Recommender Systems as Dynamical Systems: Interactions with Viewers and Creators*

Sarah Dean[1], **Evan Dong**[1], **Meena Jagadeesan**[2] **Liu Leqi**[3, 4]

[1]Cornell University, [2]UC Berkeley, [3]Princeton University, [4]UT Austin
sdean@cornell.edu, edong@cs.cornell.edu, mjagadeesan@berkeley.edu, leqiliu@princeton.edu

Abstract

The design and evaluation of recommender systems often takes the perspective of supervised machine learning, treating viewer preferences and the content catalogue as static. However, in reality, recommender systems interact with and shape the behavior of viewers and content creators. In this position paper, we argue that due to these interactions, recommendation systems are more accurately characterized as dynamical systems, impacting the environment in which they operate. Towards this goal, we propose a unified framework of a recommender system as a dynamical system, and we formulate existing mathematical models of interactions between recommender systems, viewers, and creators from prior work within this framework. This framework allows us to identify the similarities and differences between these models, which we hope aids future development of mathematical models for recommender system dynamics.

1 Introduction

Personalized recommendation algorithms were born of the decentralized web, motivated to help individuals navigate a "deluge" of articles, songs, and videos sent over email lists [28, 18, 39]. Today, recommendation algorithms are an ubiquitous part of the online experience. They mediate interactions on many online platforms, where individuals both view and create content. Recommendations have impact—not just on the immediate satisfaction of a viewer, but also on the formation of interests and popularity of content [5].

We thus argue that recommender systems should be analyzed as dynamical systems that account for the interactions with viewers and creators. We are motivated by how the recommender systems shapes and is shaped by interactions with viewers and creators; in fact, these interactions can drive shifts in viewer preferences [32] and shifts in the landscape of content available on the platform [36]. The societal ramifications of these interactions make it crucial to account for dynamics when designing and evaluating recommendation algorithms.

Towards this ambitious goal, an emerging line of work has proposed different mathematical models of interactions among a recommender, viewers, and creators. These models offer different formalizations of how a recommender system shapes (and is shaped by) viewer preferences and behavior as well as creator behavior. However, these models adopt different perspectives of impact ranging from dynamical systems to game theory to behavioral psychology, which makes it challenging to compare and integrate these models.

The goal of our work is to propose a unified framework that allows us to formulate models from prior work in a common language. We cast interactions between the recommender system, viewers, and creators as a dynamical system with measurements and states that can vary over time (Section 2). Then, we formulate several models on viewer dynamics and creator dynamics within this common mathematical framework, which allows us to compare the formalizations and perspectives taken in different papers (Sections 3-4). We hope that our unification and analysis of prior work aids the development of mathematical models for recommender system dynamics and furthers the design and evaluation of recommender systems.

2 Background and Framework

We propose the following unified model for interactions between a platform which hosts and recommends content, individuals who view the content, and individuals who create the content. We refer to these three distinct types of stakeholders as the *recommender*, *viewers*, and *creators* respectively. The overall system consist of a single platform, $m \in \mathbb{N}_+$ viewers, and $p \in \mathbb{N}_+$ content creators.

At time t, there are various observable quantities, which we refer to as *measurements* and denote by y_t. We further distinguish measurements pertaining to the three entities of interest, so $y_t = (y_t^r, y_t^v, y_t^c)$. Here, y_t^r is the most

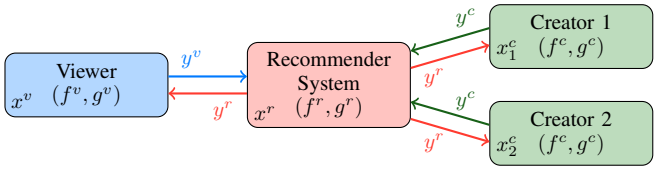

Figure 1: Illustration of recommendation systems as the interconnection of three distinct entities, each described as a dynamical system.

*Authors listed alphabetically

recent recommendation, $y_t^v = (y_{1,t}^v, \cdots, y_{m,t}^v)$ contains the most recent observed viewer behaviors (e.g., likes, clicks, watch time, replies) for each of the m viewers, and $y_t^c = (y_{1,t}^c, \cdots, y_{p,t}^c)$ contains the observed creator behavior (e.g., attributes of the created content) for each of the p creators.

To describe models of long term impact and dynamics, we additionally consider an internal state x_t. Similarly to measurements, we distinguish the state variables of the three types of stakeholders. The recommender's state is x_t^r, the viewers' states are x_t^v, and the creators' states are x_t^c. These states may not be directly observable to the recommender. However, the state at time t determines the measurement at t, subject to corruption by the measurement noise v_t. Additionally, the state contains sufficient information to predict its own evolution, aside from process noise w_t. That is, the state at $t+1$ depends on the state at t and the measurement at t, along with the unobserved w_t. We now detail how the interactions between the recommender, viewers, and creators shape the states and measurements over time.

Recommender. Recommendations are selected according to a recommendation function g_r, which takes as input the recommender system's internal state. Without loss of generality, we simply define the recommender state as $x_t^r = [y_1, \ldots, y_{t-1}]$, which is the history of observed measurements (but not unobservables). This state definition captures the fact that most recommendation algorithms are trained on historical data of viewer and creator behaviors. In addition to the state, we allow recommendations to depend on a noise variable v_t to model randomness in the selection or training algorithms. In particular, for $i \in [m]$, the output $g_r(i; x_t^r, v_t)$ describes the content shown to viewer i. The content comes from an available content pool, which is defined with respect to creator outputs y_t^c. Then the internal state x_{t+1}^r is updated by appending the current measurement y_t to the previous internal state x_t^r.

Viewers. In any recommendation system, the recommendations y_t^r impact the observed viewer behaviors y_t^v, due to the simple fact that viewers react to the recommended content. Many of the papers we survey consider an additional level of impact: effects on the viewers' internal state x_t^v. Such effects are motivated by a variety of phenomena, ranging from boredom to opinion formation (Section 3). The formal model is as follows: for each viewer $i \in [m]$, the recommendations impact the evolution of the viewer state, which evolves according to $x_{i,t+1}^v = f_v(x_{i,t}^v, y_t, w_t)$. Here, w_t is process noise capturing unmodelled effects or randomness. Then the viewer behavior $y_{i,t}^v = g_v(x_{i,t}^v, y_{i,t}^r, v_t)$ is governed by the unobserved viewer state $x_{i,t}^v$ and noise v_t. For example, many recommendation algorithms personalize by modelling each viewer with a parameter vectors that is learned from data; these parameter vectors can be seen as modelling a static internal viewer state.

