Behavioral Debt: The Cognitive Cost of Algorithmic Preference Substitution

Hugo Thack

Inverso Research Institute

hello@hugothack.com

Doctoral Dissertation — Independent Research Submission

Abstract

Algorithmic recommendation systems have evolved from instruments of preference surfacing into engines of preference shaping. This dissertation introduces the concept of **Behavioral Debt** — the accumulated cognitive cost incurred when algorithmic recommendation systems progressively substitute user-generated preferences with system-optimized alternatives. Drawing on a synthesis of cognitive psychology, behavioral economics, human-computer interaction research, and critical technology studies, the framework addresses a specific lacuna in

existing scholarship: while the attention economy (Davenport & Beck, 2001), surveillance capitalism (Zuboff, 2019), and the filter bubble (Pariser, 2011) describe the economic logic and informational consequences of algorithmic mediation, none adequately theorizes the cognitive residue that accumulates within users as a result of sustained preference substitution. Behavioral Debt is modeled through three interacting sub-constructs: (1) **Preference Substitution**, the systematic replacement of autonomous preference-formation processes with algorithmically generated alternatives optimized for platform objectives; (2) **Autonomy Erosion**, the progressive degradation of an individual's capacity for independent evaluative choice through sustained disuse of deliberative cognitive processes; and (3) the **Attention Ledger**, a conceptual accounting mechanism that tracks the cumulative displacement of self-directed cognitive engagement by algorithmically captured attention. The dissertation provides formal definitions and mathematical notation for each sub-construct, develops a taxonomy of debt accumulation pathways, and differentiates Behavioral Debt from adjacent theoretical frameworks. A proposed mixed-methods empirical study design is presented, incorporating validated psychometric instruments, behavioral experiments, and qualitative interview protocols suitable for testing the framework's core propositions. Conceptual analysis is applied across three domains — social media feeds, streaming recommendation systems, and e-commerce platforms — to demonstrate the framework's explanatory scope. Implications for platform design, regulatory policy, educational practice, and individual digital literacy are

discussed. The dissertation concludes with a call for empirical research programs to validate and refine the framework's components. (Word count: 278)

Keywords: algorithmic recommendation, preference formation, cognitive externalities, behavioral economics, digital autonomy, attention economy, choice architecture, persuasive technology, platform capitalism

---

1. Introduction

1.1 The Transformation of Digital Choice Environments

The architecture of digital attention has undergone a fundamental transformation over the past two decades. What began as a retrieval problem — how to connect users to relevant content in a vast information space — has become a production problem: how to construct user states that maximize platform-defined engagement metrics (Wu, 2016; Srnicek, 2017). This shift from preference *surfacing* to preference *shaping* is neither incidental nor invisible to researchers. Yet the dominant frameworks available to describe it remain calibrated to the economic extraction side of the interaction. They explain what platforms gain. They are less precise about what users lose — and specifically about what accumulates inside users over time as a result of sustained exposure to preference-substituting systems.

The contemporary digital ecosystem is defined by what Srnicek (2017) calls "platform capitalism" — an economic model in which digital platforms serve as intermediaries that extract value from the interactions they facilitate. Within this model, recommendation algorithms function as the primary interface between users and content, mediating access to music, news, entertainment, social connections, consumer products, and increasingly, professional and educational resources (Jesse & Jannach, 2021). These systems are not neutral conduits. As Tufekci (2015) argued in her analysis of algorithmic gatekeeping, recommendation algorithms function as subjective decision-makers of consequence, exercising editorial power comparable to — but structurally different from — traditional media gatekeepers. Their distinguishing features include individualized operation (each user receives different outputs), opacity (the decision-making process is hidden), and continuous optimization toward platform-defined objectives.

The scale of this mediation is historically unprecedented. A substantial majority of content consumption on platforms such as YouTube, TikTok, Spotify, Netflix, and Amazon is algorithmically recommended rather than user-initiated (Chaney et al., 2018). Users may scroll through dozens or hundreds of algorithmically curated items in a single session, each interaction generating behavioral data that further refines the system's model of their preferences. The result is a recursive loop in which the algorithm's model of the user progressively shapes the user's model of themselves (Burr et al., 2018).

1.2 The Gap in Existing Frameworks

The attention economy framework, developed by Davenport and Beck (2001) and elaborated by Wu (2016), describes the commodification of human attention as a scarce resource. Surveillance capitalism, as theorized by Zuboff (2019), maps the extraction and commercial deployment of behavioral data at scale. The filter bubble concept (Pariser, 2011) identifies the epistemic narrowing that results from algorithmic personalization of information environments. Each of these frameworks has made genuine contributions to the field's understanding of human-algorithm interaction. Each is also, in a precise sense, incomplete for a specific purpose: none of them adequately describes the cognitive residue that accumulates within the user's own cognitive architecture as a consequence of sustained algorithmic preference substitution.

Fogg's (2003) work on persuasive technology established that computers could be designed to change attitudes and behaviors, introducing the concept of "captology" and identifying the mechanisms through which digital systems influence users. His Behavior Model (Fogg, 2009) specifies that behavior change occurs at the intersection of motivation, ability, and triggers. While foundational, Fogg's framework focuses on discrete behavioral change events rather than the long-term cumulative cognitive consequences of sustained technology-mediated preference displacement.

Similarly, Sunstein and Thaler's (2008) influential work on nudge theory and choice architecture demonstrates that the design of choice environments profoundly influences decision outcomes. Their framework establishes that defaults are powerful, that framing shapes

evaluation, and that choice architects wield significant influence over choosers. Research by Herberz et al. (2021) has confirmed through meta-analysis that choice architecture interventions produce meaningful behavioral effects across multiple domains. However, nudge theory is primarily concerned with the moment of choice and with optimizing choice outcomes; it does not systematically address the long-term consequences for the chooser's cognitive capacity when choice is repeatedly mediated by external agents.

What is missing from this constellation of frameworks is a theory of *cognitive accumulation* — an account of what happens, over time, inside the user who has had their preferences mediated by algorithmic systems across thousands of interactions, over months and years, in domains spanning entertainment, information consumption, social connection, and purchasing behavior.

1.3 Contribution and Research Questions

This dissertation introduces the concept of **Behavioral Debt** to fill that gap. The term is deliberately borrowed from software engineering, where "technical debt" — coined by Ward Cunningham in a 1992 experience report and subsequently formalized in the literature (Cunningham, 1992; Nugroho et al., 2011) — refers to the long-term maintenance cost incurred when developers implement expedient short-term solutions rather than architecturally sound ones. The debt compounds: each shortcut makes future development marginally more expensive. The critical feature of technical debt is its *invisibility* during accumulation. Systems function. Code runs. The cost is deferred and obscured — until it is not.

Behavioral Debt operates by an analogous mechanism, but at the level of human cognition rather than software architecture. Each time a recommendation system makes a choice on behalf of a user — each playlist built by an algorithm, each feed curated by an engagement optimizer, each search result ordered by a relevance model trained on population-level behavior rather than individual preference — a small unit of cognitive work is either displaced or never performed. The user does not form a preference from scratch; instead, they accept or reject a preference offered by the system. Over thousands of such interactions, the capacity to form autonomous preferences may atrophy through disuse, even as the user remains unaware that anything has been lost.

This dissertation addresses three primary research questions:

RQ1: How can the cumulative cognitive costs of algorithmic preference substitution be formally defined and differentiated from existing frameworks such as the attention economy, surveillance capitalism, and the filter bubble?

RQ2: What are the mechanisms — cognitive, behavioral, and systemic — through which Behavioral Debt accumulates, and how can these mechanisms be formalized through a multi-component model?

RQ3: What empirical methods are appropriate for measuring Behavioral Debt, and what implications does the framework carry for platform design, regulatory policy, and individual digital literacy?

The dissertation proceeds as follows. Section 2 reviews the relevant literature across six domains: the attention economy, surveillance capitalism, nudge theory and choice architecture, algorithmic

governance, preference formation, and existing critiques of recommendation systems. Section 3 develops the Behavioral Debt theoretical framework in detail, including formal definitions, mathematical notation, a taxonomy of accumulation pathways, and differentiation from adjacent concepts. Section 4 proposes a mixed-methods empirical study design. Section 5 applies the framework through conceptual analysis across three domains. Section 6 discusses implications for design, policy, education, and individual practice. Section 7 addresses limitations and future research directions. Section 8 concludes.

---

2. Literature Review

2.1 The Attention Economy

The conceptual foundations of the attention economy were laid by Herbert Simon (1971), who observed that "a wealth of information creates a poverty of attention" (p. 40). This insight was systematized by Davenport and Beck (2001) in their influential *The Attention Economy*, which framed human attention as the scarcest resource in information-rich environments and argued that understanding attention allocation was essential for business strategy.

Wu (2016) provided a historical account of the "attention merchants" — the industries that have, since the nineteenth century, engaged in the business of capturing human attention and reselling it to advertisers. His analysis traces the evolution of attention capture from

the penny press through radio, television, and into the digital age, establishing that the fundamental business model has remained constant even as its mechanisms have become dramatically more sophisticated. The digital attention economy, Wu argues, differs from its predecessors primarily in the granularity and continuity of attention capture: where television captured attention in half-hour blocks, digital platforms capture it in micro-moments, continuously, and with real-time feedback loops that enable rapid optimization.

Fernández-Rovira et al. (2020) extended this analysis to examine how the struggle for human attention in digital environments produces measurable effects on users' psychological and social functioning. Their work documented the mechanisms through which attention-capturing interfaces generate compulsive use patterns and contribute to psychological distress — effects that accumulate over time rather than manifesting in single interactions.

The attention economy framework provides an essential macroeconomic lens for understanding why recommendation systems behave as they do — they are attention-capture machines operating in an attention-scarce market. However, the framework's analytic focus remains on the economic transaction: attention is the currency, and the question is how it is captured and spent. What is not addressed is what happens to the cognitive architecture of the person whose attention is repeatedly captured and directed. This is the gap that the Behavioral Debt framework addresses.

2.2 Surveillance Capitalism

Zuboff's (2019) *The Age of Surveillance Capitalism* represents the most comprehensive critique of the political economy of algorithmic systems. Zuboff argues that a new form of capitalism has emerged in which human experience is claimed as free raw material for behavioral data extraction. This data, processed through machine intelligence, yields "behavioral surplus" — predictions about future behavior that are sold in "behavioral futures markets." The system's success depends on increasingly intimate knowledge of users, which in turn requires increasingly invasive data collection.

Zuboff's contribution to the present framework is twofold. First, her analysis of "instrumentarian power" — the power to shape behavior at scale through computational means — provides the political-economic context within which Behavioral Debt accumulates. Recommendation systems are not neutral tools; they are instruments of a specific economic logic that benefits from sustained user engagement and, by extension, from sustained preference substitution. Second, her concept of the "Big Other" — a ubiquitous digital apparatus that observes, records, and shapes behavior — establishes the environmental context: users of recommendation systems operate within an apparatus designed to make their behavior legible and predictable.

