

Reducing echo chamber effects: an allostatic regulator for recommendation algorithms

N. I. N. Henry, M. Pedersen, M. Williams, J. L. B. Martin & L. Donkin

To cite this article: N. I. N. Henry, M. Pedersen, M. Williams, J. L. B. Martin & L. Donkin (2025) Reducing echo chamber effects: an allostatic regulator for recommendation algorithms, Journal of Psychology and AI, 1:1, 2517191, DOI: [10.1080/29974100.2025.2517191](https://doi.org/10.1080/29974100.2025.2517191)

To link to this article: <https://doi.org/10.1080/29974100.2025.2517191>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 01 Jul 2025.



[Submit your article to this journal](#)



Article views: 517



[View related articles](#)



[View Crossmark data](#)

RESEARCH ARTICLE



Reducing echo chamber effects: an allostatic regulator for recommendation algorithms

N. I. N. Henry ^a, M. Pedersen ^b, M. Williams ^c, J. L. B. Martin ^d
and L. Donkin ^b

^aSchool of Psychological Science, University of Western Australia, Crawley, Perth, Australia;

^bDepartment of Psychology and Neuroscience, Auckland University of Technology, Auckland, New Zealand; ^cSchool of Psychology, Massey University, Albany, Auckland, New Zealand;

^dDepartment of Physics, School of Physical and Chemical Sciences, Christchurch, New Zealand

ABSTRACT

Objective: Recommendation systems are prevalent on the Internet but are prone to feedback loops that cause “echo chamber” effects. These effects can have negative consequences for users’ well-being, diversity of information, and social cohesion. Therefore, there is a need for novel techniques to combat echo chamber effects and promote healthier online experiences.

Method: We present an allostatic regulator for recommendation systems based on opponent process theory. This regulator can be applied to the output layer of any existing recommendation algorithm to dynamically restrict the proportion of potentially harmful or polarised content recommended to users, based on the users’ recent content history. We implement our prototype algorithm as a code wrapper for a supervised K-Nearest Neighbors algorithm for movie recommendations and evaluate its performance using simulated user data.

Results: Our results show that allostatic regulation is effective at reducing echo chamber effects in a simulated population. The method can be used for regulating the entire range of possible online content and can adapt to evolving user behaviours.

Conclusions: The allostatic regulator is a promising technique for mitigating echo chamber effects, providing app developers with a flexible tool to help users self-regulate their online experiences.

ARTICLE HISTORY Received 29 November 2024; Accepted 28 May 2025

KEYWORDS Allostasis; opponent process; recommendation system; echo chamber; machine learning; social media

CONTACT N. I. N. Henry  nathan.henry@uwa.edu.au

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/29974100.2025.2517191>.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1. Introduction

Recommendation systems, or recommenders, are personalised algorithms that tailor digital content to a user's preferences. They often employ machine learning algorithms to enhance accuracy and improve user engagement (Portugal et al., 2018), by recording content a user has previously liked in a database and using that information to predict how much the user will like new content. This helps rank and recommend new items to users (Milano et al., 2020).

Recommenders are widely used in social media, e-commerce, video streaming services, news aggregators and gaming platforms (Cinelli et al., 2021; Feng et al., 2020; Ge et al., 2020; Hasan et al., 2018; Jiang et al., 2019; Vu & Bezemer, 2021). However, as recommenders become more accurate, there are growing concerns over their potential effects on users' behaviours, preferences and emotions (Christiano, 2022; Reisach, 2021). These concerns include issues such as doom-scrolling (Mannell & Meese, 2022), heightened rates of anxiety and depression among teenage users (particularly females) (Haidt & Allen, 2020), increased exposure to pornographic content (Lewczuk et al., 2022), and shortened attention spans (Lorenz-Spreen et al., 2019). Recommenders have also faced criticism for causing social divisions, promoting political biases and extremism, as well as spreading harmful content like hate speech and misinformation (Christiano, 2022; Persily & Tucker, 2020; Reisach, 2021; Saurwein & Spencer-Smith, 2021). These systems are prone to feedback loops that give rise to "echo chamber" or "filter bubble" effects, where a user's beliefs and preferences are positively or negatively reinforced by selective, repeated exposure to a limited subset of content (Cinelli et al., 2021; Jiang et al., 2019; Nguyen et al., 2014). This promotes confirmation bias – the tendency to seek and believe information aligning with one's existing opinions (Cinelli et al., 2021).

There have been widespread calls to regulate these algorithms, but the degree to which regulation should be enforced is under debate. Several methods for detecting and reducing polarisation in social networks have been proposed in the literature, such as inserting or deleting edges or nodes, modifying edge weights, adjusting network design, and modifying embedded recommendation systems (Interian et al., 2023). More broadly, several research pathways to mitigating echo chamber formation are being studied (Hartmann et al., 2025; Terren & Borge-Bravo, 2021), including increasing awareness of echo chambers through gamification (Jeon et al., 2021), providing visible metrics indicating levels of bias and communication between networks (Cota et al., 2019; Kalluçi & Peshkopia, 2024a, 2024b; Liao & Fu, 2014), and using random dynamical nudges to introduce more diverse content into one's feed (Currin et al., 2022). However, one method in particular stands out: that of moderating the amplification of singular

opinions or content. For example, Lim and Bentley (2022) found that amplifying opinions on social media, even slightly, can drive extreme polarisation on social media, but this can be mitigated by placing limits on amplification or by promoting balanced viewpoints.

This principle of moderation may align with psychological well-being in general. The “digital Goldilocks hypothesis” has proposed that moderate technology use benefits mental well-being by improving social connectivity, identity and skill development through platforms such as online gaming and social media (Etchells et al., 2016; Przybylski, 2014; Przybylski & Weinstein, 2017). However, too much technology use might replace meaningful activities. This is an example of a hormetic relationship, where low doses of technology use are beneficial, but high doses are harmful. Przybylski & Weinstein explicitly tested this theory in a preregistered study of a large sample of English adolescents ($n = 120,115$) and found that the relationship between screen time and mental well-being was best described as a hormetic curve (Przybylski & Weinstein, 2017) – a result that has since been replicated by Brannigan et al. (2023). Hence, digital content consumed within the hormetic limit may have a beneficial effect on mental well-being, while exceeding this limit can negatively impact mental well-being.

Solomon & Corbit’s opponent process theory posits that in humans, a stimulus or behaviour can trigger an “a-process”, the initial euphoric response to a behaviour or substance, followed by a compensatory “b-process” marked by negative feelings such as depressed mood, cravings and withdrawal (Solomon & Corbit, 1974). Once the b-process subsides, homeostasis is regained. The pleasurable a-process is a motivating force (Solomon & Corbit, 1974), generally accompanied by release of dopamine or other neurochemicals (George et al., 2012). However, continued activation of the b-process can lead to hedonic allostasis – a state in which an individual’s hedonic set point undergoes a persistent downward shift from homeostatic levels due to neural adaptations to repeated external stimuli. Thus, the person returns to homeostasis only when the b-process is no longer activated (Ahmed & Koob, 2005; Solomon & Corbit, 1974).

Henry et al. (2023) have proposed that “digital hormesis” may be detected by performing a Behavioral Frequency Response Analysis (BFRA), in which the frequency of a behaviour is varied, and the affective response of the system is measured as a function of behavioural frequency. Such an approach can be used to find the hormetic limit of the system, in terms of the maximum healthy frequency of content consumption. This raises the question of whether allostatic regulation could be applied *in the reverse direction* to ensure that recommenders deliver digital content in a way that prevents users from exceeding healthy hormetic limits for content consumption. In the case of a recommender, this could be achieved by representing digital content as a pleasurable a-process that leads to a subsequent b-process

consisting of low levels of craving, withdrawal, and depressed mood. One could model the allostatic load that builds up from consuming harmful digital content too frequently and use it as a punishment function that reduces the recommender's tendency to suggest such content.

This method could provide a personalised, dynamic solution to reduce exposure to harmful content on many online platforms. For example, it could be used to dynamically reduce extreme content recommendations on video-streaming sites, or encourage content diversity by showing educational articles on news sites when overconsumption of entertaining or political content is detected. It could also be used to reduce the negative impacts of social media on mental health – for example, by dynamically reducing the frequency of negative social comparisons (i.e. limiting exposure to highly curated, unrealistic portrayals of life) (Najar & Shabir, 2023, Esther Ajewumi et al., 2024), or by regulating notification frequency to curb compulsive phone-checking behaviours. It could even be applied to online gambling and gaming platforms, by introducing adaptive spending limits in response to excessive in-game purchases.