Creators. Traditional recommendation algorithm design considers a fixed catalog of content. In this survey, we include a growing body of work which models the process of content creation, often through a strategic lens. In these papers, the creators choose to create content on the basis of several factors: the recommendations made by the recommender, the observed behavior of other creators, and the consumption patterns of viewers.

Most works take the perspective of game theory and mechanism design, defining utility functions rather than explicit dynamics. Instead, we represent this implicit dependence through a model of memoryless state evolution. More formally, for each $j \in [p]$, the internal creator state evolves as $x_{j,t+1}^c = f_c(y_t, w_t)$ where w_t is process noise capturing unmodelled effects or randomness. Then the content that is available at time t, described by the observable attributes $y_{j,t}^c$, is determined by the internal state $y_{j,t}^c = g_c(x_{j,t}^c, v_t)$ and possibly noise v_t.

3 Models of Viewer Dynamics

Of the works included in our survey, the majority are concerned with how viewers are impacted by the recommendations they receive. In these papers, the content pool is modelled as unchanging, or changing over time independently from the recommendations and viewer behaviors. In other words, x_t^c is not dependent on y_t. Therefore, we focus our discussion on the representation of x_t^v, its evolution as describe by the model f_v, and its relationship to measurements y_t^v. We split our discussion into three categories: preference shifts, transient phenomena, and behavioral shifts.

Preference shifts

Many works consider settings in which recommendations change viewer preferences. In these settings, the primary state representation $x_{i,t}^v$ models a viewer preference, which determines the measured behavior $y_{i,t}^v$, e.g. rating or click. Below, we present an overview of these works and provide detailed mathematical definitions in Appendix A.

State Definition. Many papers model the viewer preference as a vector containing affinities for each discrete piece of content (or type of content) in the catalog [23, 1, 6, 29, 13]. Others model the preference more continuously, as a vector in a latent space [12, 34, 11, 8, 24]. In this latter category, the recommended content also has a latent representation in the same vector space. In either case, the preference state indicates how much a viewer enjoys, agrees with, or is drawn to particular pieces of content, types of content, or content with particular characteristics. In particular, the preference determines viewer response to recommendations: a continuous valued rating [12, 34, 11], the choice of an item from the recommended set [23, 1, 6, 8] or from the entire catalog [29], or the acceptance or rejection of a single recommendation [24, 13].

State Update. Viewer preferences change in response to the recommendations, as a result of either exposure or interaction. In some papers, the preference is influenced by the mere exposure to content in a slate [23, 29, 8, 13] or single recommendation [24, 13, 11, 34], while in others, the update depends on the resulting viewer selection [1, 6] or behavior [12].

In many works, the preference shifts towards the recommended content [24, 8, 29, 11], increasing the viewer's

affinity for similar content. Lu et al. [34] consider both "attraction" and "repulsion". In Jiang et al. [23], affinity increases for selected content and a decreases for the recommendations that are not selected. Dean and Morgenstern [12] adopt a model of "biased assimilation" in which preferences shift towards recommendations that the viewer enjoys and away from recommendations that they dislike. Evans and Kasirzadeh [13] model a similar effect in a simplified example of political polarization. The operant conditioning model of Curmei et al. [11] includes a scalar "memory" of previous ratings which determines a "surprise" factor that modulates the magnitude and direction of the preference shift. In addition to preference, Krueger, Maharaj, and Leike [29] model viewer "loyalty" based on recommendation quality. Some papers consider a general class of relationships between past recommendations and affinity [1, 6].

Goal. The works surveyed in this section have a variety of goals. Several papers are motivated by concerns and critiques of recommender systems, like echo chambers and polarization [34, 23, 24, 12]. These works propose particular dynamics models, demonstrate a failure mode of traditional recommendation algorithms, and propose new algorithms to address these problems. Curmei et al. [11] present dynamics models inspired by psychology literature and propose a framework for empirical validation. Other works consider general classes of preference shifts [1, 6] and develop algorithms with provable guarantees. In a similar vein, some works are interested the general phenomenon of manipulation, and present algorithms for generic reinforcement learning settings [29, 8, 13]. In these works, particular preference models are introduced as an example in simulation experiments.

Transient phenomena

A recent line of work studies how viewer preferences may evolve in a transient fashion, showing as boredom and habituation effects of recommendations. In these settings, depending on how recommenders present content to the viewers, their preferences may shift and later return to their initial state. In all these papers, the viewer state $x_{i,t}^v$ represents an internal state that governs viewer preference (or viewer preference itself) and the measurement $y_{i,t}^v$ is either viewers' feedback (e.g., ratings or clicks) or viewers' utilities. We offer a high-level overview of this literature and provide detailed mathematical definitions of viewer models in Appendix B.

State Definition. The viewer state $x_{i,t}^v$ can sometimes be summary statistics of viewers' past consumption history. For example, the viewer state is modeled as the number of times each recommendation category is presented [33]; as the time elapsed since the last recommendation of a recommendation category [27, 37, 9]; as switches among recommendation categories [30, 14]; or as the entire past interaction history a viewer has experienced [38]. In other settings, the viewer's internal state is unobserved to recommenders. For example, the internal viewer state represents an unknown viewer type [2]; or unobserved viewer satiation/memory towards different categories [31, 25].

State Update. Viewer preference models in this line of work capture different aspects of transient preference shifts. One growing line of work models boredom or satiation in recommender systems under a multi-armed bandits setting [33, 27, 37, 31, 25]. The updates of viewers' internal states capture the accumulation of memory that the viewer has over past recommendation categories. The expected measurement $\mathbb{E}[y_{i,t}^v]$ decreases (indicating the disinterest of the viewer towards the recommended category) as viewer boredom accumulates towards the recommended category. Under the same multi-armed bandits setting, the updates in Laforgue et al. [30], Foussoul et al. [14] capture that viewer preference may be seasonal or periodic; the dynamics in Ben-Porat et al. [2] model the phenomena that viewers may leave the platform upon being recommended unsatisfying content. Finally, in the most general setting, the state updates in Saig and Rosenfeld [38] appends the viewer's current interaction to the viewer's interaction history with the recommender system.

Goal. The overall goals for these works consist of two components: (1) uncover ways of estimating viewer preference dynamics [31, 37]; and (2) find optimal recommender policy g_r for viewers with these transient preference shifts [31, 27, 33, 37, 30, 14, 2, 38, 25]. The objective for finding the optimal recommender policy is often defined as a function depending on $(\{y_{i,t}^v\}_{i\in[n]})_t$. Depending on the semantic meaning of $y_{i,t}^v$, the optimal recommender policy could be either viewer-optimal (e.g., maximizing viewers' utility) or platform-optimal (e.g., maximizing viewers' click-through rates).

