EJLT European Journal of Law and Technology

Social Media Addiction, Behaviourism, and the Limits of the Digital Services Act

Will Mbioh*

Abstract:

This paper examines the Digital Services Act's behaviourist framing of 'platform addiction' and its reliance on stimulus–response assumptions about interface design. It argues that this model is both internally inconsistent and causally misattributed, privileging autoplay, infinite scroll, and notifications while downplaying intrapersonal, psychosocial, and structural drivers of compulsive use. To address this issue, it contends that proportionate design governance, e.g., the European Parliament's proposed 'right not to be disturbed' should be paired with upstream measures: early identification and evidence-based care in education settings, embedded mental-health counselling with clear crisis pathways, community and anti-loneliness programmes, and policies that reduce precarity. This dual approach both blunts engagement features and reduces the conditions that amplify risk, making problematic, compulsive use less likely.

Keywords: Digital Services Act; social media; addiction; behaviourism; mental health

1. Introduction

Over the past decade, compulsive social media use has increasingly been conceptualised as a significant public health concern,[1] with policymakers, regulators,

* University of Kent, Kent Law School.

[1] David S Bickham, 'Current Research and Viewpoints on Internet Addiction in Adolescents' (2021) 9 Current pediatrics reports 1; Rosamma Alexander and Clarissa F Delariarte, 'A

and researchers-particularly those influenced by the surveillance capitalism scholarship-[2] framing problematic digital engagement within an addiction paradigm,[3] drawing parallels between compulsive social media use and behavioural addictions such as gambling, and framing it, largely, as an effect of platform design informed by behaviourist mechanisms of: (a) classical conditioning (the process by which a neutral stimulus becomes associated with a reflexive response through repeated pairing with an unconditioned stimulus) and (b) operant conditioning (the process by which voluntary behaviour is strengthened or weakened through reinforcements and punishments rather than passive associations) associated with Ivan Pavlov and B.F. Skinner, respectively.[4] That is, social media addiction is understood as a consequence of interface design (e.g., notifications, visual alerts, and content previews in the case of classical conditioning) and features that stimulate or condition habitual engagement (e.g., likes, shares, comment interactions, and algorithmic personalisation in the case of operant conditioning) through externally mediated stimuli and rewards.[5] As such, the dominant theoretical model underpinning regulatory interventions has been informed by Skinnerian and Pavlovian behaviourism,[6] which conceptualises compulsive engagement as a function of external stimuli rather than internal cognitive, affective, or psychosocial determinants-such as intrinsic motivations, psychological distress, and broader social contexts-which are largely excluded from analysis, as they are considered unobservable mental states within the behaviourist paradigm.[7]

I will contend that this behaviourist paradigm has informed the EU's regulatory response to social media/platform addiction, as evidenced by the language, provisions, and implicit assumptions of the Digital Services Act (DSA) (2022). While

---

Qualitative Study on the Stages of Social Media Addiction among Adolescents in India.' (2022) 10 IAHRW International Journal of Social Sciences Review 442.

[2] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019) 213–218, 293–297.

[3] A Alter, *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked* (Penguin 2017).

[4] Alter (n 3); Zuboff (n 2).

[5] Vikram R Bhargava and Manuel Velasquez, 'Ethics of the Attention Economy: The Problem of Social Media Addiction' (2021) 31 Business Ethics Quarterly 321; Christian Montag and others, 'Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories' (2019) 16 International journal of environmental research and public health 2612; Alejandro L Mujica and others, 'Addiction by Design: Some Dimensions and Challenges of Excessive Social Media Use' (2022) 10 Medical Research Archives 1-18.

[6] Burrhus Frederic Skinner, 'The Evolution of Behaviour 1', *Behaviour analysis and contemporary psychology* (Routledge 2022) 33–40.

[7] John A Bargh and Melissa J Ferguson, 'Beyond Behaviorism: On the Automaticity of Higher Mental Processes.' (2000) 126 Psychological bulletin 925; Jay Moore, 'The Basic Principles of Behaviorism', *The philosophical legacy of behaviorism* (Springer 1999); Maroua Rogti, 'Behaviorism as External Stimuli: Improving Student Extrinsic Motivation through Behavioral Responses in Algerian College Education' (2021) 21 Global Journal of Human-Social Science 29; Skinner (n 6).

the legislation does not ban or require the removal of engagement-maximising features on major social media platforms, the DSA requires that platforms identify features and stimuli that "stimulate" and "cause"-the DSA uses these terms interchangeably (Recitals 81, 83)-that is, condition behavioural addiction in users, and to implement risk-mitigation measures. In practice, this generally involves offering users discretionary controls, such as optional toggles or settings menus, to modify or disable specific engagement-maximising features-including autoplay, infinite scroll, or notifications-on the assumption that reducing exposure to these stimuli will correspondingly reduce compulsive engagement (see DSA, Recitals 81, 83, 86, and Articles 34 and 35). The underlying regulatory logic, therefore, presumes that granting users the ability to modify engagement-maximising features will be sufficient to interrupt conditioned engagement cycles and mitigate compulsive social media use.

However, I will contend that the fundamental issue with the DSA is not that it recognises these dynamics, but that it operates under contradictory and incoherent assumptions about the nature of digital engagement and compulsive social media use. On the one hand, it assumes that users are being passively conditioned into compulsive engagement, akin to Pavlov's dogs or Skinner's rats, where external platform features directly "stimulate" addiction. On the other hand, it also assumes that users can exercise rational agency and self-regulation when given the option to disable or modify these same engagement-maximising features. These two positions, I will argue, are logically inconsistent: if users are so deeply conditioned by platform stimuli that they cannot disengage, then optional toggles are unlikely to be an effective intervention on its own. Conversely, if users are capable of rational decision-making and self-regulation, then compulsive engagement cannot be reduced to mere behavioural conditioning, as the DSA's reasoning suggests. This tension risks undermining the coherence of the Act's regulatory logic from the outset.

Beyond this contradiction, the deeper limitation of the DSA, I will argue, lies in its misattribution of the causes of compulsive engagement. The regulation assumes that modifying external engagement features will meaningfully reduce problematic use. Yet a substantial body of research shows that, for a defined subset of users-particularly those with pre-existing mental-health vulnerabilities (e.g. anxiety or depression), chronic social disconnection or loneliness, low self-esteem or identity distress, and comparable risk factors-social media functions less as a mere habit and more as an adaptive coping mechanism.[8] Such users do not engage compulsively

[8] Cecilia Cheng, Peizhen Sun and Kwok-Kei Mak, 'Internet Addiction and Psychosocial Maladjustment: Avoidant Coping and Coping Inflexibility as Psychological Mechanisms' (2015) 18 Cyberpsychology, Behavior, and Social Networking 539; Michelle L McNicol and Einar B Thorsteinsson, 'Internet Addiction, Psychological Distress, and Coping Responses among Adolescents and Adults' (2017) 20 Cyberpsychology, Behavior, and Social Networking 296; Jorge J Varela and others, 'Wellbeing, Social Media Addiction and Coping Strategies among Chilean Adolescents during the Pandemic' (2023) 14 Frontiers in Psychiatry 1211431; Esen Yildirim Demirdöğen and others, 'Social Media Addiction, Escapism and Coping Strategies Are Associated

simply because autoplay exists; they also engage because platforms offer accessible avenues for affect regulation, validation, and connection that may be unavailable elsewhere.[9] For this subset, modifying platform features is unlikely to remove the underlying drivers of engagement and will therefore do little to change compulsive use in a meaningful way.

Accordingly, I contend that while there is a place for targeted, algorithmic-facing measures, such as the European Parliament's proposed "right not to be disturbed"[10] with default-off settings for attention-seeking features, the main gains are likely to come from upstream interventions that reduce the very conditions under which compulsive use takes hold. That means resourcing accessible mental-health support (early identification in schools; clear referral pathways; evidence-based therapies for co-morbid anxiety, depression, and ADHD), rebuilding social infrastructure that counters isolation and loneliness (community hubs, youth services, credible social-prescribing routes), and tackling structural precarity (housing insecurity, unpredictable hours, income volatility) that functions as a persistent stressor. In regulatory terms, engagement features should be treated as risk multipliers that interact with baseline vulnerability, not as sole causes.

To develop this argument, I will first outline the Digital Services Act and the behaviourist paradigm under which it operates. I will then outline its internal contradictions and present existing academic evidence suggesting that problematic social media engagement is often more closely linked to internal motivations and broader social and structural factors that operate alongside the conditioning of engagement-maximising features but also as separate and distinct mechanisms. Then, I will set out a brief programme of action pairing proportionate design controls with upstream mental-health and social measures to address the drivers of compulsive use.

2. Regulating Digital Design: How the DSA Targets Engagement-Maximising Features

The European Union's Digital Services Act (DSA) (Regulation (EU) 2022/2065) represents a new approach to regulating digital platforms in Europe-far exceeding not

with the Problematic Internet Use of Adolescents in Türkiye: A Multi-Center Study' (2024) 15 Frontiers in Psychiatry 1355759.