However, Zuboff's framework is primarily concerned with what happens to behavioral data *outside* the user — how it is extracted, processed, and sold. Behavioral Debt is concerned with what happens *inside* the user — how the cognitive architecture through which preferences are generated, maintained, and revised is affected by

sustained algorithmic mediation. The two frameworks are complementary, not competing: surveillance capitalism describes the extraction apparatus, and Behavioral Debt describes the cognitive residue left by that apparatus within the individuals it processes.

2.3 Nudge Theory and Choice Architecture

Thaler and Sunstein's (2008) *Nudge* popularized the concept of choice architecture — the idea that the design of environments in which choices are made profoundly influences the choices people make. Their framework of "libertarian paternalism" argued that it is possible to influence behavior in welfare-enhancing directions without restricting freedom of choice, by manipulating defaults, framing, and the salience of options.

The choice architecture literature has established several findings directly relevant to the Behavioral Debt framework. First, defaults are extraordinarily powerful: people accept default options at rates far exceeding what conscious preference would predict (Herberz et al., 2021). In algorithmic recommendation contexts, the algorithm's output functions as a highly salient default — it is presented first, often with social validation signals, in an interface designed to maximize acceptance. Second, choice architecture effects are robust across domains and populations, as confirmed by meta-analytic evidence showing that nudges produce meaningful effects in health, financial, environmental, and consumer behavior (Herberz et al., 2021; Beshears & Kosowsky, 2020).

The relevance to Behavioral Debt is that recommendation systems function as automated choice architects operating at unprecedented scale. Unlike traditional nudges, which are typically designed for specific behavioral outcomes in specific contexts, algorithmic recommendation constitutes a *continuous* form of choice architecture that mediates preference expression across virtually all domains of digital activity. The user does not encounter an occasional nudge in a specific decision environment; they navigate a comprehensively architected preference landscape in which every presented option has been selected, ordered, and framed by an algorithm optimized for engagement.

Susser et al. (2019) developed a critical extension of this analysis, arguing that online manipulation — the covert use of information technology to influence decision-making by targeting and exploiting decision-making vulnerabilities — represents a fundamentally different category from nudging. Where nudges operate transparently and in the interest of the chooser, online manipulation operates covertly and in the interest of the manipulator. The boundary between algorithmic recommendation and manipulation, Susser et al. argue, is crossed when the system covertly engineers the choice environment to steer decisions toward outcomes that benefit the platform at the expense of the user's autonomous preferences.

2.4 Algorithmic Governance and Gatekeeping

Tufekci (2015) provided a foundational analysis of algorithms as gatekeepers, arguing that computational processes increasingly function as subjective decision-makers in domains of significant consequence —

from news curation to credit scoring to criminal sentencing. Her analysis identified three properties of algorithmic gatekeeping that distinguish it from traditional editorial gatekeeping: individualization (different users see different outputs), opacity (the decision process is hidden), and continuous optimization (the system evolves in response to behavioral data).

The FAccT (Fairness, Accountability, and Transparency) research community has developed an extensive body of work examining algorithmic systems through the lenses of fairness and accountability (Young et al., 2022; Gansky & McDonald, 2022). This scholarship has focused primarily on discriminatory outcomes, procedural fairness, and institutional accountability for algorithmic decision-making. While this work is essential, it has been oriented predominantly toward questions of social justice and group-level outcomes rather than toward the cognitive effects of algorithmic mediation on individual users.

Rakova and Chowdhury (2019) contributed an important bridge between the algorithmic governance literature and the cognitive effects perspective, analyzing human self-determination within algorithmic sociotechnical systems. Their concept of a "barrier-to-exit" metric — the effort a user must expend for a recommendation system to recognize a change in their preferences — is directly relevant to the Decision Outsourcing component of Behavioral Debt. High barriers to exit create a structural incentive for users to accept algorithmic recommendations rather than invest the cognitive effort required to override them.

2.5 Preference Formation and Cognitive Science

Understanding Behavioral Debt requires grounding in the cognitive science of preference formation. Preferences are not static, pre-existing entities that algorithms merely discover; they are actively constructed through cognitive processes that are themselves shaped by experience, context, and environment (Kahneman, 2011; Lichtenstein & Slovic, 2006).

Kahneman's (2011) dual-process theory distinguishes between System 1 (fast, automatic, intuitive) and System 2 (slow, effortful, deliberative) cognition. Preference formation in novel or important domains engages System 2 — it requires conscious evaluation, comparison, and judgment. Algorithmic recommendation systems, by pre-selecting and presenting options, may shift preference expression from System 2 to System 1 processing: instead of deliberating about what they want, users respond intuitively to what is offered. This shift is efficient in the moment but may, over time, reduce the engagement of deliberative processes that maintain preference-formation capacity.

Sweller's (1988) cognitive load theory, developed initially in the context of educational instruction, provides a complementary framework. The theory establishes that working memory capacity is finite and that cognitive processes are subject to capacity-dependence effects. When recommendation systems perform the work of preference-organization on behalf of users, they reduce the cognitive load experienced in the moment of choice. This reduction is the source of their subjective appeal. But cognitive load theory also suggests that the processes thus relieved may not be maintained simply through observation — they require exercise to remain available.

The research on choice overload contributes additional theoretical grounding. Iyengar and Lepper's (2000) landmark studies demonstrated that extensive choice sets can demotivate choice and reduce satisfaction — a finding subsequently examined through meta-analysis by Scheibehenne et al. (2010), who found near-zero mean effects but substantial variance, suggesting important moderating conditions. In the context of algorithmic recommendation, the relevance is that recommendation systems simultaneously resolve the choice overload problem (by reducing the effective choice set) and create a new problem (by substituting algorithmically determined selections for autonomous preference formation). The user is protected from the anxiety of excessive choice but at the cost of ceding evaluative control to the algorithm.

Schwartz's (2004) distinction between "maximizers" (who seek the best possible option) and "satisficers" (who seek an option that is good enough) is similarly relevant. Algorithmic recommendation may convert maximizers into satisficers by making the cognitive cost of maximum search prohibitively high relative to the ease of accepting a recommendation. Over time, this behavioral shift may become dispositional — not merely a strategy for managing the current choice environment but a general approach to preference expression.

2.6 Habit Formation and Behavioral Automaticity

Wood and Neal (2007) established that habits are not simply repeated behaviors — they are associations in memory between contextual cues and responses that are activated automatically when those cues are

encountered, independently of conscious intention. Subsequent work by Wood et al. (2021) confirmed that habitual responses engage regardless of concurrent goal states, and that two distinct features of habit memory — rapid activation and resistance to change — determine when habitual rather than goal-directed behavior controls action.

In algorithmically mediated preference contexts, the formation of a habit of acceptance represents a structural displacement of the deliberative process. Once the cue (a recommendation interface) reliably triggers the response (acceptance), the deliberative process that would otherwise evaluate the recommendation against genuine preference is bypassed. This is not laziness but the efficient operation of a cognitive system designed to automate frequently repeated behaviors. The irony is that the efficiency of habit formation — evolved for adaptive purposes — becomes the mechanism by which Behavioral Debt accumulates.

The speed at which these habits form in digital contexts is significant. Neal et al. (2012) demonstrated that habit strength is predicted by the consistency of the context-behavior association and the frequency of repetition. In the case of algorithmic recommendation interfaces, both conditions are maximally satisfied: the context (opening an app, scrolling a feed, being presented with a recommendation) is highly consistent, and the repetition frequency is extraordinarily high — potentially dozens or hundreds of interactions per day. Lally et al. (2010) found that habit formation follows an asymptotic curve, with automaticity increasing rapidly in the first few weeks of repeated

behavior. Applied to the algorithmic context, this suggests that the habit of accepting algorithmic recommendations may crystallize within weeks of regular platform use.

The resistance-to-change feature of habits documented by Wood et al. (2021) has particular significance for Behavioral Debt. Once the habit of acceptance is established, deliberate override requires effortful self-regulation — a limited cognitive resource (Baumeister et al., 1998). Users who wish to exercise autonomous preference formation must not only generate preferences but also actively override the habitual acceptance response, imposing a double cognitive burden that makes autonomous choice disproportionately costly relative to algorithmic acceptance.

2.7 Cognitive Offloading and Technology-Mediated Cognitive Change

Sparrow et al. (2011) demonstrated the "Google effect" on memory — that when people expect to have future access to information, they have lower rates of recall of the information itself and enhanced recall for where to access it. This finding established that technology can restructure the allocation of cognitive resources in predictable ways: when external storage is available, the brain offloads the storage function.

The cognitive offloading literature has expanded significantly since this foundational study. Gilbert (2024) developed a formal model of cognitive offloading as value-based decision-making, showing that the decision to offload is driven by the expected utility of internal versus external memory. Grinschgl et al. (2021) demonstrated empirically that cognitive offloading boosts immediate performance while diminishing subsequent memory for offloaded information — a finding with direct

relevance to Behavioral Debt. If preference-formation work is offloaded to recommendation algorithms, immediate choice performance may be maintained or improved while the underlying cognitive capacity for autonomous preference formation is diminished.

The neuroplasticity literature provides additional support. Wilmer et al. (2019) found associations between mobile technology engagement and differences in frontostriatal white matter connectivity, suggesting that chronic digital engagement may be associated with reward-system-dominant rather than executive-function-dominant cognitive architecture. Ward et al. (2017) demonstrated that the mere presence of one's own smartphone reduces available cognitive capacity, though subsequent meta-analytic review by Parry (2023) found mixed evidence for this specific "brain drain" effect. Camerini et al. (2021) conducted a scoping review of structural and functional brain correlates of screen time in adolescence, finding evidence that intensive screen-based media consumption is related to less efficient cognitive control systems.

2.8 Recommendation Systems and Their Effects

The technical literature on recommendation systems is vast. For present purposes, the most relevant findings concern the feedback loops and long-term effects that characterize recommendation-user interactions.

Chaney et al. (2018) demonstrated through simulation that algorithmic recommendation systems homogenize user behavior without increasing utility — specifically because the feedback loop between recommendations and behavioral data confounds the system's model of user preference. This finding is central to the Convergence Drift

component of the original working paper on Behavioral Debt. More recent empirical work by Coppolillo et al. (2024) modeled user-recommender system interactions across long-term scenarios and found measurable "bias-amplification deriving from the feedback loop between algorithmic suggestions and users' choices."

Jiang et al. (2019) provided a theoretical analysis distinguishing the echo chamber effect (driven by user dynamics) from the filter bubble effect (driven by algorithmic filtering), noting that the feedback loop between the two produces "degenerate" system behavior that reduces content diversity over time. Noordeh et al. (2020) simulated the recommendations given by collaborative filtering algorithms and found that "prolonged exposure to system-generated recommendations substantially decreases content diversity, moving individual users into echo-chambers characterized by a narrow range of content."