Behavioral shifts

A handful of papers move beyond viewer preferences and investigate broader forms of viewer behavioral shifts. These works consider several different mechanisms influencing behavior, including behavioral weaknesses of viewers [26] and strategic reasoning [35, 17, 10]. Below, we present a high-level overview of these models and defer further details to Appendix C.

State Definition. To encapsulate all of these nuances, the viewer state is represented as more than just a preference. Different papers, which examine different phenomena, define the state in different ways. Kleinberg, Mullainathan, and Raghavan [26] propose a dual-process model of viewer behavior reminiscent of psychology and behavioral economics, where viewers are modelled as following one of two consumption processes. In this model, the state is represented as the consumption process that the viewer is currently operating in. Mansour, Slivkins, and Syrgkanis [35] explore strategic settings in which we can decompose the internal viewer state $x_i^v = (x_i^{v,P}, x_i^{v,B})$ into viewers' preference $x_i^{v,P}$ and viewer's information set or belief $x_i^{v,B}$ in a Bayesian, game theoretic sense. Haupt, Hadfield-Menell, and Podimata [17], Cen, Ilyas, and Madry [10] study viewers strategically deciding what content to engage with to curate their feed: in this model, the viewer state x^v represents a vector-valued decision about which content to engage with.

State Update. The state update also varies across papers. For example, Kleinberg, Mullainathan, and Raghavan [26] model viewers as probabilistically switching between consumption processes depending on the content properties. Mansour, Slivkins, and Syrgkanis [35] consider independent individual viewers who sequentially interact with the system once, but make choices after Bayesian updates based on recommendation $y^r_{i,t}$. Haupt, Hadfield-Menell, and Podimata [17], Cen, Ilyas, and Madry [10] focus on game theoretic equilibria, which can sometimes be reached through best-response dynamics.

Goal. The overall goal of these works is to examine viewer behavioral shifts that are not captured by preferences alone. More specifically, these works formalize negative outcomes which arise from these behavioral nuances, and design optimal system policies that prevent these negative outcomes. Negative outcomes are defined with respect to a variety of different values, such as fairness and bias [17], viewer welfare [26][17, 10], or content exploration [35].

4 Models of Creator Interactions

A growing body of work has studied how the recommendation algorithm can shape creator behaviors, with a focus on how the recommendation algorithm implicitly shapes the landscape of content available on the platform. Although these papers take the perspective of game theory, we recast these papers within the dynamical systems framework in Section 2. The underlying model is that creators—who compete to win recommendations—strategically design their content to appear in as many recommendations as possible.

When casting these game-theoretic models within our dynamical system frameworks, the state corresponds to the creator *action* and the state update is implicitly specified by a creator *reward function* and *recommender function*. More specifically, the internal creator state $x^c_{j,t} = a^c_{j,t}$ is specified by the creator's actions $a^c_{j,t} \in \mathcal{A}$ about the content that they intend to create. (In these models, the vector $[a^c_{1,t}, \ldots, a^c_{p,t}]$ of internal states of the p creators determine the content landscape at time step t.) The creator's state evolves as a best-response f_c to the creator utility function. In these models, the creator utility depends on (1) the creator reward function h which depends on actions (i.e., $a_{j,t}$) and measurements (i.e., y^r_t and y^v_t) and (2) the recommender function g_r which depends on the recommender state. Different papers differ in the state definition (as specified by the action space) as well as state updates (as specified by the reward function and recommender function). Below, we present a high-level overview of these models and defer further details to Appendix D.

State Definition (Action Space). The geometry of the action space $\mathcal{A}$ varies from 1-dimensional, to finite, to high-dimensional across different papers. For example, Ghosh and McAfee [16] take $\mathcal{A} = [0, 1]$ to capture quality, and Buening et al. [7] take $\mathcal{A} = [0, 1]$ to capture feedback probability. Ben-Porat and Tennenholtz [4] take $\mathcal{A}$ to be any finite space which captures variation in content type. Jagadeesan, Garg, and Steinhardt [22], Hron et al. [19] take $\mathcal{A}$ to be the d-dimensional space $\mathbb{R}^d$ or nonnegative orthant $\mathbb{R}^d_{\geq 0}$. The dimensions correspond to content attributes and capture the embeddings learned by (nonnegative) matrix factorization. Yao et al. [41] consider general subsets of $\mathbb{R}^d$.

State Update (Reward Function). The specifics of the creator reward function h vary across papers, with different papers emphasizing different aspects of exposure, engagement, costs, and nonmyopicness. Ben-Porat and Tennenholtz [4], Hron et al. [19] take the creator reward to be the number of recommendations won by the creator. Jagadeesan, Garg, and Steinhardt [22] additionally incorporate production costs which scale with content quality. Immorlica, Jagadeesan, and Lucier [21] additionally incorporate that viewers don't necessarily consume recommendations (and modifies the creator reward accordingly). Yao et al. [40] take the creator reward to be engagement, and Yao et al. [41] considers more general score-based functions. Ghosh and McAfee [16] provide monetary prizes to creators based on the creator's rank along magnitude (quality) and takes the reward to be the prize minus production costs. Ghosh and Hummel [15], Hu et al. [20] focus on non-myopic creators accounting for the reward from future time steps when computing their best response, and Hu et al. [20] captures that creators can take different actions at each time step.

State Update (Recommender Function). The specifics of the recommendation function g_r also differ across papers, with some works focusing on standard recommendation algorithms and other works taking a platform design perspective. For example, motivated by matrix factorization, Jagadeesan, Garg, and Steinhardt [22], Hron et al. [19] focus on recommendations specified by the inner product of content and viewer preference vectors, and Hron et al. [19] incorporate noise. Yao et al. [40] consider recommendations which output a slate of the top K pieces of content according to an engagement metric. To capture interactions across many time steps, Ghosh and Hummel [15], Hu et al. [20] consider stateful recommendation functions based on multi-armed bandit algorithms. Taking a design perspective, Ben-Porat and Tennenholtz [4], Yao et al. [41] construct recommendation functions which achieve fairness and welfare properties.

Goal. Since these papers take the perspective of game theory, the primary focus is on the *fixed points* of the dynamical system (which corresponds to equilibria of the game between creators), rather than on dynamics. Different papers study different aspects of the fixed points: some works [4, 19] focus on the existence of fixed points (i.e., equilibrium existence); other works [22, 7, 21] examine the structure of the fixed points (i.e., the equilibrium structure), focusing on the potential for specialization by creators [22] and the potential for clickbait [7, 21]; other works study viewer welfare at the fixed points [15, 20, 41]. Finally, a handful of works (e.g. [3, 19]) start to bring in best-response and better-response dynamics and study convergence.