The following sections describe our mathematical framework for performing allostatic regulation. First, we show that we can regulate the entire possible range of digital content, from harmless to illegal, by varying the parameters of a simplified allostatic model containing only b-processes. We then apply this regulator to a simulated movie recommender and quantify the effects of varying different allostatic parameters. Finally, we demonstrate how more complex regulation can be achieved in certain contexts by adding the a-process to the model. Overall, our findings demonstrate the counter-acting effect of allostatic regulation on positive feedback loops in recommender systems, suggesting its potential to mitigate echo chamber effects.

2. Methods

2.1. Allostatic regulation model

Our allostatic regulator was mathematically defined using a behavioural dose-response model based on the concept of “intervention dose”, in which dose-response effects scale upward with increasing exposure to an intervention or behaviour (Heerman et al., 2017; Henry et al., 2023; Kirk et al., 2019; Manojlovich & Sidani, 2008; McVay et al., 2019; Voils et al., 2012). This model characterises a human behaviour through four main attributes: *Potency* (level of engagement in a behaviour, with arbitrary units), *Amount* (time spent on the behaviour, in seconds), *Frequency* (how often the behaviour occurs in a given time, in Hz), and *Duration* (the observation period, in seconds). Then, the dose for a single behaviour can be calculated as:

$$Dose_{individual\ behavior} = Potency \times Amount \quad (1)$$

and the total dose of cumulative behaviours performed within a specific period is:

$$Dose_{total} = Frequency \times Duration \times \overline{Dose_{individual\ behaviors}} \quad (2)$$

where $\overline{Dose_{individual\ behaviors}}$ represents the mean individual behavioural dose for all behaviours over the *Duration* in which $Dose_{total}$ is assessed. To simplify our analysis, we performed a Behavioral Frequency Response Analysis (BFRA) by treating *Potency*, *Amount* and *Duration* as constants and varying the *Frequency* at which the dose is delivered (Henry et al., 2023).

We assumed that the subject's response to the behavioural dose took the form of opponent processes with exponential decay, a widely used framework in neuroscience and behavioural science (e.g., see Koob & Le Moal, 2008; Pirih et al., 2022; Solomon & Corbit, 1974; Uribe-Bahamonde et al., 2019). In this model, the subject's hedonic response to the behavioural dose consisted of exponentially decaying a- and b-processes with decay constants λ_a and λ_b , with units seconds^{-1} , and the *Potency* of the behaviour was modelled as the initial intensity of the process, in hedons. We chose to represent the a- and b-processes as Hawkes processes with exponential decay kinetics, obeying the laws of superposition. This parallels the strategy employed by researchers studying the simulation of social network dynamics, in which user activity is represented using stochastic point process models, often with an exponential decay component (Dai et al., 2017; Du et al., 2015; Farajtabar et al., 2018; Pan et al., 2016; Tan et al., 2016; Trivedi et al., 2017; Wang et al., 2016, 2017, 2018).

Given a behaviour (in this case, watching a movie) performed at times $T = \{t_1, \dots, t_{|T|}\}$, where $|T|$ is the total number of repetitions of the behaviour, we can define the hedonic state $h(t)$ of the subject at time t by convolving $f(t)$, consisting of Kronecker delta functions denoting the times when the behaviour occurred and its potency, with $g(t)$, the sum of the a- and b-processes:

$$f(t) = P \sum_{i=1}^{|T|} \delta(t - t_i) \quad (3)$$

$$a(t) = I_a e^{-\lambda_a t} t \geq 0 \quad (4)$$

$$b(t) = -I_b e^{-\lambda_b t}, t \geq 0 \quad (5)$$

$$g(t) = a(t) + b(t) = I_a e^{-\lambda_a t} - I_b e^{-\lambda_b t} \quad (6)$$

$$h(t) = f(t) * g(t) \quad (7)$$

which can be condensed to the general formula:

$$h(t) = P \sum_{t_i \leq t}^{|T|} \left(I_a e^{-\lambda_a(t-t_i)} - I_b e^{-\lambda_b(t-t_i)} \right) \quad (8)$$

where t is time; P is the *Potency* of the behaviour (hedons); $\delta(t - t_i)$ is the delta function at time t_i (seconds); I_a and I_b are the initial intensities of the a- and b-processes (arb. units); and λ_a and λ_b are the decay constants of the a- and b-processes (seconds^{-1}). By setting $\text{Duration} = t_{sim}$, and by fixing *Amount* and *Frequency* as constants, we can observe the effect of modifying the values of $|T|$, *Potency*, λ_a and λ_b on the user's predicted hedonic state $h(t)$ over time.

2.2. Applying allostatic regulation to a movie recommender

To apply allostatic regulation in a real-world setting, we trained a movie recommender on the MovieLens-1 M dataset, containing 6,000 user ratings on 4,000 movies, and their genres (Harper & Konstan, 2016). We wrote a program in Python 3.8 (Van Rossum & Fred, 2009) to simulate the movie-watching behaviours of virtual subjects. In this hypothetical scenario, each subject employed allostatic regulation to reduce the proportion of movies recommended with the “Horror” genre, to lower the amount of anxiety-provoking material that they were exposed to (this was an arbitrary choice; as another example, the subject could have chosen to regulate “Romance” or “War” movies to reduce exposure to sexual or violent content). It was assumed that these subjects enjoyed watching Horror movies, but only infrequently, as too many Horror movies can lead to excessive anxiety and difficulty sleeping (Martin, 2019). Thus, the aim was to regulate the frequency of Horror movies watched to keep the user within their hormetic limit.¹ The code and figures for the simulation can be found in the Supplementary Materials.

The proportion of movies of a single genre watched by all virtual subjects during the simulation time t_{sim} can be described by the equation:

$$\Omega_j = \frac{\sum_{i=1}^n \sum_{j=1}^x m_{ij}}{\text{Total}} \quad (9)$$

where Ω_j is the proportion of movies of genre j watched by all virtual subjects over t_{sim} , n is the number of subjects, x is the total number of genres, m_{ij} is the number of movies of genre j watched by subject i during t_{sim} , and *Total* represents the total number of movies watched by all subjects over t_{sim} .

The initial movie watched was “From Dusk Till Dawn (1996)” which contained the “Action”, “Comedy”, “Crime”, “Horror” and “Thriller” genres. Each virtual subject watched one movie every 72 hours, randomly

selected from a list of the 20 top recommendations generated by a K-Nearest Neighbors (KNN) algorithm from the Scikit-learn package (Pedregosa et al., 2011). The KNN algorithm, a supervised machine learning method, assigned labels to new data points based on the labels of their closest neighbours in the training data. It identified these neighbours by calculating distances between data points and selecting the K nearest ones. The new data point was then assigned the label most common among its neighbours. In this scenario, the algorithm recommended movies similar to the most previously watched movie only, using genres and mean ratings as the features for training the model. Mean ratings were calculated for all movies with at least 10 ratings, with the remainder of movies excluded from the dataset. Previously watched movies were excluded from the latest recommendations. Although the recommender's optimisation was modest, it effectively served our objectives.

Allostatic regulation was applied by modifying the KNN algorithm using the "AdjustedKNN" class, which can be found in the second Supplementary Materials file. Specifically, the Euclidean distances for Horror movies were dynamically increased in proportion to the inverse of the user's hedonic score, $h(t)$. Hence, if the user watched multiple Horror movies in quick succession, the regulator's allostatic load would increase rapidly, causing the proportion of recommended Horror movies to decrease in response. For this model, a-processes were excluded, meaning that the value of $h(t)$ was a pure model of allostatic load for b-processes only. Parameter exploration involved testing different b-process parameters within plausible ranges ($I_b = \{0.08 - 0.16\}$, $\lambda_b = \{0.0035 - 0.0069s^{-1}\}$), with higher values of I_b and lower values of λ_b leading to stronger regulation.

All code and figures for our simulations can be found in the Supplementary Materials.

3. Results

We represented algorithmic behaviours performed by a person at times $T = \{t_1, \dots, t_{|T|}\}$ as delta functions, in the graph of $f(t)$. We then convolved $f(t)$ with pre-generated opponent processes $g(t)$ that represent the effect of the behaviour on the person's mental state over time. This is demonstrated in Figure 1. Hedonic allostasis can be observed in the graph of $h(t)$, representing the person's time-varying hedonic state as their set point decreases over time. Thus, the value of $h(t)$ can then be used to dynamically regulate a recommender's outputs over time, preventing excessive exposure to specific types of digital content based on the evolving user experience.

