



Biswas, Mriganka and Ara, Safina Showkat (2026) The adaptive machine zone: distributed agency and the locus of responsibility in the convergence of gambling and video games. *Addiction Research & Theory*. pp. 1-20. ISSN 1606-6359

Version: AAM

Available under license: Creative Commons: Attribution 4.0

Downloaded from: <https://sure.sunderland.ac.uk/id/eprint/20685/>

Which version am I reading (articles, book chapters and books)

Pre-print (or submitted version): the version of an article or chapter before any peer-review. Can be submitted to a publisher or added to a pre-print server.

AAM: Author Accepted Manuscript - The version of the article formally accepted for publication after the peer-review process. It includes the changes made by the author(s) in line with peer-reviewers' comments. It has not undergone the publisher's formatting, type-setting and copyediting.

VOR: Version of Record - The version published on the publisher's website which includes formatting, typesetting and copyediting.

Usage guidelines

Please refer to the usage guidelines at <http://sure.sunderland.ac.uk/policies.html> or alternatively contact sure@sunderland.ac.uk.

The Adaptive Machine Zone: Distributed Agency and the Locus of Responsibility in the Convergence of Gambling and Video Games

Mriganka Biswas^{1*} and Safina Ara¹

¹ *Faculty of Technology, University of Sunderland, Sunderland, United Kingdom*

*Corresponding author: Mriganka.biswas@sunderland.ac.uk; ORCID: 0000-0001-7573-4816

Abstract

Gambling disorder and gaming disorder are now both recognised behavioural addictions, in the DSM-5 and ICD-11 respectively, and their products are converging: loot boxes, gacha, social casino games, and skins and esports betting blur the line between playing and wagering. The empirical literature finds these gambling-like features reliably but moderately associated with gambling-related harm, while no causal link has been established. That associational, non-causal evidence raises a conceptual problem in its own right, which this paper takes as its starting point: what does the convergence mean for responsibility? How should responsibility be conceived when a designed, commercial product is reliably associated with harm yet cannot be shown to cause it? Its answer is the adaptive machine zone, a state of sustained engagement in which a monetised system uses behavioural data about the user to personalise offers, pacing, friction and prompts so as to narrow the opportunity to disengage. The original contribution is not to identify gambling-like mechanics but to specify behavioural personalisation as the threshold at which responsibility shifts from individual self-control to accountable adaptive design. The paper argues that agency and responsibility are, to that extent, distributed across the human-machine system, and that a commercial-determinants framing can license precautionary, design-level regulation without individual-level causation. The framework is deliberately bounded, applying only where personalised, adaptive design meets concentrated harm. A proof-of-concept audit of ten high-grossing titles surfaces an observability gradient: the precondition features are publicly verifiable, but the adaptive features that define the framework are not, so that current transparency regimes make the relevant question unanswerable from outside.

Keywords: gambling–gaming convergence; loot boxes; behavioural addiction; distributed agency; commercial determinants of health; responsibility

Plain-language summary

Many popular video games now contain gambling-like features, such as loot boxes, and many gambling products now look and feel like games. Research shows these features are linked to gambling harm, but it cannot yet prove that they cause it. This paper argues that waiting for causal proof asks the wrong question. Modern games and gambling apps track each player and adjust their offers, timing and prompts to keep that person engaged; we call the resulting state the adaptive machine zone. When a product works this way, responsibility for harm should not rest on the player alone: it is shared with the companies that design and run the system and the regulators who permit it. Telling people to ‘play responsibly’ is not enough when the product is engineered around their own data. We show that the personalised features that matter most are exactly the ones outsiders cannot currently inspect, and we argue that regulators should require transparency and safer design, so that protection does not depend on the willpower of the people these systems are built to influence.

1. Introduction

Two developments have reframed the study of gambling and gaming harm. The first is clinical and conceptual: gambling disorder was reclassified in 2013 as a behavioural addiction in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013), and gaming disorder was added to the eleventh revision of the International Classification of Diseases in 2019 (World Health Organization, 2019; King and Potenza, 2019). For the first time, both a wagering activity and a form of play are recognised as addictive conditions in their own right. The second development is technological and commercial: the products themselves are converging. Loot boxes, gacha mechanics, social casino games, skins and esports betting, and other simulated-gambling features increasingly blur the boundary between playing a game and placing a bet (Zendle and Bowden-Jones, 2019; Macey, Hamari and Adam, 2024). The convergence runs in both directions: games are gamblified through the incorporation of chance-based monetisation, and gambling is gamified through the incorporation of game-like features, missions, levels and social play. The focus of this paper is the former, gamblified games, although the framework developed in Section 5 bears on adaptive gambling products generally, a point returned to there.

This convergence is now a public-policy priority across multiple jurisdictions, which have moved variously to ban, restrict, disclose or formally study gambling-like monetisation in games (Cerulli-Harms et al., 2020). The United Kingdom offers an illustrative case: the statutory levy on gambling operators introduced in April 2025 funds an independent research programme, and the relationship between gambling and video games has been singled out for

dedicated investment, with policy notably careful about evidence. Current evidence is taken to show that gambling-like features in video games are associated with gambling-related harms, while it is explicitly acknowledged that no causal link has been established and that research is needed to clarify the specificity, direction and real-world magnitude of these associations (Department for Culture, Media and Sport, 2023). This caution is well founded: the most rigorous syntheses report positive but moderate associations between loot box engagement and problem gambling, of the order of a correlation of around 0.3, with the direction of any causal arrow unresolved (Spicer et al., 2022; Garea et al., 2021).

The condition of associational-but-not-causal evidence is usually treated as a limitation to be overcome by better studies. This paper treats it differently: the evidential situation is one thing, and the question it gives rise to is another. The situation is reproducible association without established causation; the question that emerges from it, and that this paper takes as its subject, is how responsibility for harm should be conceived when a designed, commercial product is reliably associated with harm but cannot (yet, or perhaps ever, at the individual level) be shown to cause it. The question matters because the answer that is implicit in both clinical classifications, that the disorder, and hence the responsibility, resides in the individual, does substantive work in policy, where it underwrites the doctrine of ‘responsible gambling’ and, increasingly, ‘responsible play’. If that individualist ontology is mistaken for convergent digital products, then the evidential caution about causation, far from counselling regulatory inaction, points towards a different and arguably stronger basis for intervention. The implication is worth stating plainly: if the disorder is not located solely in the individual, then interventions aimed solely at the individual address only part of the mechanism, and the persistence of individual-responsibility framing itself demands explanation, one that critical scholarship locates in the political economy of the industries concerned (van Schalkwyk, Hawkins and Petticrew, 2022; Reith and Wardle, 2022).

Put directly, the thesis is this: the harms associated with gambling–gaming convergence are best understood not only as individual addictive behaviour but also as the output of an adaptive socio-technical harm system, and responsibility should shift accordingly, from individual ‘responsible play’ towards design-level, platform-level and regulatory accountability. The convergence is treated here as a case, the sharpest available, of a more general problem in the philosophy of technology: where an adaptive system does not merely present options but participates in shaping the conditions under which a user chooses, the locus of agency and responsibility is itself put in question. Contemporary convergent products are

best understood not as fixed stimuli encountered by vulnerable individuals but as adaptive socio-technical systems in which the addictive process is co-produced and, increasingly, actively sustained. I develop the concept of the adaptive machine zone to capture what is distinctive about them: not merely the exploitation of documented cognitive vulnerabilities, but the personalised, continuously updated reconfiguration of the choice environment in response to the user's own behaviour. On this account agency is distributed across the human-machine system, and so is responsibility. Crucially, a distributed, commercial-determinants framing dissolves the apparent tension between evidential caution and regulatory action: one need not establish individual-level causation to hold that a product designed to exploit known vulnerabilities, and reliably associated with harm, is an appropriate object of precautionary regulation.

This is primarily a conceptual paper, complemented by a small proof-of-concept audit. Its method is a critical synthesis rather than a systematic review, organised around a single analytic question, the locus of responsibility for convergent harm. It draws together five literatures that rarely speak to one another: the philosophy of addiction; the cognitive science of gambling; the anthropology and political economy of machine gambling and commercial determinants of health; the empirical convergence literature on loot boxes, social casino and esports betting; and scholarship on algorithmic manipulation and social interface design. Sources were included where they contributed to one of four analytic functions, clinical classification, mechanism of harm, socio-technical production, or responsibility allocation, and were selected for their centrality to each field and their bearing on the responsibility question rather than by exhaustive search. The purpose is conceptual integration, not prevalence estimation or causal adjudication; the contribution is a framework within which the empirical study of convergent products can pose sharper questions.