Acknowledgements

We thank Micah Carroll for many insights into preference dynamics models. This work was supported in part by NSF CCF 2312774, NSF OAC-2311521, an Open Philanthropy AI fellowship, a LinkedIn Research Award, a PCCW Affinito-Stewart Grant, and a gift from Wayfair.

References

[1] Agarwal, A.; and Brown, W. 2023. Online Recommendations for Agents with Discounted Adaptive Preferences. *CoRR*, abs/2302.06014.

[2] Ben-Porat, O.; Cohen, L.; Leqi, L.; Lipton, Z. C.; and Mansour, Y. 2022. Modeling attrition in recommender systems with departing bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 6072–6079.

[3] Ben-Porat, O.; Rosenberg, I.; and Tennenholtz, M. 2020. Content Provider Dynamics and Coordination in Recommendation Ecosystems. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

[4] Ben-Porat, O.; and Tennenholtz, M. 2018. A Game-Theoretic Approach to Recommendation Systems with Strategic Content Providers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 1118–1128.

[5] Boutilier, C.; Mladenov, M.; and Tennenholtz, G. 2023. Modeling Recommender Ecosystems: Research Challenges at the Intersection of Mechanism Design, Reinforcement Learning and Generative Models. *CoRR*, abs/2309.06375.

[6] Brown, W.; and Agarwal, A. 2022. Diversified recommendations for agents with adaptive preferences. *Advances in Neural Information Processing Systems*, 35: 26066–26077.

[7] Buening, T. K.; Saha, A.; Dimitrakakis, C.; and Xu, H. 2023. Bandits Meet Mechanism Design to Combat Clickbait in Online Recommendation. *CoRR*, abs/2311.15647.

[8] Carroll, M. D.; Dragan, A. D.; Russell, S.; and Hadfield-Menell, D. 2022. Estimating and Penalizing Induced Preference Shifts in Recommender Systems. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, 2686–2708. PMLR.

[9] Cella, L.; and Cesa-Bianchi, N. 2020. Stochastic bandits with delay-dependent payoffs. In *International Conference on Artificial Intelligence and Statistics*, 1168–1177. PMLR.

[10] Cen, S. H.; Ilyas, A.; and Madry, A. 2023. User Strategization and Trustworthy Algorithms. *CoRR*, abs/2312.17666.

[11] Curmei, M.; Haupt, A. A.; Recht, B.; and Hadfield-Menell, D. 2022. Towards psychologically-grounded dynamic preference models. In *Proceedings of the 16th ACM Conference on Recommender Systems*, 35–48.

[12] Dean, S.; and Morgenstern, J. 2022. Preference Dynamics Under Personalized Recommendations. In *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022*, 795–816.

[13] Evans, C.; and Kasirzadeh, A. 2021. User Tampering in Reinforcement Learning Recommender Systems. *CoRR*, abs/2109.04083.

[14] Foussoul, A.; Goyal, V.; Papadigenopoulos, O.; and Zeevi, A. 2023. Last Switch Dependent Bandits with Monotone Payoff Functions. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, 10265–10284. PMLR.

[15] Ghosh, A.; and Hummel, P. 2013. Learning and incentives in user-generated content: multi-armed bandits with endogenous arms. In Kleinberg, R. D., ed., *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, 233–246. ACM.

[16] Ghosh, A.; and McAfee, R. P. 2011. Incentivizing high-quality user-generated content. In Srinivasan, S.; Ramamritham, K.; Kumar, A.; Ravindra, M. P.; Bertino, E.; and Kumar, R., eds., *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*, 137–146. ACM.

[17] Haupt, A. A.; Hadfield-Menell, D.; and Podimata, C. 2023. Recommending to Strategic Users. *CoRR*, abs/2302.06559.

[18] Hill, W.; Stead, L.; Rosenstein, M.; and Furnas, G. 1995. Recommending and evaluating choices in a virtual community of use. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 194–201.

[19] Hron, J.; Krauth, K.; Jordan, M. I.; Kilbertus, N.; and Dean, S. 2022. Modeling Content Creator Incentives on Algorithm-Curated Platforms. *CoRR*, abs/2206.13102.

[20] Hu, X.; Jagadeesan, M.; Jordan, M. I.; and Steinhardt, J. 2023. Incentivizing High-Quality Content in Online Recommender Systems. *CoRR*, abs/2306.07479.

[21] Immorlica, N.; Jagadeesan, M.; and Lucier, B. 2024. Clickbait vs. Quality: How Engagement-Based Optimization Shapes the Content Landscape in Online Platforms. *CoRR*, abs/2401.0980.

[22] Jagadeesan, M.; Garg, N.; and Steinhardt, J. 2022. Supply-Side Equilibria in Recommender Systems. *CoRR*, abs/2206.13489.

[23] Jiang, R.; Chiappa, S.; Lattimore, T.; György, A.; and Kohli, P. 2019. Degenerate Feedback Loops in Recommender Systems. In Conitzer, V.; Hadfield, G. K.; and

Vallor, S., eds., *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*, 383–390. ACM.

[24] Kalimeris, D.; Bhagat, S.; Kalyanaraman, S.; and Weinsberg, U. 2021. Preference amplification in recommender systems. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 805–815.

[25] Khosravi, K.; Leme, R. P.; Podimata, C.; and Tsorvantzis, A. 2023. Bandits with Deterministically Evolving States. *CoRR*, abs/2307.11655.

[26] Kleinberg, J. M.; Mullainathan, S.; and Raghavan, M. 2022. The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization. In *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022*, 29. ACM.

[27] Kleinberg, R.; and Immorlica, N. 2018. Recharging bandits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, 309–319. IEEE.

[28] Konstan, J. A.; Miller, B. N.; Maltz, D.; Herlocker, J. L.; Gordon, L. R.; and Riedl, J. 1997. Grouplens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3): 77–87.

[29] Krueger, D.; Maharaj, T.; and Leike, J. 2020. Hidden Incentives for Auto-Induced Distributional Shift. *CoRR*, abs/2009.09153.

[30] Laforgue, P.; Clerici, G.; Cesa-Bianchi, N.; and Gilad-Bachrach, R. 2022. A last switch dependent analysis of satiation and seasonality in bandits. In *International Conference on Artificial Intelligence and Statistics*, 971–990. PMLR.

[31] Leqi, L.; Kilinc Karzan, F.; Lipton, Z.; and Montgomery, A. 2021. Rebounding bandits for modeling satiation effects. *Advances in Neural Information Processing Systems*, 34: 4003–4014.