[9] Bickham (n 1); Martin Mihajlov and Lucija Vejmelka, 'Internet Addiction: A Review of the First Twenty Years' (2017) 29 Psychiatria Danubina 260; Montserrat Peris and others, 'Psychological Risk Factors That Predict Social Networking and Internet Addiction in Adolescents' (2020) 17 International journal of environmental research and public health 4598; Arti Soni, 'The Relationship Between Social Media Addiction And Mental Health: A Comprehensive Review' (2024) 10 EPRA International Journal of Multidisciplinary Research (IJMR) 60.

[10] European Parliament, 'European Parliament Resolution of 12 December 2023 on Addictive Design of Online Services and Consumer Protection in the EU Single Market (2023/2043(INI))' (2024) ELI: http://data.europa.eu/eli/C/2024/4164/oj.

only the scope of the UK's Online Safety Act (OSA 2024) but also going well beyond anything currently seen in the United States, which does not have, at the federal level, anything equivalent to the DSA or OSA. While the OSA primarily targets harmful and illegal content (see Online Safety Act 2024, ss 9, 27, 31), the DSA takes a broader approach, addressing not only content moderation but also the design and operational features of platforms.

The DSA is a horizontal, tiered regime for intermediary services; that is, mere conduits, caching, hosting, online platforms, and, at the apex, very large online platforms and search engines (VLOPs/VLOSEs). Its objectives are to secure a safer online environment, safeguard fundamental rights, and preserve the internal market. The structure couples baseline duties for all providers (points of contact, notice-and-action, reason-giving, trusted flaggers, transparency reporting) with escalating, risk-based obligations for the largest services. *Importantly*, the substantive *subject matter* of "systemic risks" is set out in the recitals: they enumerate the categories that must be considered in risk assessments-dissemination of illegal content; adverse effects on fundamental rights (expression, privacy, non-discrimination, and child protection); and the foreseeable negative impact of service design and recommender systems on public security, civic discourse, electoral processes, and public health. The recitals also flag risks linked to manipulative or addictive interface features, such as infinite scroll, autoplay, and notification architectures, with particular salience for minors (e.g., Recitals 81, 83).

Therefore, the recitals do more than "set the scene": in the DSA they supply the operative catalogue of "systemic risks" that Article 34 assessments must use and Article 35 must mitigate, because those categories are not exhaustively defined in the Articles themselves. In EU law, this is orthodox. Where the enacting terms are open-textured, the Court of Justice reads them in light of purpose and context; recitals are a primary guide in that teleological analysis. The Court has relied on recitals to ascertain an instrument's aims and structure[11] and has held that ambiguous provisions be construed to give effect to the Directive/Regulation's objectives rather than a narrow literalism.[12] Put differently: recitals cannot create free-standing obligations or override clear operative text, but when the legislature uses capacious standards (e.g., "systemic risks", "reasonable, proportionate and effective") the recitals clarify the subject-matter to which those standards attach and therefore carry significant interpretive weight for regulatory supervisors, platforms and courts. In the DSA, that means the recital list- illegal content, fundamental-rights impacts, harms linked to interface/recommender design (including addictive/manipulative features, with special attention to minors)- sets the substantive horizon of what risks must be identified (Article 34), evidenced and mitigated (Article 35) in the annual risk-management cycle.

[11] *Tomasz Ziolkowski v Land Berlin; Barbara Szeja and Others v Land Berlin* (2011) I 14035 [42–43].

[12] *The Queen v Secretary of State for Trade and Industry, ex parte Broadcasting, Entertainment, Cinematographic and Theatre Union (BECTU)* (2001) I 4881 [36–38].

For example, unlike the OSA, the DSA recognises, particularly in Recitals 81 and 83 and Articles 34 and 35, that platform design elements such as infinite scroll, autoplay, notifications, and recommendation algorithms can 'stimulate' or 'cause' (the DSA uses the terms interchangeably in Recital 81 ('cause addictive behaviour') and Recital 83 ('stimulate behavioural addictions')) user addiction, especially among minors (Recital 81), but also with regard to all users, generally (Recital 83). Thus, to target these stimuli, Articles 34 and 35 of the DSA obligate very large online platforms (VLOPs) (e.g., Instagram, TikTok, YouTube) to, firstly, identify stimuli that stimulate 'behavioural addiction' and take steps to mitigate this risk and modify the operation of these features. Specifically, Article 34 requires that VLOPs conduct risk assessments at least annually to analyse the "systemic risks" identified in the recitals arising from their services, including those stemming from the design or functioning of their algorithmic systems and online interfaces, as well as to consider how their 'recommender systems and any other relevant algorithmic system' (Article 34(2)(a)) influence these risks. Following the identification of such risks, Article35 imposes an obligation on VLOPs to implement 'reasonable, proportionate and effective mitigation measures' to address them, which may include 'adapting the design, features or functioning of their services, including their online interfaces' (Article35(1)(a)) and 'testing and adapting their algorithmic systems, including their recommender systems' (Article 35(1)(d)).

However, Recital 86 of the DSA is legally significant because it introduces a clear interpretative constraint on how Articles 34 and 35, which require systemic risk assessments and mitigation measures, must be interpreted and applied. Specifically, Recital 86 imposes a key limitation on what constitutes a reasonable and effective mitigation measure: the obligation to "avoid unnecessary restrictions on the use of their service." This constraint materially restricts how Articles 34 and 35 should be understood and applied in ways, I contend, that significantly narrow the scope of permissible mitigation measures. This has direct consequences for how platforms are expected to address risks such as behavioural addiction. In that, while it is true that recitals do not create independent legal obligations, they are key for interpreting ambiguous legal provisions. Articles 34 and 35 use broad language such as "reasonable, proportionate, and effective" mitigation measures. Recital 86 clarifies what these terms mean - it defines the boundary of what is legally permissible under these articles.

For example, in light of Recital 86, it is clear that the DSA did not intend to ban or prohibit any interface-design features per se. That is, if a systemic risk is identified under Article 34 (e.g., autoplay contributes to addiction), not any or all mitigation measures, including outright bans, can automatically be justified in light of Recital 86. This is because Article 34 only requires risk assessments - it does not specify any particular outcome. Risk identification does not equal an automatic licence to ban features. Article 35 then requires "reasonable, proportionate, and effective" mitigation measures, but Recital 86 defines what is proportionate by prohibiting unnecessary restrictions on service use. Necessity, in other words, must be evaluated

in light of Recital 86's explicit constraints, which prohibit excessive restrictions on how users interact with the service and require an impact assessment on freedom of expression. As such, to "avoid unnecessary restrictions on the use of their service" implies that mitigation must be targeted, minimal, and must not disrupt user engagement beyond what is strictly necessary. However, a feature ban is the most restrictive measure possible - it eliminates an entire form of interaction. This is, by definition, an unnecessary restriction since less intrusive options exist, such as: (a) providing user controls (e.g., letting users toggle autoplay on or off); (b) introducing nudges or transparency labels instead of outright bans; (c) allowing opt-out/in models rather than universal prohibitions. In that, since Recital 86 explicitly forbids unnecessary restrictions, and since feature bans are unnecessary by definition if other options exist, it follows that *automatic* feature bans would be difficult to justify under the DSA.

This is also supported by a literal, linguistic interpretation of Recital 86's obligation on the EU institutions and Member States to, when regulating platforms and making sense of what mitigation measures are required, "avoid unnecessary restrictions on the use of their service" - a phrase that, when examined linguistically, carries precise constraints on the interpretation of Articles 34 and 35. For example, the key word here is "avoid", which, linguistically, carries a preventative meaning rather than a corrective one, where 'preventative' implies stopping an issue from occurring at all, whereas 'corrective' would suggest responding to an issue after it has already materialised. If the DSA had said Member States must "limit unnecessary restrictions" or "minimise unnecessary interference" with platform use, that would mean that some restrictions could be introduced, but they should be kept to a minimum or managed carefully. However, "avoid" does not allow for this flexibility - it suggests that restrictions must not happen in the first place unless strictly necessary. In other words, the burden of proof shifts: instead of assuming that restrictions are acceptable as long as they are kept minimal, the default position under this wording is that restrictions are impermissible unless strictly justified. This is significant because it fundamentally shapes how Articles 34 and 35 should be interpreted: Regulators cannot impose restrictions unless they can establish that there is no reasonable alternative. Thus, the threshold for justifying feature bans is quite high- they must be *the last resort*, not the first option.

This means, if a platform can demonstrate that a non-restrictive measure (e.g., user controls, transparency labels, nudges, or opt-in settings) could achieve a similar risk reduction outcome, regulators cannot justify imposing a feature ban or forcing fundamental alterations to the platform's design. For example, if autoplay is identified as a risk, regulators would have to first consider giving users more control (e.g., toggle options or time limits) rather than banning autoplay outright. And, because Recital 86 prevents unnecessary restrictions on service use, regulators must tailor interventions to only those parts of the platform that are directly contributing to risk - they cannot apply blanket regulations across an entire service. This means that platform-wide restrictions are difficult to justify; instead, interventions must

focus on particular functionalities (e.g., adjustments to recommendation algorithms or features rather than the outright removal of them).

