the term "deindividuation", which refers to a psychological process in which self-awareness is decreased within specific contexts (mostly through group pressure), leading to a weakening of the grip that individual conviction and other types of social inhibition have on individuals. Deindividuated, the prison guards became violent, uncompassionate and cruel, and the prisoners became compliant, despondent and resentful. What makes it particularly interesting is that, according to Zimbardo, among the variables that led to a deindividuated state was anonymization. Of course, given that the prison guards were identified by name tags, and the prisoners by sequences of numbers, one could argue that there was no true anonymity here, but that is not the point of Zimbardo. Rather, the uniforms and numbers were anonymizing factors because they concealed the student's identities and reduced them to their roles.[71] However, it is important to stress out that, to Zimbardo, anonymity is only one aspect of the greater problem, which he believes, is still identity and how deindividuation affects it. According to him, several factors in the experiment worked as identityreducing forces, and they were the key cause for the events that transpired. For instance,

[70] Philip G. Zimbardo, *"Lucifer effect.",* The Encyclopedia of Peace Psychology (2011).

[71] Zimbardo (2011). The author specifically tackles this in chapter 10, deconstructing the changes for each individual student in the prison and developing an explanation to how the experiment affected them, but chapter 13 provides insights that go beyond the experiment itself into other cases, such the atrocities committed in Abu Ghraib prison U.S. army personnel.

the sunglasses of the guards hid their eyes, the outfits made the individuals blend among their peers and further made them fit into their roles. Having a sequence of numbers instead of a name was clearly a factor in changing the perception the guards had of the prisoners, and how the prisoners perceived each other and themselves. In other words, the setting of the prison changed the description that these individuals presented, not only making the guards and prisoners different enough to make them less identifiable as the young students who entered the experiment, but by tampering with their self-perception and making them act in a way that was not consistent with how they acted when they were merely students. Zimbardo believed these conditions made individuals feel like they were not morally bound by the same social standards that applied to their day-to-day student life because they were no longer students. The experiment therefore set a precedent for superficial identity and identifying traits, or the lack thereof, to be considered as a variable when analyzing moral problems.

Secondly, because of its unexpected, divisive and polemic results, Zimbardo's study raised questions that were studied by moral psychology for the coming decades, and generated entire theories axed around the findings of his experiment, be it to confirm them or prove them wrong. Among such theories, one that holds particular interest to us: The Social Identity model of Deindividuation Effects (SIDE), first introduced by Lea and Spears in 1991. Its creation followed years of general dissatisfaction with Zimbardo's classical deindividuation model, as the explanatory power of the model seemed limited, especially when applied to situations of mediated communication and other non-face-to-face situations. Several studies were conducted with the goal of showing the influence of this variable,[72] culminating in the paper produced by Lea and Spears.

[72] Stephen D. Reicher, *"Social influence in the crowd: Attitudinal and behavioural effects of de-individuation in conditions of high and low group salience."*, British Journal of Social Psychology 23.4 (1984): 341-350.

According to their findings, individuals were altering their behavior not because their selfawareness was being numbed or decreased by the situational forces around them, but quite the contrary: variables like mediated communication exacerbated self-awareness to the point of affecting social inhibitions. In other words, according to the SIDE theory, individuals online are not altering their behavior because they feel less self-aware and submerged within groups, but because they feel even more salient than before. Anonymity, in particular, seems to exacerbate some of these effects. In one study, participants were more likely to extend more extreme opinions and agree to such extreme opinions when they were anonymous than when they were identified.[73] Likewise, in a study where people were assigned a leadership role, the anonymous participants complied more easily with an anonymous peer, and were less prone to do so when they were identified.[74]

Even if it is logical that the SIDE theory would replace the classical model of deindividuation given that it is effectively better at predicting online behavior than its predecessor, what truly gave it traction was the rise of mediated communication through the internet. Through a plethora of studies testing its effects, it not only became evident that mediated communication changed the social game online, but that anonymity interplayed with it to radically alter behavior. We will now look in more detail at the ways in which anonymity affects the moral decisions of individuals in the internet, by identifying other variables that might interplay with anonymity like mediated communication does. Our intuition is that these are mostly social forces, introduced by the sheer numbers of users online or the highly specific social norms they give themselves.

---

[73] Baumeister, Roy F., Sarah E. Ainsworth, and Kathleen D. Vohs. *"Are groups more or less than the sum of their members? The moderating role of individual identification."*, Behavioral and Brain Sciences 39 (2016).

[74] Kimberly M. Christopherson, *"The positive and negative implications of anonymity in Internet social interactions: "On the Internet, Nobody Knows You're a Dog"."*, Computers in Human Behavior 23.6 (2007): 3038-3056.

2.2 Social variables of the online world: crowds and communities

2.2.1 Crowds and anonymity

We will begin with crowds because, as a social phenomenon, they are deeply tied to the concept of anonymity, as we will show, but also because it is a concept that has more direct ties to anonymity than that of groups. Large groups of people are a unique social phenomenon where the will of the many converge into, at best, an organized manifestation and at worst, an angry mob. There has been a long struggle to find the logic behind the will that drives the masses when they gather, mainly due to the fact that most philosophers discarded them as irrational, unpredictable and dangerous forces up until the 19th century,[75] according to crowd theory historian John McClelland. The first attempts on a theorization that went beyond simple intuition and superficial observation of the behavior of the masses was provided by authors like Kierkegaard and Le Bon, who attempted to give a more psychologically-minded view of crowds as composed of individuals under certain noxious conditions,[76] with Le Bon coining the concept of "herd mentality" and the idea that large groups of individuals move as one unthinking mass. Indeed, Le Bon paints a rather unflattering picture of crowds, showing them as groups that can easily forsake rationality for emotion and turn into mobs.[77] Later analyses, like that of Canetti, oppose Le Bon's view, finding in crowds a much more complex and varied phenomenon extremely influenced by context, and

[75] McClelland, J. S. (2010). The Crowd and the Mob (Routledge Revivals): From Plato to Canetti. Routledge. Pgs , 151-156.

[76] These situational effects of crowds can consist of things ranging from political rhetoric to physical closeness. It must be noted we are not aiming to provide a deep explanation of the way in which all situational forces affect people, but rather to understand how this influence changes behavior.

[77] Gustave Le Bon, "The crowd: a study of the popular mind.", London: Ernest Benn (1895, trans. 1947).

that are far from empty-minded.[78] As McClelland explains, Canetti believes that crowds vary greatly according to many variables, they all have a set of essential attributes that make it a crowd: growth, equality, density and direction.[85]

For the sake of our study, we will refrain from taking a position regarding whether crowds naturally incline humans to be unthinking and irrational, as the "herd mentality" tradition following Le Bon would suggest, or if they are complex psychological ensembles driven by a fundamentally rational purpose, as proposed by Canetti. We will, however, work under the common assumption that both theories make, namely, that crowds have a profound effect on individuals as a force that alters their behavior. We will share Le Bon's cautious pessimism as to the tendencies of crowds, and veer more towards the beliefs of Canetti in considering that the specific effects crowds will have on individuals are hard to predict until they have been immersed within the crowd. We also believe that Canetti's typology is coherent with the general understanding that has been produced by crowd theory, irrespective of the traditions. Let us use the attributes of his typology to see if certain groups of people online can be considered crowds.

We can easily identify the attributes of *growth* and *direction*. Internet groups, just like physical ones, are not required to, but can have a specific objective and a desire to increase in number exponentially, as numbers give them strength. To find proof of this, one needs only to look at the countless examples of online phenomena that echo the actions of a crowd: petitions on websites such Change.org, the numerous challenges that have taken the world by storm after being issued by influencers (Ice Bucket Challenge, etc.), the sudden subscriber and/or follower surges and plummets in the profiles of internet entertainers following situated events like dissatisfaction

[78] John S. McClelland, *"The Crowd and the Mob (Routledge Revivals): From Plato to Canetti.",* Routledge (2010). [85] Ibid. p 236.

with changes in their channel's model or drama and gossip[7980] and the witch-hunt-like cases of mass harassment campaigns against individuals surrounded in controversy.[81] Just as we saw that anonymity equalizes individuals in the freedom of their expression and circulation, so does being online allows us to come together as one for a single common purpose. Then there is the matter of *equality*. As we established in the previous chapter, one of the main reasons for permitting anonymity online is its equalizing power. Through it, individuals can come together even more easily than with physical crowds, as an anonymized congregate of individuals that are not impeded by physical barriers of prejudice or space. Even without anonymity, given that most users have exactly the same amount of technical options in the limited number of actions they can take, an amount established by the website-medium in question (for example, Twitter users have 280 characters to say what they think), there is a technical condition of equality shared among them, regardless of whether they are identifiable or not. Finally, there is the attribute of *density*, which could be seen as the one point of contention, given that the internet is by its very nature imposing a physical barrier of distance between individuals. This can be easily addressed by noting that in terms of pure density of speech acts and movements, the internet is unparalleled. The internet is home to some of the most densely populated environments in the world due to its accessibility and

[79] There have been many occurrences of this on YouTube, for example. See the cases of The Fine Bros, who lost [80] .000 subscribers in only three days (https://www.washingtonpost.com/news/theintersect/wp/2016/02/02/after-youtube-outrage-the-fine-bros-decide-not-to-trademark-react/) and James Charles, who lost 2.6 million in the same amount of time (https://www.bbc.com/news/newsbeat-48243905). It is interesting to note that, despite being only three years apart and having a similar number of subscribers at the time of their respective scandals, the difference in numbers of the subscriber loss is significant (five times higher). This could mean these internet movements are becoming more effective or explosive as the years pass. Of course, this is only an educated guess, as it might be the case that something else, such as the content of the channels or the demography of their audiences might be more significant factors.

[81] One recent example is that of the judge of the "People v. Turner" case, Aaron Persky, who lost his job as a judge, then his next job as a high school tennis coach over the controversial ruling, found by many to be too soft for the gravity of the rape allegations surrounding Turner. For more information on this, see https://www.nytimes.com/2016/06/07/us/outrage-in-stanford-rape-case-over-dueling-statements-of-victim-andattackers-father.html and https://www.cnn.com/2019/09/12/us/brock-turner-case-judge-fired-coachingjob/index.html

traffic, even if we should be aware of the absence of unique variables specific to physical crowds.[82] A group of likeminded people online would therefore fulfill this last attribute in that they congregate closely and tightly in their limited and shared pursuit. With this, we have reasonable reasons to believe that internet groups can behave like crowds. In summary, if crowds are essentially amassed throngs of individuals rendered anonymous by the sheer presence of their peers, then we can expect that amassed groups of internet users that possess the attributes of a crowd might be influenced as if they were anonymized by crowds, regardless of whether they are anonymous.

As we have seen, anonymity weaves in and out of crowds. These throngs of users moving in unison are a mix that ranges from highly identifiable to anonymous. But as a member of the crowd, an individual will be as an anonymized bit, hiding the size of the message it composes. They assume a role that is minuscule if compared to the whole and yet fundamental to producing the final picture, and in doing so, become prime targets for the effects induced both by crowds and anonymization. Without taking our eyes from social variables, we will now shift our focus towards a more persistent social form online: internet communities.

2.2.2 Virtual communities

In the first chapter of this thesis, we spoke of Anonymous, the self-proclaimed hacktivist group that became known internationally. What is less known about them is that they started as an offshoot branch of the community within a particular, infamous website: 4chan. This platform dates back to the early 2000's, and is divided into a series of different boards oriented to the

[82] Fight-of-flight reflexes, eruption of violence and/or police retaliation, among many other things.

interests of its users, where they can comment and share images and videos. 4chan stands out among other websites for two main reasons. Firstly, it is extremely liberal in the types of content it permits. From innocuous, family-friendly content to fetishist pornography, from informative lessons in history to blatantly racist, misogynistic and morally dubious content, there is a board for everything. And, if one is feeling brave, they can visit the /b/ board, the random board, and spin the wheel to see if they will bear witness to the brutal death of another human being or watch a litany of kittens play around. 4chan has it all, and while each board has its own rules, the website itself sets the bar quite low, restricting only extreme cases (such as child pornography). Secondly, 4chan is a purely anonymous platform. This means that all users are not required to create a username; instead, they are given the handle "Anonymous" followed by a number. For this reason, the community of 4chan was originally referred to as Anonymous, and it was only later that the hacktivist group claimed the name as they sought to make themselves independent and distinct from their platform of origin. 4chan users, or "anons" as they call themselves, only provide personal information at their own risk. Even the moderators of the website, sanctioned volunteers who enforce the rules of conduct and posting, do so as anonymous observers that can be in any board, at any given time. Of course, anonymity also causes complications for controlling the content of the website, as banned users can often just create new accounts and get back on it. These two characteristics make 4chan a unique website in the way it developed and grew to become not only a unique social environment, but the point of origin of widespread internet practices, subcultures and trends. There is a growing literature dedicated to documenting the history of the website and of its growth in influence since the early years of the 21st century that paints the website as an avatar of a significant portion of the internet's lifeworld, one of the many colorful communities that take shape within particular websites and construct cohesive social spaces where

people spend significant portions of their day.[83] And while social media is occurring within websites, it is often no more than an extension and translation of existing real-world communities into the World Wide Web. On the contrary, 4chan, Reddit, Tumblr, YouTube and many other websites like these, while often not designed to be mainly social spaces in the way of social media, end up becoming just as fertile to the creation of internet-exclusive communities where identifiability is not a given. They are all examples of what we've been referring to as "internet environments", and each one of these is a unique coordination of variables, resulting in a unique production of social behaviors and effects that model and are modeled by the users engaged in them. The 4chan environment, for instance, is a perpetually ephemeral environment shaped by meme-culture, weekly news and fashionable subjects (its very founder, Christopher Poole, has described it as a site that "has no memory")[84], but it is also home to localized, specialized discussions between aficionados of many subjects, that occur independently from the news of the day and other distractions. This description could be extended to similar websites, like Reddit, noting however that there are some differences, key among them being the fact that, on 4chan, almost all of these exchanges take place between anonymous users. Despite this, there are clear indications in the literature that point to there being certain social mechanisms endemic to all of these online communities. These effects of the community would then be a variable in our study that transcends the context of specific websites. But before we delve into them, we must establish what we understand by communities.

[83] W. Phillips gives a very good description of the formation of 4chan through the lens of trolls, but there are other, more focused and detailed books on the subject. See Cole Stryker, *"Epic win for anonymous: How 4chan's army conquered the web.",* Abrams (2011). We must note that it was Stryker who originally presented the opposition between murder and kittens on the website, as a way of showing the diversity of its content.