McLaughlin and Spiess (2024) formalized a related dynamic in their analysis of recommendation-dependent preferences, showing that recommendation systems alter preferences rather than merely predict them, and that decision-makers become "overly responsive to the recommendation" in ways that reduce decision quality over time. Haider et al. (2024) found empirically that algorithmic recommendations enable passive consumption practices rather than active search and evaluation, with users exhibiting "high trust" in recommendation systems while systematically "delegating search" to algorithmic intermediaries.

The research on dark patterns provides a complementary perspective. Gray et al. (2024) developed an ontology of dark patterns comprising 65 synthesized types across three levels of abstraction. The

dark patterns literature documents how interface design can manipulate user behavior through deceptive design elements — confirshaming, forced continuity, hidden costs, and other techniques that exploit cognitive biases (Mathur et al., 2019). While dark patterns represent intentional deception, the Behavioral Debt framework addresses a subtler phenomenon: the cognitive cost that accumulates even when recommendation systems function as designed, without deceptive intent, but with preference-substituting effect.

2.9 Digital Autonomy and the Ethics of Algorithmic Mediation

Susser et al. (2019) developed a rigorous philosophical account of online manipulation, arguing that it is the "covert subversion of another person's decision-making power" through information technology. Their analysis distinguishes manipulation from persuasion (which operates through transparent rational argument), coercion (which restricts options), and nudging (which alters choice architecture transparently). Online manipulation, they argue, threatens both the "competency" and "authenticity" conditions for genuine autonomy — the capacity to deliberate effectively and the capacity to act on one's own reasons.

Burr et al. (2018) provided a formal analysis of the interaction between intelligent software agents and human users, framing these interactions as instances in which an agent whose reward depends on actions performed by the user steers the user's behavior toward outcomes that maximize the agent's utility. Their analysis facilitates distinguishing subcases of interaction — deception, coercion, trading, and nudging — and identifies second-order effects including the possibility for adaptive

interfaces to induce behavioral addiction and change in user belief. Their central argument — that "the nature of the feedback commonly used by learning agents to update their models and subsequent decisions could steer the behaviour of human users away from what benefits them" — provides direct theoretical grounding for the Behavioral Debt framework.

The research on automation complacency, primarily from the human factors domain, provides additional relevant evidence. Parasuraman and Manzey (2010) documented how automation in high-stakes environments reduces human vigilance and produces over-reliance on system outputs, even when the system is demonstrably fallible. The pattern in consumer recommendation contexts operates along similar lines: algorithmic recommendations function as automation of preference expression, and the complacency dynamics documented in aviation and industrial settings may apply, in attenuated form, to the consumer domain.

---

3. Theoretical Framework: Behavioral Debt

3.1 Formal Definition

Behavioral Debt is defined as the accumulated cognitive cost incurred by an individual as a result of sustained interaction with systems that substitute algorithmically generated preferences for user-generated preferences. This cost manifests as a degradation of the individual's

capacity for autonomous preference formation, evaluative judgment, and self-directed choice — capacities that are collectively termed *preference capital*.

The concept is formally characterized by four properties:

1. **Cumulativity**: Each discrete substitution event imposes a small cost. The liability accumulates across interactions over time, producing a total cost significantly greater than any individual substitution would suggest. Formally, if c_i represents the cognitive cost of the *i*th substitution event, the total Behavioral Debt D at time t is:

$$D(t) = \sum_{i=1}^{n(t)} c_i \cdot w_i$$

where $n(t)$ is the total number of substitution events up to time t, and w_i is a weighting factor reflecting the salience and domain-specificity of the *i*th event (with higher weights for substitutions in high-importance domains).

2. **Latency**: The cost is not experienced at the point of incurrence. Users typically perceive individual algorithmic recommendations as conveniences. The debt is experienced, if at all, in degraded performance in preference-formation tasks undertaken in the absence of algorithmic assistance.

3. **Opacity**: The accumulation process is invisible to the individual debtor. There is no notification, no balance, no moment of reckoning that signals the growing liability. The most structurally significant feature of Behavioral Debt may be precisely this: debtors do not know they carry it.

4. **Compounding**: Like financial debt, Behavioral Debt compounds. As preference-formation capacity degrades, the user becomes increasingly dependent on algorithmic assistance, which further reduces the exercise of autonomous preference-formation processes, which further degrades capacity. Formally, this is modeled as a positive feedback loop:

$$\frac{dD}{dt} = \alpha \cdot R(t) + \beta \cdot D(t)$$

where $R(t)$ represents the rate of algorithmic interaction at time t, α is the per-interaction debt accumulation rate, and β is the compounding coefficient representing the degree to which existing debt increases vulnerability to further debt accumulation.

3.2 The Three Sub-Constructs

Behavioral Debt operates through three interacting sub-constructs, each addressing a distinct dimension of the phenomenon.

3.2.1 Preference Substitution (PS)

Definition. Preference Substitution is the systematic replacement of autonomous preference-formation processes with algorithmically generated alternatives optimized for platform-defined objectives. It describes the *mechanism* through which the algorithm's output displaces the user's own evaluative process.

Formal notation. Let $P_a(x, t)$ represent the autonomous preference of individual x at time t — the preference they would form through unmediated deliberation. Let $P_s(x, t)$ represent the

substituted preference — the preference that the individual expresses as a result of accepting an algorithmic recommendation. The degree of Preference Substitution at time t is defined as:

$$PS(x, t) = \frac{\sum_{d=1}^{D} |P_a(x, t, d) - P_s(x, t, d)|}{\sum_{d=1}^{D} \text{range}(d)}$$

where D is the number of preference domains (e.g., music, news, shopping, social connections) and range(d) is the range of possible preferences in domain d. This metric captures the normalized divergence between what the individual would choose and what the algorithm leads them to choose, aggregated across domains.

Preference Substitution is distinct from preference *influence*. All environments influence preferences. Preference Substitution describes a specific mechanism: the displacement of the cognitive process through which preferences are formed, not merely the alteration of the preferences themselves. The distinction is between an environment that shapes what you prefer (influence) and an environment that shapes whether you form a preference at all (substitution).

Theoretical grounding. The concept draws on McLaughlin and Spiess's (2024) formal analysis showing that recommendation systems produce "recommendation-dependent preferences" — preferences that would not exist without the recommendation — and that this dependence reduces decision quality. It also draws on Haider et al.'s (2024) empirical finding that algorithmic recommendations enable passive rather than active engagement with content, shifting users from evaluative mode to receptive mode.

3.2.2 Autonomy Erosion (AE)

Definition. Autonomy Erosion is the progressive degradation of an individual's capacity for independent evaluative choice through sustained disuse of deliberative cognitive processes. Where Preference Substitution describes what happens in the moment of interaction, Autonomy Erosion describes the cumulative consequence of many such moments: a reduction in the cognitive capacity that would be required to form preferences independently.

Formal notation. Let $A(x, t)$ represent the Autonomy level of individual x at time t, defined as their capacity to form and act on preferences in the absence of algorithmic mediation. Autonomy Erosion is modeled as a decay function:

$$A(x, t) = A_0 \cdot e^{-\lambda \int_0^t PS(x, \tau) \, d\tau} + \gamma \int_0^t U(x, \tau) \, d\tau$$

where A_0 is the individual's baseline autonomy capacity, λ is the decay rate associated with Preference Substitution, and the second term represents recovery: $U(x, \tau)$ is the rate of unmediated (autonomous) preference expression at time τ, and γ is the recovery coefficient. This formulation captures the key hypothesis that autonomy degrades with sustained algorithmic mediation but can be partially restored through deliberate exercise of unmediated preference formation.

Theoretical grounding. Autonomy Erosion is grounded in three bodies of evidence. First, the habit formation literature (Wood & Neal, 2007; Wood et al., 2021) establishes that once an automated response

pattern is established, the deliberative process that the habit displaces is no longer routinely engaged. Second, the cognitive offloading literature (Sparrow et al., 2011; Grinschgl et al., 2021) establishes that when cognitive work is delegated to external systems, the capacity for performing that work independently may diminish. Third, the automation complacency literature (Parasuraman & Manzey, 2010) establishes that sustained reliance on automated systems reduces human vigilance and the exercise of independent judgment.

The critical insight is that autonomous preference formation is a *skill-like* capacity. Like other cognitive skills — reading comprehension, mathematical reasoning, musical performance — it is maintained through exercise and attenuates through disuse. When recommendation systems consistently pre-answer the question "What do I want?", the cognitive routines that support answering that question independently may become underexercised and progressively less accessible.

3.2.3 The Attention Ledger (AL)

Definition. The Attention Ledger is a conceptual accounting mechanism that tracks the cumulative displacement of self-directed cognitive engagement by algorithmically captured attention. It represents the *opportunity cost* dimension of Behavioral Debt — the cognitive engagement that *could have been* directed toward autonomous preference formation, exploration, and deliberation but was instead captured and directed by algorithmic systems.

Formal notation. Let $T_{total}(x, t)$ represent the total attention budget of individual x during time period t. This budget is allocated between self-directed attention $T_{self}(x, t)$ and algorithmically directed attention $T_{algo}(x, t)$:

$$T_{total}(x, t) = T_{self}(x, t) + T_{algo}(x, t)$$

The Attention Ledger balance at time t is defined as the cumulative ratio of algorithmically directed to self-directed attention:

$$AL(x, t) = \int_0^t \frac{T_{algo}(x, \tau)}{T_{self}(x, \tau)} \, d\tau$$

An increasing Attention Ledger balance indicates a progressive shift from self-directed to algorithmically directed cognitive engagement — the attention-related dimension of Behavioral Debt accumulation. The Attention Ledger draws conceptually on the work of Fernández-Rovira et al. (2020) regarding the struggle for human attention in digital environments, and on the broader attention economy literature (Wu, 2016; Davenport & Beck, 2001) while shifting the unit of analysis from economic value to cognitive consequence.

Interaction dynamics. The three sub-constructs interact in a self-reinforcing cycle:

1. **Preference Substitution** displaces the cognitive work of preference formation.

2. **Autonomy Erosion** reduces the capacity to perform that work, making future algorithmic assistance more necessary.

3. The **Attention Ledger** accumulates as more cognitive engagement is directed by algorithms, reducing opportunities for the autonomous engagement that would maintain preference-formation capacity.

This cycle is depicted conceptually in Table 1.

Table 1

The Behavioral Debt Cycle: Interactions Among Sub-Constructs

Stage	Sub-Construct	Mechanism	Consequence
1	Preference Substitution	Algorithm provides pre-formed option; user accepts	Evaluative process displaced
2	Attention Ledger	Attention allocated to consuming algorithm's output rather than exploring independently	Reduced self-directed cognitive engagement
3	Autonomy Erosion	Deliberative preference-formation capacity atrophies through disuse	Higher dependence on algorithmic assistance
4	Feedback to Stage 1	Increased dependence raises acceptance rate for future substitutions	Cycle intensifies

3.3 Taxonomy of Behavioral Debt Accumulation Pathways

Behavioral Debt does not accumulate uniformly. The framework identifies four distinct pathways through which debt accumulates, varying in intensity and domain-specificity.