Furthermore, the phrase "avoid unnecessary restrictions on the use of their service" in Recital 86 is particularly significant because it imposes a broad and categorical limitation on regulatory interventions by tightly constraining both what can be restricted (use) and the justification required for any restriction (necessity). In that, the choice of "use of their service" is not confined to basic access to a platform; it extends to how users interact with its features. If the phrase was "access to their service," the protection would only apply to whether users can reach the platform at all. Instead, "use" explicitly safeguards the way users engage with platform functionalities such as autoplay, infinite scroll, notifications, and recommendation algorithms. This means that restrictions cannot just avoid blocking access; they must also avoid altering how users engage with the service. Relatedly, "service" protects the integrity of the platform as a whole. That is to say, the term "service" is broader than any individual feature; it covers the entire way the platform is designed and functions. A restriction is impermissible if it unnecessarily interferes with any part of how the platform provides its core functionality to users. This means that even if a regulation does not ban a platform outright, it should not fundamentally disrupt the way the service operates unless it is strictly necessary.

This reading helps to explain why, to date, the Commission has not, strictly speaking, prohibited a particular *feature* (such as autoplay or infinite scroll) on the ground that it is, by itself, inherently "addictive". The one prominent intervention, the TikTok Lite's "rewards hub", was different in kind: it introduced monetary incentives for time-on-platform and specific interactions, i.e., a pay-to-engage scheme layered on top of existing features.[13] The Commission's immediate concern was not to outlaw that scheme per se but to stop its roll-out until TikTok had complied with the DSA's ex ante duties: a documented Article 34 systemic risk assessment and Article 35 mitigation plan proportionate to the identified risks. In other words, the action targeted *how* a financially incentivised engagement programme was launched without prior safeguards, rather than condemning embedded interface patterns like autoplay or infinite scroll as unlawful in themselves. That distinction matters. A pay-to-watch mechanism is an *extrinsic inducement*, i.e, a reward external to the activity itself (for example, money or points convertible into benefits) that sits above ordinary use, creating a *transactional motive* distinct from the platform's native design cues-whereas autoplay or infinite scroll are interface-level design choices that shape attention from within the service. An inducement scheme can be paused pending procedural compliance; by contrast, a feature-level ban on interface design would be a substantive restriction on "use of the service" and, for reasons tied to Recital 86, far

[13] European Commission, 'Commission Opens Proceedings against TikTok under the DSA Regarding the Launch of TikTok Lite in France and Spain, and Communicates Its Intention to Suspend the Reward Programme in the EU' (European Commission 2024) Press release <https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_24_2227/IP_24_2227_EN.pdf> accessed 16 October 2025.

harder to justify absent proof that no less intrusive alternative could achieve the same risk reduction (and it is distinct from measures aimed at limiting minors' exposure to harmful content).

3. The DSA's Behaviourist Blueprint: Addiction as a Stimulus-Response Mechanism

In addition, because the Act uses the terms "stimulate" and "cause" interchangeably in Recitals 81 and 83, its underlying theory of addiction is one in which compulsive engagement is directly triggered ("stimulated" or "caused") by external platform features. This linguistic choice is significant because, by collapsing stimulation (which typically refers to an external trigger that elicits a reaction) and causation (which implies a direct relationship between cause and effect) into the same meaning (i.e., treating stimulation as a sufficient causal mechanism rather than just one contributing factor), the DSA, I contend, advances a stimulant-based model of addiction. In this thinking, the presence of a stimulus such as autoplay or notifications is presumed to directly produce a compulsive behavioural response in users, without requiring any other mediating psychological or social factors.

The fact that Recital 81 acknowledges that risks "may arise, for example, in relation to the design of online interfaces which intentionally or unintentionally exploit the weaknesses and inexperience of minors or which may cause addictive behaviour" does not materially challenge my argument but rather reinforces it. The key linguistic choice in Recital 81 is that design itself is framed as the source of risk, not just an exacerbating factor. The phrase "may cause addictive behaviour" follows a direct causal structure, implying that platform interfaces are not merely a contributory influence but a sufficient trigger for addiction. This is consistent with my argument that the DSA collapses the distinction between stimulation and causation, treating external digital stimuli (e.g., autoplay, notifications, infinite scroll) as mechanistic triggers rather than contextual factors in a broader psychosocial framework. Had Recital 81 introduced qualifiers such as 'may contribute to,' 'may increase the likelihood of,' or 'may be one factor in' the development of addiction, this would indicate a multi-factorial understanding of compulsive engagement. However, the unqualified use of 'may cause addictive behaviour' reinforces the view that addiction is understood as a direct consequence of design elements themselves, which is consistent with my argument about the DSA's stimulant-based model of addiction.

Furthermore, the reference to "exploiting the weaknesses and inexperience of minors" in Recital 81 does not introduce a fundamentally different, user-centric model of addiction but instead reaffirms the behaviourist, stimulus-driven logic that underpins the DSA's regulatory approach. That is, the phrase "exploit the weaknesses and inexperience of minors" might seem to suggest user-centric model-i.e., that addiction arises because minors have cognitive vulnerabilities. But this is not what the text actually says. Instead, it fits within a stimulus-response model for two key reasons. Firstly, minors' susceptibility is not framed as *the cause* of addiction-the

platform features are. The sentence does not say that addiction happens *because* minors have cognitive weaknesses. Instead, it says that platform features can "exploit" these weaknesses to enhance their addictive effect. This keeps the causal mechanism external (platform design), rather than internal (user psychology). Secondly, it assumes that addiction occurs regardless of user traits - minors just experience it more intensely. The phrase "intentionally or unintentionally" discussed above reinforces this point: the regulatory concern is about the effect of the external stimulus, not whether the addiction is contingent on the user's psychological state. If minors were uniquely at risk because of their cognitive weaknesses, then the regulation would not need to hold platforms accountable for addiction in general; it would only apply to addiction in minors. However, the DSA does not frame addiction as a unique risk for minors. It simply says minors are more strongly affected (see Recital 83, for example, where this is clear). This means the causal mechanism remains external (platform design), and user traits only determine the *intensity* of impact, i.e., individual differences do not disrupt the model. Instead, they simply make the stimulus-response mechanism stronger for certain groups. Different subjects may have different response intensities to stimuli, but the cause of behaviour is primarily external. This linguistic framing effectively treats social media addiction as a mechanistic outcome, where compulsive engagement occurs not as a result of internal cognitive, affective, or social processes, but as a reflexive, Pavlovian, and Skinnerian reaction to external design elements.

In this Pavlovian conceptualisation of addiction (common explicitly or implicitly in surveillance capitalism and persuasive technology literature and also implied by the DSA's language),[14] based on the idea of classical conditioning, the assumption is that compulsive engagement arises from associative learning mechanisms, wherein engagement-maximising features function as conditioned stimuli (CS) that become repeatedly paired with unconditioned stimuli (US) such as dopamine release, social validation, or cognitive distraction, ultimately leading to conditioned responses (CR) in the form of habitual or compulsive platform use.[15] This model presupposes that users do not engage with social media out of volition or intrinsic motivation but rather because platform design features (e.g., infinite scroll, autoplay, and notifications) have been conditioned to elicit an automatic, reflexive response analogous to how Pavlov's dogs, in his experiments, learned to salivate in response to a bell after repeated pairings with food.[16] Over time, these digital features, originally neutral,

[14] Daniel Black, 'The Behavior Economy: The Creation of Behavior as an Object of Online Surveillance' [2024] The Information Society 1; Sidi Athmane Maria Wiam, 'Investigating Digital Totalitarianism Dystopia From the Commodification of Pleasure-Driven Behaviors in"Brave New World"to Capitalism Surveillance in"The Circle"'; Yevhen Laniuk, 'Freedom in the : Lessons from Shoshana Zuboff' (2021) 11 Ethics & Bioethics 67; Alexandra Rutherford, 'BF Skinner and Technology's Nation: Technocracy, Social Engineering, and the Good Life in 20th-Century America.' (2017) 20 History of psychology 290; Zuboff (n 2).
[15] Zuboff (n 2); Alter (n 3); Wiam (n 14).
[16] Zuboff (n 2); Thomas Klikauer, *Media Capitalism: Hegemony in the Age of Mass Deception* (Palgrave Macmillan 2021); Rutherford (n 14).

acquire predictive value, signalling the likelihood of rewarding or reinforcing outcomes, which leads to anticipatory engagement even in the absence of the original unconditioned stimulus.[17] In this way, compulsive use is not driven by internal cognitive, affective, or psychosocial processes but by the learned association between external digital cues and behavioural reinforcement where the platform itself conditions the user's engagement patterns through repeated stimulus-response pairings.[18]

Similarly, in its Skinnerian version, based on the idea of operant conditioning, the assumption is that compulsive social media use is shaped and reinforced through reward-schedule structures, wherein engagement-maximising features serve as reinforcing stimuli that increase the likelihood of repeated engagement.[19] Unlike classical conditioning, where behaviour is elicited through passive stimulus-response associations, operant conditioning posits that behaviour is actively emitted and then strengthened or weakened based on its consequences. In this model, platform engagement is conceptualised as an instrumental behaviour that is maintained through variable reinforcement schedules, where users receive unpredictable, intermittent rewards in the form of likes, comments, and shares, algorithmic amplification (recommendation boosts), and anticipatory engagement (notifications and algorithmic cues promising new content). This also encompasses what is often labelled "gamification" or, at the extreme, "gamblification": points, levels, streaks, badges, progress bars, daily challenges and limited time rewards; variable "loot box"–style drops or surprise unlocks; social leaderboards and rank resets; and recommendation boosts that intermittently amplify visibility after posting.[20] These mechanics layer intermittent, outcome-contingent rewards onto ordinary interaction, increasing response rates and session length in the same way contingent schedules do in laboratory settings.