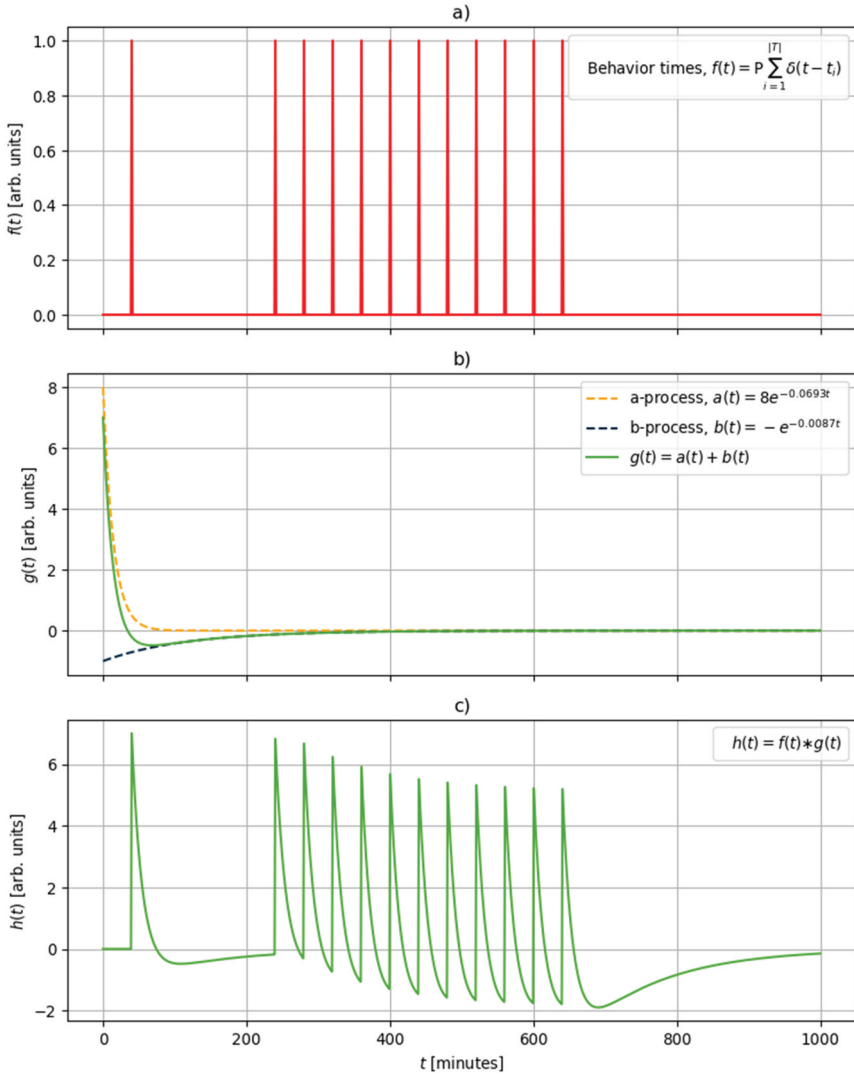


Figure 1. Demonstration of allostasis using the convolutional opponent process model. The delta function spikes of $f(t)$ represent the times when the behaviour occurred, while $g(t)$ represents the opponent process dynamics that are caused by the behaviour, and $h(t)$ represents the hedonic state of the subject, which can be calculated by convolving $f(t)$ with $g(t)$.

3.1. Regulating the entire possible range of content with b-processes

In opponent process theory, the b-process persists longer than the a-process. Therefore, when behaviours occur at time intervals shorter than the critical decay duration (CDD) of the b-process (defined here as the time it

takes for the b-process to decay to insignificance) but longer than the CDD of the a-process, then only the b-process contributes to allostasis. Under such conditions, it is unnecessary to model the a-process to simulate allostatic effects. Thus, to achieve basic regulation, we removed the a-process from the allostatic model and treated $h(t)$ as a scalar weighting factor that modified the recommendation probability of different types of content. Then, the opponent process model $g(t)$ simplifies to:

$$g(t) = b(t) = -I_b e^{-\lambda_b t}, t \geq 0 \quad (10)$$

And $h(t)$ more closely models the allostatic load generated by the b-processes (i.e. how far the hedonic state deviates from homeostatic levels, on average):

$$h(t) = P \sum_{t_i \leq t}^{|T|} \left(-I_b e^{-\lambda_b (t - t_i)} \right) \quad (11)$$

Figure 2 demonstrates how exponential b-processes can take a wide range of shapes which, when combined with an allostatic paradigm, allows us to regulate any type of content. One practical example is the regulation of online hate speech. This is a contentious issue due to the need to balance freedom of expression with prevention of harm and illegal activities (Banks, 2010). Here, we demonstrate how one may perform allostatic regulation of the entire spectrum of possible online discourse, ranging from positive to hate speech. Let us consider these two extremes.

First, we take the case of online hate speech, which in certain contexts and jurisdictions, is illegal (Alkiviadou, 2019). Given a behavioural dose represented as a delta function $\delta(t)$ delivered at $t = 0$ seconds with *Potency* = 1 hedons, and convolving it with a b-process $b(t) = -I_b e^{-\lambda_b t}$ with initial intensity I_b , time t in the interval $[0, \infty)$ and decay constant $\lambda_b = 0$ yields the person's hedonic state over time:

$$\begin{aligned} h(t) &= P\delta(t) * -I_b e^{-0t} \\ &= -I_b \theta(t) \end{aligned} \quad (12)$$

where $\theta(t)$ is the Heaviside step function:

$$\theta(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (13)$$

This results in a step function that does not decay. Hence, using $\lambda_b = 0$ seconds⁻¹ in combination with a large value of P (for example, $P = \infty$) creates a large step function that could be used to regulate – or even permanently block – harmful content.

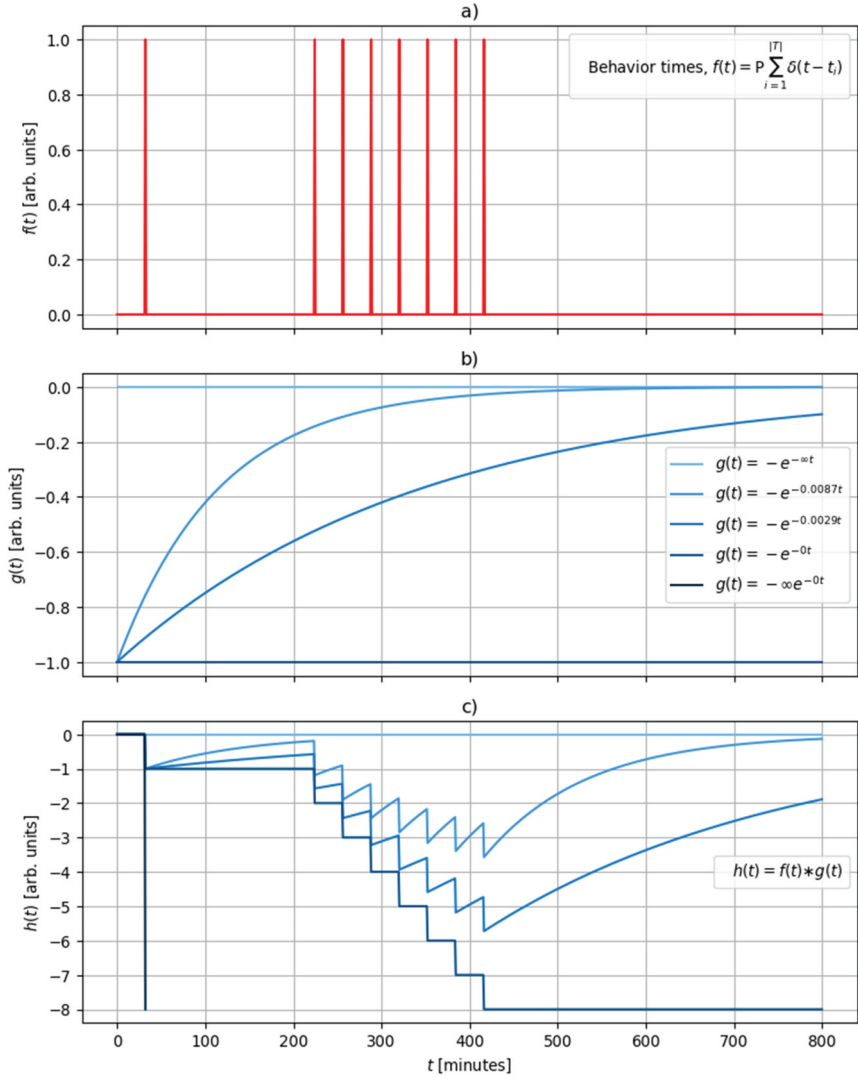


Figure 2. Demonstration of allostasis with different exponential kernels with λ_b ranging from ∞ to 0 minutes^{-1} , and I_b ranging from -1 to $-\infty$ (arb. units). The graph of $g(t)$ shows that varying λ_b between $\infty \text{ seconds}^{-1}$ (immediate decay to a hedonic state of 0) and 0 seconds^{-1} (no decay) can cover the entire range of possible content moderation.