The framework's contribution lies in what it adds to several adjacent and now well-developed concepts, and it is important to state that contribution modestly. Several adjacent literatures already capture much of the picture. Gambling-like features, gambification and gambified experiences are mapped by Macey, Hamari and Adam (2024); predatory monetisation by King and Delfabbro (2018); dark patterns and the hypernudge by Yeung (2017); the machine zone by Schüll (2012); the capture of behavioural data by Zuboff (2019); habit-forming product design by Eyal (2014); and the commercial determinants of health by van Schalkwyk et al. (2021). Between them, these accounts document the gambling-like structure of game mechanics, their cost-obscuring monetisation, their personalised choice

architecture, the co-production of compulsion at the human–machine interface, the economic logic of behavioural data, and the industrial production of harm. This paper does not claim to displace them. Its contribution is an application and extension of these concepts to the convergence domain, with two specific additions: it isolates behavioural personalisation, the adaptation of the system to the individual user in order to sustain engagement, as the property that distinguishes the adaptive machine zone from fixed gambling-like design, and it uses that property to reframe the allocation of responsibility under conditions of causal uncertainty. Stated sharply, the original contribution is not to identify gambling-like mechanics, which others have done, but to specify behavioural personalisation as the threshold at which responsibility shifts from individual self-control to accountable adaptive design. The originality lies in that threshold claim, in the responsibility argument that follows from it, and in a proof-of-concept demonstration that the resulting audit instrument is operable, not in the discovery of new mechanics. Sections 2 and 3 set out the individualist settlement in the clinic and in philosophy; Section 4 develops the co-production and commercial-determinants alternative and reviews the convergence evidence; Section 5 introduces the adaptive turn and defines the adaptive machine zone; Section 6 demonstrates the instrument through a proof-of-concept transparency audit; Section 7 advances the distributed account of responsibility, answers objections and weighs the case against it; Section 8 draws out implications; Section 9 states limitations; and Section 10 concludes.

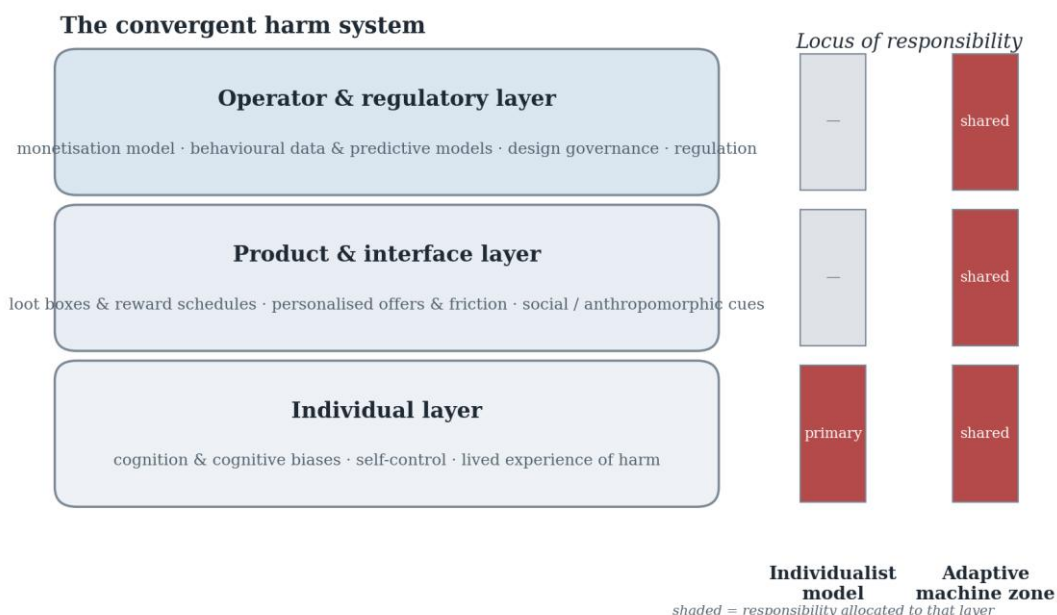


Figure 1. The convergent harm system as three coupled layers, individual, product/interface, and operator/regulatory. Under the individualist model the locus of responsibility is concentrated on the individual; in the adaptive machine zone it is distributed across all three layers.

2. Two behavioural addictions and a shared individualism

The reclassification of gambling disorder consolidated decades of evidence that disordered gambling resembles substance dependence in its diagnostic features, comorbidities and reward-related neurobiology (Clark, 2014). The components common to behavioural addictions, salience, mood modification, tolerance, withdrawal, conflict and relapse, were applied to gambling as readily as to substances (Griffiths, 2005). The subsequent recognition of gaming disorder in ICD-11 extended the same logic to play, defining the condition through impaired control, escalating priority given to gaming, and continuation despite harm (World Health Organization, 2019; King and Potenza, 2019), with meta-analytic prevalence estimates of around 3 per cent worldwide underscoring its clinical significance (Kim et al., 2022). The recognition of gaming disorder remains more contested than that of gambling disorder, with critics warning against pathologising high but non-harmful engagement and against a wave of false-positive diagnoses (Aarseth et al., 2017); the disputed constructs are chiefly withdrawal and tolerance, whose applicability to gaming is less settled than to substances or gambling. A further distinction matters for this paper: the debate spans offline and online games, but only connected, online products can be adaptive in the sense developed in Section 5, so the clinical controversy and the design question intersect precisely there. Its inclusion in ICD-11 nonetheless makes the comparison clinically and conceptually significant. Both conditions are predominantly understood, and screened for, at the level of the individual.

The cognitive science of gambling both illuminates and quietly destabilises this individualism. Clark and colleagues showed that near-miss outcomes, losses that fall just short of a win, recruit reward circuitry overlapping with that of genuine wins and increase the motivation to continue, an effect amplified when the player exercises apparent control over the gamble (Clark et al., 2009; Clark, 2010), consistent with the long-documented illusion of control (Langer, 1975). Yet near-misses, variable-ratio reinforcement and the manufacture of apparent control are not features of brains; they are features of games, deliberately engineered, with a documented history in the design of gambling machines (Parke and Griffiths, 2004). The neuroscience identifies a vulnerability; but the vulnerability is activated, and is designed to be activated, by an artefact. A neuroscientific account that halts at the individual's reward system, identifying the vulnerability but stopping there, describes one arc of a causal loop and treats its other arc, the designed product that triggers and trains that vulnerability, as inert.

This matters doubly in the convergence case, where the 'product' is a video game played by most of the population and especially, though not exclusively, by children and young people

(Ofcom, 2024; Department for Culture, Media and Sport, 2023), and gambling-like mechanics are embedded in titles accessible to minors. An ontology that makes the disordered individual the unit of analysis not only under-describes a co-produced harm; in the convergence setting it also risks attributing to children a failure of self-management in the face of professionally engineered monetisation.

3. The philosophy of addiction and the unit of analysis

Philosophical accounts of addiction have moved between a disease or compulsion pole, addiction as a chronic, relapsing brain disease (Leshner, 1997), and a choice pole, on which addictive behaviour remains reasons-responsive and in a meaningful sense chosen (Heyman, 2009). The high rate of unaided recovery, often prompted by changing circumstances, sits awkwardly with a pure disease model and indicates that neurobiology underdetermines behaviour. Several philosophers dissolve the opposition. Frankfurt's (1971) distinction between first-order desires and second-order volitions locates the impairment not in desire as such but in the harmony of the will. Kennett (2013) and Levy (2013) frame addiction as a disorder of extended, diachronic self-control rather than a momentary absence of choice.

Two contributions are pivotal here. Foddy and Savulescu (2010) characterise addiction as, at bottom, a strong appetite: people experiencing addiction pursue what they strongly want, intelligibly, given their circumstances (person-first phrasing is used here in line with recommendations against stigmatising descriptors in gambling research; Hing, Russell and Gainsbury, 2020). This takes agency seriously, but brackets the question of who manufactured, priced and presented the object of appetite, the very question the convergence case forces. Pickard's (2017) framework of responsibility without blame, developed from clinical practice, holds that we can acknowledge the agency and choices of a person experiencing addiction, which is often experienced as respectful and empowering, while detaching responsibility from the hostile affect of blame; responsibility is forward-looking, and because agency comes in degrees, so does responsibility (Pickard and Pearce, 2013). Phenomenological and economic approaches complete the field: Flanagan (2013) emphasises the first-person authority of the testimony of people experiencing addiction; Ainslie (2013) models the person as transient interests bargaining over time under hyperbolic discounting; Ross (2013) applies this piceoeconomic apparatus to gambling specifically; and Sinnott-Armstrong (2013) argues that people experiencing addiction retain enough control to be partly, gradedly responsible.

For all their disagreement, these accounts share a unit of analysis: the individual agent, whose will, capacities and responsibility are at issue, and whose environment, even where it is acknowledged, figures as a source of cues, temptations or stressors rather than as a co-constituent of the addictive process itself, so that the individual remains the primary bearer of agency and responsibility. The point here is not to adjudicate between these theories, each captures something real about the structure of addictive agency, but to identify the assumption they hold in common, and which the convergence case puts under pressure. Where the object is a person and a chemical of fixed pharmacological properties, that framing is largely adequate. But neither gambling nor a video game is a chemical. Each is a designed, commercial and increasingly adaptive artefact, and that, as the next sections argue, changes the philosophical problem.