Pathway 1: Passive Accumulation. The most common pathway. Users accept algorithmic recommendations as a matter of convenience, without conscious awareness that preference-formation work is being displaced. Each individual acceptance event is inconsequential; the debt

accumulates across thousands of such events. This pathway characterizes routine interactions with music streaming, social media feeds, and e-commerce recommendation engines.

Pathway 2: Comfort-Driven Accumulation. Users are aware that algorithmic assistance is shaping their choices but actively prefer the reduced cognitive load. The choice overload literature (Iyengar & Lepper, 2000; Schwartz, 2004) provides the motivational substrate: in environments of overwhelming choice, algorithmic curation is experienced as relief rather than substitution. Users may describe themselves as "letting the algorithm decide" with a sense of welcome delegation.

Pathway 3: Lock-In Accumulation. As Behavioral Debt increases, the cost of exercising autonomous preference rises (because the cognitive capacity has diminished), while the cost of accepting algorithmic recommendations remains constant. This creates a lock-in dynamic analogous to Rakova and Chowdhury's (2019) "barrier-to-exit" concept: the accumulated debt makes it progressively more difficult and effortful for the user to override algorithmic recommendations, even when they might prefer to.

Pathway 4: Social Accumulation. When algorithmically shaped preferences become the basis for social interaction — when shared playlists, recommended articles, and curated feeds form the substance of social exchange — departing from algorithmic recommendations carries a social cost. Users whose preferences diverge from the algorithmically shaped consensus may experience social friction, creating an additional incentive to accept substituted preferences.

3.4 Differentiation from Adjacent Frameworks

The Behavioral Debt framework is related to but distinct from several established concepts. Table 2 provides a structured comparison.

Table 2

Behavioral Debt Compared with Adjacent Theoretical Frameworks

Framework	Unit of Analysis	Primary Concern	Temporal Focus	User Model
Filter Bubble (Pariser, 2011)	Information environment	Content narrowing	Present state	Passive recipient
Surveillance Capitalism (Zuboff, 2019)	Political economy	Data extraction	Ongoing extraction	Data source
Attention Economy (Wu, 2016)	Economic transaction	Attention capture	Moment of capture	Attention resource
Nudge Theory (Thaler & Sunstein, 2008)	Choice moment	Behavioral outcome	Moment of choice	Boundedly rational agent
Dark Patterns (Gray et al., 2024)	Interface design	Deceptive manipulation	Interaction event	Exploited user
Behavioral Debt	Cognitive architecture	Capacity degradation	Cumulative over time	Adaptive cognitive system

The critical differentiating feature of Behavioral Debt is its explicitly *temporal* and *cumulative* character. Each individual interaction with a recommendation system may be inconsequential. The debt accumulates in the aggregate, across thousands of interactions, over months and years, largely below the threshold of user awareness. This temporal structure — the gap between the point of incurrence and the point of

recognition — is precisely what makes Behavioral Debt both consequential and resistant to individual mitigation. Users cannot opt out of debt they cannot perceive. This is the fundamental consent problem that distinguishes Behavioral Debt from the harms addressed by existing frameworks.

Relationship to the Filter Bubble. The filter bubble describes what the user is *shown*; Behavioral Debt describes what happens to the user's cognitive architecture as a result of sustained algorithmic mediation. A user could escape their filter bubble — by deliberately seeking diverse sources — and still carry Behavioral Debt, if their capacity for autonomous preference formation has atrophied. Conversely, a user could remain within a filter bubble but maintain low Behavioral Debt if they actively engage in preference formation even within the constrained information environment.

Relationship to Surveillance Capitalism. Zuboff's framework describes the economic and political logic by which behavioral data is extracted from users. Behavioral Debt describes the cognitive residue left within users by the extraction apparatus. The two frameworks are complementary: surveillance capitalism provides the systemic explanation for *why* preference substitution occurs at scale, and Behavioral Debt provides the individual-level account of *what accumulates* as a result.

Relationship to Dark Patterns. Dark patterns involve intentional deception in interface design (Gray et al., 2024; Mathur et al., 2019). Behavioral Debt accumulates even in the absence of deceptive intent — it is a consequence of the *structural features* of recommendation systems,

not of malicious design choices. A well-intentioned recommendation system that genuinely helps users find content they enjoy can still contribute to Behavioral Debt by displacing autonomous preference formation.

3.5 The Debt Accumulation Model: An Integrated Formulation

The three sub-constructs can be integrated into a unified model of Behavioral Debt accumulation. Define the Behavioral Debt state vector for individual x at time t as:

$$\mathbf{BD}(x, t) = \langle PS(x, t), AE(x, t), AL(x, t) \rangle$$

The total Behavioral Debt is a weighted composite:

$$D_{total}(x, t) = w_1 \cdot PS(x, t) + w_2 \cdot AE(x, t) + w_3 \cdot AL(x, t)$$

where w_1, w_2, and w_3 are weights reflecting the relative contribution of each sub-construct to overall cognitive cost. These weights are hypothesized to vary by individual (reflecting differences in cognitive style, baseline autonomy, and susceptibility to habit formation) and by domain (reflecting differences in the stakes and complexity of preference formation across domains).

The dynamics of the system are governed by a set of coupled differential equations:

$$\frac{dPS}{dt} = f(R, A, \text{interface design})$$

$$\frac{dAE}{dt} = -\lambda \cdot PS + \gamma \cdot U$$

$$\frac{dAL}{dt} = g(T_{algo}, T_{self})$$

where R is the algorithmic recommendation rate, A is current autonomy level, U is the rate of unmediated preference expression, and f and g are functions whose specific forms would be determined through empirical calibration. The key structural feature of this system is its positive feedback loop: increased PS drives decreased A (via the AE equation), which in turn increases susceptibility to future PS (via the PS equation's dependence on A).

Figure 1 (described). A conceptual diagram of the Behavioral Debt accumulation model. Three boxes represent the sub-constructs (Preference Substitution, Autonomy Erosion, Attention Ledger), connected by arrows indicating causal relationships. An outer loop shows the positive feedback mechanism: PS → AE → increased susceptibility to PS. Environmental inputs (algorithmic recommendation rate, interface design, domain complexity) feed into the PS and AL boxes. Individual moderators (baseline autonomy, need for cognition, digital literacy) are shown as moderating the relationships between sub-constructs. The diagram illustrates the system's tendency toward escalation in the absence of countervailing inputs (unmediated preference exercise, digital literacy interventions).

---

4. Methodology: Proposed Empirical Study Design

4.1 Overview and Rationale

The Behavioral Debt framework is a theoretical contribution that requires empirical validation. This section proposes a mixed-methods study design suitable for testing the framework's core propositions. The design integrates three components: a quantitative survey and behavioral assessment component, a controlled experimental component, and a qualitative interview component. The mixed-methods approach is appropriate given the framework's complexity, the latent nature of the construct, and the need to capture both behavioral indicators and subjective experience.

4.2 Research Hypotheses

The following hypotheses derive from the theoretical framework:

H1: Individuals with higher cumulative exposure to algorithmic recommendation systems will show lower scores on measures of autonomous preference-formation capacity than individuals with lower cumulative exposure, controlling for relevant individual difference variables.

H2: In unmediated choice environments (environments in which no algorithmic recommendation is present), individuals with higher Behavioral Debt will exhibit (a) longer decision latency, (b) lower choice confidence, (c) reduced preference diversity, and (d) higher rates of choice deferral, compared to individuals with lower Behavioral Debt.

H3: The relationship between algorithmic exposure and preference-formation capacity will be mediated by the three sub-constructs: Preference Substitution, Autonomy Erosion, and Attention Ledger balance.

H4: Individual difference variables — specifically need for cognition (Cacioppo & Petty, 1982), maximizing tendency (Schwartz et al., 2002), and digital literacy — will moderate the rate of Behavioral Debt accumulation.

H5: Periods of reduced algorithmic exposure ("algorithmic abstention") will produce measurable recovery in preference-formation indicators, with the rate of recovery moderated by the duration and depth of prior algorithmic exposure.

4.3 Quantitative Component: Survey and Behavioral Assessment

4.3.1 Participants

A target sample of \(N\) = 600 adults (18-65), recruited to ensure adequate representation across age groups, educational levels, and levels of technology use. Stratified quota sampling will target approximately equal distribution across three algorithmic exposure levels (low, moderate, high) as determined by a screening instrument.

4.3.2 Algorithmic Exposure Assessment

Participants will complete a custom instrument, the **Algorithmic Interaction Inventory (AII)**, assessing:

- Platform usage patterns across recommendation-intensive platforms (social media, streaming, e-commerce, news aggregation)
- Daily time spent in algorithmically mediated environments
- Self-reported reliance on algorithmic recommendations across domains
- Frequency of algorithmic override behavior (selecting options other than those recommended)

The AII will be supplemented, where participants consent, by objective usage data collected via a browser extension over a two-week observation period.

4.3.3 Behavioral Debt Indicators

The following behavioral and psychometric measures will operationalize the three sub-constructs:

Preference Substitution Indicators:

- *Preference Diversity Index (PDI):* A measure of the range and heterogeneity of preferences expressed in unmediated choice tasks. Participants will complete preference selection tasks in music, news, and food domains both with and without algorithmic recommendations. The divergence between mediated and unmediated selections provides a measure of substitution.
- *Algorithmic Override Rate (AOR):* The frequency with which participants actively select options other than those

recommended, measured during a standardized recommendation interaction task.

Autonomy Erosion Indicators:

- *Decision Latency in Unmediated Environments (DLUE):* Response time in preference-formation tasks completed without algorithmic assistance, measured in seconds.
- *Self-Reported Choice Confidence Scale (SRCCS):* A Likert-scale instrument measuring participants' subjective confidence in their ability to identify and act on their own preferences. Items include: "I find it difficult to choose what to watch/listen to without recommendations," "When I have to make choices without algorithmic help, I feel uncertain," and "I trust my own preferences more than algorithmic recommendations."
- *Choice Deferral Rate (CDR):* The proportion of unmediated choice tasks in which participants defer or abandon the choice rather than selecting an option.

Attention Ledger Indicators:

- *Self-Directed Exploration Time (SDET):* The proportion of digital engagement time spent in self-directed browsing or search versus consuming algorithmically recommended content, measured via the browser extension.
- *Attention Allocation Diary:* A brief daily diary completed over the observation period, recording instances of self-directed versus algorithmically directed content engagement.

4.3.4 Moderating Variables

- *Need for Cognition Scale* (Cacioppo & Petty, 1982): 18-item validated instrument.

- *Maximization Scale* (Schwartz et al., 2002): 13-item validated instrument.
- *Digital Literacy Scale:* Items drawn from existing validated measures of digital media literacy.
- *Demographic variables:* Age, gender, educational attainment, and technology use history.