On the other end of the spectrum, harmless content, regardless of frequency, can be given a decay constant $\lambda_b = \infty$. Then, convolving the delta function with $g(t)$ yields:

$$h(t) = P\delta(t) * -I_b e^{-\infty t} = 0 \quad (14)$$

Thus, it is possible to classify and regulate the entire range of content – from positive speech to hate speech – by varying λ_b and I_b between ∞ and 0. Figure 2 demonstrates this range of modulation. Setting $\lambda_b = \infty \text{ seconds}^{-1}$ has no regulatory effect on the algorithm's output as it causes $g(t)$ to immediately decay to zero. Such a function can be used for harmless positive content; meanwhile, setting $\lambda_b = 0 \text{ seconds}^{-1}$ generates a b-process that doesn't decay, but grows each time the behaviour is performed. Hence, it is the value of λ_b that is most important, as it defines the rate of allostasis.

The true utility of this method lies in adaptively regulating hormetic content. For instance, reading new articles online in small doses lets the reader stay informed and learn different perspectives. However, excessive news consumption can lead to addiction, doom-scrolling, and anxiety due to negative content (Anand et al., 2022; Shabahang et al., 2021). To counter this, one could use allostatic regulation to dynamically restrict their frequency of consuming news articles with negative content. Figure 2 shows how a b-process with decay constant $\lambda_b = 0.0029 \text{ seconds}^{-1}$ (a half-life of four hours) can reduce the number of negative articles recommended. If the person read multiple articles in a row (from $t = 240$ seconds onwards), then rapid allostasis occurs, reducing the number of negative article recommendations. However, if only a single negative article is read in isolation ($t = 40$ seconds), then the b-process is allowed to decay back to homeostatic levels. Users should be allowed to modulate λ_b to match their own hormetic limit of article consumption.

3.2. Simulating allostatic regulation in a movie recommender

Figure 3 shows the simulated proportion of movie genres watched by all subjects over time. Simulations with (Figure 3b–d) and without (Figure 3a) regulation are presented. Horror movies are presented in green in the centre of each Figure. Applying the regulator reduced the proportion of Horror movies recommended, and watched, during the simulation. Interestingly, even at high levels of regulation (Figure 3c), periods with no Horror movie recommendations were temporary. During these periods, the high allostatic load initially prevented any Horror movie recommendations, but then subsequently decayed to allow Horror movies to be recommended in moderation.

3.3. Quantifying the effects of different allostatic parameters

A comparison of (Figure 3c, d) shows the effect of doubling I_b had approximately the same regulatory effect as halving λ_b . This can be explained by calculating the integral of the b-process over the time interval $[0, \infty)$, which is proportional to the contribution of the b-process to allostasis:

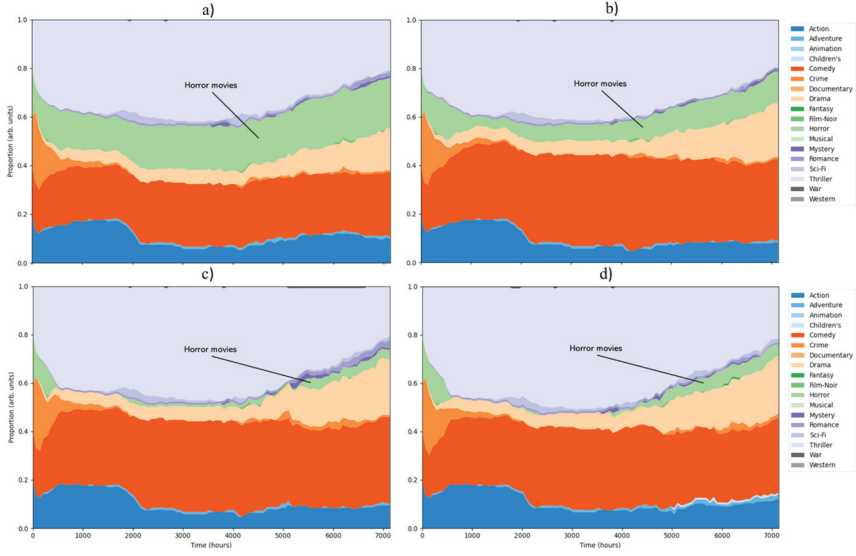


Figure 3. Simulations of 50 participants each watching 50 movies, with no allostatic regulation applied. The area of each colour represents the proportion of movies watched with that genre. Subplots used different simulation parameters, outlined here – a: no regulation applied; b: allostatic regulation with $I_b = 0.08$ and $\lambda_b = 0.0069$ hours⁻¹; c: $I_b = 0.16$ and $\lambda_b = 0.0069$ hours⁻¹; and d: $I_b = 0.08$ and $\lambda_b = 0.0035$ hours⁻¹.

$$\int_0^\infty -I_b e^{-\lambda_b t} dt = -\frac{I_b}{\lambda_b} \quad (15)$$

Therefore, I_b and λ_b are inversely proportional, in terms of their contribution to allostasis. Given that the half-life of the b-process, $t_{1/2b}$ (seconds) is related to the decay constant by the equation

$$t_{1/2b} = \ln(2)/\lambda_b \quad (16)$$

then we can also conclude that $t_{1/2b}$ is proportional to I_b ; hence adjusting either variable will have similar effects on Ω_{Horror} , the proportion of Horror movies watched. However, the rate of allostasis initially will be higher for the b-process with the lowest value of $t_{1/2b}$ (or the highest value of λ_b), as demonstrated in Figure 4: while both b-processes have the same integral and the same steady-state, the initial rate of allostasis is highest for the b-process with the highest value of λ_b .

To test the hypothesis that adjusting $t_{1/2b}$ would have similar effects to adjusting λ_b over an extended time, we conducted 150 simulations by varying I_b between 0 and 0.3, and 150 more simulations varying $t_{1/2b}$ between 0 and 300 hours, to determine the relationship between these variables and Ω_{Horror} . Figure 5 demonstrates the relationship between I_b ,