These features function analogously to Skinner's operant chamber (Skinner Box), where rats or pigeons learned to repeatedly press levers or peck at a target to receive food pellets.[21] Just as the unpredictable nature of variable ratio schedules (where rewards are delivered after an unpredictable number of responses) led to persistent lever pressing in Skinner's experiments, social media's reward structures such as likes appearing at irregular intervals or notifications triggering bursts of engagement reinforce compulsive checking, scrolling, and posting behaviours.[22]

[17] Rutherford (n 14); Zuboff (n 2); Alter (n 3).
[18] Zuboff (n 2); Alter (n 3).
[19] Rutherford (n 14); Black (n 14); Laniuk (n 14); Klikauer (n 16); Joel E Dimsdale, *Dark Persuasion: A History of Brainwashing from Pavlov to Social Media* (Yale University Press 2021).
[20] Michele Ciancimino, 'Private Law and Consumer Protection Paradigms Facing (Digital) "Addictions": A Starting Point for Reflection' (2024) 32 European Review of Private Law 803.
[21] Rutherford (n 14); Zuboff (n 2); Klikauer (n 16).
[22] Alter (n 3).

As with classical conditioning, the assumption is that social media addiction is sustained not by intrinsic cognitive or social processes, but by externally engineered reward patterns that shape behaviour. In this Skinnerian view, platforms operate as digital operant-conditioning environments: users are not merely reacting to cues, they are trained through cycles of reinforcement and avoidance of loss. On this reading, the DSA's regulatory logic treats harm reduction as a matter of breaking those reward schedules, for example, by letting users switch off engagement-maximising features-on the expectation that weakening the underlying reinforcement patterns will lower the frequency of compulsive use.

4. Between Pavlov and Free Will: The D Conditioning and User Agency

However, the first critique of the DSA is that it holds within its logic contradictory assumptions about the nature of compulsive social media engagement. For example, the Act implicitly relies on a behaviourist model of addiction, wherein users are conceptualised as being passively conditioned by digital stimuli in a manner that overrides conscious volition and executive control akin to Skinner's demonstration of reinforcement learning in animal behaviour or Pavlov's associative conditioning and renders their behaviour, largely, into a reflexive response to external cues. However, at the same time, the DSA paradoxically assumes that these same users retain the capacity for self-regulation and behavioural autonomy-presuming that, when presented with an opt-out mechanism (e.g., disabling autoplay or notifications), they will exercise rational agency in modifying their engagement patterns. This assumption is incompatible with the foundational tenets of behaviourist conditioning. If users are indeed subject to compulsive reinforcement mechanisms akin to those seen in conditioned behavioural responses, then providing them with an optional toggle does not constitute an effective intervention.

That is to say, conditioned responses do not simply extinguish in response to the mere presence of an alternative choice.[23] If compulsive engagement were solely driven by external reinforcement, then expecting users to independently disable engagement-maximising features assumes a level of metacognitive control (i.e., the ability to consciously reflect on, regulate, and alter one's own behavioural patterns and cognitive processes, despite the presence of automatic reinforcement mechanisms that typically operate below the level of conscious awareness) that the very logic of behavioural conditioning denies.[24] On a strict behaviourist account, behaviour changes when reinforcement contingencies change, not because a person introspects and overrides them. So, expecting users to "opt out" assumes metacognitive self-control that a behaviourist model does not treat as an operative cause of behaviour while reinforcement is still in place. Further, to be clear, this is *not* an *external*

[23] Skinner (n 6); Burrhus Frederic Skinner, *Contingencies of Reinforcement: A Theoretical Analysis*, vol 3 (BF Skinner Foundation 2014).
[24] Skinner (n 23).

objection but a consequence of the DSA's own logic: having cast engagement features as stimuli that 'cause' or 'stimulate' addictive responses (recitals) and required platforms to mitigate those risks (Articles 34–35), the Act presumes that users can nonetheless neutralise those contingencies through opt-outs, an assumption at odds with the behaviourist model it invokes.

However, if one were to argue that the DSA does not strictly adopt a behaviourist model, but rather takes the position that certain platform features exert some influence on user engagement without fully determining it, this would shift the theoretical foundation of the regulation. Instead of assuming that features such as infinite scroll, autoplay, and notifications operate as stimuli that 'condition' users into compulsive behaviour, the Act could instead be interpreted as acknowledging that these design choices contribute to engagement patterns without entirely overriding individual agency or social and psychological factors.

However, if this were the case, i.e. if the *DSA* merely assumes that these features have some modulating effect rather than directly "stimulating" addiction in the behaviourist sense, then the rationale for regulatory intervention becomes much weaker. If platform features do not 'cause' and/or 'stimulate' addictive behaviour but merely shape or nudge user behaviour in subtle ways, why should they be a primary focus of regulatory attention as a systemic risk? The fundamental question, then, becomes: what exactly is the justification for regulating features that are not addictive but are simply correlated with increased engagement?

That is, if compulsive engagement (by which I mean not simply heavy or frequent use, but a pattern of engagement that becomes difficult or impossible for the user to regulate, characterised by diminished self-control and continued use despite adverse consequences) is not driven by platform design alone but rather emerges from (as I will later argue) a complex interplay of psychological, emotional, and social factors, then regulating specific features as if they are the primary drivers that 'cause' and 'stimulate' addiction is a misdirection of effort. Regulatory intervention typically rests on a reasonably clear causal account: if platform features do not function as addictive stimuli in *a strict behaviourist* sense, i.e., a model that explains behaviour *primarily* as a response to external stimuli while tending to *minimise or bracket* internal motivations, cognitive processes, and wider social context, then regulating them presupposes an unwarranted level of influence. This would mean that the DSA, in its attempt to mitigate compulsive engagement, is targeting *correlates of engagement*, that is, observable patterns or byproducts of use that accompany but do not themselves constitute the underlying causal mechanisms, rather than its causes, raising questions about its effectiveness.

More than that, if the DSA strongly regards these features as a serious public health risk, as implied by its language, then making disengagement, in effect, an optional choice is open to ethical challenge. Regulatory interventions in other domains, such as gambling laws or consumer protections against addictive substances, do not merely rely on individual willpower to prevent harm but impose systemic safeguards

to limit exposure and mitigate risk. For example, gambling regulations in the EU and UK set strict age restrictions, typically at 18, rather than simply offering individuals the option to self-exclude.[25] Similarly, in the case of alcohol and tobacco regulation, governments impose minimum age restrictions, warning labels, and restrictions on where products can be sold or consumed, and taxation as deterrents rather than assuming that individuals alone should manage their consumption.[26] In other words, in regulatory settings where a product or design is said to 'cause' or 'stimulate' addiction, as the DSA's behaviourist framing implies, the usual corollary is to internalise the risk at the system level by imposing provider-facing, structural safeguards that limit exposure by default, rather than outsourcing harm-avoidance to individual willpower. The DSA, by contrast, stops short of imposing meaningful restrictions on engagement-maximising features and instead places the onus on users to self-regulate, despite acknowledging that these features can contribute to compulsive behaviour. This raises an inconsistency: if these design mechanisms pose a significant public health risk comparable to other regulated industries, why does the DSA rely on voluntary disengagement rather than structural intervention?

5. From Opt-Outs to Defaults: Parliament's 'Right the Prospective Digital Fairness Act, a

This is, moreover, a point the European Parliament has recognised. The Parliament's resolution on 'addictive design' expressly urges the Commission to propose a digital 'right not to be disturbed' by turning attention-seeking features off by design (default-off), with any activation left to a user opt-in as a clear call for structural safeguards rather than voluntary disengagement.[27] In parallel, the Commission's Digital Fairness Fitness Check[28] concludes that the current consumer-law toolkit is inadequate for the digital environment and points towards new horizontal rules (often discussed under a prospective 'Digital Fairness Act')[29] to address manipulative/addictive design practices.

---

[25] Deniz Cemiloglu and others, 'Towards Ethical Requirements for Addictive Technology: The Case of Online Gambling', *2020 1st Workshop on Ethics in Requirements Engineering Research and Practice (REthics)* (IEEE 2020); Virve Marionneau and others, 'Limit-Setting in Online Gambling: A Comparative Policy Review of European Approaches' (2025) 22 Harm Reduction Journal 1.

[26] Mirte AG Kuipers and others, 'Tobacco Retail Licencing Systems in Europe'; Anu Linnansaari and others, 'Towards Tobacco-Free Generation: Implementation of Preventive Tobacco Policies in the Nordic Countries' (2023) 51 Scandinavian journal of public health 1108.