4.4 Experimental Component

4.4.1 Design

A pretest-posttest control group design with random assignment to one of three conditions:

1. **High Algorithmic Assistance (HA):** Participants receive personalized algorithmic recommendations for all choice tasks during a two-week intervention period.

2. **Low Algorithmic Assistance (LA):** Participants receive randomized (non-personalized) content suggestions during the intervention period.

3. **No Algorithmic Assistance (NA):** Participants navigate choice tasks without any algorithmic recommendations during the intervention period.

4.4.2 Measures

Pre-intervention and post-intervention assessment of:

- Decision latency in unmediated choice tasks
- Preference diversity in unmediated selections
- Self-reported choice confidence

- Choice deferral rate

Change scores between pre- and post-intervention assessments will serve as the primary outcome variables.

4.4.3 Follow-Up Assessment

A follow-up assessment at four weeks post-intervention will test H5 (recovery from Behavioral Debt following reduced algorithmic exposure) and will be administered to all conditions.

4.5 Qualitative Component: Semi-Structured Interviews

4.5.1 Sample

A purposive subsample of 30 participants drawn from the quantitative sample, selected to represent the full range of Behavioral Debt indicator scores.

4.5.2 Interview Protocol

Semi-structured interviews (45-60 minutes) exploring:

1. **Phenomenology of algorithmic mediation:** How do participants experience and describe their interactions with recommendation systems? What language do they use to describe the role of algorithms in their preference formation?

2. **Awareness of preference change:** To what extent are participants aware of changes in their preference-formation processes? Can they identify moments or periods in which their reliance on algorithmic recommendations increased? What prompted those changes?

3. **Unmediated choice experience:** How do participants describe the experience of making preference-based choices without algorithmic assistance? Do they experience difficulty, uncertainty, or discomfort? In what domains?

4. **Agency and identity:** Do participants perceive a difference between preferences they have "chosen" and preferences they have "been given"? How do they relate their algorithmically mediated preferences to their sense of identity and self-knowledge?

5. **Strategies and resistance:** Have participants developed strategies for maintaining autonomous preference formation? If so, what motivated those strategies and how effective do they perceive them to be?

4.5.3 Analysis

Interview data will be analyzed using reflexive thematic analysis (Braun & Clarke, 2006), with initial coding informed by the theoretical framework but remaining open to emergent themes not anticipated by the theory.

4.6 Ethical Considerations

The study raises specific ethical considerations that must be addressed:

- **Informed consent:** Participants must be informed that the study investigates the effects of algorithmic recommendation on their preference formation, though the specific hypotheses will not be disclosed to avoid demand effects.
- **Harm minimization:** The experimental conditions involve either increased or reduced algorithmic exposure, neither of

which exceeds normal variation in everyday technology use. No condition involves exposure to harmful content.

- **Debrief and debtor notification:** Following study completion, all participants will receive a comprehensive debrief that includes an explanation of the Behavioral Debt framework and practical strategies for maintaining autonomous preference formation.

Figure 2 (described). A flowchart depicting the mixed-methods study design. The diagram shows three parallel tracks: (1) Quantitative survey and behavioral assessment (N = 600), (2) Experimental manipulation with three conditions (HA, LA, NA), and (3) Qualitative interviews (n = 30). All tracks converge at the analysis stage, with cross-referencing arrows indicating integration points between quantitative and qualitative findings. Timeline annotations show a 2-week baseline, 2-week intervention, immediate post-intervention assessment, and 4-week follow-up.

---

5. Analysis and Discussion: Cross-Domain Application

5.1 Conceptual Analysis Approach

The Behavioral Debt framework was developed as a domain-general theory applicable to any context in which algorithmic recommendation systems mediate preference expression. To demonstrate its explanatory scope and generate testable predictions, this section applies the

framework through conceptual analysis across three prominent domains of algorithmic recommendation: social media feeds, streaming entertainment, and e-commerce.

5.2 Case Study 1: Social Media Feeds

Social media feeds — particularly those of platforms such as Instagram, TikTok, and the former Twitter/X — represent perhaps the most intensive site of Behavioral Debt accumulation due to the frequency of interaction, the breadth of domains mediated (social connection, news, entertainment, commercial engagement), and the sophistication of the optimization algorithms employed.

Preference Substitution in social media. The algorithmic feed determines not only what content the user sees but the *order* in which they see it, the *proportion* of content from different sources, and the *type* of content that is emphasized. As Allcott et al. (2020) demonstrated in their landmark randomized experiment on Facebook deactivation, users who deactivated Facebook for four weeks reported increased subjective well-being and showed a persistent reduction in post-experiment Facebook use — suggesting that the algorithmically mediated experience had been substituting for, rather than enhancing, activities that users valued more. Notably, deactivation reduced post-experiment valuations of Facebook itself, suggesting that sustained exposure inflated users' perceived value of the platform.

In terms of the Behavioral Debt framework, social media feeds generate high rates of Preference Substitution because (a) the feed is algorithmically determined, with the user exercising virtually no control

over what appears; (b) the refresh rate is rapid, creating a continuous stream of substitution events; and (c) the engagement optimization targets attention capture rather than preference satisfaction, meaning the content presented may diverge substantially from what the user would seek independently.

Autonomy Erosion in social media. The qualitative research literature suggests that social media users often describe difficulty articulating what they "actually want" from their social media experience, as distinct from what the algorithm provides. This difficulty may represent a specific manifestation of Autonomy Erosion: the preference-formation routines that would normally generate a clear sense of informational and social needs have been displaced by the habit of scrolling through algorithmically curated content.

Attention Ledger effects. Social media platforms are designed to maximize time-on-platform — the Attention Ledger balance is, by design, shifted toward algorithmically directed attention. The research on variable reward schedules (Alter, 2017) and intermittent reinforcement explains the mechanism: the feed intermittently delivers highly engaging content amid a background of moderate content, sustaining engagement through uncertainty about when the next "reward" will arrive.

5.3 Case Study 2: Streaming Entertainment Recommendations

Music and video streaming platforms — Spotify, Netflix, YouTube — provide a particularly clear demonstration of the Behavioral Debt mechanism because they involve sustained, long-term preference mediation in domains closely tied to identity and self-expression.

Preference Substitution in streaming. The person who has had their music chosen for them by an algorithm for two years may find, when asked to select music without algorithmic assistance, that the process is unexpectedly difficult — not because their taste has changed in a way they endorse, but because the cognitive routines that once supported musical preference formation have been displaced by the habit of algorithmic delegation. Romero Meza and D'Urso (2024) documented precisely this phenomenon in their qualitative study of Netflix users, finding that recommendation systems produce a "user's dilemma" characterized by high reliance on recommendations coupled with frustration when recommendations fail to satisfy. Users exhibited prolonged search times, heightened choice effort, and moderate satisfaction levels — paradoxically depending on a system that frequently disappointed them.

Convergence Drift in streaming. Chaney et al.'s (2018) simulation demonstrated that recommendation systems homogenize user behavior across a population, and Noordeh et al. (2020) confirmed that prolonged exposure to collaborative filtering substantially decreases content diversity. In the streaming context, this means that individual musical or viewing preferences, which might otherwise reflect the full range of human cultural diversity, are progressively shaped toward the attractors identified by the recommendation algorithm — typically content that maximizes engagement across the largest user segments.

5.4 Case Study 3: E-Commerce Recommendations

E-commerce recommendation systems operate in a domain where preference expression has direct economic consequences — purchasing decisions — making the stakes of Behavioral Debt accumulation materially significant.

Preference Substitution in e-commerce. When an e-commerce platform recommends products, the recommendation functions as a highly salient default that exploits the choice architecture dynamics documented by Thaler and Sunstein (2008). The user who encounters a "recommended for you" section may accept the recommendation not because it matches their genuine preference but because the cognitive effort of searching independently exceeds the perceived benefit — particularly in the context of the choice overload that characterizes large-scale e-commerce environments (Iyengar & Lepper, 2000).

The commercial amplification of Behavioral Debt. E-commerce platforms have a direct financial incentive to maximize acceptance of recommendations, particularly when those recommendations are influenced by supplier advertising or margin optimization. This creates a misalignment between the platform's objective (maximizing revenue) and the user's objective (satisfying preferences), with Behavioral Debt serving as the mechanism that resolves the misalignment in the platform's favor: users whose autonomous preference-formation capacity has been eroded are more likely to accept recommendations that serve the platform's economic interests.

5.5 Case Study 4: News and Information Recommendation

While the previous three case studies address preference domains where the stakes are primarily personal (entertainment, purchasing, social connection), algorithmic recommendation of news and information operates in a domain with significant public consequences. News recommender systems — as deployed on platforms including Google News, Apple News, Facebook, and aggregation services — determine which information citizens encounter, shaping not only individual beliefs but collective public discourse.

Preference Substitution in news. When algorithmic systems determine which news stories a user encounters, they substitute the user's autonomous information-seeking behavior with algorithmically determined content selection. Research by Eslami et al. (2015) found that 62.5% of Facebook users in their study were unaware that the platform's algorithm curated their News Feed at all — they believed they were seeing everything their contacts posted. This lack of awareness means that the Preference Substitution occurring in news consumption is frequently invisible to the user, a condition that maximizes debt accumulation.

Bail et al. (2018) provided experimental evidence that algorithmically filtered political content on social media can produce unexpected effects on attitudes — in their study, Republicans who followed a liberal Twitter bot became substantially more conservative in their views. This finding underscores that algorithmically mediated information exposure does not produce predictable or easily modeled cognitive outcomes. The Behavioral Debt framework suggests an additional dimension: beyond the attitudinal effects of specific content,

the sustained delegation of information selection to algorithmic systems may erode the user's capacity for independent information evaluation and news judgment.

Autonomy Erosion in information consumption. The concept of "information diet" — the idea that individuals should consciously curate their information intake — presupposes a capacity for autonomous information-seeking that Behavioral Debt may progressively undermine. When users become habituated to consuming algorithmically curated news feeds, the cognitive routines involved in independent news-seeking — identifying reliable sources, evaluating credibility, seeking multiple perspectives, and forming independent judgments about newsworthiness — may atrophy. Dogruel et al. (2020) found that internet users' awareness of algorithms was closely related to their perceived autonomy: when users felt in control of their interactions online, they were paradoxically less aware of the algorithms governing those interactions, suggesting that algorithmic mediation becomes most invisible precisely when it is most pervasive.

Democratic implications. The Behavioral Debt framework adds a dimension to existing concerns about algorithmic effects on democratic discourse. Beyond the well-documented concerns about filter bubbles (Pariser, 2011) and political polarization (Bail et al., 2018), Behavioral Debt identifies a structural threat: if citizens' capacity for autonomous information evaluation degrades through sustained algorithmic news curation, the cognitive infrastructure required for democratic participation — the ability to independently assess competing claims, evaluate evidence, and form reasoned political judgments — may itself be

weakened. This is not a claim about specific political outcomes (as in the filter bubble hypothesis) but about the meta-cognitive capacity that underlies all political reasoning.

