Utrecht University (2016 – 2017)

FILTER BUBBLES AND PERSONAL AUTONOMY

A Philosophical Inquiry into How Algorithmic Personalization Undermines Frankfurt's Necessary Conditions for Personal Autonomy

THESIS PRESENTED BY
A. M. Pereira Daoud (4132963)

[1] Odin H. Standal, *Morning Light Reflected in a Bubble over the Fjord*, 2012, digital image, available from: 500px, https://500px.com/photo/5322307/morning-light-reflected-in-a-soap-bubble-over-the-fjord-by-odin-hole-standal/ (accessed May 9th, 2017).

FILTER BUBBLES AND PERSONAL AUTONOMY

A Philosophical Inquiry into How Algorithmic Personalization Undermines Frankfurt's Necessary Conditions for Personal Autonomy

THESIS PRESENTED BY

Ana M. Pereira Daoud
(4132963)

IN COMPLETION OF

BA. Liberal Arts and Sciences
Specialization in Philosophical Ethics

Universiteit Utrecht

Faculty of Humanities

Department of Philosophy and Religious Studies

1st Supervisor: Dr. H. L. C. Logister

2nd Supervisor: Dr. J. H. van Lith

June 16th 2017

8 047 words*
*excl. bibliography

Acknowledgements

I am pleased to present my philosophical inquiry into Filter Bubbles and its undermining impact to the Frankfurt-inspired model of personal autonomy as final requirement for the graduation of the bachelor Liberal Arts & Sciences at the University of Utrecht.

Writing this thesis would not have been possible without the insights of professors and friends that were kind enough to read my arguments and help me clarify the ideas that follow. I would like to take this opportunity to thank Sharida Mohamedjoesoef for investing her time and energy in making this endeavor successful. Also, I would like to show my appreciation and gratitude to my supervisor, Dr. H. L. C. Logister, for his sincere guidance, unending patience and valuable feedback, without which I could not have pursued this project.

The present thesis would not have been possible without your kindness and expertise, for which you have my deepest gratitude.

Abstract

Due to the need for filtering of the increasing amount of information provided online, users are becoming blindsided by the façade of algorithmic personalization. The one-sided information provided by personalized algorithms leads to the intellectual isolation represented by the term "filter bubble". In the present treatise, I argue that this phenomenon is particularly threatening to the user's capacity of self-governance, i.e., personal autonomy, by relying on existing literature and conceptual analysis. In the first chapter I offer an extensive analysis of one of the most prominent accounts of personal autonomy, namely Frankfurt's hierarchical theory of personal autonomy. From this, I discern three conditions deemed necessary to self-governance, namely 1) self-evaluation, 2) perseverance, and 3) critical reflection. In chapter 2, I proceed to inquire whether algorithmic personalization undermines these conditions. It is my view that these conditions are, in fact, weakened due to the lack of transparency and control of algorithmic personalization. Finally, I conclude that filter bubbles are a threat to the necessary conditions of personal autonomy and I suggest that the ensuing implications are broader than they initially seem. Political and moral theories based on the concept of personal autonomy are, for instance, as much under fire as accounts of personal autonomy. I end with a call for action: to safeguard personal autonomy, it is essential to raise awareness among users about the workings of the filtering process; such insights may help users gain more influence and control over algorithmic personalization.

Table of Contents

Introduction

In the aftermath of Brexit and Trump's victory in the US election, many who expected contrary results were left perplexed. How was this possible when nearly everyone they knew had chosen the other side?[2] In 2011, Eli Pariser, a political and internet activist, had already revealed one of the reasons for this disbelief: filter bubbles. In his book *The Filter Bubble: What the Internet is Hiding From You*, Pariser lays out how personalized algorithms isolate users from opposing views while reinforcing already-established beliefs.[3]

Cloud services, such as Facebook and Google, often deal with large amounts of data, making it difficult for the user to select and process relevant information.[4] In order to overcome this information overload, cloud services employ personalized algorithms. These algorithms record the user's online activity under a single identity, i.e. a user-model. Next, these user-models are used to predict what information is relevant for the user, in other words, what information dovetails with the specific needs, interests and preferences of the user-model.[5] Information that is considered irrelevant, i.e. contrary to the user-model, is filtered out before it ever reaches the user. In short, personalized algorithms collect user behavior data and use this knowledge accordingly to select and present only information that matches the user's interests. It is this process that gives rise to filter bubbles.

Though personalized algorithms may well be necessary to prevent information overload, they can be damaging as well. In Pariser's words, "it's not just serendipity that's at risk. By definition, a world constructed from the familiar is a world in which there's nothing to learn. If personalization is too acute, it could prevent us from coming into contact with the mind-blowing, preconception-shattering experiences and ideas that change how we think about the world and ourselves."[6] At the heart of the problem, is the fact that personalized algorithms often operate without the knowledge and control of the user for whom it is meant.[7] It is this silent algorithmic filtering taking place without the user's knowledge that leads to intellectual isolation. Consequently, "the danger of these filters is that you think you are getting a representative view of the world and you are really, really not"[8]. The user's unawareness of filter

---

[2] Jasper Jackson, "Eli Pariser: Activist whose Filter Bubble Warnings Presaged Trump and Brexit," *The Guardian* (January 2017). URL: https://www.theguardian.com/media/2017/jan/08/eli-pariser-activist-whose-filter-bubble-warnings-presaged-trump-and-brexit (accessed April 15th, 2017).
[3] Ibid.
[4] Engin Bozdag and Job Timmermans, "Values in the Filter Bubble: Ethics of Personalization Algorithms in Cloud Computing," (paper presented at the 1st International Workshop on Values in Design-Building Bridges between RE, HCI and Ethics, Lisbon, Portugal, September 2011), 7, URL: http://mmi.tudelft.nl/ValuesInDesign11/proceedings.pdf.
[5] Ibid., 8.
[6] Eli Pariser, introduction to *The Filter Bubble: What the Internet is Hiding From You* (London: Viking e-book, 2011), 26.
[7] Engin Bozdag and Job Timmermans, "Values in the Filter Bubble," 8.
[8] Jasper Jackson, "Eli Pariser: Activist whose Filter Bubble Warnings Presaged Trump and Brexit."

bubbles not only deludes her from a well-informed course of events but also prevents her from actively engaging with contradictory viewpoints. As a consequence, the user could be trapped in an echo chamber full of information she already agrees with. Cloud services have responded to these allegations. Facebook, for instance, released a study where researchers "conclusively establish that […] individual choices more than algorithms limit exposure to attitude-challenging content."[9] In other words, it is you who is at fault.