[27] European Parliament (n 10).

[28] European Commission, 'Commission Staff Working Document: Fitness Check of EU Consumer Law on Digital Fairness' (European Commission 2024) Staff Working Document <https://commission.europa.eu/document/download/707d7404-78e5-4aef-acfa-82b4cf639f55_en?filename=Commission+Staff+Working+Document+Fitness+Check+on+EU+consumer+law+on+digital+fairness.pdf> accessed 17 October 2025.

[29] European Commission, 'Digital Fairness Act' (*Better Regulation Portal – Have your say (Initiative 14622)*, 17 July 2025) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14622-Digital-Fairness-Act_en> accessed 17 October 2025.

The Digital Fairness Act (DFA)[30] is a consumer law instrument the Commission floated after its 2024 Fitness Check on digital fairness. The aim is to close gaps left by the DSA/GDPR by specifically addressing manipulative interface design and related commercial practices (e.g., dark patterns, so called "addictive" design, exploitative personalisation, influencer marketing, and hidden in-game costs) with firmer duties where children are involved. Instead of relying on opt-outs, as the DSA implies, a DFA is expected to mandate default-off or friction-by-design for certain patterns, prohibit a narrow set of practices outright, and require evidence-based testing, clearer disclosures, and stronger consumer law enforcement. Stakeholder discussions are ongoing ahead of any formal proposal.[31]

Relatedly, we can also compare the DSA with Australia's regulatory approach. Unlike the DSA, its Online Safety Amendment (Social Media Minimum Age) Act 2024 treats the risk as sufficiently coercive to justify structural limits by default: a categorical 16 plus threshold, a statutory duty on "age-restricted social media platforms" to take reasonable steps to prevent under-16s from holding or creating accounts, supporting rules and guidance on how providers are expected to meet that duty, and enforcement by the eSafety Commissioner. Those obligations are designed to internalise the risk at provider level rather than outsource it to user self-help. In other words, the premise that design can condition behaviour yields mandatory age gating (age assurance, account closure workflows, oversight and penalties) rather than optional toggles. By contrast, the DSA's Article 34–35 toolkit remains risk assessment and mitigation led and, read with Recital 86's admonition against unnecessary restrictions on the service, has so far channelled enforcement towards process duties (pre-launch risk assessments, transparency, commitments) rather than prohibiting engagement mechanisms outright.

That is not to say, of course, that the Australian model entirely escapes the contradiction in the EU's logic. In that, if the underlying premise is that these features condition users into (or 'stimulate' or 'cause') behavioural addiction, then this does not simply cease to be true at the age of sixteen. In this sense, while Australia demonstrates greater consistency in its application of behaviourist logic by imposing structural restrictions rather than relying on voluntary disengagement, it, in the end, arrives at a similar contradiction as the DSA simply shifting the boundary of assumed vulnerability rather than resolving the fundamental tension between assumptions of Skinnerian and Pavlovian reflexive, conditioning and assumed user agency and metacognition.

[30] ibid.
[31] ibid.

6. Beyond Platform Design: The Role of Anxiety, Identity, and Social Isolation in Problematic Social Media Use

My second critique of the DSA is that, if it does adopt a behaviourist stance, as its language suggests, its account of social media addiction is thin when set against evidence from clinical psychology, psychiatry, digital anthropology, media studies and sociology. That literature shows that, for a substantial *subset of vulnerable users*, i.e., adolescents and young adults navigating identity and peer status; people with pre-existing depression, anxiety or ADHD; individuals experiencing loneliness, bereavement or trauma; users with high sensitivity to social comparison and body-image concerns; and those under chronic stress or with limited offline support- problematic use is less about exposure to engagement-maximising features and more about social media functioning as a coping strategy for underlying psychological, emotional and social needs. In this sense, the claim is not that all social media use is pathological. For *most* users, everyday engagement is mixed: broadly manageable, with periods of distraction or doom-scrolling alongside genuine benefits - information, entertainment, contact with friends.[32] The focus here is the subset experiencing addiction or clinically problematic use; within this group, the balance of evidence indicates that sustained, harmful patterns are driven less by interface stimuli alone as the Skinnerian/Pavlovian model would suggest, and more by how the medium is recruited to manage unmet needs and adverse life conditions.[33]

For example, emerging evidence from clinical psychology, psychiatry, and epidemiology suggests that what is often characterised as social media addiction or problematic digital engagement is, in many cases, better understood as a form of maladaptive or compensatory affect regulation.[34] That is, rather than simply representing a conditioned behavioural loop, engagement with social media frequently serves as an adaptive, albeit sometimes dysfunctional, coping mechanism; a means through which individuals attempt to manage psychological distress, dysregulated affect, intrusive cognitions, and feelings of social disconnection when other regulatory mechanisms are either ineffective or inaccessible.[35]

---

[32] Yalin Sun and Yan Zhang, 'A Review of Theories and Models Applied in Studies of Social Media Addiction and Implications for Future Research' (2021) 114 Addictive behaviors 106699; Soni (n 9); Jasmine Fardouly, 'Potential Effects of the Social Media Age Ban in Australia for Children Younger than 16 Years' (2025) 7 The Lancet Digital Health e235.

[33] Sun and Zhang (n 32); Xiongfei Cao and others, 'Understanding the Mechanism of Social Media Addiction: A Socio-Technical Systems Perspective' [2023] Information Development 02666669231218685.

[34] McNicol and Thorsteinsson (n 8); Poh Chua Siah and others, 'Dark Triad and Social Media Addiction among Undergraduates: Coping Strategy as a Mediator.' (2021) 13 Contemporary Educational Technology; Yildirim Demirdöğen and others (n 8).

[35] Jai Meynadier and others, 'Relationships Between Social Media Addiction, Social Media Use Metacognitions, Depression, Anxiety, Fear of Missing Out, Loneliness, and Mindfulness' [2025] International Journal of Mental Health and Addiction 1.

This phenomenon is particularly salient among individuals with clinical anxiety disorders, depression, obsessive compulsive disorder (OCD), attention deficit hyperactivity disorder (ADHD), generalised anxiety disorder (GAD), acute stress disorder, and other mood and neurodevelopmental conditions, all of which are characterised by difficulties in emotional self-regulation, cognitive control, and distress tolerance.[36] For these populations, social media does not function merely as a passive stimulant but instead as an active tool through which they attempt to modulate internal states. For example, individuals with generalised anxiety disorder (GAD) or acute stress disorder may engage in compulsive scrolling behaviours as a form of cognitive avoidance, a strategy that temporarily numbs or suppresses overwhelming worry, hyperarousal, or physiological stress responses.[37] Similarly, individuals with OCD often display compulsive reassurance seeking behaviours, repeatedly checking notifications, messages, or online interactions as part of a ritualistic attempt to alleviate uncertainty or intrusive thoughts.[38]

For individuals with ADHD, whose neurocognitive profiles include impulse-control deficits, difficulties with attentional regulation, and reward processing dysregulation, social media engagement is often driven by dopaminergic reinforcement mechanisms.[39] That is, the short form, high-stimulus nature of digital content can be consistent with attentional dysregulation patterns in ways that can result in engagement loops wherein brief bursts of novelty seeking behaviour provide immediate gratification, reducing boredom and compensating for executive dysfunction.[40] Similarly, individuals experiencing clinical depression may turn to social media as a means of maintaining interpersonal connections, validating self-worth, or disengaging from dysphoric mood states. For these individuals, digital engagement

[36] Daria J Kuss and Olatz Lopez-Fernandez, 'Internet Addiction and Problematic Internet Use: A Systematic Review of Clinical Research' (2016) 6 World journal of psychiatry 143; Umay Bilge Baltaci, Melike Yilmaz and Zeliha Tras, 'The Relationships between Internet Addiction, Social Appearance Anxiety and Coping with Stress.' (2021) 14 International Education Studies 135; Fatemeh Teymouri Farkush, Mohsen Kachooei and Elahe Vahidi, 'The Relationship between Shame and Internet Addiction among University Students: The Mediating Role of Experiential Avoidance' (2022) 27 International Journal of Adolescence and Youth 102; Po-Ching Huang and others, 'Associations between Social Media Addiction, Psychological Distress, and Food Addiction among Taiwanese University Students' (2023) 11 Journal of Eating Disorders 43.

[37] McNicol and Thorsteinsson (n 8).

[38] Andrea Guazzini and others, 'Obsessive-Compulsive Disorder (OCD) Types and Social Media: Are Social Media Important and Impactful for OCD People?' (2022) 12 European Journal of Investigation in Health, Psychology and Education 1108; Martine J van Bennekom, Pelle P de Koning and Damiaan Denys, 'Social Media and Smartphone Technology in the Symptomatology of OCD' (2018) 2018 Case Reports bcr.

[39] Chao-Ying Chen and others, 'Problematic Use of Internet Associates with Poor Quality of Life via Psychological Distress in Individuals with ADHD' [2024] Psychology Research and Behavior Management 443; GL Burns and JA Walsh, 'The Influence of ADHD–Hyperactivity/Impulsivity Symptoms on the Development of Oppositional Defiant Disorder Symptoms in a 2-Year Longitudinal Study' (2002) 30 Journal of abnormal child psychology 245 EP.