5.6 Variable Reward Schedules and Engagement Engineering

The role of variable reward schedules in sustaining engagement with recommendation systems warrants dedicated analysis in the context of Behavioral Debt. Skinner's foundational work on variable-ratio reinforcement schedules, applied to digital platform design (see Alter, 2017, for an accessible synthesis), identified that variable and unpredictable reward delivery produces more robust and extinction-resistant behavioral engagement than predictable schedules.

Recommendation systems that intermittently surface highly engaging content amid a background of moderate content are, in effect, deploying variable-ratio reinforcement. The relevance to Behavioral Debt is that variable reward schedules not only sustain engagement; they also produce compulsive checking behaviors and attentional capture that crowd out the deliberative space in which autonomous preference formation would otherwise occur. The user who compulsively refreshes a social media feed is not engaging in preference formation — they are responding to a reinforcement schedule that maintains behavioral engagement independently of preference satisfaction.

Eyal's (2014) "Hook Model" explicitly codifies the mechanisms through which digital products create habit-forming user experiences: trigger, action, variable reward, and investment. Each cycle of this loop, from the Behavioral Debt perspective, represents a unit of preference

substitution: the trigger captures attention, the action directs it toward algorithmically selected content, the variable reward reinforces the behavior, and the investment deepens the commitment to the platform's recommendation ecosystem. The cumulative effect of thousands of such cycles constitutes Behavioral Debt accumulation.

The Behavioral Debt framework does not claim that all engagement with recommendation systems is harmful or that all algorithmic mediation constitutes preference substitution. The framework identifies the *conditions* under which debt accumulates: sustained algorithmic mediation, absent deliberative engagement with the mediation process, over extended periods. Occasional use of recommendations as convenient starting points for preference exploration generates minimal debt. Sustained, habitual, unexamined reliance on algorithmic preference substitution across multiple life domains generates substantial debt.

5.7 Cross-Domain Comparison

Table 3

Behavioral Debt Characteristics Across Three Application Domains

Characteristic	Social Media	Streaming	E-Commerce
Interaction frequency	Very high (continuous)	High (daily)	Moderate (episodic)
PS intensity	High	Moderate-High	Moderate
AE vulnerability	High	High	Moderate
AL accumulation rate	Very high	High	Moderate
Domain stakes	Social, informational	Cultural, identity	Economic

Characteristic	Social Media	Streaming	E-Commerce
Feedback loop strength	Very strong	Strong	Moderate
User awareness	Low	Low-Moderate	Moderate
Recovery potential	Moderate	Moderate-High	High

The comparison reveals that social media feeds generate the highest rates of Behavioral Debt accumulation due to their combination of high interaction frequency, broad domain coverage, and low user awareness. Streaming platforms generate high debt in specific preference domains (music, video) with significant identity implications. E-commerce generates more episodic debt accumulation but with direct economic consequences.

Figure 3 (described). A three-panel comparison chart showing the relative intensity of each Behavioral Debt sub-construct (PS, AE, AL) across the three application domains. Each panel uses a radar/spider chart with three axes representing PS, AE, and AL intensity on a 1-5 scale. Social media shows the largest overall area (highest combined debt), streaming shows a moderate area concentrated on the PS and AE axes, and e-commerce shows the smallest area but with notable PS intensity.

---

6. Implications

6.1 Implications for Platform Design

If Behavioral Debt is real and measurable, platform designers face a choice analogous to that faced by industrial producers regarding environmental externalities. They can continue to optimize for engagement metrics that maximize short-term platform value while the cognitive costs of preference substitution accumulate in users — invisible, uncompensated, and uncounted. Or they can treat Behavioral Debt as a design variable — a cost that competent and responsible design seeks to minimize.

Environmental economics provides an instructive analogy. Industrial processes that externalize pollution costs — allowing the costs of contamination to be borne by the public rather than the producer — were, for much of industrial history, treated as the natural order of production. Regulatory frameworks emerged to require producers to internalize those costs. The analogy to algorithmic preference substitution is imperfect but instructive: recommendation systems that erode the preference-formation capacity of users are externalizing a cognitive cost that is currently borne entirely by the user.

Specific design interventions that might reduce Behavioral Debt accumulation include:

1. **Preference Articulation Prompts:** Systems that explicitly prompt users to articulate preferences *before* presenting recommendations, requiring the engagement of deliberative preference-formation processes.

2. **Diversified Recommendation Formats:** Interfaces that present multiple divergent options rather than a single "best" recommendation, preserving evaluative choice even within the recommendation context.

3. **Algorithmic Abstention Modes:** Periodic modes in which the recommendation system deliberately withholds recommendations, inviting users to navigate without algorithmic assistance. This design pattern is analogous to the concept of "training without assistance" in exercise science — periods in which the scaffolding is deliberately removed to strengthen underlying capacity.

4. **Transparency Disclosures:** Clear indication of the degree to which a recommendation reflects individual preference data versus population-level engagement optimization. This disclosure would enable users to make informed decisions about which recommendations to accept and which to override.

5. **Debt Awareness Dashboards:** User-facing analytics that track indicators of Behavioral Debt accumulation — algorithmic reliance rates, preference diversity trends, and override frequency — providing the metacognitive information that opacity currently conceals.

None of these interventions is technically infeasible. What has been lacking is the conceptual framework that identifies Behavioral Debt as a design outcome worth minimizing.

6.2 Implications for Regulatory Policy

The Behavioral Debt framework suggests that current regulatory approaches to algorithmic systems, which focus primarily on data privacy (GDPR), content moderation (Digital Services Act), and

discriminatory outcomes (various AI fairness regulations), may be incomplete. These regulatory frameworks address important concerns but do not address the cognitive externalities of sustained algorithmic preference substitution.

The consent problem is particularly acute. Users of recommendation systems nominally consent to the use of their data and to algorithmically mediated content delivery. They do not — because they cannot — consent to the cognitive consequences of sustained algorithmic preference substitution. Behavioral Debt, as defined in this framework, is not a consequence of any individual interaction to which consent could meaningfully be given. It is the consequence of the *pattern* of interactions over time — a consequence that neither the user nor the platform may be tracking. Informed consent to preference substitution at scale would require disclosure not just of what data is collected, but of what cognitive capacity may be eroded. Current disclosure frameworks do not approach this standard.

Potential regulatory directions include:

1. **Cognitive Impact Assessments:** Requirements for platforms to conduct and disclose assessments of the potential cognitive effects of their recommendation systems, analogous to environmental impact assessments.

2. **Algorithmic Exposure Limits:** Regulatory exploration of whether exposure limits — analogous to occupational exposure limits for workplace hazards — might be appropriate for intensive algorithmic recommendation environments, particularly for children and adolescents.

3. **Right to Unmediated Access:** A user right to access platform content without algorithmic recommendation — the ability to opt out of personalization without losing access to the platform's content or functionality.

4. **Audit Requirements:** Requirements for independent audit of recommendation systems' effects on preference diversity, decision independence, and related indicators of Behavioral Debt.

6.3 Implications for Education and Digital Literacy

Digital literacy frameworks have historically emphasized skills such as information evaluation, privacy management, and critical media consumption. If Behavioral Debt is a genuine phenomenon, a new competency deserves inclusion: *debt awareness* — the ability to recognize, monitor, and counteract the accumulation of one's own algorithmic preference substitution.

Debt awareness as a pedagogical objective would include:

- **Recognition skills:** The ability to identify when one is interacting with an algorithmically mediated environment and when a presented option is a recommendation rather than the result of one's own search.
- **Metacognitive monitoring:** The practice of periodically checking whether one's choices reflect genuine preferences or accepted defaults — asking "Do I actually want this, or am I accepting what was offered?"
- **Autonomous preference practice:** The deliberate cultivation of preference-formation habits in low-stakes domains as training for higher-stakes contexts. This might include practices such as selecting music without algorithmic

assistance, choosing what to read by browsing rather than following recommendations, or navigating to specific products rather than accepting curated suggestions.

- **Critical evaluation of convenience:** Understanding that the subjective experience of convenience — the relief of having choices made easier — may coexist with the objective consequence of reduced preference-formation capacity.

6.4 Implications for Individual Practice

For individuals concerned about Behavioral Debt, the framework suggests several practical strategies:

1. **Periodic algorithmic fasting:** Deliberately interacting with digital platforms without algorithmic assistance for defined periods — using chronological feeds, browsing without recommendations, and navigating content libraries without algorithmic curation. The experimental evidence from Allcott et al. (2020) suggesting that Facebook deactivation produced persistent changes in post-experiment behavior indicates that even temporary disruption of algorithmically mediated habits can produce lasting effects on usage patterns.

2. **Preference journaling:** Maintaining a record of preferences formed through deliberation rather than recommendation acceptance, in order to maintain awareness of the distinction. This practice serves a metacognitive function: it preserves awareness of the difference between autonomous and substituted preferences, which opacity otherwise erodes.

3. **Override practice:** Deliberately selecting options other than those recommended, even when the recommendation seems adequate, in order to maintain the evaluative processes that override requires. This is the cognitive equivalent of physical exercise — the practice itself is the point, not the specific outcome of any individual override.

4. **Domain diversification:** Ensuring that at least some preference domains remain relatively unmediated — maintaining areas of life in which preferences are formed through direct experience, social recommendation, or independent exploration rather than algorithmic curation.

5. **Mindful consumption audits:** Periodically reviewing one's digital consumption patterns to identify domains in which algorithmic mediation has become the default mode of engagement. The Attention Ledger concept provides a framework for this audit: tracking the ratio of self-directed to algorithmically directed engagement across platforms and domains.

These strategies are not technophobic prescriptions. They are recognition that cognitive capacities require exercise and that environments engineered to displace that exercise serve the interests of the environment's designers, not the interests of the person within it.

6.5 Counterarguments and Responses

A rigorous theoretical framework must address the strongest counterarguments to its central claims. Several important objections to the Behavioral Debt framework warrant consideration.

Counterargument 1: Recommendation as cognitive enhancement, not substitution. One might argue that algorithmic recommendation systems enhance rather than substitute for preference formation — that they expose users to options they would never have discovered independently, thereby expanding rather than constraining the preference space. This objection has empirical support: recommendation systems do introduce users to novel content, and satisfaction with recommended content is often high.

Response: The Behavioral Debt framework does not deny the benefits of recommendation. It argues that these benefits coexist with a cognitive cost that current frameworks do not account for. The relevant question is not whether recommendations are useful in any individual instance but whether the *cumulative pattern* of reliance on recommendations degrades the capacity for autonomous preference formation. An analogy: a calculator enhances mathematical performance in any individual calculation, but a student who uses a calculator for all arithmetic from age six may develop less fluent mental arithmetic than one who practices without assistance. The enhancement and the cost can coexist.