It remains to be seen whether there is a scenario in which individual choices and personalized algorithms *can* be traded off, since they happen together: "users select from what the algorithm already filtered for them. It is a sequence."[10] That said, it is not the aim of this thesis to dwell on such questions. Instead, I will depart from the assumption that personalized algorithms indeed give way to "filter bubbles". Moreover, and given the fact that personalized filtering is gaining importance and continues to be used by multiple cloud services on a daily basis, I will assume that filter bubbles are – and will remain – a reality we must learn to deal with. After all, even if the alleged scope of filter bubbles and their respective effects are as minimal as Facebook suggests, they still impact the lives of 22% of the global population.[11]

This thesis aims to inquire which necessary conditions for personal autonomy are undermined by filter bubbles. As stated before, I depart from the assumption that filter bubbles do, in fact, threaten certain necessary conditions for personal autonomy and I aim only to inquire which ones are at stake. Needless to say, two concepts – "personal autonomy" and "filter bubbles" – require further interpretation if I am to successfully tackle the issue at hand. Therefore, this thesis consists of two chapters, each addressing a concept.

Chapter 1 focuses on the notion of personal autonomy, for it is necessary to first decide on a clear-cut classification of this concept in order to uncover its vulnerabilities with regard to filter bubbles. Among the many accounts of personal autonomy, I have opted for the most influential ones as my point of departure, namely: hierarchal accounts of personal autonomy. More particularly, I will offer an analysis of the Frankfurt-inspired model of personal autonomy and discern which conditions are deemed necessary for this conception.

To be fair, the nature and normativity of personal autonomy is still heavily debated among philosophers.[12] However, it is not the aim of the present thesis to resolve an ever-lasting

---

[9] Christian Sandvig, "The Facebook 'It's Not Our Fault' Study," in *The Social Media Collective (SMC): Research Blog* (May 2015), URL: https://socialmediacollective.org/2015/05/07/the-facebook-its-not-our-fault-study/ (accessed May 7th, 2017).
[10] Ibid.
[11] "Percentage of Global Population Using Facebook as of June 2016, by Region," Statista, URL: https://www.statista.com/statistics/241552/share-of-global-population-using-facebook-by-region/ (accessed May 15th 2017).
[12] John Christman, "Autonomy in Moral and Political Philosophy," in *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), ed. Edward N. Zalta, URL: https://plato.stanford.edu/archives/spr2015/entries/autonomy-moral (accessed April 15th, 2017).

philosophical crisis. I wish only to understand the causal relation between filter bubbles and undermined autonomy, and, to do this, a strict framework from which I can start my research is required.

Once the necessary conditions for Frankfurt's inspired-model of personal autonomy have been set, it becomes possible to conceptualize how they are undermined by filter bubbles. This will be the aim of chapter 2. Before I do this, however, I will take a step back and analyze the mechanics of the particular phenomenon at hand. After all, it is important to comprehend the process behind filter bubbles thoroughly if we are to understand how they undermine some of the necessary conditions for personal autonomy.

Finally, I conclude that filter bubbles pose a threat to the Frankfurt-inspired model of personal autonomy. In particular, they undermine, at least, three conditions deemed necessary for this theory, namely: self-evaluation, critical reflection and, although to a lesser degree, perseverance. To be fair, however, I admit that further investigation is required if we are to conclusively establish the impact of filter bubbles to conceptions of personal autonomy. Moreover, given the significant role such conceptions play in moral and political theory, it also seems worthwhile to explore the ensuing implications of undermined autonomy in a broader context. I end with a call for action: although it is difficult to foresee the exact consequences that algorithmic personalization may bring, we can and should begin by educating the 2.5 billion social media users around the globe[13] about its impact. After all, such insights may actually help users gain more influence and control over the ensuing implications of algorithmic personalization.

[13] "Number of Social Media Users Worldwide from 2010 to 2020 (in billions)," Statista, URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (accessed May 15th 2017).

Chapter 1: The Autonomous Agent

The concept of autonomy derives from the Greek words *autós* ("self") and *nomos* ("law" or "rule"); it was initially used to describe Greek city states exerting their own laws.[14] In modern western tradition, however, this concept is used in a remarkably broad fashion, as it is often employed as an equivalent to notions of liberty, freedom of the will, sovereignty or self-governance.[15] Due to the broad usage of the concept of autonomy, and to avoid any misunderstandings, it is worthwhile to explicitly mention that the present chapter will interpret this concept in terms of *self-governance*. Particularly, it will focus on the *individual capacity* for self-governance; i.e., *personal* autonomy.

Personal autonomy refers to the individual capacity of guiding oneself in accordance with one's commitment to personal projects, relationships, and ideals.[16] It is often defined as a necessary condition for "theoretical accounts of persons, conceptions of moral obligation and responsibility, the justification of social policies and in numerous aspects of political theory,"[17] and, therefore, plays various roles of irrefutable value. The idea of personal autonomy invites skeptics from several disciplines, particularly due to the idea that in order to successfully legitimize political authority and moral obligation on the basis of personal autonomy, this defense cannot be based upon individual contingencies, such as place, culture and social relations. As a result, skeptics bring into question what conditions are deemed necessary to develop and maintain this capacity.[18] Without doubt, the majority of contemporary work on autonomy has centered exactly on this very issue.