[84] Parmy Olson, *"How 4Chan's Chris Poole Runs His Web Giant On A Shoestring"*, Forbes, October 18 (2012).

The community, as a concept, is far more encompassing and multi-pronged than the crowd, as it entails a more complex connection between individuals than the coincidental characteristics that would determine a mere group or set of individuals (like people with the name John, or people born on a Tuesday). In order to avoid straying too far from our subject, we will not go into an archeological research of the concept or try to provide our own definition. We will, instead, refer to the typology of McMillan and Chavis, which elucidates the key aspects of communities that we believe we can use to better show the concept's online form. These are membership, influence, integration (and fulfillment of needs) and shared emotional connection.[85] *Membership* often entails that the user has invested time and effort into becoming a part of the community, and so their belonging has been earned.[86] Be it through rites of passage, a minimum amount of time spent in the company of their fellows or more unilateral criteria (like the ownership of a certain form of currency), individuals have to earn their place within a community to truly belong to it. This is especially true on the internet. Online gaming communities, for instance, often expect a rigid adherence to their core rules by newcomers and extend little to no benefit to them until they've put in some time, and there are often different, socially impermeable layers within websites that require a certain amount of social credit to be accessed.[87]

Influence, McMillan and Chavis explain, is a two-way road. On the one hand, individuals often seek to become a part of a community in the hopes of having influence within it. Joining a community can fill individuals with a sense of purpose, especially if they are attributed a role that

[85] McMillan, David W., and David M. Chavis, *"Sense of community: A definition and theory."*, Journal of community psychology 14.1 (1986): 6-23.

[86] Of course, membership can also be something that is there from birth, say by being born into a specific country or religion. However, we chose to focus on the more active aspect of membership because it better fits internetspecific communities.

[87] Reer, Felix, and Nicole C. Krämer, *"Underlying factors of social capital acquisition in the context of online-gaming: Comparing World of Warcraft and Counter-Strike.",* Computers in Human Behavior 36 (2014): 179-189.

they feel is essential to the community.[88] On the other hand, communities tend to put pressure on individuals to conform to the norm. This influence does not necessarily need to be direct, and might be a matter of the individual perceiving this pressure as being there. Asch's conformity experiments, mentioned earlier, stand as a prime example of perceived group pressure on individuals leading to a tendency to accept the consensus regardless of its truth value in our eyes, mainly out of a desire for cohesion and belongingness. We understand by "norm" any kind of explicit or implicit understanding between members of an online community regarding what is acceptable or expected in terms of attitude or behavior. These norms gain their power from a series of factors, ranging from in-group/out-group norm dichotomies[89] (the value of the norm lies in that our group has it and the others don't) to norm longevity and norm functionality ("if it isn't broken, don't fix it" style of approaches). If we look for the existence of such norms online, we will find community guidelines within every website and even within segmented sections of websites. On top of that, there will be unspoken rules known only to members of the community, that in a way act as a social indicator within the community.

The integration and fulfillment of needs is simplified as "reinforcement" by the authors. The gist of it is that being within communities can be a rewarding experience, and that it is often in a community's interest to give its members something to strive for and continually reinforce their connection to the community. In early stages this might be the membership itself, but more often than not the validation comes from the interaction with fellow members, and the sharing of

[88] There are many psychology studies that highlight this predisposition to engage with a group if one is given a distinct role or task to accomplish. See the study by Beaumeister et al., *"Are groups more or less"* (2016).

[89] There is an extensive amount of literature and experiments in psychology that has made evident that individuals act differently when they are made to feel like part of a group A that is in competition with, opposed to or even distinct from another group B. Henri Tajfel's 1974 paper on inter-group conflict is widely considered the foundational work on the study of intra-group bias. See Henri Tajfel, *"Social identity and intergroup behaviour.",* Information (International Social Science Council) 13.2 (1974): 65-93.

perceived values and ideals with them. In other words, individuals remain in the community because of their positive engagement with their fellows. Internet communities usually have a rather positive outlook of themselves, especially in comparison with others, and this belief is only made stronger by continued, immersed interaction with fellow community-members. As we will see in the next section of this chapter, the internet's infamous echo chambers are a possible outcome of this reinforcement mechanism.

Finally, a community has a *shared emotional connection*, a bond born out of a common history and shared experiences, even though they do not necessarily need to have lived through this history to share in it. This emotional nexus is a glue that holds members together within a certain space of thought and opinion. Online communities have the unique characteristic of being able to share the moment of birth of the community, with the launch of a website or game for example, as well as the moment of death of a website and, often, of the community that went with it.[90]

The elements we have described above, and which constitute the definition of community provided by McMillan and Chavis, should not be taken independently but understood as constituents that all come together within communities to varying degrees. Let us look at the earlier example of 4chan to apply this definition. Each individual has a specific set of interests that will lead them to the website, and within it to one specific board. Within 4chan, each board distinguishes itself from the others, and while some users may travel comfortably between all boards, most seem to gravitate around specific sections and users that match their preference

[90] The Gaming forum NeoGAF, considered one of the internet's biggest, imploded over a sexual harassment scandal involving its CEO that eventually led to the resignations of most staff and a diaspora of most of its community. For more information, see (https://www.polygon.com/2017/10/24/16536652/neogaf-resetera-forumscandal-malka).

(*integration and fulfillment of needs*). While the website has cross-board rules, such as the ban on child pornography, each board has a specific set of agreed-upon norms. For example, all pornography is explicitly forbidden in the /wsg/ "work-safe gif" board. In what regards implicit rules, within /wsg/ it is logical that users implicitly expect all their fellows to only post things that are reasonably acceptable to be viewed at work and in good taste, and avoid posting content that could dodge the restrictions but be out of tune with the feeling of the board, and therefore undermine the spirit of the community's environment (*influence*). After spending enough time on the platform, we could see users gain the comprehension of the in-group lingo and subcultural references, and bear these as proof of membership, with the capacity to understand older references translating into more influence within the community (*membership*), and eventually grow to appreciate their fellow members much more significantly than those of other boards because of their common grounds (*shared emotional connection*). We could have chosen any other website to display how communities have formed and grown within them as these mechanisms usually follow a similar path to what we've described above, but we chose the example of 4chan because of the predominance of anonymity within it, and to possibly suggest that anonymity interplays with social forces in a more complex manner because it can enhance different, maybe even unique social forms of engagement. Now that we have spoken of the two main social forces influencing online choice, these beings crowd-effects and communities, let us focus on a third set of variables: the uniqueness of internet structures.

2.3 Effects of internet characteristics

As an environment, we've already established that the internet is unique. Communication is different because it is mediated, and social forces, while remaining similar to their real-world counterparts in nature, manifest very differently in form. We will consider as a third variable a variety of aspects unique to the internet that, in their own way, could have an effect on behavior alongside anonymity. For this, we will use the categories established by Dean Cocking and Jeroen Van Den Hoven in the first half of their book *Evil Online*. While their focus is on explaining how immoral behaviors manifest themselves online and this central part of their work holds great interest for the last chapter of our thesis, we will for now limit ourselves to borrowing the features[91] that the authors point out, and that make online spaces unique. However, given that keeping them all separate would only slow our pace down without a guarantee of extra insights, we will narrow them into three main categories: online choice of experience, online privacy and online habitformation.

2.3.1 Online structure specificities

Online choice of experience

When online, individuals have the choice of how they want to engage with others, but most importantly, they get to choose who they want to engage with, allowing them to become especially selective of their interactions, and for their experience to become highly customized to their preferences. Indeed, the way we browse online shapes the future content we will encounter, as websites tailor our experience and lead us to similar products, ideologies and news to those we've

[91] Cocking, Dean, and Jeroen Van den Hoven *"Evil online."*, Vol. 15. John Wiley & Sons (2018): 39-58.

been interested in before.[92] When fully encased in a mold perfectly made to fit our interests, we can be said to be in a "filter bubble",[93] where all content we encounter has been curated to fit our previous behavior. Likewise, websites will make it easy for us to encounter and exchange with like-minded people, and often discourage exchanges with people we might disagree with. If we were to become completely surrounded by peers that never dispute our claims and whose ideas we never dispute, in an environment dangerously spiraling towards progressively more extreme views, we can be said to be within an online echo chamber.[94] Coupled with the two-headed phenomenon of filter bubbles and echo chambers is also the fact that all the information we encounter and intake while online is the product of someone's interpretation, and while this hermeneutical problem is not unique to the internet, it seems to be accentuated within it.[95] All of these factors that, by limiting, twisting or denaturing the information we encounter online and our exchanges with others contribute to what we will for now qualify as a potential collapse of the boundary between opinion and fact online.

Online privacy

As we explained above, online experiences are often tailored to fit our interests, but whether these interests remain private or are put on display is an entirely different matter. Or, at least, it should

[92] It should be noted that this phenomenon occurs regardless of whether we are anonymous or not. Websites will tailor our experience based on whatever limited information we provide them through data tracking, and often manage to identify us through algorithms in a way that human beings cannot. This is one of the greatest privacy issues raised by the International Grand Committee on Big Data, Privacy and Democracy, as we shall see below, in our section about online privacy.

[93] Please refer to the TED talk by Eli Parisier.

[94] Studies are still being made to determine the extent of the power of echo chambers. For more information, see Flaxman, Seth, Sharad Goel, and Justin M. Rao, *"Filter bubbles, echo chambers, and online news consumption."*, Public opinion quarterly 80.S1 (2016): 298-320. While some researchers seem to downplay the power of echochambers in the face of alarmism, we believe there is not enough data on the subject to have a fully informed view on the scale of the problem.

[95] We will refrain from going in depth on this specific form of hermeneutical bias, as we will address it in the final chapter of this thesis.

be. However, being online entails a relative degree of exposure within what is a fundamentally public space. In order to enter certain spaces, some information must be disclosed. Often, displaying one's information can barely be considered a matter of choice, as entire platforms can revolve around a careful management of our self-image. Think, for example, of any social media website, where exposing one's personal information is the norm. There is then a tension between our "voluntariness" to expose ourselves and disclose information that we have seen as pertinent to share, and the information that inadvertently bleeds out whenever we enter our online environments and simply exist within them. This was echoed in the concerns raised by the International Grand Committee on Big Data, Privacy and Democracy regarding the obsolescence of traditional conceptions of consent online as these are circumvented by tech giants to acquire as much information as possible from individuals.[96] Essentially, what is at stake here is the privacy of users, which is more and more easily invaded by new technologies through perniciously designed forms of agreement. We will describe this current situation as a potential collapse of the border between the private sphere and the public sphere online.

Online habit-formation

It could be said that we have gone beyond simply domesticating the internet to serve an instrumental purpose. It has now become an everyday tool, something we can hardly do without, and any studies now shed light on how addictive being online is, or how being online facilitates addictions that are already a problem offline. This ranges from social media addiction fueled by a

[96] The International Grand Committee is a collective of parliamentary representatives from various countries put in place to engage in dialogue and gain insights into the inner works of giant Tech companies in the wake of the Cambridge Analytica Scandal. The Committee last assembled in May 2019.

constant need of social interaction or being "in-the-know"[9798] to online video-games using gambling mechanisms to hook their players.[99] These habits are only made more consistent by how easily virtual environments can be accessed, and despite the fact that most online interactions now have repercussions in the real world. Online shopping is facilitated by the way online shopping has direct access to one's bank accounts, and can quickly spiral into large sums; social media posts can rapidly lead to negative consequences, from personal problems to dismissals.[100] Not that most of these problems didn't exist before; rather, they simply changed in format, and found a new home in the online world. We will describe this bleeding of the online into the offline as a potential collapse of the border between certain online and offline habits.

We are aware that some of these variables hearken back to our description of internet structures in chapter one, but we felt that it was important to make these issues into a distinct variable, manifesting in three different, and yet similar ways as a force corroding the barriers between offline and online. With this, we have CMC, online social normativity forces and online-offline corrosion forces as the main three variables we have found that might interplay with anonymity to affect behavior.

---

[97] Andreassen, Cecilie Schou, Ståle Pallesen, and Mark D. Griffiths, *"The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey."*, Addictive behaviors 64 (2017):
[98] -293.

[99] Drummond, Aaron, and James D. Sauer, *"Video game loot boxes are psychologically akin to gambling."*, Nature human behaviour 2.8 (2018): 530.

[100] For specific examples, see the parents who lost hundreds to thousands of dollars through their children's compulsive buying of online game currency
("My son spent £3,160 in one game" at https://www.bbc.com/news/technology-48925623) or policemen losing their jobs over careless posts ("Ocasio-Cortez: Officers sacked for post suggesting lawmaker be shot", at https://www.bbc.com/news/world-us-canada-49081786).

2.3.2 Anonymity interplay with internet environment variables

Having determined the three main groups of variables that interplay with anonymity, we will now see some examples of these variables coming together with anonymity to produce changes in the behaviors and moral perceptions of individuals.

Tweeting as a computer-mediated social activity facilitated by anonymity

Twitter is unique as a social media platform as it allows for individuals to interact using minimal amounts of information: there is a limit to the amount of text one can write with, only an email is required as an identifier and users are not forced to give a detailed description of themselves, thus allowing anyone to browse unidentified. Peddinti, Ross and Cappos conducted a landmark study with a remarkable one hundred thousand user sample of Twitter users to try to better understand the anonymous section of the website's user base. Through the use of Amazon Mechanical Turk services, Peddinti et al. managed to get an accurate categorization of a sizeable portion of their initial sample, with 5.9% of users being classified as being fully anonymous (meaning that, as a normal user, there was no reliable way to link them to any particular personal data set) and 20% as partially anonymous (meaning there might have been some distinguishing traits, but not enough to get a clear picture of them).[101] This means that about 25% of the population of the website surfs with some degree of anonymity. While we have no way to determine how statistically significant this is beyond Twitter's ecosystem, we can trust their sample to estimate that about one in every four users on Twitter is nonidentifiable to researchers actively trying to identify them. The study also showed that, on average, anonymous "lurker" users (meaning users that don't actively engage with the environment through tweets or retweets) had less friends and activity than highly

[101] Peddinti et al., *"On the internet"* (2014).

identifiable users (celebrities, influencers, etc.) but had more social activity than the identifiable "lurker" users (people who have built their Twitter profile as they would their Facebook, and use it as another form of social media). These results are coherent with the results found in earlier experiments backing the SIDE, as these anonymous individuals seem to have a higher social salience than their identifiable peers and are more likely to socially engage others in the platform thanks to the way communication is relayed.