Counterargument 2: Preferences were never autonomous. One might argue that human preferences have always been socially and environmentally influenced, and that algorithmic recommendation is simply a new form of an ancient process — no different in kind from peer influence, advertising, or cultural context.

Response: The framework acknowledges that preferences are always contextually situated. The claim is not that preferences were previously formed in a vacuum but that there is a qualitative difference between environmental influence (which leaves the deliberative process intact) and preference substitution (which displaces the deliberative process). When a friend recommends a restaurant, the recommendation is processed through the user's evaluative system — weighed against knowledge of the friend's taste, one's own past experiences, current mood, and so on. When an algorithm presents a restaurant recommendation, the evaluative process may be bypassed entirely, particularly when the habit of acceptance has been established. The difference is not in the existence of external influence but in the extent to which the user's deliberative system is engaged.

Counterargument 3: Individual variation makes the framework untestable. Given that individuals vary enormously in their susceptibility to Behavioral Debt, one might argue that the framework makes no strong general predictions and is therefore unfalsifiable.

Response: Individual variation is a feature of virtually all psychological frameworks, not a disqualification. The framework generates clear, testable predictions: that higher algorithmic exposure should predict lower preference-formation capacity (controlling for confounds); that experimental manipulation of algorithmic exposure should produce measurable changes in preference-formation indicators; and that individual difference variables should moderate these effects in

predictable directions. The proposed empirical study design in Section 4 operationalizes these predictions with specific measures and statistical tests.

Counterargument 4: Natural cognitive offloading is adaptive. One might argue that cognitive offloading to technology is an adaptive response — an extension of the natural human tendency to use external tools to augment cognitive capacity, consistent with the "extended mind" thesis (Clark & Chalmers, 1998).

Response: This is the most philosophically sophisticated objection and deserves careful engagement. The extended mind thesis argues that cognitive processes extend beyond the brain into the environment, and that tools that reliably augment cognition become part of the cognitive system. If recommendation algorithms are part of the extended cognitive system, then "preference substitution" may be reconceived as "distributed preference formation" — not a loss but a redistribution of cognitive function. The Behavioral Debt framework's response is empirical rather than philosophical: if the extended mind thesis applies, then removal of the algorithmic tool should produce disruption analogous to the loss of any other cognitive resource, but *not* lasting degradation. If, instead, sustained use of algorithmic tools produces measurable and persistent reduction in preference-formation capacity — even after extended periods of non-use — this would suggest atrophy rather than distribution. The reversibility studies proposed in Section 4 are designed to adjudicate this question.

Figure 4 (described). A conceptual diagram illustrating the key counterarguments and the framework's responses. Four boxes on the left represent the counterarguments (Enhancement, Non-autonomy, Untestability, Adaptive Offloading). Arrows connect each to a response box on the right, with a central column indicating the type of evidence that would resolve each dispute (empirical, philosophical, methodological). The diagram emphasizes that the Behavioral Debt framework generates falsifiable predictions that distinguish it from the counterarguments.

Figure 5 (described). A timeline diagram showing the hypothesized trajectory of Behavioral Debt accumulation for three user profiles: (1) a heavy user with no mitigation practices (exponential accumulation), (2) a moderate user with periodic algorithmic fasting (sawtooth pattern with partial recovery), and (3) a mindful user with regular override practice and domain diversification (slow linear accumulation that plateaus). The x-axis represents time in months (0-36), and the y-axis represents total Behavioral Debt (arbitrary units). The diagram illustrates the framework's prediction that Behavioral Debt is not inevitable but depends on the interaction between exposure patterns and mitigation practices.

---

7. Limitations and Future Research

7.1 Limitations of the Present Framework

The Behavioral Debt framework, as presented in this dissertation, carries several significant limitations that must be acknowledged.

Theoretical status. The framework is a theoretical contribution. Its components are grounded in adjacent literatures — cognitive load theory, habit formation research, cognitive offloading, automation complacency — but have not been tested as a unified framework in empirical research. The mathematical formulations presented are intended to provide precision and testability, not to claim measurement accuracy in the absence of empirical calibration. The framework's value at this stage is conceptual: it identifies a phenomenon, provides a vocabulary for discussing it, and generates testable hypotheses.

Directionality and causation. The proposed relationship between algorithmic exposure and preference-formation capacity may not be straightforwardly causal. Individuals with lower preference-formation capacity may gravitate toward algorithmic assistance, creating a selection effect that would inflate cross-sectional correlations between exposure and capacity. The experimental component of the proposed study design is intended to address this limitation, but longitudinal confounds remain a concern.

Individual differences. The framework treats Behavioral Debt as a general phenomenon but acknowledges that individual differences in cognitive style, personality, and baseline preference-formation capacity will moderate debt accumulation substantially. Some individuals may be highly resistant to Behavioral Debt, while others may be highly susceptible. The framework does not yet specify these moderating pathways with the precision required for strong predictive claims.

Cultural and contextual variation. The framework has been developed primarily with reference to Western, English-language digital platforms and their users. The generalizability of the framework to other cultural contexts — in which the relationship between individual autonomy, social influence, and technological mediation may differ substantially — remains an open question.

The benefits-costs balance. The framework focuses on the costs of algorithmic recommendation without giving systematic attention to the benefits. Recommendation systems provide genuine value: they reduce search costs, expose users to content they would not otherwise discover, and facilitate navigation of complex information environments. A complete assessment of the welfare effects of algorithmic recommendation would need to weigh Behavioral Debt against these benefits. The present dissertation focuses on the cost side not because the benefits are unimportant but because they are already well-documented, while the specific cognitive costs identified by the Behavioral Debt framework have not been systematically theorized.

Measurement challenges. Operationalizing Behavioral Debt faces formidable measurement challenges. Baseline establishment requires prospective longitudinal designs initiated before significant algorithmic exposure begins — practically difficult for adult populations in contemporary digital environments. The three sub-constructs may be empirically correlated and difficult to disentangle. Ethical constraints limit the intensity and duration of experimental manipulations. These challenges are genuine, not rhetorical, and they define the research agenda that the framework generates.

7.2 Future Research Directions

Empirical validation. The most urgent priority is empirical testing of the framework's core propositions, using the mixed-methods design proposed in Section 4 or alternative designs that address the same hypotheses.

Instrument development. Validated psychometric instruments are needed for each sub-construct. The proposed indicators (PDI, DLUE, SRCCS, AOR, SDET) require psychometric validation through standard scale development procedures including item analysis, factor analysis, test-retest reliability assessment, and convergent/discriminant validity testing.

Neuroimaging studies. The strongest form of evidence for Autonomy Erosion would come from neuroimaging studies demonstrating structural or functional changes in preference-formation-related brain regions associated with sustained algorithmic exposure. Building on the work of Wilmer et al. (2019) and Camerini et al. (2021), targeted studies could examine whether sustained use of recommendation systems produces detectable changes in prefrontal-striatal circuits associated with evaluative judgment and decision-making.

Longitudinal designs. Given the cumulative nature of Behavioral Debt, longitudinal study designs following the same individuals over extended periods of varying algorithmic exposure are essential. Natural experiments — such as the introduction of recommendation systems into

previously unmediated environments, or the temporary removal of algorithmic recommendation (as in Allcott et al., 2020) — provide valuable opportunities for quasi-experimental evidence.

Reversibility studies. Testing the recovery hypothesis — whether periods of reduced algorithmic exposure produce measurable recovery in preference-formation indicators — is critical for both theoretical and practical reasons. If Behavioral Debt is reversible, the implications for intervention design differ substantially from those that follow from irreversibility.

Cross-cultural research. Testing the framework across cultural contexts in which the relationship between individual autonomy and technological mediation may differ — particularly in collectivist cultures where the individual-autonomy framing of the framework may require adaptation. The framework's emphasis on individual preference formation as a cognitive capacity may require modification in contexts where collective decision-making and social consensus are valued more highly than individual autonomy.

Computational modeling. Agent-based modeling of Behavioral Debt accumulation in simulated recommendation environments could provide insights into the dynamics of the debt cycle, the conditions under which debt escalates versus stabilizes, and the potential effects of design interventions. Tang et al. (2025) have demonstrated the value of agent-based approaches for modeling the co-evolution of user preferences and algorithmic recommendations, providing a methodological foundation for computational approaches to Behavioral Debt research.

Developmental trajectories. The framework is particularly urgent for children and adolescents, whose preference-formation capacities are still developing. Research by Orben and Przybylski (2019) found small but detectable associations between digital technology use and adolescent well-being, though the specific mechanisms remain unclear. Behavioral Debt may provide a framework for understanding how sustained algorithmic mediation during critical developmental periods affects the maturation of preference-formation capacities. Children who grow up with algorithmically curated content as the default mode of engagement may never fully develop the cognitive routines for autonomous preference formation that earlier generations developed through unmediated exploration.

Platform-specific research. Different platforms employ different recommendation architectures, optimization objectives, and interface designs. Research is needed to identify which platform features generate the highest rates of Behavioral Debt accumulation and which design modifications are most effective at reducing debt. This platform-specific research would complement the domain-level analysis provided in Section 5 with finer-grained technical detail.

Interaction with other cognitive constructs. The relationship between Behavioral Debt and related cognitive constructs — including decision fatigue (Baumeister et al., 1998), cognitive depletion, information overload (Eppler & Mengis, 2004), and analysis paralysis — requires systematic investigation. These constructs may represent either

contributing factors to or consequences of Behavioral Debt, and their interrelationships may reveal additional mechanisms through which algorithmic mediation affects cognition.

Table 4

Research Priority Matrix for Behavioral Debt Empirical Program

Research Direction	Priority	Feasibility	Expected Impact	Recommended Timeline
Instrument development and validation	Critical	High	Foundation for all empirical work	Year 1
Cross-sectional survey study	High	High	Initial evidence for associations	Year 1
Experimental manipulation study	Critical	Moderate	Causal evidence	Year 1-2
Qualitative interview study	High	High	Phenomenological depth	Year 1-2
Longitudinal cohort study	High	Low-Moderate	Temporal dynamics	Years 2-5
Neuroimaging study	Moderate	Low	Biological mechanisms	Years 3-5
Cross-cultural replication	Moderate	Moderate	Generalizability	Years 2-4
Agent-based computational modeling	Moderate	High	System dynamics	Year 2-3
Developmental study (children/adolescents)	High	Moderate	Critical population	Years 2-4

8. Conclusion

Behavioral Debt names a phenomenon that many users of recommendation systems may experience but lack the vocabulary to describe. The person who, when asked what music they want to listen to, reaches instinctively for their phone and opens Spotify — not because Spotify has the music they want but because they are no longer certain what music they want — is experiencing something real. The vocabulary of filter bubbles does not quite name it. The vocabulary of surveillance capitalism does not quite name it. The vocabulary of the attention economy does not quite name it.