[32] Leqi, L.; Zhou, G.; Kilinc-Karzan, F.; Lipton, Z.; and Montgomery, A. 2023. A Field Test of Bandit Algorithms for Recommendations: Understanding the Validity of Assumptions on Human Preferences in Multi-armed Bandits. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16.

[33] Levine, N.; Crammer, K.; and Mannor, S. 2017. Rotting bandits. *Advances in neural information processing systems*, 30.

[34] Lu, W.; Ioannidis, S.; Bhagat, S.; and Lakshmanan, L. V. 2014. Optimal recommendations under attraction, aversion, and social influence. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 811–820.

[35] Mansour, Y.; Slivkins, A.; and Syrgkanis, V. 2020. Bayesian Incentive-Compatible Bandit Exploration. *Operations Resesearch*, 68(4): 1132–1161.

[36] Meyerson, E. 2012. YouTube Now: Why We Focus on Watch Time.

[37] Pike-Burke, C.; and Grunewalder, S. 2019. Recovering bandits. *Advances in Neural Information Processing Systems*, 32.

[38] Saig, E.; and Rosenfeld, N. 2023. Learning to suggest breaks: sustainable optimization of long-term user engagement. In *International Conference on Machine Learning*, 29671–29696. PMLR.

[39] Shardanand, U.; and Maes, P. 1995. Social information filtering: Algorithms for automating "word of mouth". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 210–217.

[40] Yao, F.; Li, C.; Nekipelov, D.; Wang, H.; and Xu, H. 2023. How Bad is Top-K Recommendation under Competing Content Creators? In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, 39674–39701. PMLR.

[41] Yao, F.; Li, C.; Sankararaman, K. A.; Liao, Y.; Zhu, Y.; Wang, Q.; Wang, H.; and Xu, H. 2023. Rethinking Incentives in Recommender Systems: Are Monotone Rewards Always Beneficial? *CoRR*, abs/2306.07893.

A Additional details for Preference Shifts in Section 3

We discuss each paper, describe how the states and measurement variables are defined, and provide the formal models of dynamics and measurement functions.

Notation: S^{d-1} is the sphere in d dimensions.

- Lu et al. [34] define for each viewer i the preference state $x^v_{i,t} \in \mathbb{R}^d$, the recommended content $y^r_{i,t} \in \mathbb{R}^d$, and the measured rating $y^v_{i,t} \in \mathbb{R}$ for some latent dimension d. The dynamics capture three possible behaviors: stationary, attraction, and aversation:

$$x^v_{i,t+1} \begin{cases} \sim \mu_{i,0} & \text{w.p. } \alpha_1 \\ = \sum_{\tau=1}^t w_{t-\tau} y^r_{i,\tau} & \text{w.p. } \alpha_2 \\ = -\sum_{\tau=1}^t w_{t-\tau} y^r_{i,\tau} & \text{w.p. } \alpha_3 \end{cases} \quad \mathbb{E}[y^v_{i,t}] = \langle x^v_{i,t}, y^r_{i,t} \rangle$$

- Curmei et al. [11] define for each viewer i the preference state $x^v_{i,t} \in \mathbb{R}^d$, the recommended content $y^r_{i,t} \in \mathbb{R}^d$, and the measured rating $y^v_{i,t} \in \mathbb{R}$ for some latent dimension d. The define dynamics for an effect "mere exposure"

$$x^v_{i,t+1} = (1-\alpha)x^v_{i,t} + \alpha y^r_{i,t} \quad \mathbb{E}[y^v_{i,t}] = \langle x^v_{i,t}, y^r_{i,t} \rangle .$$

  They propose an additional model of "operant conditioning" with state variable augmented to include a memory variable $x^v_{i,t} = (p_{i,t}, m_{i,t}) \in \mathbb{R}^d \times \mathbb{R}$

$$\begin{bmatrix} m_{i,t+1} \\ p_{i,t+1} \end{bmatrix} = \begin{bmatrix} \delta(m_{i,t} + (p^v_{i,t})^\top y^r_{i,t}) \\ (1-\alpha|s_{i,t}|)p_{i,t} + \alpha_1 s_{i,t} y^r_{i,t} \end{bmatrix}, \quad s_{i,t} = \arctan\left(\frac{1}{\sum_{\tau=1}^{t-1} \delta^\tau} m_t - (p_{i,t})^\top y^r_{i,t} \right), \quad \mathbb{E}[y^v_{i,t}] = \langle x^v_{i,t}, y^r_{i,t} \rangle$$

- Dean and Morgenstern [12] define for each viewer i the preference state $x^v_{i,t} \in \mathcal{S}^{d-1}$, the recommended content $y^r_{i,t} \in \mathcal{S}^{d-1}$, and the measured rating $y^v_{i,t} \in \mathbb{R}$ for some latent dimension d. The dynamics capture "biased assimilation," with

$$x^v_{i,t+1} \propto x^v_{i,t} + \eta_t \langle x^v_{i,t}, y^r_{i,t} \rangle \cdot y^r_{i,t}, \quad \mathbb{E}[y^v_{i,t}] = \langle x^v_{i,t}, y^r_{i,t} \rangle .$$

- Kalimeris et al. [24] define for each viewer i the recommendation including both content and predicted score $y^r_{i,t} = (s_{i,t}, c_{i,t}) \in \mathbb{R} \times \mathcal{C}$ and the measured click $y^v_{i,t} \in \{0,1\}$. The viewer state $x^v_{i,t} \in [0,1]$ is memoryless and determined by the predicted score:

$$x^v_{i,t} = \begin{cases} \sigma(s_{i,t}) + (1 - \sigma(s_{i,t})\gamma r(s_{i,t})) & s_{i,t} > 0 \\ \sigma(s_{i,t})(1 + \gamma r(s_{i,t})) & s_{i,t} \le 0 \end{cases}, \quad \mathbb{E}[y^v_{i,t}] = x^v_{i,t}$$

- Jiang et al. [23] define for each viewer i the preference state $x^v_{i,t} \in \mathbb{R}^p$ the recommended length k slate of content $y^r_{i,t} \in [p]^k$, and the measured binary feedback $y^v_{i,t} \in \{0,1\}^k$ for p discrete piece of content. They study a class of dynamics models wherein the preference for an item increases when it is clicked, and decreases when it is recommended but not clicked:

$$\forall\, j \in y^r_{i,t}, \quad \begin{cases} x^v_{i,t+1}[j] > x^v_{i,t}[j] & y^v_{i,t}[j] = 1 \\ x^v_{i,t+1}[j] < x^v_{i,t}[j] & y^v_{i,t}[j] = 0 \end{cases} \quad \mathbb{E}[y^v_{i,t}[j]] \text{ is increasing in } x^v_{i,j}[j]$$

- Agarwal and Brown [1] define for each viewer i the preference state $x^v_{i,t} \in \Delta(p)$, the recommended length k slate of content $y^r_{i,t} \in [p]^k$, and the clicked item $y^v_{i,t} \in y^r_{i,t}$ for p discrete piece of content. They study a class of dynamics models:

$$x^v_{i,t+1} = \frac{1}{\sum_{\tau=0}^t \gamma^\tau} (\gamma x^v_{i,t} + e_{y^v_{i,t}}) \quad y^v_{i,t} = c \text{ w.p. } \propto f_c(x^v_{i,t})$$

  Brown and Agarwal [6] study a similar model for the case that $\gamma = 1$.