4. Co-production, commercial determinants, and the convergence evidence

The sharpest challenge to individualism comes from the anthropology of machine gambling. In her ethnography of Las Vegas, Schüll (2012) argues that addiction is not an internalised psychological state but a product of the interaction between human and machine, an ‘asymmetrical collusion’ in which the machine’s algorithms, ergonomics and reinforcement schedules are calibrated to maximise ‘time on device’, while the player seeks the trancelike ‘machine zone’ the machine is built to elicit. The regulatory question Schüll poses is exactly the one the convergence raises: does the harm stem from the consumer, from the product, or from the interplay between them? The individualist settlement answers ‘the consumer’; the ethnography points to the interplay, and to its asymmetry.

The political economy of this arrangement has been theorised by Reith, who shows that ‘problem gambling’ emerged as a category through a convergence of clinical discourse and consumer capitalism in which the disordered individual performs ideological work, isolating a minority of ‘failed’ consumers so as to normalise the activity for everyone else (Reith, 2007; 1999; 2018). On this reading the doctrine of responsible gambling is the mechanism by which a structurally harmful product is reconciled with an ethic of individual self-management; commercial-determinants scholarship argues that such individual-responsibility framing is not accidental but a documented industry strategy, locating blame downstream with the consumer in order to forestall regulation of the product (van Schalkwyk, Hawkins and Petticrew, 2022; Maani, Petticrew and Galea, 2023), which is part of the explanation for the persistence of the individualist ontology described in Section 2. This analysis has matured into the commercial determinants of health, defined in the Lancet series as the systems, practices and pathways

through which commercial actors drive health and equity (Gilmore et al., 2023). Applied through a public-health lens, the analysis holds that harm is shaped by the availability, design and marketing of products and by the conduct of the industries that supply them, so that prevention requires action upstream of the individual consumer (Wardle et al., 2019; van Schalkwyk et al., 2021). This framing is especially apposite where causation is complex and distributed: because it locates responsibility at the level of products, markets and populations rather than individual exposures, it can ground preventive action on the basis of reliable population-level association and a plausible mechanism, without requiring proof of individual-level causation for each consumer. van Schalkwyk and colleagues argue specifically that ‘responsible gambling’ aligns with industry narratives and diverges from a public-health perspective that favours system-level intervention.

The convergence literature can be read as this same co-production logic migrating into play. The defining mechanic is the loot box, a purchasable, chance-based reward, which King and Delfabbro (2018) place within a wider class of ‘predatory monetisation schemes’ that disguise or withhold the true long-term cost of engagement until the player is financially and psychologically committed, and for which they propose social-responsibility design measures (King and Delfabbro, 2019). The empirical findings are consistent and carefully bounded. Loot boxes are psychologically akin to gambling in structure (Drummond and Sauer, 2018); loot box spending is reliably associated with problem gambling and with gambling-related cognitions (Zendle and Cairns, 2018; Brooks and Clark, 2019; Li, Mills and Nower, 2019), and the association generalises beyond loot boxes to a range of gambling-like practices in games (Zendle, 2020). Systematic review and meta-analysis put the association at a moderate magnitude, a mean correlation of roughly 0.27 between loot box use and problem gambling, while explicitly leaving open whether loot boxes recruit existing gamblers, act as a gateway, or stand in a bidirectional relationship with gambling (Spicer et al., 2022; Garea et al., 2021). Longitudinal evidence has since begun to bear on direction: Brooks and Clark (2023) report migration from loot box spending to gambling in a longitudinal study of young adults, which is consistent with a gateway pathway, though the argument developed here is deliberately built for the case in which direction remains unresolved and does not depend on it. Parallel concerns attach to social casino games, where youth expenditure is linked to later gambling risk (King et al., 2016), and to esports and skins betting, which draw young people into wagering through the affordances of games and streaming (Macey and Hamari, 2018).

The gacha mechanic, dominant in mobile gaming, extends and intensifies this logic. Beyond the single chance-based draw, gacha systems layer repeated draws, ‘pity’ mechanisms that guarantee a reward only after a threshold of spending, time-limited ‘banners’ that manufacture scarcity, and the binding of randomised rewards to collectible characters in which players invest identity and social comparison, a configuration that encourages escalating, sunk-cost-laden expenditure and that has been associated with harm among young players (Shibuya et al., 2019). These designs sit at the centre of the convergence and have prompted consumer-protection scrutiny of chance-based monetisation directed at the young (Cerulli-Harms et al., 2020; King and Delfabbro, 2019).

These associations must be read with care, because at least three alternatives compete with the inference that gambling-like features cause gambling harm. The relationship may run in reverse, with people already disposed to gamble seeking out loot boxes and gacha; it may be confounded by common underlying traits, such as impulsivity and sensation-seeking, that dispose individuals to both; or it may reflect selection, whereby products with gambling-like mechanics simply attract heavier and more vulnerable users. A correlation of roughly 0.27 also corresponds to under a tenth of the variance, a modest association, and the present evidence cannot decisively rule out any of these alternatives (Spicer et al., 2022; Garea et al., 2021). It would be tempting to treat this as fatal to any case for intervention. But it is fatal only to an argument that depends on showing that the product causes disorder in the individual. The argument developed below deliberately does not depend on that claim, nor on the size of the association; it rests instead on the design of the products and on the commercial determinants of harm, for which the unresolved direction of the individual-level association is not decisive.

Two points follow. First, the convergence evidence is, on the most honest reading, reproducible association of moderate size with unresolved direction. Second, that evidential situation is exactly what a co-production and commercial-determinants reading would predict, because it locates harm in a relationship and a designed environment rather than in a clean causal property of either the product or the person. The question is what normative weight such evidence can bear, that is, what actions, obligations and regulatory responses it can justify, to which Section 7 returns.

5. The adaptive turn and the adaptive machine zone

Schüll’s machine was sophisticated but fixed: its schedule was tuned in advance and applied uniformly. Convergent digital products differ in a way that is not merely quantitative.

They record interaction at fine granularity, build a behavioural model of the individual user, and reconfigure aspects of the experience, offers, prompts, pacing, friction, in response to that model. Engagement optimisation of this kind is documented in the games industry's own technical literature: dynamic difficulty adjustment systems have been designed explicitly to maximise player engagement across a publisher's titles (Xue et al., 2017), and a systematic examination of game patents finds monetisation designs that a consumer-protection lens would class as exploitative, patents being one of the few public windows into capabilities that products themselves do not disclose (King et al., 2019). The same predictive capability is evidenced on the gambling side, where account-based behavioural data can be used to operationalise diagnostic criteria and to predict problem gambling using machine-learning methods (Auer and Griffiths, 2022). The machine zone, in short, has become adaptive.

It is important to be precise about what is and is not established here, since over-claiming is both an analytic and, in this funding context, a credibility error. There is good evidence that engagement is algorithmically optimised, that monetisation is designed to exploit documented cognitive tendencies, and that behavioural models predicting harm-relevant states exist and are technically deployable. There is not, at present, robust public evidence that regulated products individually and dynamically alter the odds of play in response to a behavioural model; in regulated markets such outcomes are typically fixed and audited, and claims of individualised odds manipulation should be treated as a hypothesis to be tested rather than an established fact. The defensible claim is narrower and sufficient: that personalised offers, pacing, friction, marketing and social prompts are configured around the individual user, and that this configuration is increasingly automated and adaptive.

5.1 Defining the adaptive machine zone

It is worth defining the concept precisely, since loosely applied it could absorb every harmful feature of monetised digital play. The adaptive machine zone is defined here as a state of sustained engagement produced when a gambling-like or monetised system uses behavioural data about the individual user to personalise offers, pacing, friction, social prompts or reward structures in ways that narrow that user's opportunity to disengage. The name extends Schüll's (2012) 'machine zone', the absorbed, self-dissolving state her research participants described entering at the slot machine; the qualifier 'adaptive' marks what has changed, namely that the machine now adjusts itself to the individual in real time. The qualifier that does the work is personalised adaptation on the basis of the individual's own behaviour. This distinguishes the concept from three neighbouring phenomena that are necessary background but not themselves