Anonymity and social enablement in cheating in online games

As we stated earlier in this chapter, websites, platforms and other online structures have rules that, if broken, can result in punitive action. This is especially true of competitive environments like online videogames, where the overall player experience is greatly affected by the fairness of the game and how deeply the communities of the game are tied to it. J. Blackburn et al. conducted an investigation on the social media platform of Steam, the most popular gaming platform for computer-based gaming. Their focus was on cheating, and how the moral behavior of players was affected by it, and their results are very enlightening. They note that most servers are tightly controlled so as to prevent cheating, and that Steam's anti-cheat system is a merciless system that not only bans players from playing games where they have cheated, but also singles out their profiles to indicate their offence. This by no means signifies that players don't cheat. The reduced significance of the virtual punishments, coupled with the highly-competitive environments of online gaming, are only some of many structural forces that give incentive for players to cheat.

Another study also indicates that, when in relative anonymity, players will tend to act in the way they believe other players are acting.[102]

Players can interact with each other through the platform's social media, the Steam community, and see if other players are marked for cheating. It is worth noting that, in this platform, players are already largely unidentifiable, simple usernames and very limited profiles usually being the only information available on another user unless they willingly share it with select "friends" they have added. This leads to a very unique social phenomenon: as cheating players gain an identifying marker, the brand of "cheater", they will often be ostracized by the larger community but might end up congregating around other cheating players, adding them as friends and thus generating a new community. On top of that, players that are not cheaters but that have a significant proportion of cheater friends are more likely to also begin cheating.[103] While this does not necessarily entail that anonymity facilitates cheating, it does seem to indicate that the presence of identifying markers over mostly anonymous users might facilitate the formation of communities that will both expel cheaters and indirectly bring together kindred cheaters that will incentivize each-other's cheating practices.

Banalization of pornography in online environments

Before the internet, while there was consumption and sharing of pornography, but accessibility could vary greatly depending on what one sought.[104] However, with the overwhelmingly dense

[102] Chen, Vivian Hsueh Hua, and Yuehua Wu, *"Group identification as a mediator of the effect of players' anonymity on cheating in online games."*, Behaviour & Information Technology 34.7 (2015): 658-667.
[103] Blackburn, Jeremy, Nicolas Kourtellis, John Skvoretz, Matei Ripeanu, Adriana Iamnitchi, *"Cheating in online games: A social network perspective."*, ACM Transactions on Internet Technology (TOIT) 13.3 (2014): 9.
[104] Wolak, Janis, Kimberly Mitchell, and David Finkelhor, *"Unwanted and wanted exposure to online pornography in a national sample of youth Internet users."*, Pediatrics 119.2 (2007): 247-257. And Gubar, Susan, and Joan Hoff, eds. *"For adult users only: the dilemma of violent pornography."* Indiana University Press (1989).

traffic of Big Data internet, the dissemination and trade of all kinds of pornography seems to have found a comfortable and accessible medium. The anonymization of its users reduces the fear of being associated with socially reprehensible content, and the accessibility of the content leads to the normalization of sexually explicit content within internet subcultures. An object that was once niche and limited to extremely secretive groups of individuals who were aware of the stigma and afraid of consequences of their acts became another day to day object within the cyberspace. And while it certainly is not a mainstream object, pornography is nonetheless present within internet circles, lurking in the background and only a few clicks away from certain websites. Of course, this is not the case for borderline content such as gruesome murder, rape and child pornography, as those are often monitored by law-enforcement. Still, one could say that the ease with which some of this content is found online makes it more commonplace than it was years ago, not in the sense that has become more visible than old pornographic magazines were, but in the sense that it is accessible at all times and in great diversity through the internet.

Suicide pacts and other epidemic mutually-validated behavior

A rising trend in Japan, it believed by some in the literature that suicide pacts owe their rise in numbers to the increased availability of suicide-focused websites.[105] The people in these websites, through their interactions, not only exchange their thoughts, emotions and plans regarding suicide, but in doing so, influence others who, like them, are caught in a mindset that makes the idea of suicide not only palatable, but desirable. According to Ozawa-de Silva, these interactions are facilitated by the online medium and the comforting distance between users, who can project and empathize more easily with the limited information they receive from their peers, but will also be

[105] Chikako Ozawa-de Silva, *"Too lonely to die alone: Internet suicide pacts and existential suffering in Japan."*, Culture, Medicine, and Psychiatry 32.4 (2008): 516-551.

more likely to act on their decision if they believe they are not alone in their struggle.[106] Likewise, Cocking and Van den Hoven provide testimonies of conversations between users that show the importance of information being hidden, as it allows for certain elements of the conversation to become the only information being exchanged between the users, turning these forums into progressively more extreme echo-chambers. Anonymous users can find and provide company, support, fellowship and even tips within sites that have suicide, anorexia and self-maiming as their main topics, something they would never be able to do if they were identifiable, and once validated by the community, they can enact changes that will lead to them harming themselves to various degrees of severity not virtually, but physically.[107]

These were only some examples of online activity that we believe illustrate the complex ways in which anonymity synergizes with some of the other variables we listed to affect our behavior. We chose them not only because they show that anonymity is not an isolated force, single-handedly responsible for changes in behavior, but because they are deeply tied to the medium, display how unique the internet is as a social environment, and show how people can be affected in many ways, from casual browsing to life-or-death decisions.

Conclusion: The moral border between real and virtual

The greatest challenge in understanding anonymity's power is grasping how much significance it has as a variable on our behavior and how it may clash with the morality that tempers our actions. We began by looking at the historical understanding that was developed of anonymity's

[106] Ibid.

[107] Cocking et al, *"Evil Online"* (2018): 7-32

psychological effects, and whether this translated into an online medium. Having found that a variation of conditions leads to the influence of situational forces, we surmised that the virtual environment's structure changed the rules of the game, and that anonymity is one out of several forces influencing people's behavior within the medium. We set out to try and determine which of these other forces coordinated with anonymity to affect the behavior of internet users. The variables we found were the many layers of mediated communication, social variables like crowdfeel and community engagement and certain unique features of online design such as addictiveness and isolation. All of these variables either enhance or complement the effects of anonymity on individuals by fiddling with their social salience and altering their perceptions on the weight and importance of their behavior and actions online. We then laid out examples of online behavior that show these situational forces in action, and made evident the ethical problems that may result from this. Now that we have developed a more complex understanding of the way the behavior of individuals is affected by these forces, we have to determine what this implies for the moral evaluation of their actions. In other words, how can we understand and evaluate the moral behavior of anonymous individuals online?

Chapter 3: Online self vs offline self: fading moral threshold or multimodal morality?

In our daily life, we make decisions regarding all sorts of matters, and most of them require little to no thought as they have little ulterior significance for our lives. Deciding to eat plum jelly instead of blackberry jelly will probably not impact a person's life any more than deciding to go with the grey socks instead of white ones, and it shouldn't reflect too much about their inner self other than some superficial, personal preference that they have. And yet even these choices might reflect something about us to others, and we might perceive ourselves differently because of them. Say, for example, that we wore white socks with a blue suit when everyone else is wearing blue or black socks. There are also some choices that come charged with questions, and our decisions provide answers that define who we are. Refusing to eat pork because one is Muslim is not only expressing a preference or a set of values, but it is also answering a religious question. In a sense, as we provide information about ourselves, we answer more and more questions regarding our convictions, personality and preferences. But there is a set of answers that exists separate from mere preferences because we hold it in a higher priority, and that is the set of answers we provide to the questions that ethics asks of us.

By ethics we understand any rule, concept or idea that determines what is not only acceptable and encouraged, but expected of people within a society. It is a set of questions regarding our treatment of others, our respect of their dignity and their rights, and how our actions

can affect them. Ethics not only asks these questions of us, but evaluates our answers in terms of how well they fit within a society that expects its members to be social and cooperative, not out of fear of repercussions, but out of their own volition. In an ideal world, these rules of conduct should not falter against the temptations of rational calculation or the influence of exterior forces. Moreover, most people would not expect their moral principles to be affected by rational calculation or changes in context. And yet, as we have seen, people do engage in problematic behaviors despite the presence of guidelines. There is, then, a need to understand morality, not in a metaethical sense of what constitutes a truly moral action in theory, but in the applied sense of what morality does as a formative force within individuals. We must understand the place that morality occupies within the inner struggle that determines our actions and, in particular, how it is affected by anonymity. Firstly, we will see how moral theories have developed an understanding of behavior that defines both key components of morality and of the preferences of individuals as a matter of identity, and choose the theory that provides the best perspective to analyze our problem. We will then see what particularities of the online world complexify the exercise of morality, and what they entail for existing online with a morally consistent behavior. Finally, we will determine to which extent the influence of anonymity can pose a challenge to moral behavior online, how this should be understood within the scope of the current debate in moral philosophy and what implications this has for our distinction of the offline and online moral worlds.

3.1 Moral identity and individual consistency

3.1.1 Virtue and circumstance

In chapter 1, we avoided the problem of morality in order to discuss problematic acts as a social issue, related to the expectations of society with regards to the behavior of its members, choosing to see ethics as serving a mostly restrictive role. We did this in order to be able to grasp anonymity as a tool of rational agents, a tool that could allow us to circumvent many limitations of our daily life, including social rules, in order to satisfy our rationally calculated preferences. In chapter 2, when we presented the tale of Glaucon, we shed the rational choice approach, considering problematic acts from a more philosophical and psychological angle inevitably leading to an ethical analysis, but we stopped short when we realized that there was a need to better grasp the power of the situational forces affecting our behavior before we could gauge the importance of morality in guiding our choices. Now, fully aware of the rational advantages of anonymity and having gotten a better picture of anonymity and its other synchronous forces that might affect our choices online, we can turn our attention towards the moral question, and try to answer it to the best of our capacity.

Before anything else, we have to make a clarification in terminology. Ethics and morality are often used interchangeably, both outside academia and within the philosophical literature pertinent to the general subject of human rules of coexistence, their worth and our attraction to them as ideal objects (they are, after all, the model according to which we are expected to behave). While some in the literature have made the distinction between them a crucial element of their theories,[108] we have used the terms in a rather interchangeable manner. But we feel that from here on there is a need to distinguish between morality in terms of norms and in terms of preferences, so for the sake of simplicity, we will refer to the normative aspect of morality as ethics to better distinguish them

[108] Jurgen Habermas, among others, delineates clear lines between the two concepts. See Torben H. Nielsen, and Jürgen Habermas, *"Jürgen Habermas: Morality, Society and Ethics: An Interview with Torben Hviid Nielsen."* Acta Sociologica (1990): 93-114.

from moral preferences. To illustrate the reason behind our choice, let us consider for a moment the difference between ethical psychology and moral psychology (which is in no way demonstrative but rather a helpful etymological coincidence). On the one hand, ethical psychology is not a coined term, but rather the conjunction of the scientific practice of psychology with the normative concern of ethics. In other words, it is psychology performed in conformity with a standard of ethics, a psychology that respects the dignity of human beings and treats them as individuals and not as tools.[109] When we judge the actions of others and our own as right or wrong, good or bad, we are often referring to some standard by which we know this to be so. If it is an ethical standard then it means that it is a socially established standard, one we have been taught or told to follow at some point in our lives. Hospitals have a deontological code for their doctors and nurses, business offices have work ethics and schools have codes of conduct for both students and staff. An ethical standard is set by groups for individuals, so that the latter know what is acceptable behavior and what is not, and what the consequences of unacceptable behaviors are.

The way to judge whether social norms are respected is through these very same ethical standards. Ethical psychology, therefore, is the practice of psychology compliant with a normative set of ethical requirements. On the other hand, moral psychology "investigates the psychological properties of moral agents".[110] Unlike ethics, which are necessarily external to individuals given their normative nature, we understand morality as being fundamentally tied to the personal beliefs and preferences of an individual. What we view as a moral action is something that is not only

[109] An evident echo to Kant's precept of treating human beings as ends in themselves and not merely a means.

[110] John M. Doris, *"Lack of character: Personality and moral behavior."*, Cambridge University Press (2002): 3 [116] And, according to certain thinkers, prosocial moral intuitions might even be biologically innate. This position, first championed by philosophers like Rousseau and biologists like Darwin, is now maintained by researchers in moral psychology such as Jonathan Haidt, whom we will speak of in the following section of our chapter. For more information, see Rozin, Paul, Jonathan Haidt, and Clark R. McCauley, *"Disgust."*, In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), Handbook of emotions (pp. 757-776), The Guilford Press (2008). Haidt and his fellow researchers explain how disgust went from being a biological safety mechanism to being a gut moral reaction.

good and right if done by others, but principally as something perceived as being good and right if done by *us* as individuals. Yes, the details of this moral code might have been passed down to us by our family, taught to us by our society, acquired through our life experiences or even established by a philosopher of ethics,[116] but unlike a set of ethical standards, a set of moral preferences is something that we have appropriated as our own and that should guide us accordingly, not only out of obligation or preference, but out of conviction. And unlike ethics, it is far more difficult to understand how morality acts upon individuals precisely because of the fact that it is embedded deep within us. To summarize, we understand by ethics any set of normative, collective rules of moral conduct that is exterior of an individual, and when we refer to morality, we will be doing so in the perspective of personal, variable moral beliefs.

As we explained early in chapter 2, the tradition in moral philosophy leading up to the twentieth century shared the position that many other disciplines did by assuming that rational thought was what ultimately determined human decisions, and therefore assumed that ethical norms could be infallibly converted into moral preferences that would consistently guide individual's decisions. Ethics was a field that was less interested in seeing how morality came to us as individuals and occurred within ourselves than with identifying proper ethical behavior. In other words, philosophy was more concerned with what ethical behavior was, and not with how it occurred: it was a normative task focused on evaluating and prescribing behavior rather than understanding it. The bread and butter of conventional ethical theories like Kantian Deontology and Utilitarianism was this normative debate, and to this day these theories rely on rational deliberation and justification to determine what is ethical or not (Kantians use universalizable rules, Utilitarians calculate global consequences, etc.), often without trying their hand at a psychology of morality, hoping instead that their philosophical postulates will eventually translate into ethical

habits and commonplace practices of average citizens. Still, the aforementioned historical developments and findings in psychology that challenged our understanding of human behavior, while not necessarily contradicting these ethical theories, led to a critical reevaluation of ethics,[111] and opened up the possibility for further questioning. In the face of complex normative moral philosophy, there was a search for a more intuitive way of relating to morality. Some answers were found by turning to intuitions in earlier philosophy that were left behind by the rationalist tradition. Among them was the notion of emotivism, based on the ideas of David Hume, whose early intuition of regarding morality not as a rationally determined set of universalizable rules but as "an active feeling or sentiment"[112] had been the very thing Kant had sought to oppose when he built his own ethical theory. The inception of moral psychology, in a sense, was the admission by ethical theories that Hume was at least partially right and that understanding moral behavior required understanding the psychological intricacies the mind and what lay beyond rational thought.[113] In this mindset, many philosophers sought to reframe morality, and found in the works of Aristotle a conception of morality that stood in stark contrast to modern theories.