[40] Shanna Indira Fayzhia and Yunias Setiawati, 'The Interplay of ADHD, Social Media Usage, and Dopamine Receptors in Adolescents: A'.

may serve as a form of cognitive distraction, helping to momentarily suppress ruminative thoughts.[41] However, while social media may provide short term relief, it often fails to produce lasting improvements in mood, and paradoxically, prolonged engagement can exacerbate symptoms of anhedonia, social withdrawal, and negative self-appraisal.[42]

Beyond its role in affect regulation, for many others, social media functions as a tool for seeking validation, a space where users look for affirmation, recognition, and reassurance from others. This is particularly significant for individuals who experience low self-esteem, social anxiety, feelings of invisibility, or uncertainty about their self-worth.[43] Validation seeking on social media manifests in different ways, ranging from posting content to receive likes and comments, to checking messages and notifications compulsively, to curating one's online presence in ways that generate approval from peers, acquaintances, or even strangers.[44]

And, for many who struggle to understand themselves (that is, individuals experiencing identity diffusion, internal conflict regarding self-concept, or difficulties in integrating aspects of their personality into a coherent sense of self),[45] social media can be the place where they can explore, refine, and solidify their identity.[46] In practice, this means using it to find community and recognition by connecting with others who share their experiences, particularly if offline spaces feel isolating for many reasons, including, for example, not fitting into social norms. For example, if they live in a place where people do not share their interests, values, identity, or life circumstances, offline spaces can feel alienating, marginalising, and excluding and lead them to feel disconnected from those around them.[47] We see this especially, for example, with LGBTQ+ individuals who may live in socially conservative areas where their identities are not fully accepted or understood, leaving them without local

---

[41] Betul Keles, Niall McCrae and Annmarie Grealish, 'A Systematic Review: The Influence of Social Media on Depression, Anxiety and Psychological Distress in Adolescents' (2020) 25 International Journal of Adolescence and Youth 79.

[42] Zahir Vally, Mai Helmy and Louis Fourie, 'The Association between Depression and Addictive Social Media Use during the COVID-19 Pandemic: The Mediating Role of Sense of Control' (2023) 18 Plos one e0291034; Keles, McCrae and Grealish (n 41).

[43] Nora Alicia Maza, 'Examining Personal Self-Concept, Social Self-Concept, and External Validation Seeking as Predictors of Problematic Social Media Use in 18 to 22 Year Olds' (PhD Thesis, Texas A&M University-Corpus Christi 2022).

[44] Lee Humphreys, *The Qualified Self: Social Media and the Accounting of Everyday Life* (MIT press 2018); Chelsea P Butkowski and others, 'Quantifying the Feminine Self (Ie): Gender Display and Social Media Feedback in Young Women's Instagram Selfies' (2020) 22 New Media & Society 817.

[45] Maza (n 43).

[46] Sandra B Sebre and Anika Miltuze, 'Digital Media as a Medium for Adolescent Identity Development' [2021] Technology, Knowledge and Learning 1.

[47] Andre Cavalcante, 'Tumbling into Queer Utopias and Vortexes: Experiences of LGBTQ Social Media Users on Tumblr', *LGBTQ Culture* (Routledge 2020); Leanna Lucero, 'Safe Spaces in Online Places: Social Media and LGBTQ Youth' (2017) 9 Multicultural Education Review 117.

communities that reflect their experiences.[48] Similarly, disabled individuals, particularly those with mobility challenges or chronic illnesses, may struggle to participate in offline social spaces that are physically inaccessible or structured in ways that do not accommodate their needs.[49] And, we also see this, too, with immigrants, expatriates, and diaspora communities, who may feel culturally or linguistically disconnected from those around them, using social media to maintain ties with home, seek out others who share their background, or find spaces where they do not have to constantly translate themselves.[50] Similarly, racial and ethnic minorities who experience discrimination or marginalisation in their offline environments often turn to social media to build solidarity, share experiences, and affirm their identities in ways that may not be possible in their immediate surroundings.

For young people in transitional life stages such as teenagers who feel misunderstood by family or peers, university students adjusting to new environments, or young adults navigating unstable careers, social media can serve as an important space for finding guidance, validation, and reassurance from those who have had similar experiences.[51] Likewise, new parents, particularly those facing postpartum isolation, may turn to online communities for support and advice in ways that their offline environments do not provide.[52] In the same vein, caregivers of elderly relatives or disabled family members often feel cut off from traditional social life due to the demands of their responsibilities, using digital spaces as a means of connection, respite, and shared understanding.[53]

---

[48] Cavalcante (n 47); Adam Bates, Trish Hobman and Beth T Bell, '"Let Me Do What I Please with It... Don't Decide My Identity for Me": LGBTQ+ Youth Experiences of Social Media in Narrative Identity Development' (2020) 35 Journal of Adolescent Research 51.

[49] Taxiarchis Vouglanis and Athanasios Drigas, 'The Internet Addiction and the Impact on the Cognitive, Psychological and Social Side of People's Personality with Disabilities' (2022) 35 Technium Soc. Sci. J. 93; IP Volkova and others, 'Problematic Internet Usage by Adolescents with Disabilities' (2019) 21 Образование и наука 98.

[50] Ghadeer Udwan, Koen Leurs and Amanda Alencar, 'Digital Resilience Tactics of Syrian Refugees in the Netherlands: Social Media for Social Support, Health, and Identity' (2020) 6 Social Media+ Society 2056305120915587.

[51] Charles-Étienne White-Gosselin and François Poulin, 'Associations between Young Adults' Social Media Addiction, Relationship Quality with Parents, and Internalizing Problems: A Path Analysis Model.' (2024) 56 Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement 61; Keles, McCrae and Grealish (n 41); Raquel Lozano-Blasco, Alberto Quilez Robres and Alberto Soto Sánchez, 'Internet Addiction in Young Adults: A Meta-Analysis and Systematic Review' (2022) 130 Computers in Human Behavior 107201.

[52] Ashley R MacPherson, Morgan Reid and Natalie Dautovich, 'Examining the Postpartum Period through Social Media: A Content and Thematic Analysis Of# Postpartum Instagram Posts.' (2023) 12 Psychology of Popular Media 424; Brenda Baker and Irene Yang, 'Social Media as Social Support in Pregnancy and the Postpartum' (2018) 17 Sexual & Reproductive Healthcare 31; Emily Hughes, 'A New Mother's Cry: Analyzing Traditional Social Support versus Online Social Support in Relation to the Postpartum Crisis'.

[53] Jazette Johnson and others, '" It's a Lonely Disease": Cultivating Online Spaces for Social Support among People Living with Dementia and Dementia Caregivers' (2022) 6 Proceedings of

At the same time, we see this with individuals experiencing economic precarity such as those working multiple jobs, gig workers with unstable schedules, or those in rural or isolated areas where social infrastructure is limited.[54] For them, social media can provide an important sense of connection and community that their working conditions or geography make difficult to access in person.[55] Additionally, people going through personal or existential crises such as those grappling with grief or major life changes may turn to online spaces for solidarity, emotional processing, and new models of meaning that they cannot find in their immediate surroundings.[56] This includes survivors of abuse or trauma who may find it unsafe or impossible to discuss their experiences openly in their offline lives, but who can find community, validation, and resources in online spaces designed for those with similar experiences.[57]

7. Compensatory Engagement: How Social Media Fills Psychological and Structural Gaps

In light of this, when we look at all these different groups and contexts, a few foundational psychological and structural themes emerge that explain why they may engage with social media in ways that could be considered problematic. Firstly, all of these groups, whether individuals with clinical anxiety, those experiencing identity struggles, caregivers facing isolation, or workers in precarious employment, share a heightened vulnerability to some form of dysregulation. This dysregulation may be affective (emotional instability, distress, loneliness, stress, or anxiety), cognitive (rumination, intrusive thoughts, executive dysfunction, attention dysregulation), or

---

the ACM on Human-Computer Interaction 1; Kristine Newman and others, 'The Role of Internet-Based Digital Tools in Reducing Social Isolation and Addressing Support Needs among Informal Caregivers: A Scoping Review' (2019) 19 BMC Public Health 1; Maryam Tabatabaee and others, 'Loneliness in the Presence of Others: A Mixed-Method Study of Social Networks of Caregivers of Patients with Severe Mental Disorders' (2023) 69 International Journal of Social Psychiatry 190.

[54] Cristina Miguel and others, '"Alone on the Road": Loneliness among Digital Nomads and the Use of Social Media to Foster Personal Relationships' [2024] Media, Culture & Society 01634437241290087; Debbie Laliberte Rudman and Rebecca M Aldrich, 'Social Isolation, Third Places, and Precarious Employment Circumstances: A Scoping Review'; Zheng Yao and others, 'Together but Alone: Atomization and Peer Support among Gig Workers' (2021) 5 Proceedings of the ACM on Human-Computer Interaction 1.