Numerous philosophical theories have offered an analysis of the nature and normativity of personal autonomy. Among them, "hierarchal" accounts, based on the work of Harry Frankfurt and Gerald Dworkin, have proven most popular and influential. The popularity of these approaches is, at least in part, due to their "compelling idea that autonomy is a matter of being guided not merely by desires one happens to have but by the desires one reflectively endorses having."[19] In other words, such theories reconcile emotionality and rationality. Moreover, by defining personal autonomy "in terms of a psychological capacity of structuring one's desires in

---

[14] Michael Kühler and Nadja Jelinek, introduction to *Autonomy and the Self* (Dordrecht: Springer Netherlands, 2013), ix, DOI 10.1007/978-94-007-4789-0.
[15] Gerald Dworkin, "The Nature of Autonomy," *Nordic Journal of Studies in Educational Policy* 2 (2015), 8, DOI: 10.3402/nstep.v1.28479.
[16] Joel Anderson, "Disputing Autonomy: Second-Order Desires and the Dynamics of Ascribing Autonomy," *SATS: Northern European Journal of Philosophy* 9 (2008), 8, URL: http://phil.uu.nl/~joel/research/publications/Anderson-DisputingAutonomy.pdf (accessed May13th, 2017).
[17] John Christman, "Autonomy in Moral and Political Philosophy."
[18] Ibid.
[19] Joel Anderson, "Disputing Autonomy," 9.

a particular way"[20], hierarchal accounts strategically avoid the problems of content-laden and metaphysically committed models, such as Kant's theory of autonomy, while still presupposing that virtually any agent – perhaps with the exception of pathological and handicapped agents – can develop this capacity.[21]

This chapter aims to understand and analyze the concept of personal autonomy according to hierarchal accounts. More specifically, I will offer my interpretation of Frankfurt's inspired-model of personal autonomy and discern some of the underlying conditions presupposed necessary for this concept. It must be noted, however, that it was not Frankfurt's aim to discuss the concept of autonomy – but rather that of freedom of the will – in his influential article "Freedom of the Will and the Concept of a Person". Nonetheless, I believe that a conception of autonomy can be derived from it, as his discussion on freedom of the will offers a substantial outlook on the workings of self-governance.

The Frankfurt-Inspired Model of Personal Autonomy

Generally speaking, hierarchical accounts make a distinction between first-order desires, i.e. what one wants, and second-order desires, i.e. what one wants to want.[22] Within this context, personal autonomy requires *identification* between first and second-order desires; to put it differently, an agent is said to be autonomous when *she wants* what *she wants to want*.[23] It should be clear then that identification between desires of the first and second order is the hallmark of personal autonomy. However, before I can sensibly discuss the concept of identification, I must show what first and second-order desires are in the first place. This will be the aim of this section.

It is the distinction between first and second-order desires that Harry Frankfurt, a leading Humean compatibilist and major advocate of hierarchical accounts, takes to be essential in distinguishing persons from other creatures. After all, "human beings are not alone in having [first-order] desires and motives, or in making choices. They share these things with the members of certain other species, some of whom even appear to engage in deliberation and to make decisions based upon prior thought."[24] In other words, a desire of the first order is not an exclusively human capacity but one we share with a vast majority of animals. By means of

---

[20] Gareth S. Owen e.a., "Mental Capacity and Decisional Autonomy: An Interdisciplinary Challenge," *Inquiry* 52 (2009), 86, DOI: 10.1080/00201740802661502.
[21] Mark Piper, "Autonomy: Normative" in *Internet Encyclopedia of Philosophy*, URL: http://www.iep.utm.edu/aut-norm/ (accessed April 13th, 2017).
[22] Gareth S. Owen e.a., "Mental Capacity and Decisional Autonomy," 85.
[23] Mark Piper, "Autonomy."
[24] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," in *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988), 12, DOI: 10.1017/CBO9780511818172.003.

illustration, my first-order desire to eat a refreshing ice cream is one I can share with my cat – he, too, may desire to take a bite of that lovely vanilla ice.

Alongside simplistic desires, however, humans are also capable of grasping what they want to want for themselves; i.e., they are capable of forming second-order desires. They recognize that their desires are not always in line with their aspirations, and they develop second-order desires that guide them in their path to their true self. Indeed, "besides just wanting certain things, humans tend to want to want certain things."[25] So, in forming second-order desires, humans differ significantly from animals. As put by Frankfurt, "no animal other than man (…) appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires."[26]

What, then, aside from ranking, is the difference between first and second-order desires? The specific distinction is that while the former "makes no mention of any other desire,"[27] the latter necessarily presupposes additional desires. Consider, for example, the following first-order desire: I want to eat chocolate. This desire is simple – it presupposes nothing else other than one's craving for chocolate. By means of contrast, now imagine the following second-order desire: I want to want to study for my exams. It should be clear that this claim presupposes 1) that I want to study for my exams and 2) that I want to want this. Second-order desires are, as such, to be understood as the desire to have a certain desire.

More generally, and in Frankfurt's terms, we speak of second-order desires either when the agent wants "to have a certain desire or when he wants a certain desire to be his will."[28] Here, Frankfurt makes yet another distinction, namely the distinction between desires and volitions respectively. What differentiates regular desires from volitions is that the latter are, if you will, more one's own. They represent what the agent truly wants for herself, i.e., what she *really cares about.*

It must be noted that Frankfurt's conception of volitions remains somewhat mysterious. Nonetheless, for the present purposes it suffices to say that volitions are not a new kind of desires entirely, but rather "desires [that are] supported and encouraged by one's self."[29] The distinction between desires and volitions becomes particularly clear in light of identification. According to Frankfurt, volitions necessarily possess the criterion of identification, while regular desires do not: "as for a person's second-order volitions themselves, it is impossible for him to be a passive bystander to them. They constitute his activity – i.e., his being active rather than

[25] Martin Vezér, "On the Concept of Personhood: A Comparative Analysis of Three Accounts," *Lyceum Philosophy* 1 (Fall 2007), 3, URL: http://lyceumphilosophy.com/9-1/Lyceum-9-1-Vezer.pdf (accessed May 26th, 2017).
[26] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," 12.
[27] Martin Vezér, "On the Concept of Personhood," 3.
[28] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," 16.
[29] Martin Vezér, "On the Concept of Personhood," 4.

passive – and the question of whether or not he identifies himself with them cannot arise. It makes no sense to ask whether someone identifies himself with his identification of himself, unless this is intended simply as asking whether his identification is wholehearted or complete."[30] Hence, volitions, in contrast to regular desires, necessarily presuppose that the agent cares about them.