- Krueger, Maharaj, and Leike [29] define a global "loyalty" state $x^v_{g,t} \in \mathbb{R}^m$, for each viewer i the preference state $x^v_{i,t} \in \mathbb{R}^p$, the recommended content $y^r_{i,t} \in [p]$, and the consumed content $y^v_{i,t} \in [p]$. A single viewer is active at each timestep with $i_t \sim \text{softmax}(x^v_{g,t})$. Then for this viewer $i = i_t$, the loyalty and preference update, and the viewer selects a content independent of the top recommendation.

$$x^c_{g,t+1}[i] = x^c_{g,t}[i] + \alpha_1 x^v_{i,t}[y^r_{i,t}], \quad x^v_{i,t+1} = \frac{x^v_{i,t} + \alpha_1 e_{y^r_{i,t}}}{\|x^v_{i,t} + \alpha_2 e_{y^r_{i,t}}\|_2}, \quad y^v_{i_t,t} \sim \text{softmax}(x^v_{i_t,t})$$

- Carroll et al. [8] define for each viewer i the preference state $x^v_{i,t} \in \mathbb{R}^d$, the recommended content $y^r_{i,t} \in \Delta(\mathcal{C})$ for a fixed content set $\mathcal{C} = \{c_1, \ldots, c_p\} \subset \mathbb{R}^d$, and viewer selection behavior $y^v_{i,t} \in \mathcal{C}$. The state update is influenced by a viewer belief over the future available content

$$x^v_{i,t} = x \text{ w.p. } \propto \exp(\beta_2(\lambda \bar{c}^\top x^v_{i,t} + (1-\lambda)\bar{c}^\top p)), \quad \bar{c} = \frac{\sum_{c \in \mathcal{C}} (y^r_{i,t}[c])^3 c}{\sum_{c \in \mathcal{C}} (y^r_{i,t}[c])^3}, \quad y^v_{i,t} = c \text{ w.p. } \propto y^r_{i,t}[c] \exp(\beta_1 {x^v_{i,t}}^\top c)$$

- Evans and Kasirzadeh [13] define for each viewer i the belief state $x^v_{i,t} \in [0,1]^3$, recommended content type $y^r_{i,t} \in \{1,2,3\}$, and viewer response $y^v_{i,t} \in \{0,1\}$. For either $j, j' = 1,3$ (when $x^v_{i,t}[1]$ is largest element) or $j, j' = 3,1$ (when $x^v_{i,t}[3]$ is largest element), $x^v_{i,t+1}[j] = \mathbf{1}\{y^r_{i,t} = j'\} \min\{p_t x^v_{i,t}[j], 1\}$ where p_t is sampled of r.v. with $\mathbb{E}[p_t] > 1$. $y^v_{i,t} = 1$ w.p. $x^v_{i,t}[y^r_{i,t}]$

B Additional details for Transient Phenomena in Section 3

In this section, most of the papers adopt a multi-armed bandits framework where the learner is the recommender system; the arms to pull indicates the recommendation (category) to present to the viewer; and the received reward is the feedback given by the viewer or the viewer utility. We use the notation that for any vector a, $a[k]$ indicates the k-th entry of a. In cases where a is a one-hot vector representing the action taken by the learner, its non-zero entry represents the arm being pulled.