[55] Sophie Pychlau, 'Understanding Belongingness in the Gig Economy: The Uplifting and Undermining Effects of Online Communities on Lonely Gig Workers' (PhD Thesis, University of Oregon 2023).

[56] Quintin Gaus, Anna Jolliff and Megan A Moreno, 'A Content Analysis of YouTube Depression Personal Account Videos and Their Comments' (2021) 3 Computers in human behavior reports 100050; Emily Scheinfeld, Chinasa Elue and Gary B Wilcox, 'Grieving Online: A Text Mining Analysis of Narratives of Grief on Reddit' (2024) 13 The Journal of Social Media in Society 75.

[57] Stephanie Madden and Rebecca A Alt, 'Know Her Name: Open Dialogue on Social Media as a Form of Innovative Justice' (2021) 7 Social Media+ Society 2056305120984447; Jason B Whiting and others, 'Trauma, Social Media, And# WhyIDidntReport: An Analysis of Twitter Posts about Reluctance to Report Sexual Assault' (2021) 47 Journal of marital and family therapy 749.

social (marginalisation, alienation, lack of connection or belonging in offline spaces). Social media, in this context, serves as a mechanism, whether consciously or unconsciously, through which they attempt to regulate this dysregulation. Some engage compulsively to manage overwhelming affect (e.g., anxious scrolling to suppress distress), others to reinforce a sense of self (e.g., validation-seeking to counteract self-doubt), and others to stabilise a sense of connection (e.g., using digital spaces as a substitute for lacking offline social support). This reliance on social media as a regulatory tool makes disengagement particularly difficult, because even when engagement is distressing, it remains functionally necessary for coping with instability.

Secondly, every group identified lacks access to offline structures or relationships that could sufficiently meet their psychological, emotional, or social needs. Whether due to geography (rural isolation), identity (marginalisation of LGBTQ+ individuals, racial minorities, or disabled individuals), life stage (teenagers, new parents, university students), or economic precarity (gig workers, those in unstable employment), their offline environments are insufficient in providing the validation, connection, or support they need. This insufficiency, accordingly, makes social media not just a tool but a necessity. If offline alternatives were more readily available such as accessible mental health services, community spaces, stable employment, or inclusive social environments, reliance on digital engagement would likely decrease. However, in the absence of these, social media fills an important gap, creating cycles of engagement that can become difficult to moderate, even when distressing.

Thirdly, what is also clear is that certain individuals are far more susceptible to reinforcement mechanisms due to pre-existing psychological, neurological, or social factors. For example, individuals with ADHD or anxiety disorders are particularly sensitive to dopamine driven reward systems, meaning that intermittent reinforcement such as unpredictable notifications, social feedback loops, or algorithmic engagement hooks can be particularly compelling. Similarly, those experiencing identity instability, marginalisation, or loneliness are more likely to seek out social validation, making them disproportionately responsive to social approval mechanisms (likes, shares, comments). Relatedly, individuals in high stress environments (such as caregivers or precarious workers) may turn to social media for brief but engaging forms of cognitive escape, meaning their engagement is shaped by immediate relief rather than long-term intentional use. And, thus, because their psychological needs and vulnerabilities fit with the mechanics of platform design, they are more likely to engage in cycles of reinforcement that may not affect other users in the same way.

Fourthly, a common thread across all these groups is that social media is experienced with ambivalence - as both a source of relief and a source of distress, as both empowering and entrapping, and as both deeply necessary and deeply exhausting. For individuals with anxiety or OCD, social media provides reassurance but also fuels compulsivity. For those who feel socially isolated, it creates connection but also highlights exclusion. For those who use it to escape, it offers relief but also reinforces

dependence. This paradox makes disengagement particularly difficult because logging off does not simply mean quitting a habit but relinquishing a tool that has provided comfort, validation, or meaning. The more an individual recognises the contradictions in their engagement, the more frustrating and difficult their relationship with social media becomes leading to cycles of intermittent disengagement, guilt, and re-engagement.

Fifthly, a unifying factor is that social media functions as a compensatory strategy - a means through which individuals adapt to and cope with psychological, emotional, and structural deficiencies in ways that can, over time, become maladaptive. In this sense, problematic engagement is not irrational but entirely functional given the context in which it occurs. If someone has no alternative means of support, validation, regulation, or connection, their reliance on digital engagement is not simply a byproduct of external platform conditioning but an *adaptive survival strategy.* However, because these digital interactions often do not fully resolve the underlying issues (e.g., they do not eliminate anxiety, provide real stability, or replace tangible human connection), they become self-perpetuating in ways that keep individuals caught in an engagement cycle where the very tool they rely on can also deepen their distress.

Sixthly, a further, well established risk factor that often intensifies this cycle is age. Children and adolescents are in a phase of rapid neurodevelopment in which the systems that detect and value reward mature earlier than the systems that regulate behaviour.[58] During adolescence, subcortical circuits (striatum, amygdala) show heightened dopaminergic reactivity to novelty, social approval and intermittent rewards, while the prefrontal cortex, responsible for planning, inhibitory control, delay of gratification and metacognition, continues to myelinate and integrate into the mid-twenties.[59] This "maturational gap" increases susceptibility to cues that promise variable pay-offs (likes, streaks, algorithmic boosts) and reduces the capacity to disengage once arousal is triggered. At the same time, adolescents experience intensified salience of peer evaluation and social status, amplified stress reactivity, and identity work that makes feedback loops, both positive and negative, especially powerful.[60] Moreover, circadian shifts typical of puberty (that is, the puberty-related delay in the biological sleep–wake cycle that leads to later sleep onset and later wake times, i.e., a move toward an "evening" chronotype) compound the problem: evening chronotype and sleep restriction lower executive function and emotional regulation,

[58] Michaela Åström, 'The Adolescent Brain on Social-Media: A Systematic Review'.
[59] ibid; Dustin Albert, Jason Chein and Laurence Steinberg, 'The Teenage Brain: Peer Influences on Adolescent Decision Making' (2013) 22 Current directions in psychological science 114; Sarah-Jayne Blakemore, 'Brain Development during Adolescence', *Educational Theories, Cultures and Learning* (Routledge 2012); M Lamblin and others, 'Social Connectedness, Mental Health and the Adolescent Brain' (2017) 80 Neuroscience & Biobehavioral Reviews 57.
[60] Åström (n 58); Michelle Achterberg and others, 'Longitudinal Associations between Social Media Use, Mental Well-Being and Structural Brain Development across Adolescence' (2022) 54 Developmental Cognitive Neuroscience 101088.

increasing compulsive checking at night.[61] These developmental features do not act in isolation: when combined with the psychosocial factors described above (i.e., anxiety, isolation, family or school stressors) the probability that design cues will precipitate heavy, dysregulated use rises.

8. Sparks and Tinder: An INUS Account of Compulsive Social Media Use

Thus, does that mean social media "causes" addiction? Not in a simple, monocausal sense. Social media is better understood as a *risk multiplier* and *accelerant* within a broader causal ecology.[62] It is present everywhere because it is ubiquitous and unusually efficient at providing mood regulation, validation and connection; that ubiquity means it frequently participates in the pathway to dysregulated use, but it does not, by itself, generate addiction. For most people, heavy use remains just that.[63] Addiction-like patterns tend to emerge when platform design (notifications, feeds, streaks) meets pre-existing vulnerabilities (anxiety, low mood, ADHD, identity stress, loneliness), adverse contexts (thin offline support, precarity, disrupted sleep), and developmental factors (the adolescent maturational gap).[64] In those conditions, the same cues that are trivial for a well-supported peer can become strong regulators of affect and status, making disengagement hard. That is interaction, not simple causation.

Two further points guard against overclaiming: first, as explained, bi-directional or reverse causality is common (e.g., distress drives people to seek relief online);[65] second, confounding factors are many, in that contexts that elevate distress can also, under certain conditions, raise screen time (for example, bereavement or loneliness increasing the need for connection, exam pressure or precarious work disrupting routines, injury or illness limiting offline activity, or pandemic-style restrictions reducing alternative outlets).[66] So, social media is rarely a necessary or sufficient cause of "addiction." It is typically a *risk multiplier* or what philosopher John Leslie Mackie would call an *INUS factor* (i.e., an **I**nsufficient but **N**on-redundant part of an **U**nnecessary but **S**ufficient condition; insufficient on its own, but non-redundant

[61] Meyran Boniel-Nissim and others, 'Adolescent Use of Social Media and Associations with Sleep Patterns across 18 European and North American Countries' (2023) 9 Sleep Health 314; Kaitlyn Burnell and others, 'Associations between Adolescents' Daily Digital Technology Use and Sleep' (2022) 70 Journal of Adolescent Health 450.

[62] Fardouly (n 32); C Anderson and others, 'Bidirectional Relationships between Muscularity-Oriented Disordered Eating and Mental Health Constructs: A Prospective Study' [2024] Eating Disorders; C Fitzpatrick, 'Adolescent Internet Use and Symptoms of Depression: A Rejoinder' [2024] Psychological Medicine; Anderson and others.

[63] Fardouly (n 32); Anderson and others (n 62); Fitzpatrick (n 62); Anderson and others (n 62).