Aristotle's ethics were much more focused on individuals, their character and their personal qualities as the sources of moral action, rather than abstract propositions and rational prescriptions. The school of philosophy that embraced Aristotle's view, actualized it and adapted it to answer the questions asked by the twentieth century became known as virtue ethics. If we listen to proponents

[111] Kwame Anthony Appiah provides a good description of this in his recounting of Anscombe's ruthless evaluation. See Kwame A. Appiah, *"Experiments in ethics."*, Harvard University Press, (2008). The book, of course, is much more than this, as Appiah makes a case for the importance of considering situational forces, and constitutes a valuable part of the literature informing this chapter.

[112] David Hume, *"An enquiry concerning human understanding."* Seven Masterpieces of Philosophy. Routledge (2016). 191-284. Appendix 1, section II.

[113] The emotivist theories directly resulting from an adoption of Hume's principle were mostly engaged in metaethical discussion, and so they do not provide us with the answers we need. Still, Hume's intuitions of considering morality as being deeply tied to emotion remain of interest to us as a catalyst for the diffusion of moral psychology within the philosophical world.

of virtue ethics theories like Alasdair McIntyre and Bernard Williams, morality is a substantive quality present in people, much like the virtues of courage or wisdom in Aristotle's work.[114] When individuals behave morally, it is because they are moved to do so by an inner trait that encompasses them in a global manner: someone compassionate will generally provide a helping hand to others, while someone patient will not lose their cool across a variety of situations. These traits of character are not innate to people, and if we teach them to aspire for excellence of character, or virtue, then these traits can become fundamental aspects of people's personality that will determine moral behavior.

In other words, according to virtue ethics, moral behavior is not exclusively determined by rational considerations, but rather, by a combination of rational awareness and emotional determination born out of deep traits embedded within the core of an individual. As such, morality is more than a preference or a feeling, it is something humans become disposed to do, and only by encouraging and educating individuals to have these characters traits and moral dispositions can we expect them to behave morally. As such, moral character is not something easily shaken by conditional factors and situational forces: the virtuous should be able to act in a moral way regardless of the situation, intuitively knowing the right thing to do and confirming it through careful moral consideration. The evaluation provided by virtue ethics, then, is one that decries moral wrong as a fault in character, a failure of individuals to have the moral integrity and consistency to know what is right and to use that knowledge to inform their decision-making. In virtue ethics, moral exemplars are not made by blindly dictating and following rules but by having

[114] We took our understanding of these theories from McIntyre's "After virtue" and Williams' "Moral luck: philosophical papers 1973-1980", as well as through the descriptions that other authors, such as John M. Doris and Kwame A. Appiah, have made of their theories. We have done this in the hopes of forming a clear idea of what virtue ethics entails, and what its main tenets are, and not to advance an interpretation of their respective theories.

them acquire the moral dispositions that will make them avoid immorality by turning it into something perceived as unpleasant at best and unbearable at worst.

If we are to believe proponents of virtue ethics, we have a good direction for answering our question: since behavior derives almost exclusively from specific traits of character, those engaging in problematic acts are either incapable of restraining themselves by their own will, or lack the moral traits that would prevent them from performing these acts. Anonymity then serves its purpose as a force that makes it so that individuals that were already predisposed to commit these problematic acts will be more likely to commit them. After all, if their character traits already make them susceptible to engage in problematic behavior, then the strategic advantages of being unidentifiable should only add to the tally of their preferences, where moral preferences already hold a meager sway. This would align with more psychologically-oriented theories of explanation that would categorize these individuals as being antisocial and unempathetic at best, sociopathic and cruel at worst.[115] Not only that, but it would also confirm a narrative that has long been the dominant one in the media: problematic behavior online comes from problematic users online.[122] Trolls, cyberbullies and other toxic users would all be problem children that managed to repress their behavior not because of a moral drive but out of a fear of repercussions. Finding a new environment that provides them with anonymity, and thus with a freedom from accountability and an advantage over others, enables them to let loose and engage in the problematic behavior that they always have been driven towards by their character, but kept from by ethical norms and the consequences of disobeying them. The philosophical approach to the problem would therefore be one of trying to understand these acts as expressions of moral wrongdoing due to "evil" or

[115] Simon Baron-Cohen, "The science of evil: On empathy and the origins of cruelty.", Basic books (2012). [122] Phillips (2013) provides an in-depth critique of the way mainstream media in the United States of America has branded trolls as dangerous sociopaths, with particular focus on the issue within the fifth chapter of her book.

"problematic" traits within the users, and the correction could only be done through the educational methods that are so fundamental to virtue ethics.[116] However, a quick look back at what we presented in the second chapter will indicate that, while this narrative is intuitive, it rings hollow. After all, if we take the overwhelming numbers of individuals engaging in problematic behaviors online to signify that every single individual that has occasionally engaged in problematic behavior online to be morally bankrupt, then we may have a moral crisis in our hands.

However, if other factors besides character were to be considered as significant in producing a moral evaluation of behavior online, the story is different. Besides, if there's anything we've hoped to show so far in this thesis, it would be that problematic behavior online is not only greatly facilitated by the structure of the online world and the nature of the interactions of the users within it, but that these influences should have power over just about anyone from the moment they find themselves within the anonymized condition, regardless of the moral integrity of their character and the configuration of their personality. In that mindset, while we are not yet fully disposed to abandon the idea of character traits in developing our understanding of online moral behavior, we will refer to another set of theories that stands in opposition to virtue ethics: situationist ethics.

The situationist position should not necessarily be understood as a stark contrast against virtue ethics.[117] It shares an interest in the emotional side of morality, and also assumes that moral

[116] Simon Baron-Cohen does argue that there is a fundamental, biological predisposition to problematic behavior in certain individuals, such as psychopaths. In these cases, moral education would not be enough. See Baron-Cohen (2012), 67-68.

[117] The authors we mainly considered in developing our understanding of situationism were explicitly situationist authors, such as John M. Doris and Gilbert Harman, but also authors that espouse a compatible view with situationism, such as Owen Flanagan, Appiah and Haidt. We will develop on the particulars of these authors further below, when we decide which theory best fits the needs of our research.

beliefs and preferences are deeply seated within individuals, and not purely formulated with rational, scientific logic. It also does not necessarily deny the existence of character traits. Contrary to virtue ethics, however, philosophical situationism originated from the influence of psychological situationism, which in turn derives its core beliefs from the very research and findings in psychology we explored in our second chapter. Situationism, as its name indicates, is a theory of psychology according to which situational forces, rather than character or rational thought, are mostly responsible for our decision-making. Likewise, philosophical situationism places the burden of moral justification not on inner human traits, but on the influence of situational forces. In other words, in a situationist model, moral dispositions are not solid, general character traits that will remain undisturbed by exterior influence. Rather, these elements of our personality can easily be disturbed or fiddled with by external forces and contextual conditions, resulting in moral preferences and decisions that are not necessarily coherent with our dispositions.

Likewise, personality and character are, as all other inner workings of the human mind, something highly susceptible to the environment, and influencing our actions could be as easy as making us smell freshly baked bread or making us be in a hurry.[118] For situationists, while character traits are not necessarily eliminated, they are greatly reduced in power when it comes to their influence as a factor in moral decision-making, sometimes to the point of being irrelevant in understanding moral behavior. A deeper understanding of morality is therefore necessarily connected to a deeper understanding of these forces and how they influence the human mind, and philosophical work can therefore not be done without the support of moral psychology. So far, situationism might seem much more fitting to explain and to make sense of the results presented

[118] We are referring here to the Darley and Batson (1973) Good Samaritan experiment, the Isen and Levin (1972) Dime Booth studies and the Baron (1997) bread smell studies.

by our work than virtue ethics does, and so our entire presentation of virtue ethics might seem like building an unnecessary strawman, but the truth is that we need to have an understanding of virtue ethics if we are to advance a situationist argument in the treatment of our problem, and this for two reasons.

Firstly, situationism is a highly counterintuitive theory that clashes with fundamental preconceptions about the human mind and, most importantly, about moral good. Unlike virtue ethics, which substitutes the moral good of rational rules with the idea of a moral good through exemplary human qualities, situationism can be seen as substituting the idea of a human moral good with a moral uncertainty, as our capacity to independently decide becomes diminished by the possibility of our choice being influenced by exterior forces. Here, the ghost of a behaviorist approach like that of Skinner, which sought to reduce all human behavior to reinforced, learned mechanisms, can lead us down a theoretical rabbit hole that will eventually turn into a metaphysical debate on free will; in a predetermined behaviorist world, morality becomes irrelevant. The importance of not abandoning the notion of character is imperative if we are to discuss morality as something in which we have a choice, as it provides us with a theoretical core of individual moral consistency within the psychological composition of individuals.

Secondly, moral situationism has a weakness: very much like situationism in psychology, it often raises more questions than answers, as the way in which situational forces affect moral decisions is still something that we are far from fully understanding. The evidence in psychology from which situationism is based is obtained in experiments where some of the variables are controlled, and while their results are significant, they are not invulnerable to skepticism. After all, determining the weight of variables in actual situations is an entirely different matter, as the task of attributing a significance to each variable in behavior is beyond the means of most psychology.

Unlike virtue ethics, situationism can rarely provide ready-made answers on general behavior, and only point to specific instances with the hopes of identifying what variables influence, or don't influence, moral decision-making. Still, in the scope of our work, we have good reason to believe that situationism is the branch of moral psychology that enable us to understand anonymity online as a complex force influencing moral behavior. Keeping the weaknesses of situationism in mind, we will take a look at some of the situationist literature to better grasp its implications for online life.

3.1.2 The "right preference": Being and becoming moral

As we've seen, problematic behavior breaks ethical standards of behavior, and therefore, if we are to understand why this is, we must understand where it comes from. Now, while it is interesting to consider the inception of moral ideas within individuals, our focus is more on the matter of the act and what propels it. For this reason, we will give only a brief summary of what we understand as the process of "becoming" moral, to then focus on the act of "being" moral. When it comes to its origins, morality is a debated topic, as it is deeply tied to the ideas of evil and good. Much like the debate around these more abstract notions, it is a debate of nature versus nurture. Many philosophers, ranging from Jean-Jacques Rousseau[119] to Philippa Foot,[120] have suggested that human beings have a natural tendency towards good, and that they therefore have innate dispositions towards moral values that seem to transcend across cultures. These philosophical intuitions were picked up on within evolutionary ethics by defenders of nativism who support the

[119] Jean-Jacques Rousseau, " *« Discours sur l'origine de l'inégalité »*, (1755), Les Échos du Maquis, v. 1,0, 2011 (1755) : p. 36-39.
[120] Philippa Foot, *"Natural goodness."*, Clarendon Press (2003): 5.

idea that humans might share a biological moral mold.[121] In other words, morality would have a biological component, and some moral intuitions would be so fundamental to our species that they would be integrated within our very biology. Beyond the nativist debate there is also a didactic component, which is to say that morality is something we learn through observation, incorporation and repetition. While it is generally understood that our morality is not a rigid set of beliefs and that it might shift over the course of our lives, it is also correct to assume that this susceptibility to change is especially significant during the formative years of our early youth.[122] The nurture section of the debate itself goes relatively unchallenged: while they disagreed on how moral education was to be done, both Plato and Aristotle both believed in the value of education as a formative force for morality.[123] This disagreement perseveres to this day, although it is slightly different. On the one hand, there is discussion on how significant this didactic aspect of morality is, and on the other hand, on what is the proper way to teach morality. If we look towards older, rationalistic ethical theories, the didactic aspect is not only essential, but almost exclusively relevant. There is no excuse for Kantians or Utilitarians to disrespect a rule or to not maximize utility: education and rationality should always overcome biology.

One of the reasons why virtue ethics is better equipped to answer the questions of moral psychology than these two theories is that it provides a more comprehensive conception of the moral self. It is extremely intuitive for human beings to think of themselves as being "honest" or "fair" because it is a personal quality of theirs than it is for them to believe that they are simply following moral precepts dictated to them by an exterior theory, without any emotional drive to do

[121] While not only not endorsing nativism but suggesting the contrary position, Prinz gives a detailed and informative breakdown of the debate, as well as the challenges the nativist position faces. See Jesse Prinz, *"Is morality innate.",* Moral psychology 1 (2008): 367-406.
[122] Ibid, 393-397.
[123] Gordon, Peter, and Denis Lawton, *"A history of western educational ideas."* Routledge (2019): chapter 2.

so. If we act morally towards several people in one day, say by helping someone pick up fallen groceries in the morning and carrying a baby stroller up the steps of the subway in the afternoon, we probably prefer to think that we have the character trait "helpful" or "compassionate" instead of just seeing ourselves as being compelled by some moral imperative or being placed in a series of situations where most people would offer their help. However, this is what is known as the correspondence bias, or fundamental attribution error: individuals believe that they or others possess inner dispositions that justify certain behaviors, whilst ignoring the significance of situational explanations for that same behavior. For example, in the Good Samaritan experiment conducted by Darley and Batson, the attribution error would consist in saying that the subjects that didn't stop to help the confederate failed to act morally because they are selfish, instead of recognizing the external pressure that the time constraint supposed for them.[124] The correspondence bias, among other findings in psychology, raised the question regarding the consistency of character traits and their influence on morality. Despite the questions that these psychological findings may raise, it is certain that, as humans, we undergo a process of growth and change as we live our lives. This process leads us to believe that we have something within us that makes us more than mere rule-following, unthinking moral robots or behaviorist homunculi with no inner moral contents. Something that distinguishes us as individuals from others, be it through the various dimensions of our personality, particular beliefs or other unique aspects of our psychological makeup. All of this information is, in essence, our identity, and we call the process that makes this identity individuation.[125] Identity, in the most fundamental sense, is what

[124] The correspondence bias was first brought to light in an experiment by Jones and Harris (1967), in which the subjects believed confederates writing an essay in support of Fidel Castro actually empathized with Castro even when told they had been assigned the argumentative position.