Behavioral Debt names it: the accumulated cost of having had one's preferences answered for them, repeatedly, over years, by systems optimized for engagement rather than wellbeing.

The framework presented in this dissertation is theoretical and, in its current form, unfalsified. This is a limitation that the dissertation has acknowledged repeatedly and that the proposed empirical program is designed to address. What the framework offers now is a conceptual contribution: a defined construct, a formal model, a set of testable hypotheses, and a vocabulary for discussing a phenomenon that has been widely intuited but inadequately theorized.

The three sub-constructs — Preference Substitution, Autonomy Erosion, and the Attention Ledger — provide a structure for empirical investigation that is both sufficiently precise to generate falsifiable predictions and sufficiently flexible to accommodate the complexity of human-algorithm interaction across diverse domains and populations.

The mathematical formulations provide a foundation for computational modeling. The proposed empirical study design offers a roadmap for the first systematic tests of the framework's propositions.

The implications of the framework, if empirically validated, are substantial. For platform design, Behavioral Debt identifies a class of cognitive externalities that current design practices neither acknowledge nor address. For regulatory policy, it identifies a gap in the consent and disclosure frameworks that currently govern algorithmic systems. For education, it identifies a new dimension of digital literacy. For individuals, it provides a conceptual tool for understanding and counteracting a process that operates below the threshold of ordinary awareness.

Recommendation systems are among the most consequential cognitive environments ever designed. They mediate access to information, entertainment, social connection, and commerce for billions of people, continuously, across virtually all domains of digital life. The frameworks available to evaluate their human costs have not kept pace with their proliferation. Behavioral Debt is proposed as one addition to that evaluative vocabulary — an attempt to name, formalize, and begin to measure the cognitive residue that these systems leave within the people they serve.

The observation with which this dissertation began bears repeating in conclusion: each individual interaction between a user and a recommendation system is, typically, inconsequential. The debt accumulates in the aggregate, across thousands of interactions, over

months and years, largely below the threshold of user awareness. It compounds: as capacity erodes, dependence increases, which further erodes capacity. And it is invisible: debtors do not know they carry it.

The most expensive debt, in any domain, is the one the borrower does not know they carry. The purpose of this dissertation has been to make that debt visible — to give it a name, a structure, and a pathway toward measurement and, ultimately, remediation.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–676. https://doi.org/10.1257/aer.20190658

Alter, A. (2017). *Irresistible: The rise of addictive technology and the business of keeping us hooked*. Penguin Press.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–1265. https://doi.org/10.1037/0022-3514.74.5.1252

Beshears, J., & Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes*, *161*(Suppl.), 3–19. https://doi.org/10.1016/j.obhdp.2020.09.001

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Burr, C., Cristianini, N., & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. *Minds and Machines*, *28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Camerini, A. L., Marciano, L., & Morese, R. (2021). The developing brain in the digital era: A scoping review of structural and functional correlates of screen time in adolescence. *Frontiers in Psychology*, *12*, 671817. https://doi.org/10.3389/fpsyg.2021.671817

Chaney, A. J. B., Stewart, B. M., & Engelhardt, B. E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. *Proceedings of the 12th ACM Conference on Recommender Systems*, 224–232. https://doi.org/10.1145/3240323.3240370

Coppolillo, E., Mungari, S., Ritacco, E., Fabbri, F., Minici, M., Bonchi, F., & Manco, G. (2024). Algorithmic drift: A simulation framework to study the effects of recommender systems on user preferences. *arXiv preprint*. https://doi.org/10.48550/arXiv.2409.16478

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Cunningham, W. (1992). The WyCash portfolio management system. *ACM SIGPLAN OOPS Messenger*, *4*(2), 29–30.

Davenport, T. H., & Beck, J. C. (2001). *The attention economy: Understanding the new currency of business*. Harvard Business School Press.

Dogruel, L., Facciorusso, D., & Stark, B. (2020). "I'm still the master of the machine." Internet users' awareness of algorithmic decision-making and their perception of its effect on their autonomy. *Information, Communication & Society*, *24*(14), 2109–2128. https://doi.org/10.1080/1369118X.2020.1863999

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, *20*(5), 325–344. https://doi.org/10.1080/01972240490507974

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 153–162. https://doi.org/10.1145/2702123.2702556

Eyal, N. (2014). *Hooked: How to build habit-forming products*. Portfolio/Penguin.

Fernández-Rovira, C., Aldana Afanador, P. N., & Giraldo-Luque, S. (2020). The struggle for human attention: Between the abuse of social media and digital wellbeing. *Healthcare*, *8*(4), 497. https://doi.org/10.3390/healthcare8040497

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann.

Fogg, B. J. (2009). A behavior model for persuasive design. *Proceedings of the 4th International Conference on Persuasive Technology*, Article 40, 1–7. https://doi.org/10.1145/1541948.1541999

Gansky, B., & McDonald, S. (2022). CounterFAccTual: How FAccT undermines its organizing principles. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1982–1992. https://doi.org/10.1145/3531146.3533241

Gilbert, S. J. (2024). Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory. *Cognition*, *247*, 105783. https://doi.org/10.1016/j.cognition.2024.105783

Gray, C. M., Santos, C., Bielova, N., Toth, C., & Clifford, D. (2024). An ontology of dark patterns knowledge: Foundations, definitions, and a taxonomy. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, Article 304. https://doi.org/10.1145/3613904.3642007

Grinschgl, S., Meyerhoff, H. S., & Papenmeier, F. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Haider, J., Hirvonen, N., & Jylhä, V. (2024). Algorithmic recommendations enabling and constraining information practices among young people. *Journal of Documentation*, *80*(1), 218–236. https://doi.org/10.1108/JD-05-2023-0102

Herberz, M., Mertens, S., Brosch, T., & Hahnel, U. J. J. (2021). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*, *117*(52), 2107346118. https://doi.org/10.1073/pnas.2107346118

Firth, J. A., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., Alvarez-Jimenez, M., Gleeson, J., Vancampfort, D., Armitage, C. J., & Sarris, J. (2019). The "online brain": How the internet may be changing our cognition. *World Psychiatry*, *18*(2), 119–129. https://doi.org/10.1002/wps.20617

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006. https://doi.org/10.1037/0022-3514.79.6.995

Jesse, M., & Jannach, D. (2021). Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, *3*, 100052. https://doi.org/10.1016/j.chbr.2020.100052

Jiang, R., Chiappa, S., Lattimore, T., György, A., & Kohli, P. (2019). Degenerate feedback loops in recommender systems. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 383–390. https://doi.org/10.1145/3306618.3314288

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press.

Lu, W. (2024). Inevitable challenges of autonomy: Ethical concerns in personalized algorithmic decision-making. *Humanities and Social Sciences Communications*, *11*, Article 1306. https://doi.org/10.1057/s41599-024-03864-y

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), Article 81. https://doi.org/10.1145/3359183

McLaughlin, B., & Spiess, J. (2024). Algorithmic assistance with recommendation-dependent preferences. *arXiv preprint*. https://doi.org/10.48550/arXiv.2208.07626

Neal, D. T., Wood, W., Labrecque, J. S., & Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology*, *48*(2), 492–498. https://doi.org/10.1016/j.jesp.2011.10.011

Noordeh, E., Levin, R., Jiang, R., & Shadmany, H. (2020). Echo chambers in collaborative filtering based recommendation systems. *arXiv preprint*. https://doi.org/10.48550/arXiv.2011.03890

Nugroho, A., Visser, J., & Kuipers, T. (2011). An empirical model of technical debt and interest. *Proceedings of the 2nd Workshop on Managing Technical Debt*, 1–8. https://doi.org/10.1145/1985362.1985364

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.

Parry, D. A. (2023). Does the mere presence of a smartphone impact cognitive performance? A meta-analysis of the "brain drain effect." *Media Psychology*, *27*(1), 1–28. https://doi.org/10.1080/15213269.2023.2286647

Rakova, B., & Chowdhury, R. (2019). Human self-determination within algorithmic sociotechnical systems. *arXiv preprint*. https://doi.org/10.48550/arXiv.1909.06713

Romero Meza, L., & D'Urso, G. (2024). User's dilemma: A qualitative study on the influence of Netflix recommender systems on choice overload. *Psychological Studies*, *69*, 446–464. https://doi.org/10.1007/s12646-024-00807-0

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*(3), 409–425. https://doi.org/10.1086/651235

Schwartz, B. (2004). *The paradox of choice: Why more is less*. HarperCollins.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, *83*(5), 1178–1197. https://doi.org/10.1037/0022-3514.83.5.1178

Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, communications, and the public interest* (pp. 37–72). Johns Hopkins University Press.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Srnicek, N. (2017). *Platform capitalism*. Polity Press.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review, 8*(2), 1–22. https://doi.org/10.14763/2019.2.1410

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tang, M., Huang, X., & Sang, J. (2025). When algorithms mirror minds: A confirmation-aware social dynamic model of echo chamber and homogenization traps. *arXiv preprint*. https://doi.org/10.48550/arXiv.2508.11516

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal, 13*(2), 203–218.

Wang, G., & Pea, R. (2024). Algorithmic autonomy in data-driven AI. *arXiv preprint*. https://doi.org/10.48550/arXiv.2411.05210

Ward, A. F., Duke, K. E., Gneezy, A., & Bos, M. (2017). Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research, 2*(2), 140–154. https://doi.org/10.1086/691462

Wilmer, H. H., Hampton, W. H., Olino, T., Olson, I. R., & Chein, J. (2019). Wired to be connected? Links between mobile technology engagement, intertemporal preference and frontostriatal white matter connectivity. *Social Cognitive and Affective Neuroscience, 14*(4), 367–379. https://doi.org/10.1093/scan/nsz024

Wilmer, H. H., Sherman, L. E., & Chein, J. (2017). Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in Psychology*, *8*, 605. https://doi.org/10.3389/fpsyg.2017.00605

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit–goal interface. *Psychological Review*, *114*(4), 843–863. https://doi.org/10.1037/0033-295X.114.4.843

Wood, W., Mazar, A., & Neal, D. T. (2021). Habits and goals in human behavior: Separate but interacting systems. *Perspectives on Psychological Science*, *16*(3), 655–674. https://doi.org/10.1177/1745691621994226

Wu, T. (2016). *The attention merchants: The epic scramble to get inside our heads*. Knopf.

Xu, R., & Dean, S. (2023). Decision-aid or controller? Steering human decision makers with algorithms. *arXiv preprint*. https://doi.org/10.48550/arXiv.2303.13712

Young, M., Katell, M., & Krafft, P. M. (2022). Confronting power and corporate capture at the FAccT conference. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1375–1386. https://doi.org/10.1145/3531146.3533194

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Correspondence regarding this dissertation may be directed to: Hugo Thack, Inverso Research Institute, hello@hugothack.com

The author declares no conflicts of interest. No external funding was received for the preparation of this manuscript. This work represents an original theoretical contribution by the author.

Word count: approximately 15,400 words (body text, excluding references)