Caring, The Binding Foundation of Personal Autonomy

Now that I have outlined the structure and meaning of first and second-order desires (or, more particularly, volitions), it is worthwhile to consider what triggers them.

One of the most essential concepts of Frankfurt's inspired-model of autonomy is *caring*, for it is this that elicits desires of the first order. In "The Importance of What We Care About", Frankfurt argues that what a person cares about coincides, at least in part, with how this person guides herself; i.e., with how a person expresses her personal autonomy.[31] In this, he departs from the belief that the inquiry about how to govern oneself must begin somewhere. That it needs at least an initial stock of cares, needs, beliefs and goals, for otherwise there would be nothing to work from.[32] According to Frankfurt, the inquiry about self-governance begins with what a person cares about, since "caring about something necessarily involves taking its interests as reasons for acting."[33] In other words, caring is not rooted in reasons but rather goes back to the source; it is not the expression of desires but their origin. Accordingly, the development of desires and volitions "is most fundamentally a matter of his coming to care about certain things, and of his coming to care about some of them *more* than about others."[34]

It must be noted, however, that Frankfurt's interpretation of "caring" is not equivalent to "wanting".[35] After all, one can have desires – or, "wants" – without caring for them. He illustrates this by means of an example. Consider two incorrigibly addicted men: the "wanton addict" and the "unwilling addict". Both these men share the first-order desires (1) to desperately ingest the drug and (2) to refrain from it because of its harmful side effects to their well-being.[36] Nonetheless, Frankfurt argues, their degree of personal autonomy differs remarkably.

The wanton addict, on the one hand, is indifferent about the "ranking" of his first-order desires. He does not necessarily care *more* for his health than for his craving for drugs; because

---

[30] Harry Frankfurt, "Three Concepts of Free Action," *in The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988), 54, DOI: 10.1017/CBO9780511818172.005.
[31] Harry Frankfurt, "The Importance of What We Care About," 82.
[32] Richard Moran, "On Frankfurt's The Reasons of Love," in *The Philosophical Imagination: Selected Essays*, (Oxford University Press, 2017) 162.
[33] Ibid., 161.
[34] Harry Frankfurt, "The Importance of What We Care About," 91.
[35] Richard Moran, "On Frankfurt's The Reasons of Love," 161.
[36] Martin Vezér, "On the Concept of Personhood," 5.

of this, he passively allows the strongest desire to guide him without prioritizing one over the other. In sum, although he has desires – or "wants" – he does not *care* about which desire takes the upper hand.[37] The unwilling addict, on the other hand, actively pursues one desire over the other by means of what he cares about. He cares so deeply about the harmful consequences of addiction to his health, that he develops the desire to want to refrain from taking drugs. Because he cares *more* about his health than he cares about his cravings, this first-order desire is converted into a *volition*; that is, his desire becomes part of his will.[38]

With respect to personal autonomy, then, the wanton addict's success in quitting his addiction could have been accidental. The desire to quit just so happened to be stronger than the desire to continue taking drugs. Yet, in the case of the unwilling addict, it becomes clear that here it was a matter of personal *choice* to quit. Regardless of the strength of his desire for drugs, he willfully prioritized the desire he truly cares about over the other. Indeed, it is the very fact that he cared for one desire *more* than the other, that enabled him to decide which one he most fundamentally identifies with. As Frankfurt puts it, "what distinguishes the rational wanton from other rational agents is that he is not concerned with the desirability of his desires themselves. He ignores the question of what his will is to be. Not only does he pursue whatever course of action he is most strongly inclined to pursue, but he *does not care* which of his inclinations is the strongest."[39] Caring is, thus, implicitly inherent to what one most fundamentally identifies with.

Now consider the possibility that our imaginary addict had cared more about his desire for taking drugs than the desire to be healthy. In this case, the addict would have identified himself with his first-order desire of ingesting drugs rather than the other. Controversially, the autonomous choice would have been to pursue his addiction. This illustrates the problem of Frankfurt's concept of "caring", namely that what a person cares about is not always a matter of decision-making; i.e., it is not always under her volitional control. As Frankfurt himself admits, what one cares about "depends upon conditions which do not always prevail."[40] And although it is possible for an agent to actively wish to care for something else, and sometimes to successfully do so, this rarely happens.

The intangible and "complex set of cognitive, affective, and volitional dispositions and states"[41] that embody what one cares about, makes it very hard to sensibly analyze the concept of "caring" in light of filter bubbles. Nonetheless, I think it is possible to discern other, more tangible, conditions which remain necessary for what one "cares about". After all, for an agent to

[37] Martin Vezér, "On the Concept of Personhood," 5.
[38] Ibid., 6.
[39] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," 17. My italics.
[40] Harry Frankfurt, "The Importance of What We Care About," 85.
[41] Ibid.

"care about something", i.e., to guide herself, she must at least 1) know what she cares about (self-evaluation) and 2) care about it with a certain consistency (perseverance).

For starters, an agent must be capable of *self-evaluation* in order to know what she cares about. Indeed, "in order to be autonomous (…) one has to decide and act, or more broadly to live in general, according to motives that can count as expressions of one's self, i.e. of who one is (or wants to be)."[42] This is, in turn, only possible through self-awareness and evaluation. The agent is only able to discern what she desires if she knows herself and, more specifically, what she wants for herself. Without self-evaluation, caring would be unattainable.