[125] By individuation, we refer to a general philosophical concept of identity acquisition, and not to the more specific concept used by sociology or by analytic psychology.

distinguishes us from our peers: it is the combination of innate attributes and personal education, the fruit of a process of both emotional and rational distinction from others that gives us our uniqueness. Our choices as moral actors are, therefore, tied to our identity whether we like it or not. The question, of course, is how strong this connection is.

Philosophical conceptions of morality centered around character answer this easily through the explanatory power of global traits. A moral trait is directly responsible for a moral behavior, and a problematic act is derived from a weakness in our character: a noble soldier dies to protect his allies, a cowardly one runs away from battle. However, the problem with these conceptions is that this causal force is often disputed, as these theories generally do not employ experiments in psychology to prove their points, often relying instead on abstract philosophy and conjecture. This leads some in the situationist tradition to hold a rather harsh stance, namely that character traits are, at best, irrelevant to decision-making, and at worst, that they are a fiction and do not constitute a part of our identity. Gilbert Harman, for instance, believes that character traits are artifacts of a "folk morality",[126] guided by gut intuitions which he considers scientifically unjustified. In Harman's eyes, social psychology provides the proof that was necessary to dismantle the moral tradition of character traits and moral character, and to make evident that correlating moral behavior with these traits is an attribution error that not only serves no purpose, but that can lead to a misinformed and dangerous interpretation of behavior.[134] Given this, Harman recommends abandoning this conception and acknowledging the power of situations when producing moral evaluations and giving moral education to people. This transition, he believes, will not necessarily devolve into a deterministic, behavioristic system, but will instead allow people to recognize that

[126] Gilbert Harman, *"Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error."* Proceedings of the Aristotelian society. Aristotelian Society (1999):1-17, 1-4 [134] Ibid. 14-17

the responsibility they often believe to bear within problematic scenarios is not as heavy a burden as they think. Harman's theory inevitably leads to a moral view that reduces the value of shame and guilt as moralizing forces, ultimately preferring moral relativism to moral inaccuracy. While

Harman's concerns are valid, they seem to minimize the role and importance of the more nonrational, intuition-driven side of human behavior, painting it as a slave to situational forces. While we have good reason to believe this more emotional side of our identity to be the one most easily influenced by the situational forces of the online environment, this does not necessarily mean that character is a void notion with no utility or that it is incompatible with moral psychology.

For instance, social psychologist Jonathan Haidt considers that moral decision-making is mostly based on personal intuition, which is then justified by rational means, after the fact. Haidt calls this social intuitionism, and argues that findings in psychology have demonstrated that emotional decisions are not as devoid of logic as the rationalist tradition had us believe. The "gut" moral intuition, far from being an uninformed and spontaneous phenomenon, results from deep internal conviction, and surfaces with immediacy because of the way our brains are wired. It is not necessarily compatible with what we believe is our moral stance, but it does speak about our inner convictions. Moreover, the findings of researchers like Daniel Kahneman and Amos Tversky indicate that the human mind processes information in multiple ways, and that most of our decision-making in our daily lives is made by fast processes that do not compute and treat information in the traditional understanding of rational thought, but are rather like a faster system of thought informed by past experience and first impressions, with little to no reflexive work.[127] In this configuration, moral psychology does not necessarily discard the existence or power of

[127] Daniel Kahneman, *"Thinking, fast and slow."*, Macmillan (2011).

character traits, but rather gives them a much-needed broadening. Indeed, it demystifies the idea of moral decisions being rationally guided decisions backed by strong emotion, presenting them instead as a moral feeling that leads to an action, and is then justified after the facts by a biased moral system. This is only one example of the many ways in which moral psychology has challenged our notions of rationality, morality and behavior; those in the field see crossdisciplinary analysis as a priceless tool to understand ethical problems, and have progressively developed a more and more complex understanding of moral identity and behavior, allowing for the consideration of situational forces as key theoretical elements. Indeed, theories of moral psychology, such as those presented by Appiah and Flanagan, emphasize the power of situations in determining behavior,[128] and could therefore be compatible with situationism even if they do not explicitly endorse it.

3.1.3 Local morality

To judge the situation online, we would need a theory that acknowledges the intricacies of the human mind brought forward by moral psychology, a theory that understands the complexity of situational variables on the one hand, and that grasps the necessity of a framework of moral evaluation on the other. A theory that is aware of the philosophical importance of psychological facts presented by researchers like Haidt and yet disposed to not entirely forego character traits like Harman does. We didn't have to look far from the situationist tree, to find such a theoretical approach in John M. Doris' "Lack of Character", a defense of situationism that does not do away

[128] Appiah, *"Experiments in ethics"* (2008) : chapter 2, and Flanagan, Owen J., and Sophie Marnat. *"Psychologie morale et éthique.",* Presses universitaires de France (1996): 363-385.

with character as a notion but rather limits it, accentuating the moral importance of considering the power of situations. Indeed, Doris believes that the problem with the virtue ethics tradition is that it has a conception of character that is far too global: brave individuals will be brave all the time, and being so will be a core aspect of their identity that will manifest whenever it is pertinent for it to do so. Doris believes this is not the case, but doesn't believe this disproves character traits per se:

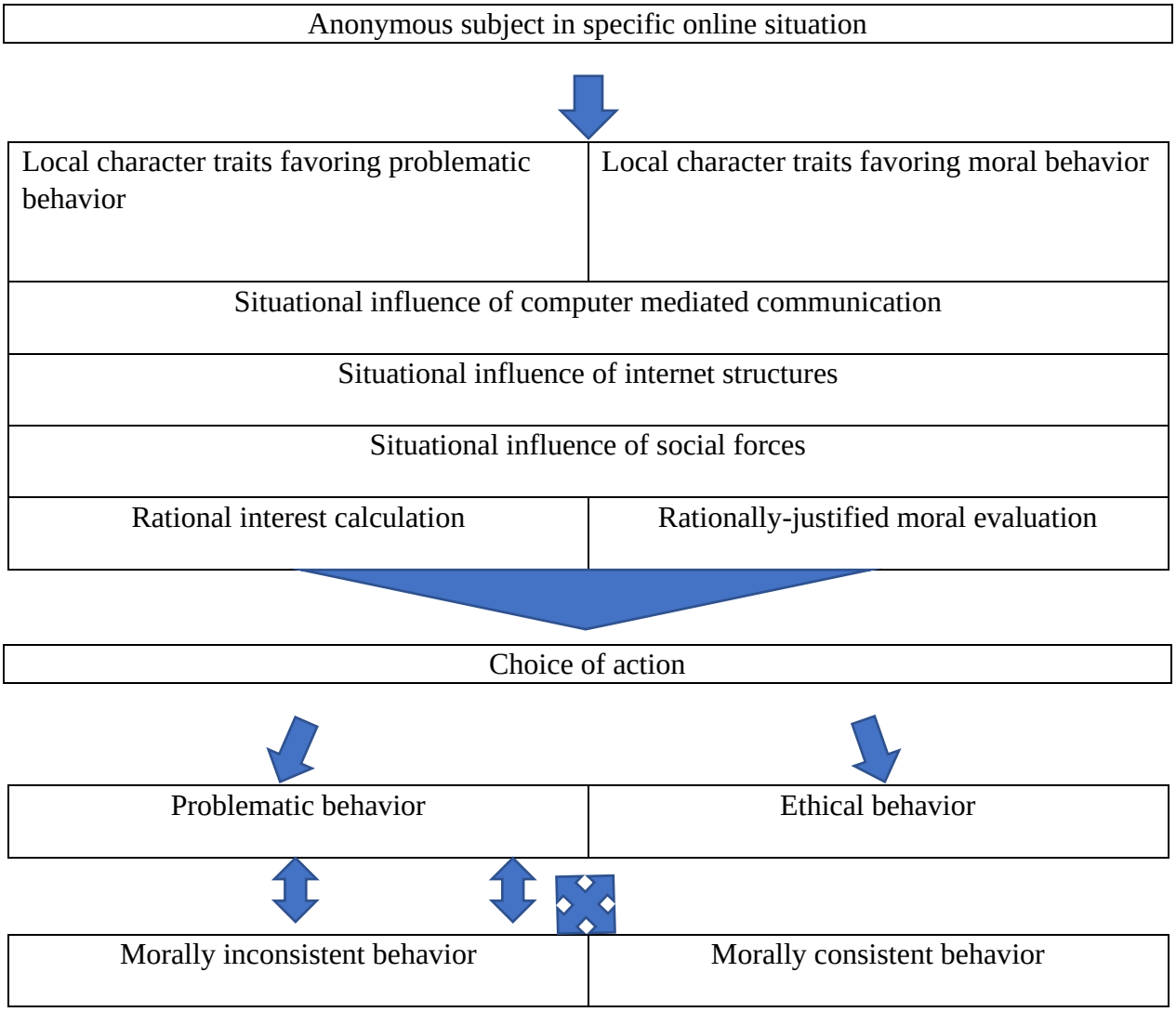
> "The situationist does not deny that people have personality traits; she instead denies that people typically have highly general personality traits that affect behavior manifesting a high degree of cross-situational consistency."[129]

To substitute the explanatory void of general or global character traits, Doris proposes the idea of "local" character traits, which are much more context-dependent and can still be consistently part of an individual's identity. These traits manifest within individuals not as widespread attitudes but in specific situations that elicit them. For example, someone who has the trait "patient with toddlers" might also have the trait "cruel to animals". Doris pushes this even further by suggesting that these traits do not even need to be consistent within a specific domain or sphere of our lives: a politician immune to bribery and corruption might be entirely capable of cheating on his wife, something incompatible with the idea of a general character trait like "integrity" or "loyalty". [130] Indeed, his conception of human personality is that it is "fragmented",[131] meaning that traits exist in wild variety instead of following a consistent, general line, and that this variety is what constitutes identity. Because of this variety, the influence of

[129] Doris, *"Lack of character"* (2002): 38-39
[130] Ibid. 21
[131] Ibid. 64, 69-74

situational forces becomes much less predictable, as these forces might affect different traits in different ways, and might be reacted to in different ways depending on the situation that individuals find themselves in. In Doris' model, instead of being floors of colored, sinuous marble, we are mosaics of character, and depending on the time of day, the angle of the sun and the brightness of the light, we might shine in entirely different ways.

The reason we believe Doris' situationism to be the theory that best deals with our problem should be evident now: it is a philosophical lens that both allows us to see individuals as being morally supple, with identities constituted of context-dependent traits, and that permits us to view problematic behaviors not only as the results of non-rational choice or weakness of character, but also as symptoms of situational influences. This would mean individuals can engage in problematic behavior without necessarily having character traits that lead to problematic behavior. In other words, what would traditionally seem like antisocial behavior would not necessarily be done by individuals who are characteristically antisocial themselves. Situationism also allows us to understand different social domains, environments and structures as being distinct enough in nature and composition so as to allow for individuals to engage in wildly disparate kinds of behavior, something that would not be possible with a conception of global moral traits like that defended by virtue ethics. Our interest is to see how these traits manifest under the influence of anonymity and to see how to evaluate them within the online world. Still, before we move on to do that, let us summarize the problem in question through the following diagram, which symbolizes the process of decision-making of the average anonymized internet user.



We've reorganized the order of things top-to-bottom to display the order in which we believe decision-making is made online while we are anonymous. Firstly, local character traits influence behavior as components of our identity. Users are then affected by situational forces, that either reinforce the trait's predisposed tendencies, alter them or weaken them. These situational forces are relatively unpredictable, but we will generally assume that they reinforce anonymity's effects. Finally, once all the more psychological influences have been exerted, there is an element of

rational choice, in which moral preferences struggle against non-moral ones within a person's inner logic. The end result of these aggregated influences is an action that will either constitute a problematic behavior or not. Then, finally, there is the question of whether this behavior fits the general pattern of behavior of that person. So far, based on the evidence we brought forward in our second chapter, we have good reasons to believe that when these actions are being performed by individuals relatively unidentifiable, there is a tendency to act in a way that is different from their moral standard in offline life. What we now seek to do is to try and understand the inner workings of users, and determine if anonymity interferes with their behavior to the point of making it hard for individuals to be consistent with their moral identity.

3.2 Existing online: new aspects of the moral identity

3.2.1 Confusing moral identity: Perspective and interpretation

In chapter 2, we presented anonymity and its effects, alongside all of the situational forces that might enhance them or otherwise modify them. The effects of anonymity identified by experiments in psychology reveal that, in great part, what it affects is our self-perception and our perception of others.[132] As we have just seen, morality is something deeply tied to our identity as individuals, and so, if anonymity fiddles with our self-perception, it is entirely possible that it is doing so by meddling with our identity. If so, we must determine why this influence is different enough in online and offline scenarios to warrant our attention. We will do so by seeing how the concealment

[132] We refer to the studies by Beaumeister et al.(2016), Gray et al. (2015) and Christopherson (2007). Also see Scott, Susan V., and Wanda J. Orlikowski (2014) and Joinson (1999).

of our identity can lead to a blurring or a reorganization of our moral preferences. Let us begin by talking about identity and perspective. As we pointed out previously, two key aspects of online structures are choice of experience and privacy. Indeed, users can easily customize their interactions when online, shaping the pathways and access points of information that they encounter as they select what content better fits their preferences. They can also choose how much information they share and how much they keep to themselves. We would consider this to be an experience over which users have a wide amount of control. As they browse, however, they can become deeply engaged in this highly customizable experience through phenomena like filter bubbles and echo chambers, and the scope of their perspective will become consistently smaller and more specific. All information received, treated and used to inform decision-making will therefore be not only partial, but most likely biased. This in turn might signify that, as their preferences become more and more adapted to the specificities of the environment, their decisionmaking will be influenced by the biased information, to the point of leading to decisions colored by this bias.