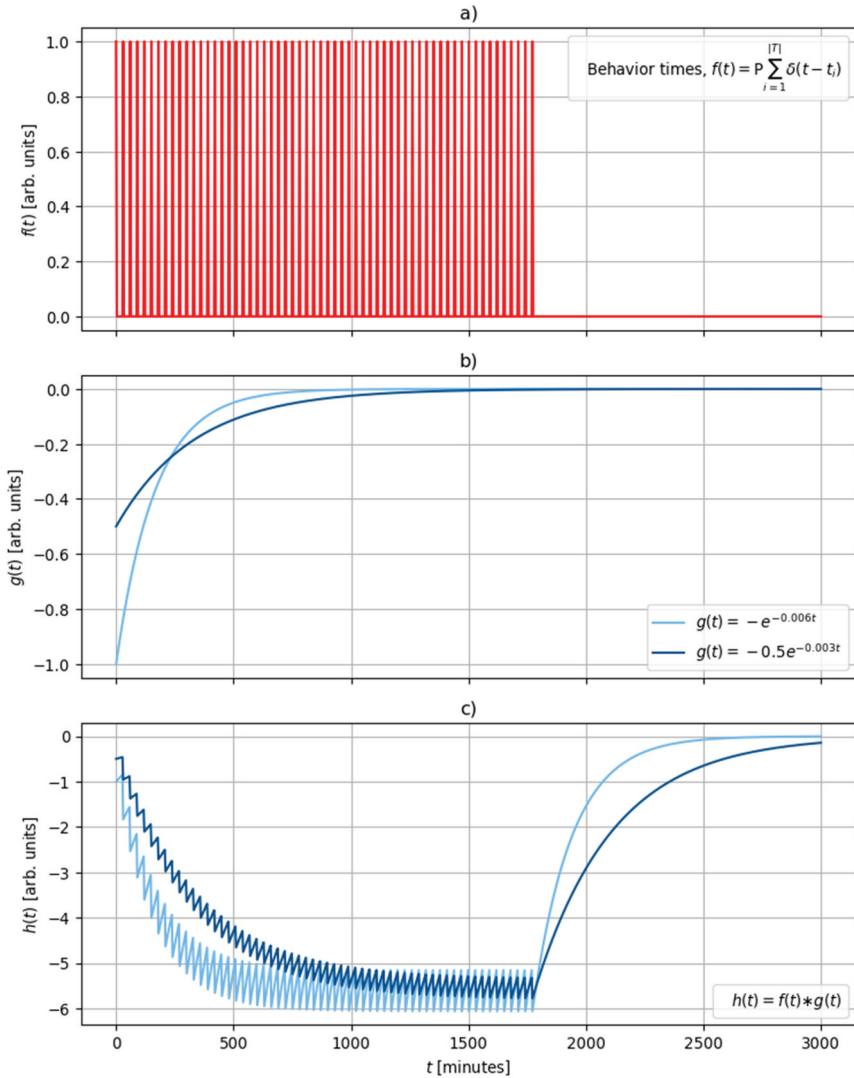


Figure 4. Rate of allostasis is higher for the b-process with the highest decay constant. However, both b-processes reach the same steady-state value during allostasis because they have the same integral value between 0 and ∞ .

$t_{1/2_b}$, and Ω_{Horror} , with trendlines derived via piecewise regression. To the right of the discontinuity in each Figure, the behaviours were performed at a time interval shorter than the CDD of the exponential, preventing the exponentials from decaying sufficiently before the next behaviour was performed. This led to increasing rates of allostasis, producing an exponential reduction in the percentage of recommended Horror movies as I_b or $t_{1/2_b}$ increased.

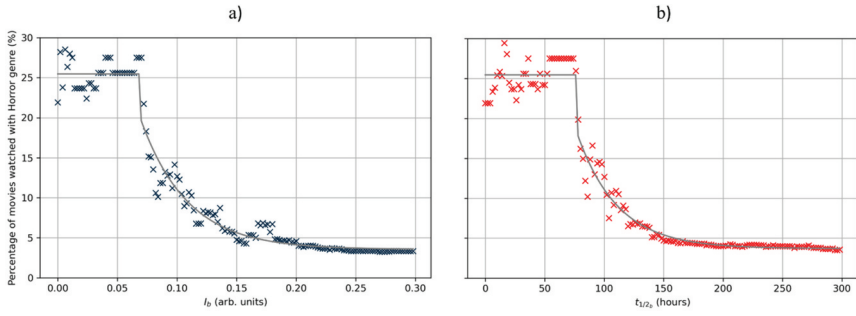


Figure 5. Relationship between I_b , $t_{1/2b}$, and percentage of Horror movies watched by all users. Half-life ($t_{1/2b}$) was set at 100 hours, while the initial intensity of the b-process (I_b) was set at 0.1. Lines were derived via piecewise regression.

However, an asymptotic limit can be seen in each Figure, because as frequency increased towards infinity, the rate of allostasis was balanced by the decay rate of newly formed b-processes. Hence, given an infinite simulation time and assuming $\lambda_b > 0$ seconds⁻¹ (i.e. $t_{1/2b} < \infty$ seconds), the percentage of Horror movies watched can never fully decrease to 0, as an equilibrium will always form between the rate of allostasis and the generation rate of new b-processes. This equilibrium represents a balance where the user is neither overexposed to nor completely deprived of this type of content.

3.4. Allowing behavioral bursts by incorporating a-processes

Our movie regulator was conservative and only included b-processes, as demonstrated by the dark blue curves in Figure 6. However, if the time interval between behaviours is shorter than the CDD of the a-process, positive allostasis (sensitisation) can occur before negative allostasis (habituation). This is demonstrated in the green opponent process curves in Figure 6, representing combined a- and b-processes. At a behavioural frequency of 0.1 minutes⁻¹ (Figure 6a), the CDD of the a-process is shorter than the time period between behaviours, and so any positive allostasis effects are negligible. However, at a behavioural frequency of 0.5 minutes⁻¹ (Figure 6b), the CDD of the a-process is longer than the behavioural period, leading to an initial rise in $h(t)$ for the combined opponent processes (transient sensitisation). Eventually, b-process allostasis exceeds that of the a-process, leading to negative allostasis.

Including a-processes in the allostatic model allows for more sophisticated regulation of different behavioural patterns, such as short behavioural bursts that may be beneficial in certain contexts. For instance, a person may wish to study a topic by watching several videos in quick succession. In this case, it may be helpful for the algorithm to suggest

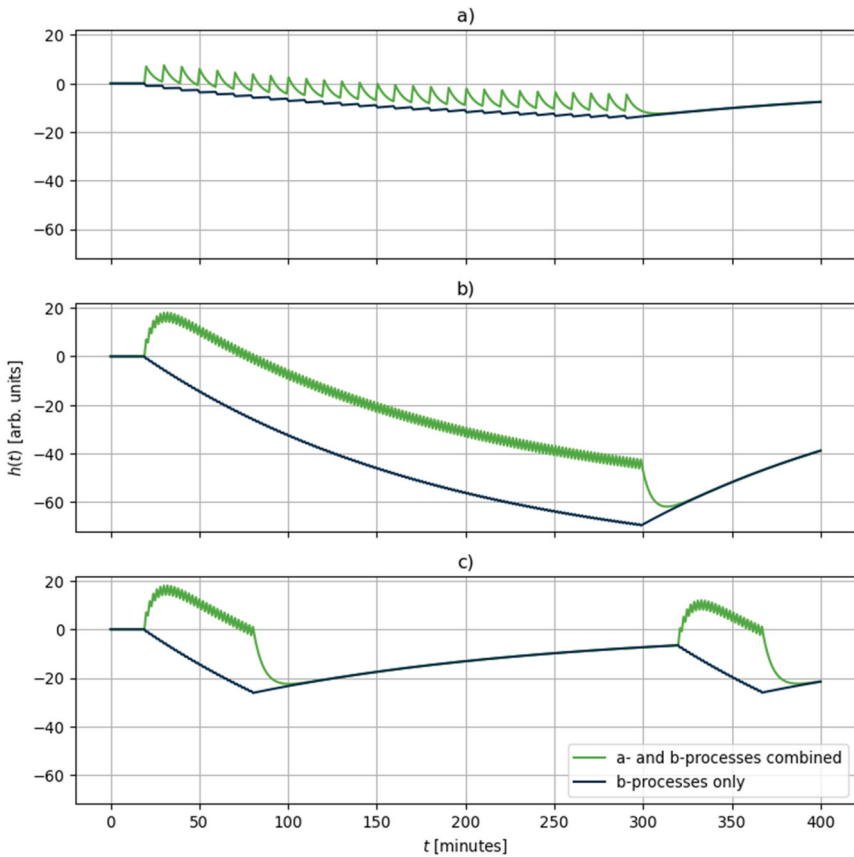


Figure 6. Simulating $h(t)$ for opponent processes (light green) and b-processes only (dark blue), at both low (plot a, 0.1 min^{-1}) and high behavioural frequencies (plots b-c, 0.5 min^{-1}). For all plots, $\lambda_a = 0.1733 \text{ minutes}$, $\lambda_b = 0.0058 \text{ minutes}^{-1}$, $I_a = 8$, and $I_b = -1$. At the higher frequency, the opponent processes demonstrate initial positive allostasis (a form of transient sensitisation) before negative allostasis occurs. Short behavioural bursts are demonstrated in plot c.

videos on the topic for an hour, before introducing some regulation to prompt a break from learning. A similar method could be applied to allow brief periods where news articles or social media posts can be accessed in moderation during work breaks. Figure 6c illustrates these “behavioral bursts”, simulating a person who consumes multiple pieces of digital content in short periods of time. In this case, regulation was applied to end the behaviour as soon as the a-process no longer exceeded the person’s baseline hedonic state. While this shows the versatility of allostatic regulation, more research is needed to verify its effectiveness in such scenarios.

4. Discussion

4.1. Summary of findings

In this article, we have introduced a prototype allostatic regulator which can be applied to recommendation systems as a code wrapper – an additional layer of code that adds functionality to the recommendation system without modifying the original code. This technique provides a dynamic, anticipatory solution for reducing echo chambers that form on social media, and can be tailored to the individual based on their preferences for regulation. While our results focused on the regulation of a movie recommender, allostatic regulation can in theory be used for regulating the outputs of any type of recommendation system using the principles of opponent process theory. For example, it could be used to limit exposure to extreme political content and conspiracy theories, reduce the frequency of sensationalised news articles, and prevent overexposure to addictive, pornographic, or violent material on social media platforms.