Caring also presupposes *perseverance* since, without it, an agent would be unable to tell impulsive desires apart from authentic ones. As argued by Frankfurt, "a person who cared about something just for a single moment would be indistinguishable from someone who was being moved by impulse. He would not in any proper sense be guiding or directing himself at all."[43] From this perspective, perseverance can best be understood in terms of a desire that endures. Indeed, caring about something for only a single moment is undistinguishable from being moved by instinct. For although "desires and beliefs have no inherent persistence"[44], so the argument goes, "the notion of guidance, and hence the notion of caring, implies a certain consistency or steadiness of behavior; and this presupposes some degree of persistence."[45] Thus, perseverance is a necessary condition for authentic desires and, more generally, personal autonomy.

In short, even though the concept of caring is too intangible to sensibly analyze in light of filter bubbles, it does offer two necessary conditions for what one cares about (and consequently, necessary conditions for self-governance), namely self-evaluation and perseverance.

Identification, The Hallmark of Personal Autonomy

Now, let us return to the concept of identification – the hallmark of personal autonomy. As I mentioned in the second section of the present chapter, it is only through identification between first and second-order desires that an agent is said to be autonomous. But what exactly does it mean to identify with one's desires?

For an agent to identify with her desires, she must 1) care for her desires and 2) rationally endorse them. To briefly recapitulate, desires of the first order arise from what the agent cares about. Like any Humean theorist, Frankfurt too presupposes that emotional dispositions come before rationality. Our passions or, in Frankfurt's terms, our "cares" are

[42] Michael Kühler and Nadja Jelinek, introduction to *Autonomy and the Self*, x.
[43] Harry Frankfurt, "The Importance of What We Care About," 84.
[44] Ibid.
[45] Ibid.

inherent to whom we truly are and, more often than not, unchangeable. What we care about, thus, shapes our first-order desires. For instance, imagine that I have a natural inclination to care about animals. This may spark the first-order desire to become vegetarian, since causing animals pain would contradict my caring for them.

Yet how does Frankfurt justify autonomy in terms of what the agent cares about, if this is not always under her volitional control? To such concerns, he argues that "even if the person is not responsible for the fact that the desire occurs, there is an important sense in which he takes responsibility for the fact of having the desire […] when he identifies himself with it. Through his action in deciding, he is responsible for the fact that the desire has become his own in a way in which it was not unequivocally his own before."[46] In other words, although one cannot directly control what she desires, she must willingly prioritize one over the other. It is my interpretation, then, that identification must consist, on the one hand, of what we care about, i.e., our "emotional dispositions", and, on the other hand, of what we want for ourselves through rational endorsement, i.e., our "rational dispositions". Indeed, "the rational self is responsible for 'identifying' with or making a 'decisive commitment' about which of her desires to act on and which to reject."[47] Hence, in order to decide whether I, for example, do actually desire to become vegetarian, I must *rationally endorse* this desire and inescapably reject others; e.g., eating meat.

So, to return to the example above, in order to ensure that my desire to become vegetarian is, indeed, my own – and not externally imposed on me –, I go on to question whether this is what I want to want. If the answer is positive in the sense that I want to want to become vegetarian, I then identify my second-order desire (what I rationally endorse) with my first-order desire (what I care about) and, consequently, act autonomously. Identification, thus, consists of caring and rational endorsement.

Rational endorsement is, however, only attainable through *critical reflection*. In making a decision, the agent necessarily establishes constraints "by which other preferences and decisions are to be guided,"[48] which, in turn, implicitly entail a certain degree of consistency and critical reflection. After all, if I am to identify with my first-order desires, I must first consider whether they correspond with who I take myself to be (or who I want to become), i.e., my second-order desires. However, to determine whether my first and second-order desires are coherent or contradictory, I must be able to critically analyze them. It is my interpretation, therefore, that identification between desires inevitably posits the capacity for critical reflection.

---

[46] Harry Frankfurt, "Identification and Wholeheartedness," in *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988), 170, DOI: 10.1017/CBO9780511818172.013.
[47] Sharli Anne Paphitis, "Questions of the Self in the Personal Autonomy Debate: Some Critical Remarks on Frankfurt and Watson," *South African Journal of Philosophy* 2 (2010), 120, DOI: 10.4314/sajpem.v29i2.57055.
[48] Harry Frankfurt, "Identification and Wholeheartedness," 175.

I believe Frankfurt would agree, since he argues that the decision-making process suggests performing an action that "essentially involves reflexivity […]. Thus, creatures who are incapable of this volitional reflexivity necessarily lack the capacity to make up their minds."[49] As a consequence, since Frankfurt's concept of identification is rooted in rational endorsement, i.e., decision-making, and since making a deliberate decision is dependent upon the capability for critical reflection, I believe to have sufficiently argued in favor of yet a third necessary condition for personal autonomy, namely *critical reflection*, for without it agents would be unable to make up their minds and govern themselves.

In a nutshell, aside from the conditions stemming from Frankfurt's concept of caring, it is also possible to discern the condition of *critical reflection* from the concept of identification. As such, Frankfurt's inspired-model of personal autonomy posits, at least, three necessary conditions for autonomy: *self-evaluation*, *perseverance* and *critical reflection*.

[49] Harry Frankfurt, "Identification and Wholeheartedness," 176.

Chapter 2: Introducing Filter Bubbles

The term "filter bubble" refers to the intellectual isolation that arises due to algorithmic personalization. Initially, the internet was "an anonymous medium where anyone could be anyone"[50] and search results were irrespective of users, offering them the freedom and transparency it promised. However, this utopian vision has become obsolete. The availability of immense computing power and large amounts generated data have compromised the transparency of search engines by making the employment of personalized algorithms a necessary evil.[51]

The World Wide Web can, in theory, actually increase the agent's degree of personal autonomy.[52] For starters, it would seem as if pluralism and viewpoint diversity are necessary for well-considered decisions regarding which desires one most truly identifies with. Up to a certain point, this is true. The amount of information available online offers the agent the possibility to choose from a wide-range of desires. However, decision-making would not be possible for the agent to do without a certain degree of filtering. After all, in order to be autonomous one must be capable of making up one's mind, which would be unbearable in the tremendous deluge of information provided.[53] Consequently, an agent may well be capable of desiring, thinking and acting but remain limited in her decision-making,[54] since her human brain is incapable of processing each and every piece of available information. In sum, without filtering, the agent may well possess the capabilities for autonomy without ever actually being able of expressing them.