[64] Fardouly (n 32); Anderson and others (n 62); Fitzpatrick (n 62); Anderson and others (n 62).

[65] Anderson and others (n 62).

[66] Francesca C Ryding and Daria J Kuss, 'The Use of Social Networking Sites, Body Image Dissatisfaction, and Body Dysmorphic Disorder: A Systematic Review of Psychological Research' (2020) 9 Psychology of Popular Media 412.

within the causal set, and one component of a set that, taken together, is sufficient for harm).[67]

This means two things in practice. First, the same design cue (autoplay, streaks, notifications) will not, on its own, produce addiction; it becomes harmful mainly when it coincides with other conditions - pre-existing distress, sleep loss, thin offline support, adolescent reward sensitivity - that together make compulsive use likely. Second, within that *combination* of factors, the interface/feature cue is still doing real work: remove it (or blunt it) *while the other conditions persist* and we can lower the probability or intensity of harm; change the surrounding conditions (treat anxiety, restore sleep, strengthen social ties, reduce precarity) and the cue can lose much of its pull.[68] So social media features should be treated as *accelerants* within a combustible setting: necessary to many "fires" *as configured*, but neither a universal match nor a self-igniting cause, and, therefore, an incomplete target for policy unless the co-factors are addressed.

Moreover, while social media is rarely a standalone cause, the *medium matters*. When vulnerabilities are enacted *through* social media, the syndrome takes on properties that are *co-produced* by the medium's affordances:[69] its 24/7 portability and reach; algorithmic, just-in-time curation; public, quantified feedback (likes, views, streaks); persistent notifications; and friction-light interfaces (infinite scroll, autoplay). These features shape the *phenomenology* of compulsive use - its tempo (rapid, continuous), cadence (intermittent rewards), triggers (anticipatory alerts), and social meaning (visibility, comparison, surveillance). In that sense, addiction-like patterns are not merely pre-existing distress "happening online"; they are *reconfigured* by the platform: episodes extend into circadian lows, relapse risk is heightened by ambient cues, and identity/status dynamics are amplified by metrics and audiences. So, even if social media is neither necessary nor sufficient, the *form* that dysregulated use takes is *distinctively digital*-an outcome of interaction between person and platform, where the stage does not just host the behaviour; it writes part of the script.

---

[67] John Leslie Mackie, *The Cement of the Universe: A Study of Causation* (Clarendon Press 1974).
[68] Bickham (n 1); Alexander and Delariarte (n 1).
[69] Maurell Audrey, Monty Satyadarma and Untung Subroto, 'The Correlation between Self-Esteem and Body Image: A Study on Female Adolescent Instagram Users', *The 2nd Tarumanagara International Conference on the Applications of Social Sciences and Humanities (TICASH 2020)* (Atlantis Press 2020); Dian A De Vries and others, 'Social Comparison as the Thief of Joy: Emotional Consequences of Viewing Strangers' Instagram Posts' (2018) 21 Media psychology 222; Chang Liu and Jianling Ma, 'Social Media Addiction and Burnout: The Mediating Roles of Envy and Social Media Use Anxiety' (2020) 39 Current Psychology 1883.

9. A Setting Can Disable Autoplay - But It Can't Disable Lo DSA's Fundamental Misattribution

So, at a foundational level, the DSA fundamentally misunderstands the nature of compulsive social media engagement because it assumes that the primary mechanism driving prolonged, distress-inducing use is external platform design rather than the underlying psychosocial and emotional needs that make digital spaces functionally irreplaceable for many users. To be clear, acknowledging this is *not* the same as claiming interface design and engagement features play no role. They plainly can and do lengthen sessions, raise cue exposure, and nudge checking frequency. Infinite scroll, autoplay, variable notifications, and streaks are good at *extending* attention and lowering the friction to re-enter a feed. That is why I broadly support the European Parliament's call for a default "right not to be disturbed":[70] keeping attention seeking features off by default is a low cost, proportionate baseline with little downside. The point, rather, is *specificity*: when the policy target is *compulsive, addictive*, or clinically problematic use, *not simply heavy use*, design governance *alone* is unlikely to be sufficient.

Instead, if the aim is to cut addictive use, the decisive interventions should be upstream and combine design with context. Specifically, this means prioritising children, adolescents and young adults within a stepped care system.[71] In education, this includes deploying validated mental health screeners with clear referral pathways; embedding counsellors and child–adolescent clinicians on site across secondary schools, further education colleges and universities with formal links to Child and Adolescent Mental Health Services (CAMHS), Adult Mental Health Services (AMHS), and clearly defined crisis pathways (including 24/7 crisis lines, crisis resolution/home treatment teams, and A&E liaison psychiatry);[72] and ensuring rapid access to evidence-based care for prevalent risk profiles (anxiety, low mood, OCD-spectrum reassurance checking, ADHD, and sleep/circadian problems).[73] Digital literacy should also broaden into 'algorithmic and affect literacy', covering how recommenders optimise for predicted engagement; why late night cue exposure has greater impact; how reassurance loops and rumination are reinforced; and practical

[70] European Parliament (n 10).
[71] AL Sanchez and others, 'The Effectiveness of School-Based Mental Health Services for Elementary School-Aged Children: A Meta-Analysis' (2018) 57 Journal of the American Academy of Child & Adolescent Psychiatry 153.
[72] E McCauley, MS Berk and JR Asarnow, 'Efficacy of Dialectical Behavior Therapy for Adolescents at High Risk for Suicide: A Randomized Clinical Trial' (2018) 75 JAMA Psychiatry 777; A Wickersham and others, 'Computerised Cognitive Behavioural Therapy for the Treatment of Depression in Adolescents: A Systematic Review and Meta-Analysis' (2022) 9 JMIR Mental Health e29313.
[73] KE Minges and NS Redeker, 'Delayed School Start Times and Adolescent Sleep: A Systematic Review of the Experimental Evidence' (2016) 28 Sleep Medicine Reviews 86; YC Liu, 'Delayed School Start Time Is Associated with Better Sleep, Mental Health, and Life Satisfaction' [2024] Sleep.

strategies to reduce harm under real-world pressures.[74] This prioritisation does not exclude others: adults facing isolation, bereavement, chronic stress, disability or precarious work also need accessible support through primary care screening, community mental health provision, social-prescribing routes, bereavement services and anti-loneliness programmes.[75]

10. Conclusion

That being so, the central claim of this article is diagnostic: regulatory discourse on 'platform addiction' is misattributed when it relies on a stimulus–response vocabulary. By construing compulsive use through a Skinner/Pavlov lens stimuli that "cause" or "stimulate" behaviour, the DSA reproduces a reductionist model in which humans are cast as organisms entrained by exogenous cues - the analogous laboratory subject, the lever-pressing pigeon in a Skinner box or the bell-salivating dog in Pavlov's lab, whose behaviour is presumed fully explicable in terms of conditioned reflexes and variable-ratio reinforcement. Such framing privileges the micro-variables of interface design (autoplay, infinite scroll, variable rewards) while backgrounding intra-personal and psychosocial determinants (motivational states, affect-regulatory aims, self-esteem dynamics, identity work), alongside the meso- and macro-level factors that social and clinical psychology repeatedly identify as primary risk vectors for platform/social media addiction or problematic use: affective dysregulation and low mood; attentional and reward-processing profiles; status anxiety and social comparison; thin local networks and role strain; identity work under conditions of surveillance and precarity; and adolescent neurodevelopment intersecting with sleep-phase delay and heightened peer salience.

Thus, the goal of this paper is to move beyond a reductive stimulus–response account and bring the full complexity of intrapersonal, social, and structural factors into view. As such, if the aim is lasting, material reductions in compulsive use, policy should combine proportionate design defaults with upstream mental-health provisions and social measures that cut the demand for digital self-medication. Accordingly, the paper asks regulators to retire the Skinner/Pavlov imaginary of trained pigeons and lever-pressing rats and to adopt a more grounded frame that matches how harm arises if the current evidence is accepted. That is, we should think less lab rats in Skinner boxes and more lived ecologies: a 15-year-old awake at 1.30 a.m. in a small bedroom, a new mother scrolling through postpartum worry, a shift worker on zero-hours contracts chasing brief relief between jobs, a student with ADHD seeking

[74] CS Gordon and others, 'Outcomes of a Cluster Randomized Controlled Trial of the "SoMe" Social Media Literacy Program for Improving Body Image-Related Outcomes in Adolescent Boys and Girls' (2021) 13 Nutrients 3637.

[75] Meynadier and others (n 35); Sri Wahyuning Astuti, 'Social Media Effect on Loneliness, Rumination and Social Comparison' (2022) 19 Webology 6430; Önder Baltaci, 'The Predictive Relationships between the Social Media Addiction and Social Anxiety, Loneliness, and Happiness.' (2019) 15 International Journal of Progressive Education 73; B Dibb and M Foster, 'Loneliness and Facebook Use: The Role of Social Comparison and Rumination' (2021) 7 Heliyon.

stimulation, a queer teenager in a hostile town looking for community. I contend that policy which addresses these settings -school, home, clinic, workplace, community - alongside interface design is more likely to reduce compulsive use than toggles alone.