Let us consider this from a situationist perspective: if Alex has the local trait "loyal to her wife" and then joins a forum on the subject, this trait will probably not only be maintained, but become even more salient in the wake of group dynamics and CMC-induced bias.[133] The problem comes from the fact that, because of their limited perspective, users may have little to no incentive to evaluate the moral validity of their position, even if the internet provides them with plenty of opportunity to do so. When engaging with users that will consistently agree or echo their own views, individuals might cease to doubt their preferences, even as these become progressively more

[133] See Postmes, Tom, Russell Spears, Khaled Sakhel, Daphne de Groot, *"Social influence in computer-mediated communication: The effects of anonymity on group behavior.",* Personality and Social Psychology Bulletin 27.10 (2001): 1243-1254.

polarized by the gradual effects of the echo-chamber. For example, Alex, who has the local trait "loyal to her wife", most likely has the preference not to cheat, or the belief that "cheating on your spouse is wrong". But within an echo-chamber, this belief can easily become "cheaters are bad people", and the latter can then devolve into "cheaters should be punished". What was originally a simple opinion guided by a moral intuition has become a moral judgement, less concerned with integrity than with retribution. Or maybe the more polarized traits were already present within Alex, and simply needed to find an outlet for expression through an anonymized condition and a favorable environment.

Of course, this is not a problem exclusive to anonymous environments, as there are plenty of Facebook conversations that show the same signs of escalation.[134] What is particularly problematic from a moral standpoint, however, is that if we are anonymous, the sobering reminder of our own identity often doesn't follow us when we make these moral assertions. Our choice range, which was already quite wide, becomes even wider as we dissociate from our identity in picking specific subjects of interest. Then, when the pressure of online forces narrows our perspective, it narrows it to an even more specific level than if we had identified ourselves as we advanced our position.

In other words, when anonymized, it is much easier for Alex to hold any amount of beliefs over any number of subjects online because, on the one hand, she is not being tied down by the social conventions that would limit her speech in a conventional offline setting, and on the other hand, because she is under more significant pressure from online conventions. Likewise, it is easy

[134] Godwin's law famously states that most internet arguments eventually lead to one of the parties reducing the other to a Nazi or equating them to Hitler, thus providing a very telling fact about the escalation of arguments online.

for her to hold an increasingly harsh position against cheating if she is not identified as a young woman who has never been betrayed by her partner because, if pressed, her own offline identity is susceptible to render her opinion worth less in the eyes of her peers. Not only that, if she had given these key facts about herself, maybe her stance would not have been as polarized because she would recognize herself as being someone who is neither a perpetrator or a victim of infidelity, and who therefore doesn't have as pertinent a reason to have a strong opinion on the matter as someone who has gone through such events. In hiding our identity, anonymity might be giving us a myopic moral sense that will make us less susceptible to critically assess whether our online behavior is consistent with our offline behavior. However, anonymity might also allow for the manifestation of different local character traits that lie relatively dormant in us. Indeed, while not significant to our offline decision-making, these traits might become relevant with the change to an anonymous, online context.

The problem, then, is on of perception. By modifying our perspective, anonymity not only changes the way how we will be influenced by situations, but the way that we will react to them. After all, when we are online, most of our exchanges are the result of a personal interpretation. And given the complex game of influences operating over our mind, there is no hermeneutical procedure to accurately determine what fuels every comment, tweet or post a user makes when we read them, thus leaving us with the full task of developing meaning and understanding.[135] This both echoes and goes beyond the hermeneutical doubt that encircles every human conversation.[136] This is because when we are faced with words without a face, we will not only assign meaning to

[135] Joseph B. Walther, *"Computer-mediated communication",* (1996): 33-38.

[136] We mean by this the idea that all communication is interpretation and that we may never truly understand what someone else is saying. For more on this, see the first chapter of Ludwig Wittgentein's *Philosophical investigations* (John Wiley & Sons (2009)).

these words, but also assign them a face. In their words, we will interpret not only a belief, but a personality, traits of character, ideals and whatever else we need to turn a virtual message into a person's voice. We will give others a face to match our expectations, and they in turn will do the same for us. This game of interpretation makes it so that the identity we have online is, to a certain extent, crafted, both by us and other users, especially when we are unidentified. But we cannot ignore that there's also the matter of our actual identity, beyond the surface interpretation we and others make of it, and how our online behavior is a manifestation of it. We might have to consider the possibility that our personality is affected by online context, and therefore, by entering an anonymous condition online.

3.2.2 Complexifying moral identity: avatars, personas and narratives

Before the internet, living a double life was possible, but it wasn't easy. A married woman cheating on her husband with a lover; a young man from a conservative family who is secretly a homosexual; a teenage girl in a deeply religious family who is an atheist. These double lives define them as people, and yet they might be deeply problematic for them and others around them. The reason why we chose these examples is because they are clear cases of two conflicting lifestyles, but also of two conflicting elements in a person's life: they are different manifestations of their identity. The cheating wife has to balance the excitement of her infidelity with the moral guilt of deceiving her husband, but also the behavior of a loving wife with that of the passionate mistress. The gay man has to reconcile the expectations of his parents (providing them with a grandson or granddaughter) with the fact that he doesn't want to find a wife. And the atheist girl is certainly aware that, were her secret to be found, a chasm might form between her and her family. A double

life requires effort and planning, and it can all come crashing down if one world comes into even small contact with the other.

Now imagine that, through some magical artifact, we had access to another world, another dimension through a door to which only we had a key. In this other world, no one knows who we are, as a thin mist covers our features and others see only a faint impression of us. The wife could find a lover who would never know of or meet her husband, the young man could live without fear of prejudice and the girl could speak freely without fear of judgement. And, when they were all done, all they would have to do was to close the door and all of it would be left behind with the world, and locked away with that key. No proof, no ties: their actions are like sounds without an echo in their world. That is in a way the problem of the internet: to all intents and purposes, the virtual world we have created is a new, separate social space in very much the same way as the parallel dimension of our example is one. Within it, individuals can exist without being tied to their identity, and that means they can behave in a different way from what they usually do.

At a first glance, a user's identity constitutes little to no visible obstacle in their life online. However, we know that their local character traits, as well as the situational forces that may affect them, constitute conflicting forces that influence behavior despite of their being, and possibly even because they are unidentifiable. And, as we've stated above, the structure of the online environment and the nature of its interactions can affect our perceptions of others and, most importantly, of ourselves. We spoke of trolls in chapter 1, and of how they separate their online lives from their offline ones. A troll might be an exemplary student, a caring partner or a dutiful employee in their day-to-day life, but as soon as they log into their account, their behavior shifts completely. Their character traits should still be the same, and yet, they are capable of engaging in problematic actions that they never would offline. If we had a global conception of character traits, we could say that

it is the removal of accountability that leads to a disinhibition of these individuals who will consequently indulge in immoral behavior. This means that if they are engaging in these problematic actions, then there is a problem at their core, and their character traits are defective in the moral department. The layman's understanding of trolling as immoral behavior is similar, as the media coverage of the phenomenon seems to show,[137] and our portrayal of them in chapter 1 was not too different: the ideal self-image of the troll is that they are amoral, self-aware troublemakers that relish the rhetorical supremacy anonymity provides them. But if we look at it from a situationist perspective, the phenomenon of trolling becomes more complex, and can be seen as something more than malicious play by uninhibited actors.

Our character traits being local entails that our character is molded by the environments where moral questions are (or aren't) asked. Situational forces that would already trump a local moral trait of an individual in their offline life could be present with more intensity online, and become exacerbated by the medium; likewise, maybe all that is required is a change in environment for a local moral trait to lose against a situational force. Think, for instance, of bullying in schools. Engaging in bullying forces the individual to be confronted to the immediate moral implications of their acts: whether they like it or not, if one is to bully others, they often have to confront them, to see them suffer and cry. Physical presence is a significant factor for human beings because of how it can affect empathy and guilt, and removing this factor removes the immediacy of an act and its consequences. Furthermore, when online, this immediacy and emotional stress are not only lessened due to CMC, but their absence is felt more deeply because the situational forces that make bullying more palatable (such as social pressure to conform) are not only present online, but might

[137] Please refer to our earlier note on Phillips in page 81.

[146] Wingate, et al. (2013). *"Sticks and stones"*:87-106.

be even stronger.[146] This is not to say that all forces will systematically go against a user's local traits, but more to point out that a shift in circumstances is all that might be required for a local moral trait to become overwhelmed by situational forces. With this in mind, we will now try to provide our interpretation of the problem.

3.3 The heart of the problem: Moral inconsistency in a digital world

3.3.1 Anonymity as a problem of moral confusion

When we began this project, we did so by basing ourselves on our observation that there is problematic behavior online. We did not go into much length as to why this behavior was wrong, choosing to pick examples that were very intuitive and to appeal to the layman's understanding of right and wrong in determining what constituted problematic behavior with things like lying, cheating and mistreating others. But we can no longer be content with that. After all, we have, on several occasions, given a description of what we understand to be the field of ethics, and what is understood as being problematic from that standard. Not only that, moral local traits, as something tied to an individual's psychological profile and identity, are not only a vector of behavior but a standard by which to judge others. In this last section of our work, we will try and produce our own judgement of what to make of the problematic behavior of anonymous actors.

In their book *Evil Online*, Dean Cocking and Jeroen Van Den Hoven present a general answer to the issue of problematic behavior online. Their work visits many of the themes and subjects we have tackled in this thesis and, although it doesn't specifically focus on anonymity, it was a fundamental source in understanding the complexity of the problem. Their hypothesis is that

when an average user who is neither a sociopath nor a saint engages in problematic behavior online, they are most likely enveloped in a "moral fog of evil".[138] For the authors, this "fog" is the result of the very structure of the online world, which they believe to be a broken social environment that is impersonal and distant, molded by systems that are much unlike the social structures and environments that human beings frequented in the pre-internet world. According to Cocking and Van Den Hoven, "our online-transformed worlds have delivered new and widespread forms of moral fog that limit and negatively shape moral imagination and understanding."[139] This fog identified by the authors could easily be seen as the particular combination of situational forces that, when exerted over the right individual with a specific set of local character traits, would lead them to act against what they could rationally judge and intuitively be disposed to consider as moral. While they refrain from taking an explicit position in the moral debate on character versus situations, they do indicate that they believe that a conception of moral virtue centered on general character traits, while not necessarily correct, is preferable in better preparing individuals for the moral challenges they will face online.[140] We take this to mean that the authors believe that the problem is, after all, indeed a matter of "fogged up" moral compasses, and that the solution is to instill both situational awareness and strong moral awareness in individuals online.

The arguments we developed in the first and second chapters, along with the experiments we cited and the evidence we brought forward, all serve to show that anonymity exerts a kind of pressure that matches what Cocking and Van Den Hoven describe. Anonymity, then, could be one of the branching tendrils of the moral fog of evil, as through its peculiar influence it changes the

[138] Cocking et al, *"Evil Online"* (2018): preface, XVI
[139] Ibid. 147
[140] Ibid. 145-149

way that users exist within the online environment and the way that they interact with each other. This does not necessarily mean that anonymity systematically predisposes individuals to engage in problematic behaviors, but rather that it puts them in a condition that tampers with their perception, and in turn their moral dispositions and the hold they have on their decision-making. Consequently, we hope to have shown that, when the balance between local character traits, situational forces and rational calculation is done, individuals provided with anonymity are likely to act in a way that is inconsistent with the dispositions they hold offline. Therefore, anonymity can be understood as a complex modifier of behavior that mainly affects the moral dispositions of individuals by heightening the effects of internet-borne situational forces. These range from particularities in the internet's structural composition to the online translations of offline social forces, and the emergence of new, internet-dependent ones. This finding was the goal that we sought at the beginning of our study, and through the lens of moral psychology and situationism, we believe that it has allowed us to better understand what drives some of the problematic behavior by anonymous actors online.

However, there are still some behaviors that the moral fog of evil cannot explain as consistently. The authors of *Evil Online* often mention the aspect of plurality in identity, and the idea that there is a certain multimodality to behavior that has to be taken into consideration.[141] It seems that, as much as we'd like to exclusively blame a moral fog for our behavior online, there is still plenty of evidence to suggest that people who are engaging in problematic behaviors online often do not recognize themselves as being in the wrong until confronted by others and pulled

[141] Ibid., 113-114

away from their own frame of perception. The problem, then, seems to also be one of consistency in their moral identity.

3.3.2 Anonymity as a problem of identity

Are we morally consistent when we are online? There are two ways to understand this question. Firstly, it could be understood in the sense that we, as human beings, are morally consistent on a daily basis. We must here note that being morally consistent does not mean that we are necessarily moral; rather, it means that our moral decision making follows a pattern that fits our identity. For example, someone with the local character trait "fair in bartering", when exposed to situational forces that are not too overwhelming, can make a decision that is rationally consistent with their local trait. Being inconsistent, then, means that we will go against both our local character traits and our rational moral preferences because we are under the influence of situational forces: that is, in essence, what the moral fog does. Assuming we act consistently on a day to day basis, we do not behave consistently online because the moral fog increases our chances of engaging in behavior that is "not like us", and we are therefore less likely to be morally consistent online than when we are out in the real world. But there is the second aspect of the question, which concerns whether we are acting consistently not only from one environment to another, but within each one specifically. In this sense, consistency does not refer to a general capacity to act morally, but to the capacity to act morally in specific ways, on specific situations. We are already working under the assumption that there are no general character traits, and that any character trait that exists is local and elicited by specific situations. We would like to suggest, at this point, the idea that onlinelocal traits and offline-local traits operate on different levels of preference, and have their own requirements of consistency.

A common trend among problematic acts online like cyberbullying, harassment and grooming is that those responsible for them are often unaware of how responsible they are, and unaware of the moral weight of their actions (these often lead to grave consequences for their victims from persistent psychological troubles to suicide)[142]. Like with the Stanford Prison experiment, it seems that some of these problematic actions are so inconsistent with the moral profile of their perpetrators that it is as if someone else was committing them. Looking back at trolls, the way they engage with the online world has often been compared to a sort of Dr. Jekyll and Mr. Hyde transformation,[143] where polite and friendly individuals become virulent practical jokers with no empathy for their victims. It is no wonder that one would assume their personality to be entirely different because, to a certain extent, it might be. To clarify, we are not about to launch into a technical discussion of the psychology of multiple personalities and identity disorders. Our interest is in how the average user's moral identity and behavior might be influenced by anonymity. But there is something to be said about the fact that being online is, to a great extent, an exercise in self-exploration and expression. CMC means that existing online requires mediation. This means there is a selection, preparation and premeditation of our self: it is, in a way, a form of role-playing.