sufficient: fixed gambling-like structural features such as near-misses and random rewards, which are a precondition rather than the thing itself (conventional advertising belongs to this fixed background too: marketing that presents chance-based products as risk-free fun shapes uptake and expectations, and bears on responsibility, without being adaptive); generic persuasive design applied uniformly, such as streaks and limited-time events, which need not be adaptive; and mere personalisation of content, which becomes part of the adaptive machine zone only when it is keyed to behavioural states and operates to sustain engagement. This last distinction is decisive, because personalisation is now ubiquitous in digital commerce. A music or retail recommender also models the user and optimises engagement, so personalisation alone cannot be the mark of harm, or the whole internet would qualify. What distinguishes the adaptive machine zone is personalisation layered onto chance-based, loss-structured monetisation: what is optimised is continued exposure to a gambling-like risk, not the next song or purchase. Personalisation is the mechanism; the gambling-like substrate is what qualifies the system for the specific class of harm at issue here, and the absence of that substrate is why most personalised products fall outside the concept. This is a threshold on a continuum rather than a binary between harmless and harmful personalisation: personalised purchase prompting can itself drive harmful spending, and sits between content recommendation and gambling-like risk on that continuum. The claim is not that personalisation without a chance-based substrate is innocuous, but that the conjunction marks the point at which the specific analysis of gambling-related harm, and the regulatory apparatus attached to it, is engaged. This is also what distinguishes the threshold from the broader literatures it draws on: where accounts of dark patterns and addictive design identify manipulative or compulsion-forming interfaces in general, and such practices are now systematically documented in gambling products themselves (McGarrigle et al., 2026), the adaptive machine zone is defined by the conjunction of behavioural personalisation with a chance-based, loss-structured substrate, and it is that conjunction, not either element alone, that does the normative work. A boundary question follows: if regulation keyed narrowly on individual-level personalisation, could operators achieve equivalent effects through increasingly granular cohort-level targeting, segmenting users into ever-finer behavioural classes without modelling any one individual as such? The framework is drawn to prevent this evasion. Its trigger is behavioural adaptation keyed to user data, and a cohort of one is simply the limiting case of a cohort: the normative analysis applies along the whole granularity gradient in proportion to how closely the segmentation tracks the individual, so regulation should target the function, behaviourally contingent engagement optimisation, rather than the implementation detail of per-user versus per-segment models. The

strongest, and least firmly evidenced, case is exploitative adaptation that targets identified risk or vulnerability states; this paper treats that as a hypothesis for audit rather than an established practice. The adaptive machine zone is accordingly proposed not as a settled empirical description of all current convergent products, but as an audit framework for identifying when and where gambling-like monetisation becomes behaviourally personalised harm production. A clarification of scope follows from the definition. Because the concept is defined by adaptive personalisation on a gambling-like substrate, it is not confined to convergent video games: contemporary in-app, account-based online gambling, which models its users and personalises inducements, bonuses and friction, falls squarely within it, and the framework therefore characterises adaptive gambling-like systems in general. Convergence is nonetheless this paper's case, and it is the sharpest case for three reasons: convergent products reach populations, including children and adolescents, that licensed gambling does not; they typically operate outside gambling-specific regulatory regimes, so the accountability gap between design power and design obligation is at its widest; and their adaptive layer is the least disclosed of any product class examined here. The construct is thus developed from and demonstrated on convergence, while applying to the wider class of adaptive gambling-like systems of which convergence is the least regulated instance.

So defined, the concept is empirically tractable, though its indicators take different forms across product types. Cross-cutting indicators, applicable to gambling and games alike, include varying the timing of prompts following losses or disengagement, differentiating the treatment of users after high-risk patterns of play or spending, and deploying social or anthropomorphic messaging to prompt return. In gambling products these are joined by product-specific indicators such as personalised deposit or withdrawal friction, individualised bonuses and 'VIP' management, and account-based risk modelling. In free-to-play games the corresponding indicators concern in-app purchase flows and premium-currency conversion, time-limited 'banners' and cooldowns, pity counters (mechanisms that guarantee a rare reward after a set number of unsuccessful chance-based attempts, disclosed to the player as progress toward that guarantee), the choreography of loot-box opening, and player segmentation through churn prediction (models that forecast which users are about to stop playing or spending) and live-ops targeting ('live operations', the continuous post-launch running of a game as a service, through which events, offers and content are scheduled and, potentially, targeted). A system exhibits the adaptive machine zone to the degree that these features are keyed to behavioural data about the individual user and operate to sustain engagement. So specified, the indicators

convert the concept from a persuasive metaphor into an audit framework: the next section applies them in a proof-of-concept audit, and Section 8 sets out a fuller empirical agenda.

Two further mechanisms complete the picture. The first is the broadly understood structure of algorithmic influence that Yeung (2017) terms the ‘hypernudge’, a networked, continuously updated, personalised choice architecture that channels behaviour by exploiting known irrationalities, and that Susser, Roessler and Nissenbaum (2019) define as online manipulation: hidden influence that targets and exploits decision-making vulnerabilities. The second is specific to the interface. Convergent products increasingly present themselves through social and anthropomorphic features (avatars, conversational hosts, social cues, simulated companionship) that recruit the well-established human tendency to treat computers as social actors and to extend them social trust (Reeves and Nass, 1996; Nass and Moon, 2000). Where a product can both model a user’s vulnerabilities and address them through a trusted, social-seeming channel, its levers extend from the cognitive into the relational. Delivering a personalised prompt in a warm, social register at the moment a user is modelled as most susceptible is not merely a hypernudge but an asymmetric exploitation of the human disposition to extend trust to apparent social partners, and it engages a relational vulnerability that the user has limited capacity to guard against precisely because the response it recruits is pre-reflective and evolutionarily deep.

The convergence of these mechanisms is best made concrete, with the explicit caveat that the following is an illustrative synthesis of separately documented features rather than a description of any specific product. (1) A free-to-play game uses variable-ratio reward schedules and chance-based loot boxes whose structural kinship with gambling is established (Drummond and Sauer, 2018; Zendle and Cairns, 2018). (2) Interaction is logged, and a behavioural model classifies the user’s engagement and spending propensity (Auer and Griffiths, 2022; Xue et al., 2017). (3) Personalised offers, limited-time events and friction are configured to that model. (4) Social and conversational features deliver prompts in a warm register (Nass and Moon, 2000). (5) The product sits within an ecosystem in which gaming and gambling mechanics blur, recruiting and habituating young users before the activity is recognised as gambling (Zendle, 2020; Macey and Hamari, 2018). The adaptive machine zone is the resulting state, in which a personalised system continuously narrows the gap between the user’s fluctuating impulse to stop and the design’s imperative that they continue. Table 1 summarises the principal mechanisms, their evidence base, their adaptive form, and their implication for the location of responsibility. The mechanisms listed are related levels of one

commercial logic rather than mutually exclusive categories: engagement optimisation names the overarching commercial objective; hypernudging (Yeung, 2017) is a specific personalised choice-architecture technique in its service; and live-ops is the operational layer through which both are deployed in practice.

Mechanism	Evidence base	Evidence strength	Adaptive / personalised form	Harm pathway	Responsibility implication
<i>Fixed / background structural features (preconditions)</i>					
Near-miss & variable-ratio reinforcement	Clark et al. 2009; Parke & Griffiths 2004	Strong in gambling; adaptive form inferred	Possible tuning of reward presentation / event timing	Sustains play past intended limits	Design choice → shared causal responsibility
Loot boxes / chance-based purchases	Drummond & Sauer 2018; Spicer et al. 2022	Strong (structural); moderate (gambling link)	Randomised, often opaque; event-timed offers	Gambling-like spend; youth exposure	Monetisation design + regulatory gap
<i>Adaptive / personalised features (the adaptive machine zone)</i>					
Behavioural tracking & risk prediction	Auer & Griffiths 2022	Moderate–strong (technical)	Real-time risk/engagement classification	Dual use: protect vs. retain	Governance of the predictive model
Engagement optimisation / live-ops	Xue et al. 2017; King & Delfabbro 2019	Moderate; inferential bridge	Personalised difficulty, offers, friction	Prolongs session and expenditure	Operator design responsibility
Social / anthropomorphic interface	Reeves & Nass 1996; Nass & Moon 2000	Indirect; audit hypothesis	Conversational hosts, social cues	Recruits social trust relationally	Relational manipulation; platform duty
Personalised targeting ('hypernudge')	Yeung 2017; Susser et al. 2019	Indirect; audit hypothesis	Continuously updated marketing	Exploits documented biases at scale	Platform + regulator responsibility

Table 1. Principal mechanisms relevant to the adaptive machine zone, grouped into fixed/background structural features (preconditions) and adaptive/personalised features, with an indicative grading of evidence strength. The grading is a summary judgement; 'indirect' marks mechanisms whose adaptive, harm-exploiting form is plausible and theorised but not yet directly evidenced in deployed products.

6. Demonstrating the framework: a proof-of-concept transparency audit

A framework that proposes an audit instrument ought to show that the instrument can be applied. This section reports a small proof-of-concept audit that codes a sample of popular monetised titles against the indicators set out in Section 5.1. Its purpose is not to estimate prevalence or to establish harm, but to demonstrate that the indicators are operable and to surface a structural finding about what can, and cannot, be observed from outside the system. The role of the audit is accordingly philosophical and methodological rather than empirical in the usual sense: it is an object lesson in what current public transparency regimes systematically foreclose, and therefore in why the morally decisive features of adaptive design lie beyond external assessment as things stand. By a transparency regime is meant the set of disclosure obligations and public information practices, statutory and self-regulatory alike, through which

a product’s mechanics can be known from outside: probability disclosures, store metadata, published privacy and data policies, and comparable instruments.