- In Levine, Crammer, and Mannor [33], there are K recommendation categories/arms; and $y^r_{i,t}$ is a one-hot vector of K dimensions representing the recommendations given to viewer i at time t. If $y^r_{i,t}[k] = 1$, arm k is pulled. The viewer state $x^v_{i,t} \in \mathbb{N}^K_+$ has its k-th entry be the number of time times arm k has been pulled so far. More specifically, $x^v_{i,t+1}[k] = x^v_{i,t}[k] + \mathbb{I}\{y^r_{i,t}[k] = 1\}$ and $x^v_{i,0}[k] = 0$. The expected measurement is the expected reward of pulling an arm: If $y^r_{i,t}[k] = 1$, $\mathbb{E}[y^v_{i,t}] = m_k(x^v_{i,t}[k])$ where m_k is an arm-dependent monotonically decreasing function of the number of arm pulls.
- In Kleinberg and Immorlica [27], the recommendation $y^r_{i,t}$ is defined the same as that of Levine, Crammer, and Mannor [33]. The viewer state $x^v_{i,t} \in \mathbb{N}^K_+$ has its k-th entry be the number of time steps elapsed since k is pulled last time. More specifically, $x^v_{i,t+1}[k] = x^v_{i,t}[k] + 1$ if $y^r_{i,t}[k] \neq 1$, $x^v_{i,t+1}[k] = 1$ if $y^r_{i,t}[k] = 1$ and $x^v_{i,0}[k] = 0$. The expected measurement is the expected reward of pulling an arm: If $y^r_{i,t}[k] = 1$, $\mathbb{E}[y^v_{i,t}] = m_k(x^v_{i,t}[k])$ where m_k is a concave function of the number of arm pulls. In Pike-Burke and Grunewalder [37], m_k is drawn from a Gaussian process. In Cella and Cesa-Bianchi [9], m_k is a monotonically increasing function.
- In Leqi et al. [31], the recommendation $y^r_{i,t}$ is defined the same as that of Levine, Crammer, and Mannor [33]. The viewer state $x^v_{i,t} \in \mathbb{N}^K_+$ has its k-th entry be the satiation that the viewer has towards arm k. More specifically, $x^v_{i,t+1}[k] = \gamma_k(x^v_{i,t}[k] + y^r_{i,t}[k])$ and $x^v_{i,0}[k] = 0$ where $\gamma_k \in (0,1)$ is the satiation retention factor. The expected measurement is the expected reward of pulling an arm: If $y^r_{i,t}[k] = 1$, $\mathbb{E}[y^v_{i,t}] = b_k - \lambda_k x^v_{i,t}[k]$ where $b_k \in \mathbb{R}$ is the base reward of arm k and $\lambda_k \geq 0$ is the exposure influence factor for arm k.
- In Ben-Porat et al. [2], the recommendation $y^r_{i,t}$ is defined the same as that of Levine, Crammer, and Mannor [33]. The internal state $x^v_{i,0} \sim \mathbf{Q}$ is a viewer type belonging to $[B]$ ($B \in \mathbb{N}_+$) sampled from a prior distribution $\mathbf{Q}$. For $t \geq 1$, $x^v_{i,t} = x^v_{i,t-1}$ if $y^v_{i,t-1} = 1$ (i.e., viewer state remains if they have clicked on the recommendation) else $x^v_{i,t} = 0$ with probability $\mathcal{L}_{k,x^v_{i,t}}$ (indicating that the viewer may leave the platform with probability $\mathcal{L}_{k,x^v_{i,t}}$). The expected measurement is the expected click rate when the recommender pulls arm k, i.e., $\mathbb{E}[y^v_{i,t}] = \mathbf{P}_{k,x^v_{i,0}} \cdot \mathbb{I}\{x^v_{i,t} \neq 0\}$ if $y^r_{i,t}[k] = 1$.
- In Saig and Rosenfeld [38], the viewer internal state $x^v_{i,t}$ is a set of tuples consisting of their previous interaction with the platform. That is, $x^v_{i,t} = \{(b_j, y^r_{i,j}, y^v_{i,j}) \mid j < t\}$ where b_j is time when j-th event happens, $y^r_{i,j}$ is the j-th recommendation viewer i obtained and $y^v_{i,j}$ is the viewer's reported rating. Depending on the subject of interest, $y^v_{i,t}$ is defined differently as a function of $x^v_{i,t}$. For example, $y^v_{i,t}$ can be $(b_t - b_{t-1})^{-1}$, which is viewer's "instantaneous [response] rate."
- In Laforgue et al. [30], the recommendation $y^r_{i,t}$ is defined the same as that of Levine, Crammer, and Mannor [33]. The internal state $x^v_{i,t}$ is a K-dimensional vector where $x^v_{i,t}[k] = \delta_k(x^v_{i,t-1}[k], y^r_{i,t-1})$ where δ_k is a transition function that tracks arm switches. The expected measurement when $y^r_{i,t}[k] = 1$ is $\mathbb{E}[y^v_{i,t} = m_k(x^v_{i,t}[k])$ where m_k is an arm-wise reward function. In Foussoul et al. [14], m_k is assumed to be monotonic.
- In Khosravi et al. [25], the recommendation $y^r_{i,t}$ is defined the same as that of Levine, Crammer, and Mannor [33]. The viewer internal state is real-valued, i.e., $x^v_{i,t} \in \mathbb{R}$. If $y^r_{i,t}[k] = 1$, $x^v_{i,t} = x^v_{i,t-1} + \lambda(b_k - x^v_{i,t-1})$ where b_k is an arm-wise constant and λ is a universal constant. If $y^r_{i,t}[k] = 1$, then $\mathbb{E}[y^v_{i,t}] = r_k \cdot x^v_{i,t}$ where $r_k \in \mathbb{R}$ is an arm-wise constant.

C Additional details for Behavioral Shifts in Section 3

The models in this section vary substantially from paper to paper.

- Kleinberg, Mullainathan, and Raghavan [26] define two systems with their own logics and utility distributions $\mathcal{U}$ and $\mathcal{V}$, which may be correlated. In system 1, item t produces utility $u_t \sim \mathcal{U}$; if $u_t > 0$, the viewer continues to next round, otherwise, the viewer's continuation to the next round is determined by System 2. With System 2, item t produces utility $v_t \sim \mathcal{V}$. Leaving the system (not consuming t) gives utility W, with net utility $W - v_t$. With probability $q > 0$, the viewer continues. With probability $1 - q$, the viewer discontinues.
- Mansour, Slivkins, and Syrgkanis [35] examine Bayesian agents who separately and sequentially choose to pull bandit arms once; each arm i has reward drawn from $\mathcal{D}(\mu_i)$ for expected reward μ_i and $\mu = (\mu_1, ..., \mu_m)$ drawn from known

prior $\mathcal{P}^0$ (which is common knowledge). The recommender system recommends σ_t to viewer t, and the viewer chooses $\arg\max_i \mathbb{E}[\mu_i|\sigma_t]$ (with some additional conditions on past viewers having followed recommendations).

- Haupt, Hadfield-Menell, and Podimata [17] model viewers having personal preference type θ (a preference distribution over content types $j \in [d]$) and observe system-chosen recommendation policy g mapping consumption frequencies to recommendation types. Viewers strategically choose a "consumption plan" (i.e. consumed distribution over $[d]$ to feed into the algorithm). They explore several recommender system interventions, including an "incognito mode" where the recommendation function is restricted. They focus on a complete-information setting where such an intervention is known, but explicitly suggest a more general incomplete information setting with limited knowledge of the recommendation function for future work.
- Cen, Ilyas, and Madry [10] similarly use a game-theoretic model. A viewer has a memoryless best-response function f^v mapping recommendations y^r to a distribution x^v, from which their behavior y^v is drawn.

D Additional details for Content Creators in Section 4

The models for creator behavior can be broken down into two categories: models which focus on a single time step and treat the system as myopic, and models which consider non-myopic interactions across multiple time steps. In all of these models, the viewers are static.

Myopic models. We describe the myopic models ([16, 4, 3, 22, 19, 40, 41, 21]) where the creator and recommender system do not factor in the future impacts.

Before diving into the papers, we introduce some formalisms. Creator j's internal state $x^c_{j,t} = a^c_{j,t}$ is specified by the creator's internal action $a^c_{j,t} \in \mathcal{A}$ about what content to create. The state update f_c captures the creator's best-response to their utility function $u_j(a; a^c_{-j,t}, y^r_t)$ which depends on the action a of a creator j, the actions of other creators $j' \neq j$, and the recommendations y^r_t. In particular, the state update is defined by the best-response

$$f_c(y_t, w_t) = \text{argmax}_{a\in\mathcal{A}} u_j(a; a^c_{-j,t}, y_t).$$

The state update implicitly depends on a creator reward function $h(j, y^r_t, y^v_t, a)$ (which captures the reward that creator j receives from the measurements y^r_t and y^v_t along with production costs of their action a) and the recommender function $g_r(i, x^r_t, v_t)$ (which maps the recommender state, which includes the content landscape, to measurements). In the myopic models, the recommender function $g_r(i; x^r_t, v_t)$ only depends on y_{t-1} and v_t, and notably *not* on $[y_1, \ldots, y_{t-2}]$. We thus introduce the simplified function $g'_r(i; y_{t-1}, v_t)$ to denote recommendations.