The model of most internet social environments is one of self-description. Individuals pick their profile pictures; they describe themselves in their own words. They paint their own picture to the world, sharing their likes and dislikes and other affinities. If they do so when they are identifiable, in social media for example, then they are extending their identity to the online medium (of course,

[142] Ibid. 101-107

[143] Phillips, *"This is why"* (2015): 21-26

even here there is a dimension of fabrication and roleplaying).[144] Doing the same thing when one is anonymous, however, is an entirely different process. Instead of engaging in an act of complementation, when we are anonymous, every act is self-defining. In every choice, every expressed preference or every thoughtless action we provide information about ourselves. It might be a result of a local character trait manifesting or it might be the influence of a situational force, but the end result is the same: we've attributed a shape, a tone or a color to something that, before, was nothing more than an aggregate of unmarked data. When we are anonymous online, we are given the option of simulating a start from nothing and building ourselves up: we can create a persona with which to interact with others. In online videogames, for instance, users often take control of a character and experience the world through that character, developing a narrative through which they will engage with other users of the virtual world. Often highly customizable experiences, users are allowed to customize fundamental aspects of their characters like gender and race, but also to narratively engage with notions of good and evil, easily switching from villainous to noble on a whim. Of course, we are not advancing some preposterous, unscientific notion like the idea that playing a terrorist in a videogame makes users more likely to engage in terrorism,[145] but we are noting that these narrative options allow users to fiddle with identity in a way that is extremely novel and probably not entirely understood. Not only that, users can become deeply entangled in these narratives in a variety of ways, to the point of becoming addicted to the life they lead within the game.[146] If a person is performing the vast majority of their social

[144] Tsikerdekis, Michail, and Sherali Zeadally, *"Online deception in social media.",* Communications of the ACM 57.9 (2014): 1-18.

[145] While there are signs of the influence of videogames in children, teenagers and young adults, research has provided much more insight in regards to how addictive video games are than on whether they induce violent behavior or desensitize players. For more information, see Kowert, Rachel, and Thorsten Quandt (edt), *"The video game debate: Unravelling the physical, social, and psychological effects of video games."*, Routledge (2015). Specifically chapters 4 and 5.

[146] Gray, et al. "More than addiction" (2015): 138-145.

interactions through a videogame, then it is only to be expected that these interactions are conforming to the norms and social conventions of that environment, and that they might become just as significant to this person as offline norms and conventions. Of course, we extend this same logic to other online social environments like those we spoke of in the second chapter; the main idea is that anonymity usually allows for a certain degree of fabrication.

Likewise, some local traits that might be obstacles to interaction, like someone who is "shy when meeting new people", might not manifest when online, or might even manifest differently while online due to continuous exposure or the change in conditions. Therefore, it might be the case that as soon as these individuals assume their online persona, their decisions become motivated by different local character traits from the ones that move them on their offline routine. In other words, when we go online, we might be accessing a separate array of traits adapted to the internet environment. This, however, doesn't by any means entail that we are different people, but rather that different local traits might be manifesting depending on context, and that our personality might in turn be more fluid than we expect. Like with the parallel universe we presented, going from online to offline shouldn't change who we ultimately are: the adulterous wife, the homosexual bachelor and the atheist teenager might all change their behavior as soon as they enter the other world, but the moment they are back home they will switch gears. However, it is unlikely that their identity will remain unaffected if they continually cross the threshold.

Will a person be as willing to decry what they perceive as immoral offline as they would online? The local traits, situational forces and rational choices that lead us to stand up to a racist in a supermarket will certainly be relatively different from those that move us to decry a racist comment on YouTube. Decrying immorality is an action that is not only evidently easier to do

online, but that is clearly not an act that is equal across these two disparate worlds.[147] If we refer to different sets of local traits when we are online, then it becomes entirely possible to understand how trolls so easily shed their ordinary behavior to become vicious practical jokers: they might not be undergoing a transformation like Dr. Jekyll, but rather reorganizing their priorities and preferences to better fit their new environment. This change in priorities is not unlike what occurs when individuals working in a business have to change the way they treat people, or the way individuals behave when they are around their close ones. In other words, when we are online and under the influence of situational forces, we might not only be engaging in problematic behavior, but we might be doing so in a way that we perceive as being consistent, maybe even morally consistent if our bias is strong enough. After all, moving into the online environment changes how we perceive others, how we perceive ourselves, how we prioritize things, and how relevant our moral preferences are amidst this complex game of choices. And having a progressively bigger gap between our ways of behaving is certain to have an influence on us in the long run. After all, if through our regular visits to the online medium our local traits have adapted to be as fluid as our moral preferences seemingly are, then the validity of our moral considerations might be at stake. As the barrier between online and offline becomes both less and less evident and increasingly delineated, we must ask ourselves if the internet is still something we can treat as having an on and off switch, and whether this means that our morality is being constantly molded by it or if we are compartmentalizing our morality because of it.

3.3.3 Persisting anonymity

[147] Shared responsibility, social pressure and social inadequacy might prevent us from speaking out.

We can at this point consider that some of the problematic behaviors online stem from the fact that individuals might be under situational forces so significant that their local traits are not strong enough to swim against the current. The idea that moral local character traits become more or less effective in influencing our behavior in the face of a shift in situational forces is a key assumption of situationism. This assumption is supported by the SIDE, as well as some of the experiments in psychology we cited,[148] and we believe that some of the examples of problematic situations we cited at the end of chapter 2 were good indicators of this phenomenon. However, some cases are more elusive, as a mere shift in conditions seems insufficient to explain the change in behavior.

Problematic behavior, as we've seen, is often a problem of our moral identity's consistency. On the one hand, because online behavior is often inconsistent with offline behavior. Users can quickly engage in problematic behavior as soon as they change environments and enter an anonymized condition, losing sight of their offline moral standards and preferences. On the other hand, because online behavior is often consistent with itself. As users maintain this consistency in their problematic behavior, they can find themselves going down a rabbit hole of decisions that, framed as they are by the internet environment, might continually be perceived as coherent and non-problematic but are, ultimately, self-justifying.

The structure of the online world gives us a sense of privacy, but its mechanisms end up profiling us. We are exposed to an increasing amount of information, but most of it is selected to fit our preferences. Our interactions with others are all disconnected from time and space, and rather than a game of communication, we all play a game of interpretation in which our biased views are the final determinant in what things really mean. We are given an identity by the internet

[148] Please refer to the studies mentioned on page 94.

and yet we may do all of the above acts unidentified. We can choose not to have a face just as the internet paints us one, as the mechanisms of the online world and the perceptions of fellow users mold us into something that they can make sense of. We are not, at heart, truly humans living in this virtual world. We are more like impressions, vague descriptions of something that doesn't really exist. Untied from our identities, from the moment we are online we become far more susceptible to disregard our offline moral dispositions because we fear being identified, or because they are inconsistent with the expectations that are projected onto us by others and ourselves while we exist online. Maybe even because the internet is another place, and because here, we do not have to remain chained to who we are outside. If anonymity is a problem, it is not only because it makes us more susceptible to the situational effects of the online world that can corrode or confuse our moral preferences, but because it more easily allows us to complexify and multiply these preferences. The local traits of our offline self (the individual that exists behind the screen) might become progressively more distinct from the local traits of our online self (the anonymous individual that exists beyond the veil of programming and data), and in doing so, our moral preferences might not only be confused, but diffused in order to better fit the conditions of these different environments.

Maybe, after this entire time, one could accuse us of doing a thesis on pseudonymity, and of disregarding the fact that all of this should only apply to those who don't identify themselves online. Both these complaints are fair. Of course, we may do all of this through the disguise of a persona, or even fully identify ourselves. We can put up profile pictures and names, we can try and abide by the same ideals that move us offline and tie ourselves to our own identity. But this doesn't ultimately change the fact that others will continue to make their own interpretations of who and what we are when we enter the virtual world. Rarely will they look beyond the first impression,

and when they do look beyond, it is only to interpret their own understanding of who we are within the online narrative. Because when we are online, to the vast majority of the world that doesn't truly know us, has never seen us or talked to us, and has no reason to be interested in the person behind the screen, to that vast majority we are, to all intents and purposes, anonymous.

Conclusion

Given the tone of our last section, one could expect the diagnosis made by this thesis to be a dire warning against the dangers of anonymity and the internet. Indeed, we began by developing a technical explanation of why anonymity facilitates problematic behavior by endowing users with excessive freedom and strategic advantages over others, thus making it highly plausible that those among us with no strong moral preferences have little to no reason to act morally online. We then presented situational forces that, thanks to the nature of the internet, can coordinate with anonymity to influence users and facilitate these problematic behaviors. Finally, we developed an argument for why character is not a moral bedrock against these problematic acts. By presenting character as a localized, highly susceptible feature of the human identity in a situationist model, we advanced that identity is not only not the main reason why people refrain from acting immorally, but that it might be a part of the reason why they engage in problematic acts in the first place. Overall, we determined that problematic behavior online is something that can be expected out of any of us, be it because we are under the overwhelming influence of coordinated situational forces that dull our moral dispositions, or because these dispositions have become progressively adapted to the online environment due to our continued existence within it, and are consequently not entirely consistent with our offline moral dispositions.

However, we believe that this presents a new direction the field of ethics can explore, and this in two ways. Firstly, we will propose that, to better understand and control the effects of the moral fog of evil, it could be interesting to use a situationist perspective. If we accept the existence and strength of situational forces, and progressively move away from folk conceptions of moral psychology, we might be able to avoid confusion in the face of problematic acts online. According to the recommendations of Doris:

> "If we take situationism to heart, we […] will be reluctant to evaluate persons in terms of robust traits or evaluatively integrated personality structures, because we will think it highly unlikely that actual persons instantiate such psychological features. Accordingly, we will be unwilling to speak in terms of general evaluative categories such as 'good person' and 'bad person'".[149]

Situationism requires us to change our frame of mind in a significant way. Even Doris recognizes how counterintuitive and difficult it is to abandon the notions of folk psychology in the real world. When we are confronted with people, we are predisposed to try and unite the elements of their identity to form a general idea of who they are in our eyes. We are therefore not suggesting that individuals must take this approach in their day to day life: more comprehensive philosophical theories like Virtue Ethics remain much more intuitive and, as Van Den Hoven and Cocking advance, it might be preferable to inform moral education. But when we are online and exposed to mere fragments of these people, it would be easier to see their actions as being the fruit of local traits influenced by unique situations, especially if these people are unidentifiable. Indeed, we could become less prone to reduce and generalize the interpretations we make of the identity of

[149] Doris, *"Lack of character"* (2002): 114

our fellow users if we can't be given their general picture from the very get go. Likewise, we can begin perceiving our own choices online as being the complex results of a work in selfpresentation, and become more aware that being anonymous online is a change in conditions that will most likely affect how we behave. Most importantly, conceiving of things in this way allows us to move away from a global character conception that reduces problematic behavior to acts of problematic people, and allows us to better understand it as a byproduct of circumstances. The idea that problematic behavior online is issued from problematic conjunctions of factors and not from morally deficient people might be harder to grasp for the average user, but we believe that it is necessary to change this perspective. After all, trying to prevent problematic behavior requires an understanding of where it comes from, and the assumption that it is issued from a few bad apples is not only incorrect, but prejudicial to the entire project of ethics online. If we are to judge and understand problematic cases online, then we have to try and see beyond the first impressions and overgeneralizations.

Secondly, we would like to broaden our interpretation of the way anonymity influences identity by suggesting that anonymity is not necessarily a problematic condition to be in. The online world has provided us with an opportunity to exist beyond the constraints of our daily lives. People can connect across social, economic and cultural divides, and just as they can learn about each other, they can choose to not learn about each other. In other words, information that could lead to bias and prejudice is evacuated, and only the things that bring users together remain. Of course, this doesn't mean that the bias will be gone, as it is part of a person's identity, and users might discover aspects of the identity of their fellows that are not pleasant or commonplace for them. But being confronted with the parallels in interests and beliefs we find with others who can be drastically different from us can be an enlightening experience. In this way, social media friends,

videogame players and forum denizens can all experience exchanges that, while limited in their information about the emitter, can be very rich in their content. They might be filtered and processed through the online medium, they might be performed under the influence of situational forces and they might even be inconsistent with a personal disposition, but this doesn't entail that these aren't earnest forms of self-expression and sharing. People that are not going online to sow disorder or to pick up fights can easily find spaces where all that matters is the matter at hand, and no other element of their identity needs come into play unless they wish to bring it forward. Their arguments can be measured as words free from their speaker, and their ideas can be seen for what they are and not as a byproduct of who they belong to. And, if they were to shed their mask and find out that the others are very unlike them in every other aspect of their identity, maybe it would show that interactions that would never have occurred offline may occur online thanks to anonymity.

If we had to describe anonymity online in one word, we would say that it is liberating. It frees us from ethical constraints and rules, but it also frees us from the judgement of others. It allows us to lie, deceive and confabulate, but it also allows to exert our imagination, to express ourselves and to unbind from a rigid identity. It is a versatile tool, and as we have seen, it is far from being innocuous. That is why, if we are to develop norms, conceive of practices and encourage attitudes towards functional ethics online, there is a need to give anonymity the attention it deserves. We must understand that, when anonymous, we are playing a different game, with different rules, and that it might make us lose sight of what we truly value, or make us believe that we value something less than we actually do. We must be aware that it can shift our perspective, and make us especially susceptible to the influences of this brave new virtual world, this complex ever-shifting environment that is so different from the world we were born into. And it is our hope that, once

we've properly adapted our understanding to better grasp the complexity of anonymity, maybe through it we can begin to better understand what the increasing amounts of time we spend online will mean for our conceptions of ethics.

Bibliography

Agatston, Patricia, Robin Kowalski, and Susan Limber. "Youth views on cyberbullying.", Cyberbullying prevention and response, Routledge (2012): 68-82.

Andreassen, Cecilie Schou, Ståle Pallesen, and Mark D. Griffiths, "The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey.", Addictive behaviors 64 (2017): 287-293.

Anne Ferry, "Anonymity": The Literary History of a Word, New Literary History 33.2 (2002), 193-214.

Appiah, Kwame Anthony. "Experiments in ethics". Harvard University Press (2008).

Arendt, Hannah. Eichmann in Jerusalem. Penguin (2006).

Ariely, Dan, "Predictably irrational.", HarperCollins (2008).

Asch, Solomon E. "Studies of independence and conformity: I. A minority of one against a unanimous majority." Psychological monographs: General and applied 70.9 (1956): 1-70.

Barak, Azy, and Orit Gluck-Ofri. "Degree and reciprocity of self-disclosure in online forums." CyberPsychology & Behavior 10.3 (2007): 407-417.

Baron, Robert A. "The sweet smell of... helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls." Personality and Social Psychology Bulletin 23.5 (1997): 498-503.