Empirical validation of this technique should involve controlled user studies, in which participants interact with recommendation systems with and without allostatic regulation, measuring engagement, content diversity, and psychological outcomes. Furthermore, A/B testing on live platforms could assess real-world effectiveness by analysing metrics such as watch time, click-through rates, and user retention. Economic analyses could also be conducted to determine whether implementing allostatic regulation aligns with the financial interests of social media companies, demonstrating feasibility and improving industry adoption.

4.2. Theoretical implications

Allostatic regulation may help to limit polarisation and echo chamber effects by moderating the extremes of content recommended to users. It achieves this in two ways. Firstly, the regulator dynamically adjusts the allostatic load based on the recency and frequency of behaviours with exponentially decaying b-processes. If the allostatic load is not allowed to decay back to homeostatic levels, then the regulator will naturally adjust the set point of the system, such that a new steady state is reached; this prevents positive feedback loops (Goldstein & McEwen, 2002). Secondly, as shown in [Figure 5](#), the regulator never eliminates Horror movies completely, but always allows some non-zero proportion of Horror movies to be recommended. The equilibrium observed between decaying b-processes and the rate of allostasis means that *content diversity is preserved without completely suppressing user preferences*. In terms of a solution for political content, an allostatic regulator can ensure that users are occasionally exposed to

differing viewpoints, even if most of the content aligns with their preferences. Our proposed approach may empower users and ensure they are not trapped in echo chambers for specific content categories. “Random nudges” could also expose the user to diverse views to reduce polarisation (Currin et al., 2022). However, this may not be sufficient; Noordeh et al. (2020) found that prolonged exposure to collaborative filtering recommendations on the MovieLens-1 M dataset reduced content diversity, leading to echo chambers that individual users could not escape by simply adjusting their own ratings. This suggests that further interventions may be required to prevent echo chamber formation.

To employ allostatic regulation, app developers must find the optimal level of regulation that produces beneficial effects while not being too restrictive (Brown, 2020). Users may reject the regulator if it is perceived as biased or overly censorious. It may be desirable to employ a user-centred approach to regulation, where app developers include simplified virtual dials that allow users to adjust their content diversity without needing to understand allostasis. For example, the developer could group potentially harmful content into different categories (such as pornographic, violent, or political content), and apply allostatic regulation to each category. Users could then adjust the dial for each category to fine-tune the amount of regulation applied to each type of content.

The allostatic model of hormesis is flexible enough to classify “grey area” content, which can be defined as content consumed in small doses or frequencies up to a hormetic limit, beyond which the content becomes harmful. An example of such content is conspiracy theories, which may have some value in stimulating critical thinking and exposing possible abuses of power. However, consuming many conspiracy theories may increase experiences of anxiety, uncertainty aversion, and existential threat (Liekfett et al., 2023). Further, people have different levels of receptivity to conspiracy theories, which their previous exposure may influence (Douglas et al., 2019). Each person may have a unique hormetic limit for conspiracy theory content, influenced by their environment, genetics, and upbringing. As such, individuals may require that their allostatic parameters be uniquely quantified to keep them within their hormetic limits.

To perform this quantification, longitudinal methods like Ecological Momentary Assessment (EMA) may prove valuable, in which participants’ emotional states are tracked over time and measured at regular intervals (Shiffman & Stone, 1998). The parameters for the a- and b-processes specific to individuals consuming certain types of content could be modelled from this EMA data once it becomes available. In the meantime, we can make some general assumptions about the shape of these opponent processes based on the current literature. For example, we can assume that stronger

stimuli lead to larger opponent processes, based on studies such as the one by Roberts and David (2023), who showed that smartphone users experience a greater flow state and higher degrees of subsequent depression when using the TikTok app compared to other mobile apps.

4.3. Limitations and future directions

One limitation of this study is that we did not explore the effects of allostatic regulation across multiple genres. Our focus on the Horror genre was arbitrary and served only as a demonstration; similar regulatory effects would likely emerge for any genre. However, future research should examine how allostatic regulation operates when applied to multiple genres simultaneously, with each genre having unique decay and potency parameters. This would better reflect real-world situations, where users can consume content with mixed genres in parallel and may need to regulate their use of multiple genres at once. Indeed, this will be one of the primary challenges in integrating such a system into real-world recommendation algorithms: finding ways to both accurately classify and regulate content with mixed genres such that users can have their content automatically regulated.

Yet even if these technical challenges are overcome, a fundamental dilemma remains: what incentive do major social media companies have to adopt such a system? Since allostatic regulation reduces the frequency of highly engaging but potentially harmful content, companies may argue that it risks decreasing user engagement and, consequently, advertising revenue. However, in the long run, platforms that integrate responsible content regulation may benefit financially by fostering healthier user behaviours, reducing burnout, and improving public perception. By positioning themselves as ethical alternatives to unregulated competitors, these platforms may attract users who prioritise digital well-being. If voluntary adoption proves unlikely, top-down regulatory intervention may become necessary, though government-imposed solutions risk being perceived as heavy-handed and could face resistance from both industry and users. A balanced approach – where platforms are encouraged to self-regulate while being held accountable for algorithmic harm – may be the most viable path forward.

Allostatic regulation can be seen as a form of adversarial training, in which the machine learning algorithm makes predictions, and the allostatic regulator provides an alternative objective function for the algorithm to optimise. We used a simple form of supervised learning – the KNN algorithm – to demonstrate that it is possible to adjust the model weights even after training. Importantly, this method is potentially scalable to more complex algorithms, including neural networks. This implementation may allow for regulation of more complex machine learning behaviours beyond recommender systems;

for example, it may provide a method of classifying and regulating behaviours in wider branches of AI.

Subsequent research should explore the integration of this mechanism *within* the training process. The simplest approach involves altering the final layer of a machine learning algorithm or neural network. A more advanced method could involve adjusting the internal functions of the algorithm using the hedonic state as a multiplicative factor. For example, consider a recurrent neural network with Long Short-Term Memory (LSTM) architecture (Hochreiter & Schmidhuber, 1997). By using allostatic regulation to shift the sigmoid function's centre in the forget gate of an LSTM network, one could regulate the "forgetfulness" of the network concerning different types of content. This could provide greater dynamic control of the network's outputs than can be achieved by modifying the final network layer. Still, there is reason to avoid modifying the inner workings of algorithms. For example, a recent study revealed a decline in OpenAI's GPT-4 model accuracy over time (Chen et al., 2023). Indeed, OpenAI's technical report on their GPT-4 model (Achiam et al., 2024) demonstrates that Reinforcement Learning with Human Feedback (RLHF) improves model performance on some metrics but worsens it on others. This highlights an advantage of implementing allostatic regulation as a code wrapper around the machine learning model, which does not directly affect the model's internal functions. Instead, it allows the model to learn user preferences while acting as a dynamic filter for the recommender's outputs. Future research should compare the strengths and weaknesses of both methods.

In conclusion, we have created a prototype allostatic regulator that can provide dynamic content regulation for social media users, improving content diversity and reducing harmful content exposure. This regulator can theoretically be applied at the output layer of even the most complex recommendation algorithms, and represents a novel paradigm shift in digital content regulation, offering a psychology-based solution that can help minimise the formation of echo chambers online. While this technique works in a simulated environment, more empirical research is needed to quantify its effectiveness in real-world situations.

Note

1. The choice of "Horror" movies was arbitrary, as this regulation method can be applied to any genre in the database. A single genre was chosen to simplify the demonstration of the allostatic regulator.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

No funding was received for conducting this study.

ORCID

N. I. N. Henry  <http://orcid.org/0000-0002-4299-1442>

M. Pedersen  <http://orcid.org/0000-0002-9199-1916>

M. Williams  <http://orcid.org/0000-0002-0571-215X>

J. L. B. Martin  <http://orcid.org/0000-0002-7204-231X>

L. Donkin  <http://orcid.org/0000-0002-0457-8788>

Data availability statement

The data for the MovieLens-1 M dataset is publicly available at <https://grouplens.org/datasets/movielens/1m/>. The Python code for the simulations performed can be found in the Supplementary Materials.

Author contributions

Nathan Henry: Conceptualisation, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualisation.

Mangor Pedersen: Conceptualisation, Methodology, Writing – Review & Editing, Supervision.