We specify each paper in terms of how it defines the action space $\mathcal{A}$, the reward function h, and the recommender system g'_r:

- Ghosh and McAfee [16] take $\mathcal{A} = [0, 1]$ (capturing quality). The recommender function g'_r awards prizes to creators based on their quality rank. The reward h captures prize for creator j according to the recommendation ranks given by y^r_t minus a 1-time cost of production $c(a)$. The model additionally incorporates participation decisions.
- Ben-Porat and Tennenholtz [4] take $\mathcal{A}$ to be a finite set. The recommender function is a randomized function g'_r mapping each viewer to a content in y^c_t or no content (denoted by $\emptyset$). The reward h captures the number of recommendations in y^r_t assigned to creator j.
- Ben-Porat, Rosenberg, and Tennenholtz [3] take $\mathcal{A}$ to be a finite set capturing content topics. When a given creator j writes on a topic $a \in \mathcal{A}$, their article is a predetermined quality $q_{a,j}$. Each viewer seeks a particular topic $a \in \mathcal{A}$ of content. The recommender function g'_r assigns a viewer seeking a topic a the content with highest quality of that topic: that is, $\text{argmax}_{j\in[p]} q_{a,j} \cdot I[a = a^c_{j,t}]$. The reward h captures the number of recommendations in y^r_t assigned to creator weighted by topic-specific weights.
- Jagadeesan, Garg, and Steinhardt [22] take $\mathcal{A} = \mathbb{R}^D_{\geq 0}$ to be D-dimensional content embeddings in the nonnegative orthant. Each viewer i has a fixed preference vector $u_i \in \mathbb{R}^D_{\geq 0}$. The recommender function g'_r assigns the viewer i the content created by creator j that maximizes the inner product $\langle u_i, a^c_{j,t}\rangle$: that is, $\text{argmax}_{j\in[p]} \langle u_i, a^c_{j,t}\rangle$. The reward h is the number of recommendations won minus the 1-time cost of production specified by $c(a) = ||a||^\beta$.
- Hron et al. [19] take $\mathcal{A}$ to be D-dimensional content embeddings in the unit sphere $\left\{x/||x||_2 \mid x \in \mathbb{R}^D\right\}$. Each viewer i has a fixed preference vector $u_i \in \mathbb{R}^D$. The recommender function g'_r assigns viewer i the content created by creator j with probability proportional to $e^{\eta\langle u_i, a^c_{j,t}\rangle}$. The reward h is the number of recommendations won.
- Yao et al. [40] take $\mathcal{A} \subseteq \mathbb{R}^D$ to be an abstract set. There is a set of viewers $X \subseteq \mathbb{R}^D$. The recommender system computes scores for each content and viewer pair and assigns the top K recommendations to be the top K scores. The reward h is the sum of the scores of the recommendations won.
- Yao et al. [41] take $\mathcal{A} \subseteq \mathbb{R}^D$ to be an abstract set. There is a set of viewers $X \subseteq \mathbb{R}^D$. The recommender system computes scores for each content and viewer pair and determines recommendations g'_r based on an arbitrary function of these scores and the viewer characteristics. The reward h captures the reward received by the creator (specified by a general reward function of the scores and the viewer) minus the production cost $c(a)$.

- Immorlica, Jagadeesan, and Lucier [21] take $\mathcal{A} = \mathbb{R}^2_{\geq 0}$ to be 2-dimensional content embeddings capturing a clickbait dimension and quality dimension. Each viewer i has a fixed type $t_i > 0$ representing their tolerance for clickbait and only engages with content if they derive nonnegative utility (viewer utility increases with quality and decreases with clickbait). The recommender function g_r optimizes an engagement metric (which is misaligned with viewer welfare) and assigns viewer i the content that maximizes viewer i's engagement. The reward h is the number of recommendations won (that viewers actually engage with) minus a 1-time cost of production.

We specify the form of the creator utility function u_j. The creator utility function u_j anticipates the impact of the creator's actions on recommendations in the next time step, which make it slightly messy to formalize in the dynamical system framework, but which we can nonetheless formalize as follows. At time step t, creators j first assesses how their actions a and the actions of other creators $a^c_{j,t}$ would affect the content landscape

$$\tilde{y}^c_t(a) := [a^c_{1,t}, a^c_{2,t}, \ldots, a^c_{j-1,t}, a, a^c_{j+1,t} \ldots, a^c_{p,t}].$$

Let the measurements with this content landscape be denoted by $\tilde{y}_t(a)$. Then, the creator assesses the impact on the downstream recommendations

$$\tilde{y}^r_{t+1}(a) := [g'_r(1; \tilde{y}_t(a)), \ldots, g'_r(n; \tilde{y}_t(a))]$$

and observed viewer behaviors

$$\tilde{y}^v_{i,t+1}(a) = g_v(x^v_i, \tilde{y}^r_{i,t+1}(a), v_t),$$

and computes their utility:

$$u_j(a; a^c_{-j,t}, y_t) := h(j, \tilde{y}^r_{t+1}(a), \tilde{y}^v_{t+1}(a), a).$$

Non-myopic models. We next turn to the non-myopic models ([15, 20, 7]). The dynamic models of content creator behavior are slightly harder to formulate, since the recommender function and creators are non-myopic. The recommender function g_r is modelled as a multi-bandit algorithm. Creators account for their (potentially discounted) cumulative reward when selecting actions.

- Ghosh and Hummel [15] consider a setup where each creator j enters the platform at a potentially different time step. The action set is $\mathcal{A} = [0, 1]$ and captures content quality. The recommender system is a stochastic multi-armed bandit algorithm. The reward h is the number of recommendations to be won at the current round and in the future minus the cost of production $c(a)$. The model additionally incorporates participation decisions.
- Hu et al. [20] consider a setup where all creators arrive at every time step and can create new content at each time step. The action set is $\mathcal{A} \in (\mathbb{R}^D_{\geq 0})^T$. The recommender system is an adversarial multi-armed bandit algorithm. The reward h is the cumulative discounted sum of the number of recommendations to be won minus the cost of production at each time step. Production costs at each time step t are specified by $c(a_t) = ||a_t||^\beta$. This model makes the simplifying assumption that each creator chooses the content that they will produce at every time step at the beginning of the game.
- Buening et al. [7] consider a setup where creators all arrive at every time and each choose the feedback rate $a \in \mathcal{A} = [0, 1]$ of their content at the beginning of the game. The recommender system is a multi-armed bandit algorithm that accounts for probabilistic feedback. The reward h is the total number of recommendations won.