

**Ethical Challenges of the Recommendation Algorithm:
*Should we make social media more boring?***

While the ubiquity of social media gives rise to a myriad of ethical challenges, much of the current debate centers around bad actors. From bullies and hackers to politicians and corporates, bad actors are individuals or groups who take advantage of features of social media to cyberbully/shame, spread misinformation, collect private information, or simply troll around. But in this essay, I discuss an ethical challenge caused by a feature intrinsic to the medium of social media: recommendation algorithms. Unlike bad actors which can be regulated through various methods (e.g., identity verification, account removal, shadow banning), recommendation algorithms cannot be separated from social media. I argue that social media algorithms can infringe on free will and undermine the individual's ability to self-consciously form their own preferences and character, preventing them from becoming the person they want to be. Specifically, I employ Frankfurt's framework of desires and volitions, and focus on two dimensions: (1) how algorithms form preferences and desires, and (2) how algorithms create political echo chambers and limit objective news.

For the scope of this paper, recommendation algorithms are defined as mathematical input/output systems that take in a variety of factors relating to the user (such as age, gender, geographic location, and engagement rate with previous posts) and output content that they believe the user will find interesting/relevant. These algorithms mean that there will be endless entertaining content that a user can consume, and such content would be unique to each user. Recommendation algorithms are arguably the main feature that distinguishes social media from traditional media. They imply that most of our time on social media is passive rather than active: instead of seeking content, we consume whatever it presents us. When we buy/borrow books, magazines, or videotapes, we consciously pick the title based on blurbs, reviews, and our own

preferences. On the other hand, when we use TikTok, Instagram, Twitter, YouTube etc., we do not know what we will encounter until we open the app. The algorithm decides what we see and think about.

We can now introduce Frankfurt's framework of desires and free will to analyze the implications of the effects of such an algorithm. Frankfurt distinguishes two manners of desires when we say, 'Person A wants to X'. If X refers to an action, the statement 'Person A wants to X' refers to a first-order desire. According to Frankfurt, actions are first-order desires for two main reasons. First, when A wants to X, it could be that X conflicts with other actions ABCD which Person A feels is more important, and so despite strongly wanting X, A may or may not act on X. Second, Frankfurt states "it may be true that he wants to X despite the fact that, when he acts, it is not the desire to X that motivates him to do what he does"¹, meaning that wanting to X is not necessarily wanting to *perform* the *actual* action. For example, suppose that A struggles with substance abuse and has the desire to take drugs. Even though taking drugs is the X in this case, A is not wanting the physical action/motions of taking drugs (e.g., the act of rolling weed and smoking it), but the effects of taking drugs.²

On the other hand, second-order desires take the form "A wants to X" where X refers to a first-order desire, meaning second-order desires can be formulated as "A wants to want to X". Second-order desires can and often do conflict with first-order desires. Continuing the example of a person fighting against drug addiction, A's second-order desire might be "I want to *not* want to take drugs" even though his first-order desire is "I want to take drugs"³. Most importantly, Frankfurt argues that only when A wants the first-order desire to be *effective*, meaning to actually

¹ Frankfurt, Harry. *Freedom of the Will and Concept of the Person* (The Journal of Philosophy, 1971), 8

² X is not necessarily single action (e.g., sleep or eat), it can also be objects or traits, as we want to *have* them.

³ *Ibid.*, 10

motivate his actions, can we say that his second-order desire is aligned with his will—and he is acting freely. For example, if my first-order desire is to be caring for my friends, then my second-order desire to want this first-order desire is *effective* since I want my daily actions to be guided by consideration towards my friends. This kind of second-order desire is called “second-order volitions” (to distinguish it from second-order desires that do not involve wanting the first-order desire to be *effective*).⁴ Frankfurt argues that such volitions are “essential to being a person”: the ability to step back and evaluate the desirability of desires allows a person to enjoy or lack free will and distinguishes a person from animals that must act according to first-order desires.⁵

After characterizing Frankfurt’s theory, we now return to recommendation algorithms. Take the prevalent example of lifestyle influencers—people who sell a certain aspirational persona, whether it be through material goods (e.g., houses, accessories, clothing, food), experiences (e.g., travelling, having a certain career, raving, certain relationships), or certain traits/skills (e.g., beauty, confidence, gaming). When we are recommended such content, we see their lifestyle being presented in a romanticized and enticing manner, and we see that it has earned them millions of views, tens of thousands of likes, and hundreds of adoring comments. We are then liable to form a desire to be like the content in some manner. Thus, the recommendation algorithm instills a variety of random first-order desires in us, as the desires all take the form “I want to have wealth” or “I want to have beauty”. The use of social media also instills the first-order desire to use it, i.e. “I want to open YouTube” or “I want to keep scrolling on TikTok”. However, we do not want the majority of these first-order desires to be effective—i.e., we do not want to want to use social media all the time, we want to *not* want to use social

⁴ *Ibid.*, 11

⁵ *Ibid.*, 14

media all the time. By creating undesirable first-order desires that conflict with second-order desires, the recommendation algorithm harms the freedom of our will.

Next, I turn my discussion to how social media recommendation algorithms impede free will in its shifts toward being a major source of news. Today, YouTube and Twitter are the websites with #2 and #3 most daily visits (behind Google), while websites like The New York Times rank at #51⁶. TikTok has 50 million daily active users in America., while cable news has 2.12 daily prime-time viewers in total.⁷ Many social media users might not ever watch traditional news sources, resulting in social media becoming the main news source for a large population. This is not a role that social media was created for. The nature of the recommendation algorithm—giving users more of what they are interested in—means that an echo chamber is inevitable. People will only see content that they already agree with and content that is popular. Thus, many perspectives and equally important issues will be left out. This is problematic in many ways, with an obvious case being in politics: democrats and republicans will be shown entirely different stories. Since social media is usually less formal and sources might be more provocative—including open slander, shaming or humiliation—this results in polarization and animosity.