Matt Williams: Conceptualisation, Methodology, Writing – Review & Editing, Supervision.

Jamin Martin: Conceptualisation, Methodology, Writing – Review & Editing.

Liesje Donkin: Conceptualisation, Methodology, Writing – Review & Editing, Supervision.

Use of artificial intelligence statement

GPT-4 was used to assist in debugging the Python code developed for this study.

References

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., and Zoph, B. (2024). *GPT-4 Technical Report* (No. arXiv:2303.08774). arXiv. <https://doi.org/10.48550/arXiv.2303.08774>

- Ahmed, S. H., & Koob, G. F. (2005). Transition to drug addiction: A negative reinforcement model based on an allostatic decrease in reward function. *Psychopharmacology*, 180(3), 473–490. <https://doi.org/10.1007/s00213-005-2180-z>
- Alkiviadou, N. (2019). Hate speech on social media networks: Towards a regulatory framework? *Information & Communications Technology Law*, 28(1), 19–35. <https://doi.org/10.1080/13600834.2018.1494417>
- Anand, N., Sharma, M. K., Thakur, P. C., Mondal, I., Sahu, M., Singh, P., Kande, J. S., Ms, N., & Singh, R. (2022). Doomsurfing and doomscrolling mediate psychological distress in COVID-19 lockdown: Implications for awareness of cognitive biases. *Perspectives in Psychiatric Care*, 58(1), 170–172. <https://doi.org/10.1111/ppc.12803>
- Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers and Technology*, 24(3), 233–239. <https://doi.org/10.1080/13600869.2010.522323>
- Brannigan, R., Cronin, F., McEvoy, O., Stanistreet, D., & Layte, R. (2023). Verification of the goldilocks hypothesis: The association between screen use, digital media and psychiatric symptoms in the growing up in Ireland study. *Social Psychiatry and Psychiatric Epidemiology*, 58(8), 1259–1264. <https://doi.org/10.1007/s00127-022-02352-5>
- Brown, N. I. (2020). Regulatory goldilocks: Finding the just and right fit for content moderation on social platforms. *Texas A&M Law Review*, 8(3), 451–494. <https://doi.org/10.37419/LR.V8.I3.1>
- Chen, L., Zaharia, M., & Zou, J. (2023). How is ChatGPT's behavior changing over time? *arXiv* (No. arXiv:2307.09009). <https://doi.org/10.48550/arXiv.2307.09009>
- Christiano, T. (2022, January). *Algorithms, manipulation, and democracy*. Cambridge Core; Cambridge University Press. <https://doi.org/10.1017/can.2021.29>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Cota, W., Ferreira, S. C., Pastor-Satorras, R., & Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, 8(1), 35. <https://doi.org/10.1140/epjds/s13688-019-0213-9>
- Currin, C. B., Vera, S. V., & Khaledi-Nasab, A. (2022). Depolarization of echo chambers by random dynamical nudge. *Scientific Reports*, 12(1), 1. <https://doi.org/10.1038/s41598-022-12494-w>
- Dai, H., Wang, Y., Trivedi, R., & Song, L. (2017). Deep coevolutionary network: Embedding user and item features for recommendation. *arXiv*. (No. arXiv:1609.03675). <https://doi.org/10.48550/arXiv.1609.03675>
- Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology*, 40(S1), 3–35. <https://doi.org/10.1111/pops.12568>
- Du, N., Wang, Y., He, N., & Song, L. (2015). Time-sensitive recommendation from recurrent user activities. In *Proceedings of the 29th International Conference on Neural Information Processing Systems*, Montreal, Canada (pp. 3492–3500).
- Esther Ajewumi, O., Magbagbeola, V., Christine Kalu, O., Ada Ike, R., Folajimi, O., & Omolola Diyaolu, C. (2024). The impact of social media on mental health and well-being. *World Journal of Advanced Research & Reviews*, 24(1), 107–121. <https://doi.org/10.30574/wjarr.2024.24.1.3027>

- Etchells, P. J., Gage, S. H., Rutherford, A. D., Munafò, M. R., & Maher, B. (2016). Prospective investigation of video game use in children and subsequent conduct disorder and depression using data from the Avon Longitudinal Study of Parents and Children. *PLoS One*, 11(1), e0147732. <https://doi.org/10.1371/journal.pone.0147732>
- Farajtabar, M., Gomez-Rodriguez, M., Wang, Y., Li, S., Zha, H., & Song, L. (2018). COEVOLVE: A joint point process model for information diffusion and network Co-evolution. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW*, Lyon, France (Vol. 18, pp. 473–477). <https://doi.org/10.1145/3184558.3186236>
- Feng, C., Khan, M., Rahman, A. U., & Ahmad, A. (2020). News recommendation systems - accomplishments, challenges & future directions. *Institute of Electrical and Electronics Engineers Access*, 8, 16702–16725. <https://doi.org/10.1109/ACCESS.2020.2967792>
- Ge, Y., Zhao, S., Zhou, H., Pei, C., Sun, F., Ou, W., & Zhang, Y. (2020). Understanding echo chambers in E-commerce recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Xi'an, China (pp. 2261–2270). <https://doi.org/10.1145/3397271.3401431>
- George, O., Le Moal, M., & Koob, G. F. (2012). Allostasis and addiction: Role of the dopamine and corticotropin-releasing factor systems. *Physiology & Behavior*, 106(1), 58–64. <https://doi.org/10.1016/j.physbeh.2011.11.004>
- Goldstein, D. S., & McEwen, B. (2002). Allostasis, Homeostats, and the nature of stress. *Stress*, 5(1), 55–58. <https://doi.org/10.1080/102538902900012345>
- Haidt, J., & Allen, N. (2020). Scrutinizing the effects of digital technology on mental health. *Nature*, 578(7794), 226–227. <https://doi.org/10.1038/d41586-020-00296-x>
- Harper, F. M., & Konstan, J. A. (2016). The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4), 1–19. <https://doi.org/10.1145/2827872>
- Hartmann, D., Wang, S. M., Pohlmann, L., & Berendt, B. (2025). A systematic review of echo chamber research: Comparative Analysis of conceptualizations, operationalizations, and varying outcomes (No. arXiv:2407.06631; Version 3). arXiv. *Journal of Computational Social Science*, 8(2). <https://doi.org/10.1007/s42001-025-00381-z>
- Hasan, M. R., Jha, A. K., & Liu, Y. (2018). Excessive use of online video streaming services: Impact of recommender system use, psychological factors, and motives. *Computers in Human Behavior*, 80, 220–228. <https://doi.org/10.1016/j.chb.2017.11.020>
- Heerman, W. J., JaKa, M. M., Berge, J. M., Trapl, E. S., Sommer, E. C., Samuels, L. R., Jackson, N., Haapala, J. L., Kunin-Batson, A. S., Olson-Bullis, B. A., Hardin, H. K., Sherwood, N. E., & Barkin, S. L. (2017). The dose of behavioral interventions to prevent and treat childhood obesity: A systematic review and meta-regression. *International Journal of Behavioral Nutrition and Physical Activity*, 14(1), 157. <https://doi.org/10.1186/s12966-017-0615-7>
- Henry, N., Pedersen, M., Williams, M., & Donkin, L. (2023). Behavioral posology: A novel paradigm for modeling the healthy limits of behaviors. *Advanced Theory and Simulations*, 6(9), 2300214. <https://doi.org/10.1002/adts.202300214>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