6.1 Method

The audit is a desk-based documentary classification rather than primary play-testing or account-level experimentation: it records what is publicly observable or documented about each title, not behaviour elicited through play, and it follows in method the content-analytic audits of loot box features and probability disclosures established in the literature (Zendle et al., 2020; Xiao, 2024). Ten titles were purposively selected to span the principal convergent categories, gacha role-playing games, sports loot-box modes, chest-based mobile games, social casino, weapon-case economies and social or user-generated-content platforms, drawn from titles consistently among the highest-grossing internationally. Each was coded against indicators in two groups. The precondition (fixed) features were chance-based monetisation, probability or pity disclosure, time-limited scarcity, and loyalty or ‘VIP’ tiers. One clarification attaches to the disclosure indicator: probability disclosure and pity disclosure are grouped because both are public disclosures of a chance mechanic, which is what the audit codes, but their likely behavioural functions differ and may even oppose one another. Probability disclosure is primarily a transparency and consumer-protection measure, informing the player of typically low odds; a pity-counter disclosure signals accumulating progress toward a guaranteed reward and may thereby encourage continued spending. The audit records their observability, not their valence; distinguishing and evaluating their behavioural effects is precisely the kind of question a fuller, access-enabled audit regime should take up. The adaptive (personalised) features were behaviourally-keyed or loss-contingent offers and prompts, and risk or churn segmentation, with collection of behavioural analytics recorded separately as the shared precondition of the adaptive group. Each indicator was classified on a four-level scheme: present and publicly verifiable (✓), from official disclosures, store listings or published privacy policies; present but only inferable (~), where the capability is documented at industry level without title-specific confirmation; not externally verifiable (ND), where the feature is covert and the public can neither confirm nor refute it; or absent (–). Coding drew on official in-game and publisher disclosures, store metadata, published privacy and data policies, and the peer-reviewed and regulatory literature, and reflects the public record as at May 2026; because monetisation features change with updates, the codings are time-stamped rather than standing claims. The exercise is explicitly proof-of-concept: a single coder,

at a single time point, using public information only, with no inter-rater reliability assessment; it is offered to demonstrate the instrument, not as a definitive measurement.

6.2 Results

Two patterns emerged, and together they form a sharp gradient (Table 2). The precondition features were, with few exceptions, present and publicly verifiable. Chance-based monetisation was near-ubiquitous across the sample, consistent with prevalence estimates in the wider literature, around 60 per cent of the highest-grossing Western mobile games (Zendle et al., 2020), rising to roughly 90 per cent in some Asian markets and persisting at high rates in the United Kingdom (Xiao, 2024; Xiao and Lund, 2025). Probability or pity disclosure was present where law requires it and patchier where it depends on self-regulation, mirroring documented compliance gradients that fall well short of full disclosure under voluntary regimes (Xiao and Lund, 2025). Time-limited scarcity and loyalty tiers were openly observable in most titles. One title in the sample, a puzzle game monetised largely without chance-based mechanics, returned mostly absent codings on the precondition group. This case matters for the instrument’s validity: a framework that classified every successful monetised game as gambling-adjacent would discriminate nothing, whereas here the indicators separate a heavily monetised but non-gambling-like product from the rest of the sample, which is precisely the bounded behaviour the framework claims for itself (Section 7.2).

The adaptive features showed the opposite pattern: across the sample they were, almost without exception, not externally verifiable. That behavioural analytics are collected is disclosed in every title’s privacy policy; how the resulting models are used, whether to personalise offers, time prompts to behavioural states, modulate friction or segment users by risk, is disclosed by none. Two cautions follow, and they bound the result tightly. First, the ND coding cannot distinguish a feature that is absent from one that is present but concealed; from outside, the two states are indistinguishable. Second, the relevant capability is well documented at the level of the industry, since dynamic difficulty has been patented explicitly to maximise engagement (Xue et al., 2017) and account-based behavioural models can predict harm-relevant states (Auer and Griffiths, 2022), but documented capability is not evidence of deployment in any audited title. The audit therefore does not show that these products personalise harmfully; it shows that, from outside, one cannot tell whether they do.

Title (category)	Chance-based	Prob. / pity disclosed	Time-limited	VIP / loyalty	Analytics	Personalised / contingent offers †	Risk / churn segmentation †
Genshin Impact (gacha RPG)	✓	✓	✓	~	✓	ND	ND

Title (category)	Chance-based	Prob. / pity disclosed	Time-limited	VIP / loyalty	Analytics	Personalised / contingent offers †	Risk / churn segmentation †
EA Sports FC – Ultimate Team (sports loot box)	✓	✓	✓	~	✓	ND	ND
Raid: Shadow Legends (gacha)	✓	~	✓	✓	✓	ND	ND
Coin Master (slot-loop mobile)	✓	ND	✓	~	✓	ND	ND
Candy Crush Saga (puzzle)	–	–	✓	–	✓	ND	ND
Clash Royale (chest-based)	✓	✓	✓	✓	✓	ND	ND
Roblox (UGC platform)	~	ND	✓	~	✓	ND	ND
Counter-Strike 2 (weapon cases)	✓	✓	✓	–	✓	ND	ND
Slotomania (social casino)	✓	ND	✓	✓	✓	ND	ND
Honkai: Star Rail (gacha RPG)	✓	✓	✓	~	✓	ND	ND

Table 2. Proof-of-concept transparency audit of ten high-grossing titles against the Section 5.1 indicators. ✓ = present and publicly verifiable; ~ = present but only inferable from documented industry capability; ND = not externally verifiable (covert); – = absent. † marks the adaptive, machine-zone-defining features. Codings reflect public information at the time of writing and require independent verification; the audit is a single-coder, desk-based proof of concept, not a definitive measurement.

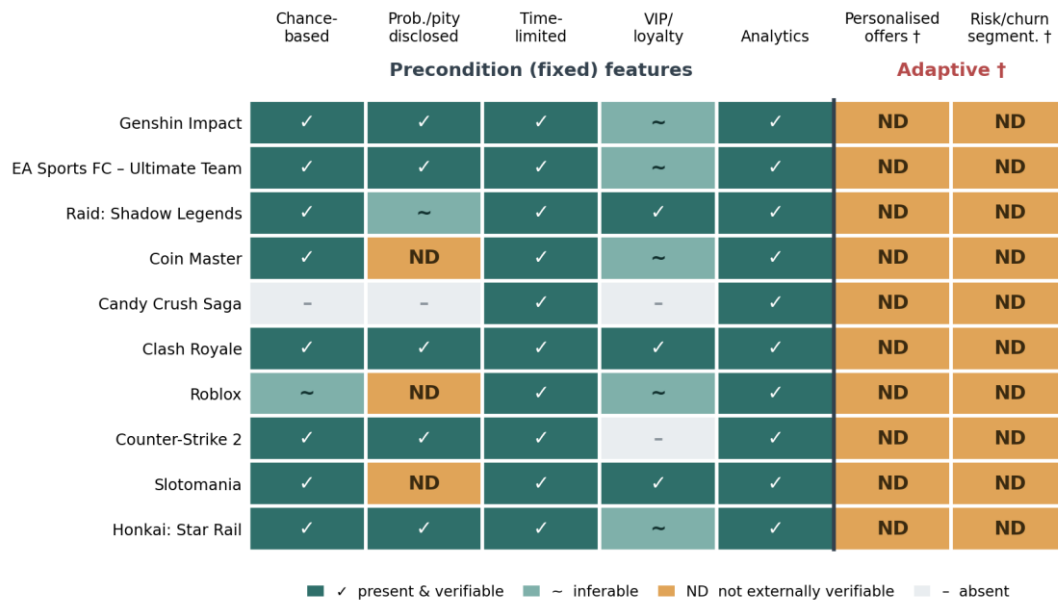


Figure 2. The observability gradient. Audit codings from Table 2 rendered as a heatmap: precondition features (left) are largely present and externally verifiable, whereas the adaptive, machine-zone-defining features (right of the divider) are uniformly not externally verifiable across the sample.

6.3 Interpretation

The single safest statement of what this section establishes is this: the pilot does not show that adaptive harm is occurring; it shows that current transparency regimes make the relevant question unanswerable from outside. With that framing fixed, the pilot does two things. It shows that the instrument is operable: every indicator could be coded, the precondition group

could be discriminated from products that lack it, and the precondition/adaptive distinction of Section 5.1 maps onto a real divide in what can and cannot be observed (Figure 2). And it demonstrates, rather than merely asserting, the limit of external auditability: the features that define the adaptive machine zone are precisely those that current transparency regimes do not reach. This observability gradient is a claim about disclosure, not about deployment. It does not show that the adaptive machine zone is pervasive, nor even that it is present in any particular title, since, as noted, the ND coding cannot separate the absent from the concealed. What it shows is narrower, and still consequential: that the presence or absence of these features cannot be established from outside. The structure of this opacity is itself informative. Three explanations are compatible with the pattern, and they are not mutually exclusive. The opacity may be commercially rational concealment, since disclosing that offers are keyed to behavioural states would invite both reputational cost and regulatory attention; it may be legitimate trade-secret protection of engagement systems that confer competitive advantage; or it may simply reflect the absence of any regulatory requirement to disclose, in which case non-disclosure is the default rather than a choice. Which explanation dominates matters for remedy: concealment calls for audit powers, trade secrecy for protected disclosure regimes of the kind used in financial and pharmaceutical regulation, and default silence for straightforward disclosure mandates. What no explanation supports is the status quo, in which the question cannot be asked. It might be objected that, since the audit cannot show these features are deployed, the framework is a solution in search of a problem. The reply is that the conjunction at issue, a documented and in places patented capability to maximise engagement dynamically (Xue et al., 2017) together with the complete covertness of its deployment, is not a missing data point but a harm in its own right: it is an active barrier to the epistemic conditions of democratic accountability, foreclosing the very inquiry by which a public could establish whether it is being exploited. A phenomenon that exists as a documented industry capability, yet is opaque as deployed practice, cannot be reached by external research or audit without mandated access, which is the condition the distributed argument identifies as a governance failure and which Section 8 develops.