Then how do issues such as political polarization relate to free will? There are many parallels between how lifestyle content instills desires and how news endorses certain agendas/beliefs and pushes certain topics to the forefront. For example, consider the recent discussions of abortion caused by the Roe v. Wade decision. Suppose that there is Person A and Person B, and based on their past activity, A might be considered more conservative and

⁶ O'Brien, Clodagh. "[How Do Social Media Algorithms Work?](#)" The Digital Marketing Institute. Jan 19, 2022.

⁷ "[Leading cable news networks in July 2022.](#)" Statista. July 2022.

Christian while B is more liberal. The algorithm would recommend pro-life content to A and pro-choice content to B. Each person would be presented with content showing how nonsensical the opposing side is, how the other side is bad people, and affirming the correctness of their own side. If A or B were initially neutral, they are now likely to believe in whichever side the algorithm selected, since they were not given a fair view of the other side. If they were already leaning to one side, then they would only feel confirmation bias and feel no need to critically evaluate their own argument or consider other perspectives. Although such news does not elicit a first-order desire, wanting to watch such news is a first-order desire which can be formulated as ‘A/B wants to watch sensational content that makes them feel morally in the right’. However, this first-order desire is not desirable—we would not want our actions guided by a want for sensational content. Meanwhile, the first-order desire ‘I want to watch objective, well-analyzed news’ is a desirable first-order desire. Thus, although A and B might both have a second-order desire ‘I want to want objective news and to be able to make my own judgment’, their first-order desires are not in alignment. Furthermore, Frankfurt also writes: “The will of the person whose will is free could have been otherwise; he could have done otherwise than to constitute his will as he did”. However, if A or B only ever received biased information favoring one side, they encounter a distorted reality and could never have formulated a differing opinion. They would not even realize that they are in an echo chamber and have no way to escape it. Thus, just like desire-instilling influencer content, algorithm-recommended news can undermine free will.

One objection to the above might be that first-order desires created by consuming recommended content are not *necessarily* in conflict with second-order desires, even if they are random and arbitrary. For example, after watching an algorithm-recommended video of a guitar player, Alice might form the first-order desire ‘I want to play guitar’, and she might have a

second-order volition: she wants her actions to be motivated by this first-order desire so she can learn and practice guitar. Alice might become a great guitar player which brings her immense joy and fulfillment. This way of picking up a hobby is just as organic and authentic as picking it up as a child or after watching a live performance. I have three replies to this objection. First, it is not necessary that *all* of the algorithms-recommended content contradicts free will for recommendation algorithms to undermine free will. Second, even if we form a second-order volition through social media, we might still feel a first-order desire to continue consuming content, which can give rise to some other volition that contradicts the original volition to play guitar. This cycle can continue forever, resulting in us agitatedly moving from volition to volition with no gain. Third, Frankfurt also wrote that second-order volitions are formed from reflection rather than “mindless indifference” towards first-order desires⁸. Recommendation algorithms do not allow us the time for such reflection as it maintains an endless ocean of content. This contrasts with becoming interested in something through watching a performance or a friend’s introduction, since these encounters can only happen so often. Thus, a second-order volition formed through social media might be less authentic than one formed through some consideration and reflection. Furthermore, this second-order volition might contradict more important pre-existing volitions. Returning to the guitar example, let us say that Alice also has the second-order volition to become a philosophy professor, and has acted in ways motivated by this desire since high school. While applying to Ph.D. programs, she is distracted by this other volition to play the guitar and becomes a good guitar player. However, she fails to get into a Ph.D. program despite valuing it significantly more than playing guitar. Thus, even though algorithms can create

⁸ Frankfurt, 13

second-order volitions, they can still limit our free will by making us want to want less important things than what we already want to want.

Finally, I discuss some solutions to the problem. One obvious one is simply for everyone to use less social media, or only use social media when looking for specific content (i.e. to use it as we use Google). However, this solution feels a little like telling a drowning person to swim. There are things that can be implemented from an algorithm perspective to alleviate the problem.

First, it would be beneficial to increase transparency by making the input and outputs of each user's algorithm visible, so the user has a better perspective on what kind of content they are being recommended and why. It may be easier for the user to align her first-order desire regarding social media ("I want to use/not use social media") to her second-order volition "I don't want to want to use social media" if the user sees the types of content she will be presented and realizes the content will cause other undesirable first-order desires.

Second, when recommendation algorithms were designed, the creators did not take into consideration the fact that social media might become the main source of news for many. Thus, the algorithm should be adjusted to not include factors such as the user's political beliefs when calculating recommendations. A politically neutral algorithm would show both sides of an issue. Of course, the user might disagree, but at least the algorithm has presented counterarguments and given the user a free choice between sides. Similarly, perhaps it would be beneficial to make social media more boring to limit daily usage time, by making the algorithm less effective.

In conclusion, I have shown how social media recommendation algorithms can undermine free will and personhood by applying Frankfurt's framework of desires. A future topic of investigation could be to consider whether tech companies now have a moral responsibility to modify algorithms to avoid (or at least decrease the extent of) affecting free will. Another might

be how recommendation algorithms may have different effects on people in different groups. Inevitably, as more people spend more time on social media, we will give more power to the algorithms that govern the digital world. The ethics of algorithms and their role in shaping our lives warrants further philosophical inquiry.