- Interian, R., Marzo, R. G., Mendoza, I., & Ribeiro, C. C. (2023). Network polarization, filter bubbles, and echo chambers: An annotated review of measures and reduction methods. *International Transactions in Operational Research*, 30(6), 3122–3158. <https://doi.org/10.1111/itor.13224>
- Jeon, Y., Kim, B., Xiong, A., Lee, D., & Han, K. (2021). ChamberBreaker: Mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2), 1–26. <https://doi.org/10.1145/3479859>
- Jiang, R., Chiappa, S., Lattimore, T., György, A., & Kohli, P. (2019). Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, Honolulu, USA (pp. 383–390). <https://doi.org/10.1145/3306618.3314288>
- Kalluçi, E., & Peshkopia, R. (2024a). Invasiveness, intrusiveness and influence: Three new metrics to measure communication between political digital echo chambers. *Social Network Analysis and Mining*, 14(1), 31. <https://doi.org/10.1007/s13278-023-01186-6>
- Kalluçi, E., & Peshkopia, R. (2024b). Measuring the communication between multiple digital political echo chambers: Multidimensional invasiveness, intrusiveness, and influence. *Journal of Computational Social Science*, 8(1), 1. <https://doi.org/10.1007/s42001-024-00336-w>
- Kirk, A., Michael, K., Bergman, S., Schorr, M., & Jameson, J. P. (2019). Dose response effects of cognitive-behavioral therapy in a school mental health program. *Cognitive Behaviour Therapy*, 48(6), 497–516. <https://doi.org/10.1080/16506073.2018.1550527>
- Koob, G. F., & Le Moal, M. (2008). Neurobiological mechanisms for opponent motivational processes in addiction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1507), 3113–3123. <https://doi.org/10.1098/rstb.2008.0094>
- Lewczuk, K., Wójcik, A., & Gola, M. (2022). Increase in the prevalence of online pornography use: Objective data analysis from the period between 2004 and 2016 in Poland. *Archives of Sexual Behavior*, 51(2), 1157–1171. <https://doi.org/10.1007/s10508-021-02090-w>
- Liao, Q. V., & Fu, W.-T. (2014). Can you hear me now? Mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, Baltimore, USA (pp. 184–196). <https://doi.org/10.1145/2531602.2531711>
- Liekefett, L., Christ, O., & Becker, J. C. (2023). Can conspiracy beliefs be beneficial? Longitudinal linkages between conspiracy beliefs, anxiety, uncertainty aversion, and existential threat. *Personality and Social Psychology Bulletin*, 49(2), 167–179. <https://doi.org/10.1177/01461672211060965>
- Lim, S. L., & Bentley, P. J. (2022). Opinion amplification causes extreme polarization in social networks. *Scientific Reports*, 12(1), 18131. <https://doi.org/10.1038/s41598-022-22856-z>
- Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., & Lehmann, S. (2019). Accelerating dynamics of collective attention. *Nature Communications*, 10(1), 1. <https://doi.org/10.1038/s41467-019-09311-w>
- Mannell, K., & Meese, J. (2022). From doom-scrolling to news avoidance: Limiting news as a wellbeing strategy during COVID lockdown. *Journalism Studies*, 23(3), 302–319. <https://doi.org/10.1080/1461670X.2021.2021105>

- Manojlovich, M., & Sidani, S. (2008). Nurse dose: What's in a concept? *Research in Nursing & Health*, 31(4), 310–319. <https://doi.org/10.1002/nur.20265>
- Martin, G. N. (2019). (Why) do you like scary movies? A review of the empirical research on psychological responses to horror films. *Frontiers in Psychology*, 10, 2298. <https://doi.org/10.3389/fpsyg.2019.02298>
- McVay, M. A., Bennett, G. G., Steinberg, D., & Voils, C. I. (2019). Dose-response research in digital health interventions: Concepts, considerations, and challenges. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, 38(12), 1168–1174. <https://doi.org/10.1037/hea0000805>
- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & Society*, 35(4), 957–967. <https://doi.org/10.1007/s00146-020-00950-y>
- Najar, M. A., & Shabir, Q. A. (2023). Social media and reality: Are social media profiles a true representation of people's lives? *International Journal of Advanced Multidisciplinary Research and Studies*, 3(4), 878–887.
- Nguyen, T., Hui, P.-M., Harper, F., Terveen, L., & Konstan, J. (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity. In *WWW 2014—Proceedings of the 23rd International Conference on World Wide Web* (pp. 686). <https://doi.org/10.1145/2566486.2568012>
- Noordeh, E., Levin, R., Jiang, R., Shadmany, H., King, A. L., Mantz, A., Morris, R. G., Stanford, S. A., & Brammer, G. (2020). Spectroscopic confirmation of a mature galaxy cluster at a redshift of 2. *Nature*, 577(7788), 39–41. (No. arXiv:2011.03890). arXiv. <https://doi.org/10.48550/arXiv.2011.03890>
- Pan, J., Rao, V., Agarwal, P., & Gelfand, A. (2016). Markov-modulated marked poisson processes for Check-in data. In *Proceedings of The 33rd International Conference on Machine Learning*, New York City, USA (pp. 2244–2253).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Persily, N., & Tucker, J. A. (2020). *Social media and democracy: The state of the field, prospects for reform*. Cambridge University Press.
- Pirih, P., Ilić, M., Meglič, A., & Belušić, G. (2022). Opponent processing in the retinal mosaic of nymphalid butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1862), 20210275. <https://doi.org/10.1098/rstb.2021.0275>
- Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205–227. <https://doi.org/10.1016/j.eswa.2017.12.020>
- Przybylski, A. K. (2014). Electronic gaming and psychosocial adjustment. *Pediatrics*, 134(3), e716–e722. <https://doi.org/10.1542/peds.2013-4021>
- Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the goldilocks hypothesis: Quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological Science*, 28(2), 204–215. <https://doi.org/10.1177/0956797616678438>
- Reisach, U. (2021). The responsibility of social media in times of societal and political manipulation. *European Journal of Operational Research*, 291(3), 906–917. <https://doi.org/10.1016/j.ejor.2020.09.020>

- Roberts, J. A., & David, M. E. (2023). Instagram and TikTok Flow States and their association with psychological well-being. *Cyberpsychology, behavior, and Social Networking*, 26(2), 80–89. <https://doi.org/10.1089/cyber.2022.0117>
- Saurwein, F., & Spencer-Smith, C. (2021). Automated trouble: The role of algorithmic selection in harms on social media platforms. *Media and Communication*, 9(4), 222–233. <https://doi.org/10.17645/mac.v9i4.4062>
- Shabahang, R., Aruguete, M. S., & Shim, H. (2021). Online news addiction: Future anxiety, fear of missing out on news, and interpersonal trust contribute to excessive online news consumption. *Online Journal of Communication and Media Technologies*, 11(2), e202105. <https://doi.org/10.30935/ojcm/10822>
- Shiffman, S. M., & Stone, A. A. (1998). Ecological momentary assessment: A new tool for behavioral medicine research. In D. Krantz & A. Baum (Eds.), *Technology and methods in behavioral medicine* (pp. 117–131). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410602770>
- Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychological Review*, 81(2), 119–145. <https://doi.org/10.1037/h0036128>
- Tan, X., Naqvi, S. A. Z., Qi, A. Y., Heller, K. A., & Rao, V. (2016). Content-based modeling of reciprocal relationships using Hawkes and Gaussian processes. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, Jersey City, USA (pp. 726–734).
- Terren, L. T. L., & Borge-Bravo, R.-B.-B. R. (2021). Echo chambers on social media: A systematic review of the literature. *Review of Communication Research*, 9, 9. <https://doi.org/10.12840/ISSN.2255-4165.028>
- Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017). Know-Evolve: Deep temporal reasoning for dynamic knowledge graphs. *arXiv*. (No. arXiv:1705.05742). <https://doi.org/10.48550/arXiv.1705.05742>
- Uribe-Bahamonde, Y. E., Becerra, S. A., Ponce, F. P., & Vogel, E. H. (2019). A quantitative account of the behavioral characteristics of habituation: The sometimes opponent processes model of stimulus processing. *Frontiers in Psychology*, 10, 504. <https://doi.org/10.3389/fpsyg.2019.00504>
- Van Rossum, G., & Fred, L. (2009). *Python 3 reference manual*. CreateSpace.
- Voils, C. I., Chang, Y., Crandell, J., Leeman, J., Sandelowski, M., & Maciejewski, M. L. (2012). Informing the dosing of interventions in randomized trials. *Contemporary Clinical Trials*, 33(6), 1225–1230. <https://doi.org/10.1016/j.cct.2012.07.011>
- Vu, N. Q., & Bezemer, C.-P. (2021). Improving the discoverability of indie games by leveraging their similarity to top-selling games: Identifying important requirements of a recommender system. In *The 16th International Conference on the Foundations of Digital Games (FDG)*, Montreal, Canada (pp. 1–12). <https://doi.org/10.1145/3472538.3472548>
- Wang, Y., Theodorou, E., Verma, A., & Song, L. (2018). A stochastic differential equation framework for guiding online user activities in closed loop. *arXiv*. (No. arXiv:1603.09021). <https://doi.org/10.48550/arXiv.1603.09021>
- Wang, Y., Williams, G., Theodorou, E., & Song, L. (2017). Variational policy for guiding point processes. *arXiv*. (No. arXiv:1701.08585). <https://doi.org/10.48550/arXiv.1701.08585>
- Wang, Y., Xie, B., Du, N., & Song, L. (2016). Isotonic hawkes processes. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, New York City, USA (Vol. 48, pp. 2226–2234).