7. Distributed agency and responsibility under causal uncertainty

If the addictive process is constituted across the human–machine system, our vocabulary for agency and responsibility must follow it there. The claim required is deliberately modest, and weaker than the constitutive thesis of the extended mind (Clark and Chalmers, 1998): in the spirit of situated and distributed accounts of cognition, the behaviour to be explained,

sustained, escalating, self-defeating engagement, is a property of the coupled person-and-system and is not adequately characterised by reference to either component alone. Schüll's ethnography already describes exactly this at the gambling machine; the adaptive case strengthens the coupling, because the system now adjusts itself to the person. Nothing in what follows depends on any stronger metaphysical thesis; the normative work is done by the analysis of responsibility to which we now turn.

Once agency is understood as distributed, responsibility can be redistributed without being dissolved. It helps to separate three notions usually run together: causal responsibility (what produced the outcome), capacity responsibility (whether an agent had the relevant abilities), and what Hart (1968) terms liability-responsibility. I depart from Hart's label deliberately, calling this accountability responsibility: Hart's term was coined for legal contexts and invites a doctrinal reading, whereas the concept needed here is the broader normative one of answerability, the obligation to prevent, justify and remedy harm, of which legal liability is only one institutional expression. The sense intended is ethical rather than administrative: an obligation to justify one's design choices to those they affect, grounded in the agent's control over, and foresight of, the harm those choices risk, and not a matter of corporate-governance procedure or public-relations 'accountability' in the looser managerial sense. Translating that obligation into the appropriate legal mechanisms, whether product safety, consumer protection, statutory duties of care or platform regulation, is a task for the legal strand of the agenda set out below and not one this paper presumes to settle. The individualist settlement reads the gambler's or player's residual capacity responsibility, their retaining some agency and choice, as settling accountability responsibility, concluding that the harm is theirs to own. The distributed account breaks this inference. The individual may retain genuine, if degraded, capacity responsibility, which Pickard's framework rightly insists we honour. But causal responsibility is shared with, and in the adaptive case substantially held by, those who design, optimise and permit the choice architecture; and accountability responsibility should track causal contribution and the capacity to have done otherwise. The party best placed to have done otherwise is frequently the one that designed the system, not the one whose behaviour it is built to shape. This is a version of what the literature on socio-technical systems calls the problem of many hands: where an outcome is produced by many contributing agents, responsibility can appear to dissipate, leaving no one accountable (Thompson, 1980; van de Poel, Royakkers and Zwart, 2015). Adaptive systems sharpen this problem rather than merely instantiating it, because the contributing 'hands' now include continuously updating, opaque algorithms that

respond to the very behaviour they are shaping, so that the locus of decision is itself moving and partly automated. This is the setting in which a responsibility gap can open, where increasingly autonomous behaviour leaves no human straightforwardly answerable for the result (Matthias, 2004). The distributed account is precisely an attempt to resist that gap, by allocating accountability across the layers in proportion to control and foresight rather than allowing the diffusion of causation, or the interposition of an algorithm, to extinguish responsibility altogether.

This step must be taken carefully, because the inference from shared causation to shared accountability is precisely where arguments of this kind are usually too quick, and a familiar reply waits for them: operators merely enable, while individuals choose, so accountability remains with the chooser. The reply has real force, but it conflates two distinct objects of accountability. On standard accounts, accountability attaches to an agent in proportion to two conditions: control over the relevant state of affairs, and the capacity to foresee the harm that flows from it (Fischer and Ravizza, 1998; Hart, 1968). These conditions should be applied separately to the act and to the architecture. For the act of playing or spending, the individual ordinarily satisfies both conditions and is accountable, and nothing in the distributed account denies this. For the architecture (the schedule of rewards, the personalisation of offers, the placement of friction, the targeting of inducements) the individual has little control and little knowledge, whereas the operator has near-total control and detailed, often superior, knowledge of its effects. Accountability for the architecture therefore sits overwhelmingly with the party that designs it; and it is the architecture, not the isolated choice, that the distributed account identifies as the principal site of harm. The ‘operators enable, individuals choose’ reply succeeds only while the architecture is treated as neutral background. Once it is seen as itself designed, chosen and controlled by the operator, the reply dissolves: the distributed account does not abolish individual accountability but locates it at the level of the act, while assigning accountability for the designed conditions of action to those who design them. This is also why the redistribution does not over-reach into every product that shapes behaviour: it follows the two conditions, and bites only where control over, and foresight of, harm-relevant design sit decisively with the operator.

This is where the evidential situation that troubled the introduction becomes an asset rather than an embarrassment. A purely individualist, causal framing needs to show that the product causes disorder in the person before it can ground intervention; under associational evidence it stalls. A commercial-determinants framing does not need this. It treats reliable association

between a designed, profit-motivated product and population-level harm, together with a plausible mechanism grounded in documented cognitive exploitation, as sufficient to ground precautionary, system-level regulation, in the same way that public health acts on the commercial determinants of harm elsewhere without first proving individual-level causation for each consumer (van Schalkwyk et al., 2021; Maani, Petticrew and Galea, 2023). Distributing responsibility and adopting a precautionary, design-focused stance are thus two aspects of one move; the move is available precisely because the harm is co-produced. Responsibility without blame can correspondingly be reconstructed at the level of the system: the individual is held responsible in Pickard’s agency-affirming, non-stigmatising sense, while designers and operators bear the fuller, accountability-bearing responsibility, and regulators bear responsibility for the permissions they grant. This reconstruction has a practical payoff worth stating explicitly. Stigma is a well-established barrier to help-seeking for gambling-related harm, and much of that stigma flows from an explanatory story in which harm evidences a defect of character or will. By relocating part of the explanation from character to design, a distributed account gives people experiencing harm a truthful, less shaming self-understanding: their difficulty is partly the intended output of a system engineered around their data. To the extent that public narratives, screening conversations and treatment framings absorb this account, the framework should reduce self-blame and public stigma and thereby lower a documented barrier to seeking help, a prevention benefit that operates alongside, and independently of, any regulatory change. The rhetoric of ‘responsible play’ mislocates responsibility, asking the weaker, less-informed party in an asymmetrical and increasingly adaptive relationship to bear what causal analysis assigns to the stronger.

7.1 Objections

Five objections deserve direct answers. First, that distributing responsibility excuses the individual and erodes the agency recovery depends on. It does the opposite: following Pickard (2017), it retains capacity responsibility in full and denies only the inference from residual capacity to sole accountability; affirming the first does not require concentrating the second on the individual.

Second, that the appeal to distributed cognition over-reaches. As stated in Section 7’s opening, the argument invokes no constitutive thesis about the mind; it needs only the descriptive claim that the behaviour is a property of the coupled person-and-system, and the normative conclusion rests entirely on the analysis of causal, capacity and accountability responsibility.

Third, that this proves too much, since most consumer products encourage use and many exploit biases. What distinguishes the convergent product is a conjunction: a business model disproportionately dependent on harmful patterns of use, a mechanism that targets documented cognitive distortions, a personalised and adaptive optimisation against the individual, disproportionate exposure of children, and outcomes of recognised severity. It is the conjunction, not persuasive design as such, that warrants treating the product as the locus of regulation.

Fourth, and most testing, that operators merely conditionally enable harm rather than causing it, especially given associational evidence. This is a real difficulty, and it should be conceded plainly that the paper does not establish that operators cause harm in any individual case; the claim is weaker and explicitly precautionary. The commercial-determinants framing is built for precisely this situation: conditional enabling by a profit-motivated party that designs the enabling conditions, foreseeably profits from the resulting harm, and is best placed to alter those conditions attracts accountability responsibility on a precautionary basis, grounded in a plausible mechanism and reliable population-level association rather than in proof of individual causation. To insist on individual causation before acting is itself a normative choice, and one that happens to favour the better-resourced party; but the converse must also be granted, that the strength of the resulting obligation should scale with the strength of the evidence, so that the framework licenses proportionate precaution rather than unconditional liability. There is also a further reason the conditional-enabling reply understates the operator's role. The objection pictures a neutral enabler standing back while an autonomous user chooses, but the adaptive interface does not stand back: through personalised, behaviourally-keyed prompts and the anthropomorphic, social-seeming channels described in Section 5, it acts upon the user at the moment of decision and in proportion to that user's modelled susceptibility. Enabling that is targeted, asymmetric and exploitative of relational trust is closer to influence than to mere provision, which strengthens rather than dilutes the operator's share of accountability.