The human need for filtering is what prompted algorithmic personalization. Personalized algorithms filter "the content and structure of a web application to adapt it to the specific needs, goals, interest and preferences of each user."[55] It all begins with a user-model created on the basis of various "user-signals", such as click history, location, personal information, etcetera. By unifying the various signals under a single identity, the system can predict what information will be of relevance for the user in question and filter out data that does not coincide with the user-model, making it easier for the user to find her way in the flood of online information.

Algorithmic personalization is not without risks though, as it can severely undermine the necessary conditions for personal autonomy proposed in chapter 1. The impact of filter bubbles to self-evaluation, critical reflection and perseverance will be discussed below.

---

[50] Eli Pariser, introduction to *The Filter Bubble*, 10.
[51] Engin Bozdag and Job Timmermans, "Values in the Filter Bubble," 7.
[52] Ibid., 11.
[53] Ibid., 9.
[54] Harry Frankfurt, "Identification and Wholeheartedness," 176.
[55] Engin Bozdag and Job Timmermans, "Values in the Filter Bubble," 8.

The Loss of Autonomy in Algorithmic Personalization

To begin with, the agent's capacity of *self-evaluation* is considerably undermined. As I have argued, for the agent to form desires she must be able to evaluate who she is and who she wants to become. However, this becomes difficult, if not impossible, when personalized filtering limits her knowledge of the world and of herself.

Consider someone who believes climate change is a hoax invented by the Chinese. "The weather hasn't changed at all," she argues, "so global warming must be a lie." Personalized algorithms take notice of these beliefs and provide information that will only further reinforce such trains of thought: for instance, they offer articles that read "IPCC is alarmist" or "increasing CO2 has little to no effect". Without knowing any better, the agent continues to drive her gas-guzzling car to work every morning, completely disregarding concerns related to climate change. More importantly, she becomes convinced of her own righteousness with little room for doubt, exactly because she is constantly surrounded with the familiar and has little or no contact with information that challenges her views.[56] Indeed, "personalization filters serve up a kind of invisible autopropaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar and leaving us oblivious"[57] of any divergent information. As a result, filter bubbles prevent users from realizing that their beliefs and desires may need rethinking; thus, heavily undermining the agent's capacity for self-evaluation.

Critics could argue that filter bubbles cannot be the only issue at fault. After all, the agent ought to actively search for reliable information and question whether her beliefs are indeed justified. More fundamentally, it is her responsibility to do so. Though I most certainly agree, I doubt this is self-evident to every person. As argued by Frankfurt, many people are "wantons" and have little regard for the desirability and justifiability of their beliefs or desires. What is known as the "confirmation bias", i.e., the human "tendency to believe things that reinforce our existing views, to see what we want to see" [58], surely illustrates this.

Nonetheless, critics may continue, like the "confirmation bias", filter bubbles are a threat one can account for. But this is not a valid comparison. For starters, while the confirmation bias is a matter of the human psyche, filter bubbles are not. In these terms, then, it seems that personalized algorithms are imposed on persons to a larger degree, rather than being a characteristic of their human make-up. Moreover, filter bubbles are a relatively recent phenomenon which many people are still unaware of, particularly because "information is filtered before reaching the user, and this occurs silently."[59] While raising awareness, as I aim to

---

[56] Eli Pariser, "The Adderall Society," in *The Filter Bubble: What the Internet is Hiding From You* (London: Viking e-book, 2011), 14.
[57] Eli Pariser, introduction to *The Filter Bubble*, 26.
[58] Eli Pariser, "The Adderall Society," in *The Filter Bubble*, 17.
[59] Engin Bozdag and Job Timmermans, "Values in the Filter Bubble," 8.

do here, may be an honorable endeavor, this too has its limits: for even if the agent is aware of algorithmic personalization, there is little she can do. The criterion upon which information is filtered is neither transparent nor customizable.[60] Moreover, it seems obvious that "for the users to control the services they are using, users must know what information is used for personalization and how their data are used."[61] Yet this is unlikely, especially considering that even programmers can no longer fully explain any given output due to the complexity of algorithmic personalization.[62] Perhaps, then, the only viable way of protecting our capacity for self-evaluation, and autonomy respectively, is to opt out. This would, however, be an unrealistic scenario since, as I have argued above, the human brain is incapable of processing the deluge of online information on its own.

The same holds true with regard to yet another condition of personal autonomy proposed by this paper, namely *critical reflection*. As I have argued, to rationally endorse a particular desire rather than others, the agent must question the validity and appropriateness of her desires; i.e., critically reflect upon what she wants (to want). This becomes difficult though without the environmental signals that prompt agents to employ this capacity. After all, in order to reflect critically on one's desires and beliefs, there must be at least some kind of contradiction, or otherwise challenging information, to initiate this process. Personalized algorithms, however, "offer a vision of a custom-tailored world, every facet of which fits us perfectly. […] We're never bored. We're never annoyed. Our media is a perfect reflection of our interests and desires."[63] Because of this intellectual isolation, and particularly due to the agent's unawareness of the filtering process, agents are never actually challenged to look beyond their personal preferences, let alone critically reflect on them. The unawareness of users, combined with the lack of control and transparency of personalized algorithms, places the capacity for critical reflection in a vulnerable position.

Again, it could be argued that the employment of critical reasoning should be self-evident for any autonomous person. I find arguments of this kind always somewhat snobbish: after all, the relevance of critical reflection is only self-evident to academics who have learned its value and necessity.

Paradoxically, algorithmic personalization does seem to reinforce the *perseverance*-condition for personal autonomy. As mentioned before, perseverance, in this particular context, can be defined as a desire that endures. In light of filter bubbles, this condition seems to be strengthened: by constantly only providing information that matches the user's interests, algorithmic personalization secures a certain degree of consistency in what one desires or

[60] Engin Bozdag and Job Timmermans, "Values in the Filter Bubble," 12.
[61] Ibid.
[62] Ibid., 10.
[63] Eli Pariser, introduction to *The Filter Bubble*, 21.

believes. For instance, consider my desire to become vegetarian. Once the system picks up on this, it will only offer comparable information, allowing for my desire to endure.