Baron-Cohen, Simon, "The science of evil: On empathy and the origins of cruelty.", Basic books (2012).

Baumeister, Roy F., Sarah E. Ainsworth, and Kathleen D. Vohs. "Are groups more or less than the sum of their members? The moderating role of individual identification.", Behavioral and Brain Sciences 39 (2016).

Benson, Phil. The discourse of YouTube: Multimodal text in a global context. Routledge, 2016.

Berners-Lee, Tim. "Answers for young people." World Wide Web Con (2011).

Blackburn, Jeremy, Nicolas Kourtellis, John Skvoretz, Matei Ripeanu, Adriana Iamnitchi, "Cheating in online games: A social network perspective.", ACM Transactions on Internet Technology (TOIT) 13.3 (2014): 9.

Chen, Vivian Hsueh Hua, and Yuehua Wu, "Group identification as a mediator of the effect of players' anonymity on cheating in online games.", Behaviour & Information Technology 34.7 (2015): 658-667.

Chesney, Thomas, and Daniel KS Su. "The impact of anonymity on weblog credibility." International journal of human-computer studies 68.10 (2010): 710-718.

Chiou, Wen-Bin. "Adolescent's sexual self-disclosure on the internet: deindividuation and impression management." Adolescence 41.163 (2006): 547.

Chu, Zi, Steven Gianvecchio, Haining Wang, Shushil Jajodia, "Who is tweeting on Twitter: human, bot, or cyborg?." Proceedings of the 26th annual computer security applications conference. ACM (2010).

Cocking, Dean, and Jeroen Van den Hoven "Evil online.", Vol. 15. John Wiley & Sons (2018).

Cohen, Daniel J. Digital history: A guide to gathering, preserving, and presenting the past on the web. University of Pennsylvania, 2005.

Coleman, Gabriella, Hacker, hoaxer, whistleblower, spy: The many faces of Anonymous. Verso books, 2014. Chapter 1.

Conti, Gregory, and Edward Sobiesk. "An honest man has nothing to fear: user perceptions on web-based information disclosure." Proceedings of the 3rd symposium on Usable privacy and security. ACM, 2007.

Cover, R. (2012). Performing and undoing identity online: Social networking, identity theories and the incompatibility of online profiles and friendship regimes. Convergence, 18(2), 177-193.

Darley, John M., and Bibb Latané. "Bystander intervention in emergencies: Diffusion of responsibility." Journal of personality and social psychology 8.4p1 (1968): 377.

Darley, John M., and C. Daniel Batson. "" From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior." Journal of personality and social psychology 27.1 (1973): 100.

DeHue, Francine, Catherine Bolman, and Trijntje Völlink. "Cyberbullying: Youngsters' experiences and parental perception." CyberPsychology & Behavior 11.2 (2008): 217-223.

Doris, John M., "Lack of character: Personality and moral behavior.", Cambridge University Press (2002)

Drummond, Aaron, and James D. Sauer. "Video game loot boxes are psychologically akin to gambling." Nature human behaviour 2.8 (2018): 530.

Ellison, P. A., John M. Govern, Herbert L. Petri, & Michael H. Figler, Anonymity and aggressive driving behavior: A field study, Journal of Social Behavior and Personality 10 (1995), 265.

Elster, Jon. Strong feelings: Emotion, addiction, and human behavior. MIT Press, 2009.

Fägersten, Kristy Beers, "The role of swearing in creating an online persona: The case of YouTuber PewDiePie.", Discourse, context & media 18 (2017): 1-10.

Fallows, Deborah. Search engine users: Internet searchers are confident, satisfied and trusting--but they are also unaware and naive. Pew Internet & American Life Project (2005).

Feigenbaum, Joan, and Bryan Ford, "Seeking anonymity in an internet panopticon.", arXiv preprint arXiv:1312.5307 (2013).

Felman, Shoshana. "Madness and philosophy or literature's reason." Yale French Studies 52 (1975): 206-228.

Field, Andrew, "Waterloo: the French perspective", Pen and Sword (2012).

Flaxman, Seth, Sharad Goel, and Justin M. Rao, "Filter bubbles, echo chambers, and online news consumption.", Public opinion quarterly 80.S1 (2016): 298-320.

Flores, Andrea, and Carrie James. "Morality and ethics behind the screen: Young people's perspectives on digital life." New Media & Society 15.6 (2013): 834-852.

Frankfurt, Harry G., "On bullshit.", Princeton University Press (2009).

Gewirth, Alan. Reason and morality. University of Chicago Press, 1978.

Gordon, Peter, and Denis Lawton, "A history of western educational ideas." Routledge (2019).

Gray, Kishonna, and Wanju Huang, More than addiction: Examining the role of anonymity, endless narrative, and socialization in prolonged gaming and instant messaging practices, Journal of Comparative Research in Anthropology & Sociology 6.1 (2015).

Griffin, Robert J., Faces of Anonymity: Anonymous and Pseudonymous Publication, 16002000, Springer, 2003

Gross, Ralph, and Alessandro Acquisti, "Information revelation and privacy in online social networks.", Proceedings of the 2005 ACM workshop on Privacy in the electronic society (2005).

Gubar, Susan, and Joan Hoff, eds. "For adult users only: the dilemma of violent pornography." Indiana University Press (1989).

Guéguen, Nicolas. "The sweet smell of… implicit helping: Effects of pleasant ambient fragrance on spontaneous help in shopping malls." The Journal of social psychology 152.4 (2012): 397-400.

Haidt, Jonathan. "Morality." Perspectives on psychological science 3.1 (2008): 65-72.

Harman, Gilbert. "Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error." Proceedings of the Aristotelian society. Aristotelian Society, 1999.

Hinduja, Sameer, and Justin W. Patchin, "Bullying, cyberbullying, and suicide.", Archives of suicide research 14.3 (2010): 206-221.

Hume, David. "An enquiry concerning human understanding." Seven Masterpieces of Philosophy. Routledge, 2016. 191-284.

Isen, Alice M., and Paula F. Levin. "Effect of feeling good on helping: cookies and kindness." Journal of personality and social psychology 21.3 (1972): 384.

Jean-Jacques Rousseau, " « Discours sur l'origine de l'inégalité », (1755), Les Échos du Maquis, v. 1,0, 2011 (1755).

Jenson, Joli. "Fandom as pathology: The consequences of characterization." The adoring audience: Fan culture and popular media (1992): 9-29.

Johnson, Deborah G. , "Ethics online", Communications of the ACM 40.1 (1997), 60-65.

Joinson, Adam. "Social desirability, anonymity, and Internet-based questionnaires." Behavior Research Methods, Instruments, & Computers 31.3 (1999): 433-438.

Jones, Edward E., and Victor A. Harris. "The attribution of attitudes." Journal of experimental social psychology 3.1 (1967): 1-24.

Kandel, Eric, "In Search of Memory: The Emergence of a New Science of Mind", WW Norton & Company (2007).

Kang, Ruogu, Stephanie Brown, and Sara Kiesler. "Why do people seek anonymity on the internet?: informing policy and design.", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2013).

Kimberly M. Christopherson, "The positive and negative implications of anonymity in Internet social interactions:"On the Internet, Nobody Knows You're a Dog".", Computers in Human Behavior 23.6 (2007): 3038-3056.

Kowert, Rachel, and Thorsten Quandt, eds. "The video game debate: Unravelling the physical, social, and psychological effects of video games.", Routledge (2015).

Landau, Susan, "Making sense from Snowden: What's significant in the NSA surveillance revelations", IEEE Security & Privacy 11.4 (2013), 54-63.

Le Bon, Gustave, "The crowd: a study of the popular mind.", London: Ernest Benn (1895, trans. 1947).

Lea, Martin, and Russell Spears. "Computer-mediated communication, de-individuation and group decision-making." International journal of man-machine studies 34.2 (1991): 283-301.

Lerner, Jennifer S., and Dacher Keltner. "Beyond valence: Toward a model of emotionspecific influences on judgement and choice." Cognition & emotion 14.4 (2000): 473-493.

Livingstone, Sonia. Children and the Internet, Polity, 2009.

Ludwig Wittgenstein, *"Philosophical investigations."*, John Wiley & Sons (2009).

MacIntyre, Alasdair. After virtue. A&C Black (1981/2013).

Mackie, Diane M. "Social identification effects in group polarization." Journal of Personality and Social Psychology 50.4 (1986): 720.

Marr, Bernard. "how much data do we create every day? The mind-blowing stats everyone should read." Forbes, May 21st, www. forbes. com/sites/bernardmarr/2018/05/21/how-muchdata-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/, accessed July 10th (2018).

Marx, Gary T., What's in a Name? Some Reflections on the Sociology of Anonymity, The Information Society 15.2 (1999): 99-112.

McClelland, John S. The Crowd and the Mob (Routledge Revivals): From Plato to Canetti. Routledge (2010).

McMillan, David W., and David M. Chavis, "Sense of community: A definition and theory.", Journal of community psychology 14.1 (1986): 6-23.

Metzger, Miriam J., Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research, Journal of the American Society for Information Science and Technology 58.13 (2007), 2078-2091.

Milgram, Stanley, "Behavioral study of obedience.", The Journal of abnormal and social psychology 67.4 (1963): 371-378.

Miller, Christian B. "Character and situationism: New directions." Ethical Theory and Moral Practice 20.3 (2017): 459-471.

Mislove, Alan, Bimal Viswatah, Krishna P. Gummadi, Peter Druschel, "You are who you know: inferring user profiles in online social networks." Proceedings of the third ACM international conference on Web search and data mining. ACM (2010).

Misoch, Sabina, "Stranger on the internet: Online self-disclosure and the role of visual anonymity.", Computers in Human Behavior 48 (2015): 535-541.

Nielsen, Torben H. and Jürgen Habermas, "Jürgen Habermas: Morality, Society and Ethics: An Interview with Torben Hviid Nielsen." Acta Sociologica (1990): 93-114.

Nissenbaum, Asaf, and Limor Shifman. "Internet memes as contested cultural capital: The case of 4chan's/b/board." New Media & Society 19.4 (2017): 483-501.

Nozick, Robert. Examined life: Philosophical meditations. Simon and Schuster, 1990. Olson, Parmy, “How 4Chan's Chris Poole Runs His Web Giant On A Shoestring”, Forbes, October 18 (2012).

Ozawa-de Silva, Chikako, "Too lonely to die alone: Internet suicide pacts and existential suffering in Japan.", Culture, Medicine, and Psychiatry 32.4 (2008): 516-551.

Palme, Jacob, and Mikael Berglund. "Anonymity on the Internet." Retrieved August 15 (2002): 2009.

Peddinti, Sai Teja, Keith W. Ross, and Justin Cappos, "On the internet, nobody knows you're a dog: A Twitter case study of anonymity in social networks.", Proceedings of the second ACM conference on Online social networks (2014).

Pedersen, Darhl M. "Dimensions of privacy." Perceptual and motor skills 48.3_suppl (1979): 1291-1297.

Philippa Foot, “Natural goodness.”, Clarendon Press (2003).

Phillips, Whitney “This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture.”, MIT Press (2015).

Postmes, Tom, Russell Spears, Khaled Sakhel, Daphne de Groot, "Social influence in computer-mediated communication: The effects of anonymity on group behavior.", Personality and Social Psychology Bulletin 27.10 (2001): 1243-1254.

Prinz, Jesse. "Is morality innate." Moral psychology 1 (2008): 367-406.

Reer, Felix, and Nicole C. Krämer. "Underlying factors of social capital acquisition in the context of online-gaming: Comparing World of Warcraft and Counter-Strike." Computers in Human Behavior 36 (2014): 179-189.

Reicher, Stephen D., "Social influence in the crowd: Attitudinal and behavioural effects of de-individuation in conditions of high and low group salience.", British Journal of Social Psychology 23.4 (1984): 341-350.

Rozin, Paul, Jonathan Haidt, and Clark R. McCauley, "Disgust.", In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), Handbook of emotions (pp. 757-776), The Guilford Press (2008).

Russell, Bertrand, “History of western philosophy: Collectors edition.”, Routledge (2013).

Scott, John, Rational choice theory. Understanding contemporary society: theories of the present, International Encyclopedia of Social Sciences 2, 2000, 126-138.

Scott, Susan V., and Wanda J. Orlikowski. "Entanglements in practice: performing anonymity through social media." MIS Quarterly: Management Information Systems 38.3 (2014): 873-893.

Sheehan, Kim B., Toward a typology of Internet users and online privacy concerns, The Information Society 18.1 (2002), 21-32.

Stryker, Cole, “Epic win for anonymous: How 4chan's army conquered the web.”, Abrams (2011)

Sunstein, Cass, "Democracy and the problem of free speech.", Publishing Research Quarterly 11.4 (1995): 58-72.

Tajfel, Henri, "Social identity and intergroup behaviour.", Information (International Social Science Council) 13.2 (1974): 65-93.

Talwar, Victoria, and Kang Lee. "Development of lying to conceal a transgression: Children’s control of expressive behaviour during verbal deception." International Journal of Behavioral Development 26.5 (2002): 436-444.

Tsikerdekis, Michail, and Sherali Zeadally. "Online deception in social media." Communications of the ACM 57.9 (2014): 72.

Walther, Joseph B. , "Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction.", Communication research 23.1 (1996): 3-43.

Williams, Bernard. Moral luck: philosophical papers 1973-1980. Cambridge University Press (1981).

Wingate, V. Skye, Jessy A. Minney, and Rosanna E. Guadagno, "Sticks and stones may break your bones, but words will always hurt you: A review of cyberbullying." Social Influence 8.2-3 (2013): 87-106.

Wolak, Janis, Kimberly Mitchell, and David Finkelhor, "Unwanted and wanted exposure to online pornography in a national sample of youth Internet users.", Pediatrics 119.2 (2007): 247-257.

Zimbardo, Philip G., "Lucifer effect.", The Encyclopedia of Peace Psychology (2011).

Other sources:

Anonymity on the Internet: why the price may be too high. Communications of the ACM, 45(4), 33-35.

Facebook. "Number of monthly active Facebook users worldwide as of 2nd quarter 2019 (in millions)." Chart. July 24, 2019. Statista. Accessed October 11, 2019. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-usersworldwide/

https://dustinstout.com/social-media-statistics/

Internet World Stats: https://www.internetworldstats.com/stats.htm

Newzoo. "Number of smartphone users worldwide from 2016 to 2021 (in billions)." Chart. September 11, 2018. Statista. Accessed October 11, 2019. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/