Fifth, that regulators cannot realistically audit proprietary adaptive systems. This is the most serious objection, and it must be partly conceded: absent mandated data access, the distributed framework remains in part aspirational, because its central empirical claims about adaptive targeting cannot be confirmed or refuted from the outside, and responsibility cannot be assigned with confidence to mechanisms that remain opaque. The right conclusion is not to abandon the framework but to treat transparency, access regimes and independent auditing not as desirable extras but as preconditions for its application, and as primary objects of the

research and policy agenda (Section 8), and to regard the present opacity of these systems as itself a governance failure rather than a neutral fact. A residual and fair objection is that ‘responsible play’ is not always wholly mislocated but incomplete: individual-level tools retain value as one component of a system-level regime, and demonstrably help some users. The distributed account accepts this, and the mechanism of its claimed improvement should be stated rather than assumed. Individual tools share a structural property: they require the user to deploy them, at the moment of need, against a system optimised in the opposite direction, which is why uptake is low and concentrated among the least affected. Design-level regulation operates upstream of that asymmetry: it removes or constrains the adaptive features themselves, so that protection does not depend on the depleted self-control of the person the system is modelling. The two are therefore complements with different failure modes, and the distributed claim is only that the upstream layer should be primary, not exclusive; relied upon as the principal response, individual tools mislocate the larger share of responsibility. A further sense of ‘responsibility’ deserves separate treatment, because the rhetoric of responsible play trades on it. Beyond the causal, capacity and accountability senses distinguished above lies an aretaic sense: responsibility as a character trait, an exhortation to be a temperate, self-governing kind of person. Industry and policy can invoke this sense even under a fully distributed account, since distributing causal and accountability responsibility does not obviously reduce the responsibility an agent ought to display. The reply is not to deny the virtue: temperance remains a virtue, and nothing here counsels abandoning it. It is rather that exhorting the virtue is subject to precisely the asymmetry just described for individual-level tools. Adaptive design operates on the conditions under which the trait can be exercised: a system that models fluctuating self-regulation and times its inducements to the moments when the capacity for temperance is weakest makes the exercise of the virtue systematically harder, for that user, at that moment. An exhortation to character cannot substitute for design accountability when the design is engineered to defeat the character being exhorted; the aretaic sense of responsibility is real, but it presupposes background conditions that adaptive systems are built to erode, and securing those conditions is exactly what the accountability sense assigns to operators and regulators.

7.2 The case for the defence

A framework of this kind should weigh the considerations that tell against it, not only the objections to its internal logic. Four are substantial. First, the great majority of people who play monetised games or who gamble are not harmed, and to describe the products as harm systems risks pathologising ordinary enjoyment. Second, these products deliver genuine goods

(entertainment, sociality, achievement, creative expression) that a harm-only framing simply omits. Third, the free-to-play economics that depend on a minority of high-spending users also subsidise free or low-cost access for the majority, so that constraints on monetisation carry distributive consequences that are not all benign. Fourth, individual-level tools such as voluntary limit-setting, self-exclusion and personalised feedback do appear to benefit some users, and to dismiss them wholesale would discard real protection.

These considerations bound the argument; they do not defeat it. The claim is not that monetised play is harmful as such, nor that enjoyment should be overridden, but that where a product crosses into the adaptive machine zone, personalising around behavioural states to sustain engagement in a domain where harm is concentrated and can be severe, the design of those features, and not only the conduct of the affected minority, is a proper object of responsibility and regulation. This also answers the worry that the argument licenses unlimited precaution against anything correlated with any harm. It does not, because the trigger is not correlation alone but a conjunction: a documented mechanism of cognitive exploitation, behavioural personalisation keyed to the individual, dependence of the business model on heavy or harmful patterns of use, and concentrated or severe harm, frequently among the young. A product that is merely correlated with harm, but that lacks a documented exploitative mechanism, is not behaviourally personalised, or is not commercially dependent on harmful use, does not meet that threshold. The framework is in this sense deliberately bounded: it does not bite on cosmetic, non-randomised or non-personalised monetisation, nor on optional purchases in single-player games that lack adaptive targeting; it does not deny the value of the activity to the unharmed majority; and it treats individual-level tools as a genuine, if insufficient, component of a system-level response rather than as worthless. Finally, because the audit cannot establish deployment, the argument is explicitly conditional: it identifies where responsibility lies if and where the adaptive machine zone is present, and establishing its presence in a given product is the task that mandated access would enable.

8. Implications for the research and policy agenda

Although framed here partly through current UK policy, the implications generalise, and other jurisdictions illustrate the range of available responses: Belgium has treated paid loot boxes as gambling under existing law, effectively prohibiting them; the Netherlands has pursued enforcement actions on similar grounds; China mandates probability disclosure; and consumer-protection analyses for the European Parliament have set out regulatory options short of prohibition (Cerulli-Harms et al., 2020; Xiao, 2024). What these divergent regimes share is

a focus on the fixed, observable features; none yet reaches the adaptive layer this paper foregrounds, which is why the argument travels. Independent, prevention-oriented research funding, meaning the regimes through which research on gambling harm is commissioned and financed independently of industry, of which the levy-funded UK Research Programme on Gambling is one example, provides an opportunity to act on a distributed analysis, because such funding is insulated from the industries whose products are at issue. Three implications follow.

First, prevention should be conceived substantially as product and design regulation, framed through commercial determinants of health, rather than as consumer education whose effectiveness when industry-funded is itself contested (van Schalkwyk et al., 2021). For convergent products this implies attention to the permissible design space: constraints on chance-based monetisation aimed at minors, transparency of drop rates and expected cost, limits on personalised inducements and on adaptive friction, age-rating and age-assurance requirements, and scrutiny of influencer marketing that channels young audiences towards wagering. The regulatory logic Yeung (2017) calls ‘regulation by design’ can be turned to protective ends as a safer-by-design standard, with the burden of demonstrating that a design does not exploit known vulnerabilities placed on the operator.

Second, the predictive capabilities operators already hold should be treated as a regulated public-interest resource rather than a proprietary engagement tool. The behavioural modelling that can predict harm (Auer and Griffiths, 2022) could, under independent governance and within secure data environments, trigger mandatory protective responses; left in commercial hands it is an instrument of the adaptive machine zone. Who holds the model, to what end it is optimised, and to whom it answers are precisely the questions a distributed analysis foregrounds. They also raise a concrete governance question that the framework cannot dodge: who can compel access to proprietary models and data? Plausible mechanisms include a statutory regulator with audit powers, conditions attached to operating licences or to levy-style research funding, and emerging platform-transparency duties; specifying and comparing these access regimes is itself a priority for the legal and policy strand of the agenda below.

Third, the analysis points to a distinctive and genuinely interdisciplinary methodological contribution. Clarifying the specificity, direction and magnitude of convergence associations, the first-order empirical priority, requires longitudinal and quasi-experimental designs from psychology and behavioural science: for example, cohort studies following adolescents and young adults over twelve to twenty-four months that measure loot box and gacha expenditure,

problem-gambling severity and problem-gaming severity, in order to adjudicate gateway, selection and bidirectional hypotheses. Understanding monetisation and governance requires economics, law and policy analysis under a commercial-determinants frame; and characterising the adaptive interface requires computer science and game design. Methods from human–computer interaction can be brought to bear in three complementary strands: a structured audit of convergent products that applies the operational indicators of Section 5.1 through a documented coding framework, mapping adaptive and social design features to the vulnerabilities they recruit; controlled experiments that manipulate social and adaptive interface characteristics, the presence of a conversational host, the timing of personalised offers, the level of friction, using established instruments for trust, anthropomorphism and perceived agency (Reeves and Nass, 1996; Nass and Moon, 2000), with pre-registered outcomes such as spending intention, persistence, loss-chasing and perceived control, so as to isolate the causal contribution of specific design features to harm-relevant behaviour; and participatory work with people with lived experience, including young people, to characterise the phenomenology of the adaptive machine zone from within and to shape the constructs and interventions used in the other strands. Such a programme of work would convert the conceptual claim advanced here into a falsifiable, policy-relevant agenda, and would answer to the wider international priority of grounding regulation in independent evidence and in the experience of those affected.