However, this seemingly perseverance in interests could just be a front, for it is not necessarily the agent who willingly decides to continuously pursue this desire. For instance, suppose I accidently clicked on an article about cruelty in the egg industry. Now, algorithmic personalization will interpret this as an interest in animal cruelty and proceed to offer similar articles while filtering out contrary information. Due to the shocking information provided, I may feel compelled to change my behavior. But this begs the question of whether my desire to become vegetarian was actually my own or induced by the system.

A similar example may shed more light on this: a study has revealed that "graphic warnings can reduce interest in smoking among occasional smokers."[64] The idea behind the shocking images on cigarette packs is similar to the argument I propose: by constantly exposing people to a particular kind of information, their beliefs and desires will be affected accordingly. With regard to algorithmic personalization, it follows that if the system continuously feeds you the same kind of information, the human brain will "act to reduce cognitive dissonance in a strange but compelling kind of unlogic – 'Why would I have done x if I weren't a person who does x – therefore I must be a person who does x.'"[65] In other words, by constantly feeding similar information, algorithmic personalization leads the brain to redefine itself as having interest x. This, therefore, raises the question whether a particular desire is, indeed, part of the agent's will or externally coerced upon her.

In short, it seems to me that although algorithmic personalization can strengthen the condition of perseverance, it does so illegitimately. Algorithmic personalization – or filter bubbles, more generally, – cannot, therefore, be considered a positive influence to the condition of perseverance. In fact, it is quite the opposite, since it offers the agent the illusion of consistency and steadiness of behavior, i.e., autonomous behavior, on false grounds.

To be fair, much more could be said about the potential implications of filter bubbles, or algorithmic personalization in particular, for the three suggested conditions of personal autonomy. I believe, however, that the present chapter offers a sufficient analysis of the threatening and undermining properties of algorithmic personalization: filter bubbles do, in fact, undermine the capacities for self-evaluation, critical reflection and, although to a lesser degree, perseverance.

---

[64] Hart Blanton e.a., "Effect of Graphic Cigarette Warnings on Smoking Intentions in Young Adults," *PLOS ONE* 9 (2014), 1, DOI: 10.1371/journal.pone.0096315.
[65] Eli Pariser, "The You Loop," in *The Filter Bubble: What the Internet is Hiding From You* (London: Viking e-book, 2011), 35.

Conclusion

The present thesis aimed to inquire which necessary conditions for personal autonomy are undermined by filter bubbles. Due to its limited scope, I departed from the assumption that filter bubbles do, in fact, undermine autonomy and I aimed only to inquire how this transpires. In order to do so, I split the present thesis in two sections, each of which focused on a key concept of my research: "personal autonomy" and "filter bubbles" respectively.

In chapter 1, I offered an analysis of the Frankfurt-inspired hierarchical account of personal autonomy, which is to be understood in terms of identification between first and second-order desires. I have chosen this theory in particular due to its prominent position in the current philosophical debate. Please note, however, that Frankfurt did not actually aim to discuss personal autonomy *an sich*. In his book *The Importance of What We Care About*, it was rather his aim to offer a conception of freedom of the will. Nonetheless, it is possible to interpret his theory in light of autonomy, as I have done by relying on secondary literature.

In the subsections of the first chapter, I offered my interpretation of two essential concepts of Frankfurt's theory, namely caring and identification. It is from these concepts that I discerned conditions deemed necessary for (Frankfurt's account of) personal autonomy. For starters, it is through caring that agents instigate their first-order desires, i.e., what one wants. This posits, at least, two necessary conditions: *self-evaluation*, for the agent cannot know what she cares about without self-assessment, and, *perseverance*, for caring about a certain desire necessarily implies a certain degree of consistency and steadiness. That is, caring for a desire presupposes that the agent cares about it for a longer period of time. In the following subsection, I argued that identification between desires is to be understood as rationally endorsing those desires. Rational endorsement entails questioning the validity and desirability of one's desires. This is particularly visible in second-order desires, i.e., what one wants to want. In sum, I argued that rational endorsement – or identification, more generally, – posits, at least, one necessary condition: *critical reflection*.

In chapter 2, I zoomed into the concept of filter bubbles and its undermining properties to the three necessary conditions mentioned above. To be clear, a filter bubble is the general term used to refer to the user's intellectual isolation due to algorithmic personalization. This isolation significantly decreases the capacities for *self-evaluation* and *critical reflection*. The "cozy world" presented by filter bubbles, prevents agents from actively challenging their desires: every piece of information already seems to be in agreement with what users believe, giving them no reason to assess themselves or the information provided. The agent's unawareness and lack of control over algorithmic personalization strengthen this tendency even further. Hence, considering opting out is not a viable possibility, users are often rendered helpless to online filtering.

On the other hand, it remains somewhat inconclusive whether *perseverance* is undermined by filter bubbles. Although I personally believe it is, others may argue that algorithmic personalization actually strengthens this condition. After all, lack of information variety prevents users from caring for new desires while simultaneously reinforcing desires they already cared about. This necessarily leads to "desires that endure". I argue, however, that this seemingly persistence may be a front, for it is the system, rather than the conscious agent, that prompts this perseverance. As such, it does not seem logically valid to claim that the constant repetition of information, beliefs and desires, serves the condition of perseverance.

Concluding, filter bubbles, or algorithmic personalization more specifically, undermine at least three necessary conditions of personal autonomy, namely *self-evaluation*, *perseverance* and *critical reflection*. As a result, the user's degree of personal autonomy decreases significantly.

Due to the limitations of this thesis, however, further research is required in order to conclusively establish the impact of filter bubbles to personal autonomy. To begin with, I did not investigate the potential difficulties of Frankfurt's inspired accounts of personal autonomy and I must admit that doing this can potentially lead to very different conclusions. Furthermore, it may be possible to discern more conditions deemed necessary for Frankfurt's inspired-model of autonomy. Yet to do so, would weaken the aim of this thesis since such a discussion would be too extensive for the present purposes. Moreover, as briefly mentioned in the introduction, given the significance of personal autonomy to moral and political theory, it also seems relevant to study the ensuing implications of undermined autonomy to moral and political accounts of justifiability, obligation and responsibility. Unfortunately, I had to postpone this discussion but I strongly advise academics to consider this issue.

In sum, much is left unsaid, but there is one thing I want to say for sure: users should be given more insight and control over the filtering process. After all, self-governance implies a certain degree of assessment and influence in the algorithms doing the filtering. This is unlikely, however. So, perhaps more importantly, it falls upon academics, and other involved parties, to inform and educate users about the workings and implications of filter bubbles. For although it may be difficult to control – or even avoid – algorithmic personalization, raising awareness may provide users with sufficient tools to limit its impact on personal autonomy.

Bibliography

Anderson, Joel. "Disputing Autonomy: Second-Order Desires and the Dynamics of Ascribing Autonomy." *SATS: Northern European Journal of Philosophy* 9 (2008), 7 – 26. URL: http://phil.uu.nl/~joel/research/publications/Anderson-DisputingAutonomy.pdf (accessed May 13th, 2017).

Blanton, Hart, Leslie B. Snyder, Eri Strauts and Joy G. Larson. "Effect of Graphic Cigarette Warnings on Smoking Intentions in Young Adults." *PLOS ONE* 9 (2014), 1 – 7. DOI: 10.1371/journal.pone.0096315.

Bozdag, Engin and Job Timmermans. "Values in the Filter Bubble: Ethics of Personalization Algorithms in Cloud Computing." Paper presented at the 1st International Workshop on Values in Design-Building Bridges between RE, HCI and Ethics, Lisbon, Portugal, September 2011, 7 – 15. URL: http://mmi.tudelft.nl/ValuesInDesign11/proceedings.pdf.

Christman, John. "Autonomy in Moral and Political Philosophy." In *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), ed. Edward N. Zalta. URL: https://plato.stanford.edu/archives/spr2015/entries/autonomy-moral (accessed April 15th, 2017).

Dworkin, Gerald. "The Nature of Autonomy." *Nordic Journal of Studies in Educational Policy* 2 (2015), 7 – 14. DOI: 10.3402/nstep.v1.28479.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." In *The Importance of What We Care About: Philosophical Essays*, 11 – 25. Cambridge: Cambridge University Press, 1988. DOI: 10.1017/CBO9780511818172.003.

Frankfurt, Harry. "Identification and Wholeheartedness." In *The Importance of What We Care About: Philosophical Essays*, 159 – 176. Cambridge: Cambridge University Press, 1988. DOI: 10.1017/CBO9780511818172.013.

Frankfurt, Harry. "The Importance of What We Care About." In *The Importance of What We Care About: Philosophical Essays*, 80 – 94. Cambridge: Cambridge University Press, 1988. DOI: 10.1017/CBO9780511818172.008.

Frankfurt, Harry. "Three Concepts of Free Action." *In The Importance of What We Care About: Philosophical Essays*, 47 – 57. Cambridge: Cambridge University Press, 1988. DOI: 10.1017/CBO9780511818172.005.

Jackson, Jasper. "Eli Pariser: Activist whose Filter Bubble Warnings Presaged Trump and Brexit." *The Guardian* (January 2017). URL: https://www.theguardian.com/media/2017/jan/08/eli-pariser-activist-whose-filter-bubble-warnings-presaged-trump-and-brexit (accessed April 15th, 2017).

Kühler, Michael and Nadja Jelinek. Introduction to *Autonomy and the Self*, ix – xxxvi. Dordrecht: Springer Netherlands, 2013. DOI 10.1007/978-94-007-4789-0.

Moran, Richard. "On Frankfurt's The Reasons of Love." In *The Philosophical Imagination: Selected Essays*, 158 – 168. Oxford University Press, 2017.

Owen, Gareth S., Fabian Freyenhagen, Geneyra Richardson and Matthew Hotopf. "Mental Capacity and Decisional Autonomy: An Interdisciplinary Challenge." *Inquiry* 52 (2009), 79 – 107. DOI: 10.1080/00201740802661502.

Paphitis, Sharli Anne. "Questions of the Self in the Personal Autonomy Debate: Some Critical Remarks on Frankfurt and Watson." *South African Journal of Philosophy* 2 (2010), 117 – 131. DOI: 10.4314/sajpem.v29i2.57055.

Pariser, Eli. "The Adderall Society." In *The Filter Bubble: What the Internet is Hiding From You*, 1 – 58. London: Viking E-Book, 2011.

Pariser, Eli. "The You Loop." In *The Filter Bubble: What the Internet is Hiding From You*, 1 – 51. London: Viking E-Book, 2011.

Pariser, Eli. Introduction to *The Filter Bubble: What the Internet is Hiding From You*, 1 – 36. London: Viking E-Book, 2011.

Piper, Mark. "Autonomy: Normative." In *Internet Encyclopedia of Philosophy*. URL: http://www.iep.utm.edu/aut-norm/ (accessed April 13th, 2017).

Sandvig, Christian. "The Facebook 'It's Not Our Fault' Study." In *The Social Media Collective (SMC): Research Blog* (May 2015). URL: https://socialmediacollective.org/2015/05/07/the-facebook-its-not-our-fault-study/ (accessed May 7th, 2017).

Standal, Odin H. *Morning Light Reflected in a Bubble over the Fjord*, 2012. Digital image, available from: 500px, https://500px.com/photo/5322307/morning-light-reflected-in-a-soap-bubble-over-the-fjord-by-odin-hole-standal/ (accessed May 9th, 2017).

Vezér, Martin. "On the Concept of Personhood: A Comparative Analysis of Three Accounts." *Lyceum Philosophy* 1 (Fall 2007). 1-15. URL: http://lyceumphilosophy.com/9-1/Lyceum-9-1-Vezer.pdf (accessed May 26th, 2017).