9. Limitations

Several limitations should be stated plainly. First, the paper is a conceptual synthesis, not a systematic review; its selection of sources is purposive, and a comprehensive evidence map of the convergence literature would refine its claims. Second, the illustrative composite in Section 5 is exactly that: a synthesis of separately documented features, not evidence that any single product combines them as described, and the strength of evidence for each adaptive mechanism varies, with individualised odds manipulation in particular treated here as an untested hypothesis. Third, the distributed and commercial-determinants framing is normatively committed; it can be resisted by those who hold that individual causation is a precondition of accountability, and the paper argues against that position rather than presupposing its falsity. Fourth, the framework applies unevenly across product types: the adaptive machine zone is likely to be most pronounced in continuous, data-rich products such as online slots, social casino and live-service gacha games, and weaker in others. In particular, it has least purchase on games whose monetisation is purely cosmetic, non-randomised, or not

behaviourally personalised, where the gambling-like and adaptive elements that the framework targets are largely absent; its explanatory reach should therefore be calibrated to the product rather than assumed uniform. Fifth, the proof-of-concept audit in Section 6 is a single-coder, desk-based classification of public information at one time point, without inter-rater reliability or primary play data; it demonstrates the instrument and the observability gradient but does not measure prevalence, and a fuller audit with multiple coders, account-level access and longitudinal sampling would be required to move from demonstration to evidence. None of these qualifications, however, disturbs the central claim: that for convergent, adaptive products the addictive process and its attendant responsibility are distributed, and that the prevailing individualist ontology and the rhetoric of responsible play do not describe them well.

10. Conclusion

Gambling and gaming are now both recognised as behavioural addictions just as their products converge, and the evidence on that convergence is reliably associational but not, or not yet, causal. Read through an individualist lens, this evidential caution counsels inaction; read through the co-production and commercial-determinants framing developed here, it counsels a different and firmer response. Contemporary convergent products are, to varying degrees, adaptive socio-technical systems that model the individual user and reconfigure themselves to sustain engagement, the adaptive machine zone, and to the extent that they do, agency and responsibility are distributed. Holding the individual responsible in a non-blaming, agency-affirming sense is compatible with, and indeed requires, assigning the fuller responsibility to those who design, operate and permit the system. For any independent, prevention-oriented regime, this implies a shift of emphasis from the responsabilisation of players towards the regulation of adaptive design, and an opening for interdisciplinary research, with human–computer interaction and game design alongside psychology, public health, law and economics, equal to the systems now in question. The normative takeaway is twofold, and ordered. Because the audit shows that the morally decisive features are currently unobservable from outside, mandated transparency and disclosure of behaviourally contingent design, personalised inducements, loss-contingent prompts, friction modulation and risk-based segmentation, is the immediate precondition; without it, accountability cannot even be assessed. But transparency is a means, not the end. Its purpose is to make the deeper standard enforceable: a safer-by-design regime in which the burden falls on operators to show that adaptive features do not exploit known vulnerabilities. Transparency is what makes design accountability possible; design accountability is what transparency is for. The task is therefore

not only to ask why some individuals struggle to stop, but also to examine how product systems may be designed in ways that make stopping harder, and to consider where responsibility should then lie.

Statements and Declarations

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests. The authors declare no competing financial or non-financial interests. The work received no funding or other support from the gambling, gaming, or related industries.

Ethics approval. Not applicable. The study involved no human or animal participants; the transparency audit reported in Section 6 is a desk-based documentary analysis of publicly available information.

Consent to participate / consent for publication. Not applicable.

Data availability. No new datasets were generated beyond the coding reported in the manuscript. The coding protocol and the public sources underlying the audit are documented in Appendix A.

Author contributions. Mriganka Biswas: Conceptualisation, Methodology, Formal analysis, Investigation, Writing (original draft), Writing (review and editing). Safina Ara: Conceptualisation, Writing (review and editing).

References

- Aarseth, E., Bean, A.M., Boonen, H. et al. (2017) ‘Scholars’ open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal’, *Journal of Behavioral Addictions*, 6(3), pp. 267–270.
- Ainslie, G. (2013) ‘Money as MacGuffin’, in Levy, N. (ed.) *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience*. New York: Oxford University Press, pp. 16–37.
- American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders*. 5th edn. Arlington, VA: American Psychiatric Publishing.
- Auer, M. and Griffiths, M.D. (2022) ‘Using artificial intelligence algorithms to predict self-reported problem gambling with account-based player data in an online casino setting’, *Journal of Gambling Studies*, 38, pp. 1273–1294.

- Brooks, G.A. and Clark, L. (2019) 'Associations between loot box use, problematic gaming and gambling, and gambling-related cognitions', *Addictive Behaviors*, 96, pp. 26–34.
- Brooks, G.A. and Clark, L. (2023) 'The gamblers of the future? Migration from loot boxes to gambling in a longitudinal study of young adults', *Computers in Human Behavior*, 141, 107605.
- Cerulli-Harms, A., Münsch, M., Thorun, C., Michaelsen, F. and Hausemer, P. (2020) *Loot Boxes in Online Games and Their Effect on Consumers, in Particular Young Consumers*. Luxembourg: European Parliament, Policy Department for Economic, Scientific and Quality of Life Policies (IMCO).
- Clark, A. and Chalmers, D. (1998) 'The extended mind', *Analysis*, 58(1), pp. 7–19.
- Clark, L. (2010) 'Decision-making during gambling: an integration of cognitive and psychobiological approaches', *Philosophical Transactions of the Royal Society B*, 365(1538), pp. 319–330.
- Clark, L. (2014) 'Disordered gambling: the evolving concept of behavioral addiction', *Annals of the New York Academy of Sciences*, 1327(1), pp. 46–61.
- Clark, L., Lawrence, A.J., Astley-Jones, F. and Gray, N. (2009) 'Gambling near-misses enhance motivation to gamble and recruit win-related brain circuitry', *Neuron*, 61(3), pp. 481–490.
- Department for Culture, Media and Sport (2023) *Video Games Research Framework*. London: Department for Culture, Media and Sport.
- Drummond, A. and Sauer, J.D. (2018) 'Video game loot boxes are psychologically akin to gambling', *Nature Human Behaviour*, 2, pp. 530–532.
- Eyal, N. (2014) *Hooked: How to Build Habit-Forming Products*. New York: Portfolio/Penguin.
- Fischer, J.M. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Flanagan, O. (2013) 'Phenomenal authority: the epistemic authority of Alcoholics Anonymous', in Levy, N. (ed.) *Addiction and Self-Control*. New York: Oxford University Press, pp. 67–93.
- Foddy, B. and Savulescu, J. (2010) 'A liberal account of addiction', *Philosophy, Psychiatry, & Psychology*, 17(1), pp. 1–22.
- Frankfurt, H.G. (1971) 'Freedom of the will and the concept of a person', *Journal of Philosophy*, 68(1), pp. 5–20.
- Garea, S.S., Drummond, A., Sauer, J.D., Hall, L.C. and Williams, M.N. (2021) 'Meta-analysis of the relationship between problem gambling, excessive gaming and loot box spending', *International Gambling Studies*, 21(3), pp. 460–479.
- Gilmore, A.B., Fabbri, A., Baum, F. et al. (2023) 'Defining and conceptualising the commercial determinants of health', *The Lancet*, 401(10383), pp. 1194–1213.

- Griffiths, M. (2005) 'A "components" model of addiction within a biopsychosocial framework', *Journal of Substance Use*, 10(4), pp. 191–197.
- Hart, H.L.A. (1968) *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Clarendon Press.
- Heyman, G.M. (2009) *Addiction: A Disorder of Choice*. Cambridge, MA: Harvard University Press.
- Hing, N., Russell, A.M.T. and Gainsbury, S.M. (2020) 'Avoiding use of stigmatising descriptors in gambling studies', *International Gambling Studies*, 20(3), pp. 369–372.
- Kennett, J. (2013) 'Just say no? Addiction and the elements of self-control', in Levy, N. (ed.) *Addiction and Self-Control*. New York: Oxford University Press, pp. 144–164.
- Kim, H.S., Son, G., Roh, E.-B. et al. (2022) 'Prevalence of gaming disorder: a meta-analysis', *Addictive Behaviors*, 126, 107183.
- King, D.L. and Delfabbro, P.H. (2018) 'Predatory monetization schemes in video games (e.g. "loot boxes") and internet gaming disorder', *Addiction*, 113(11), pp. 1967–1969.
- King, D.L. and Delfabbro, P.H. (2019) 'Video game monetization (e.g., "loot boxes"): a blueprint for practical social responsibility measures', *International Journal of Mental Health and Addiction*, 17(1), pp. 166–179.
- King, D.L., Delfabbro, P.H., Gainsbury, S.M., Dreier, M., Greer, N. and Billieux, J. (2019) 'Unfair play? Video games as exploitative monetized services: an examination of game patents from a consumer protection perspective', *Computers in Human Behavior*, 101, pp. 131–143.
- King, D.L. and Potenza, M.N. (2019) 'Not playing around: gaming disorder in the International Classification of Diseases (ICD-11)', *Journal of Adolescent Health*, 64(1), pp. 5–7.
- King, D.L., Russell, A., Gainsbury, S., Delfabbro, P.H. and Hing, N. (2016) 'The cost of virtual wins: an examination of gambling-related risks in youth who spend money on social casino games', *Journal of Behavioral Addictions*, 5(3), pp. 401–409.
- Langer, E.J. (1975) 'The illusion of control', *Journal of Personality and Social Psychology*, 32(2), pp. 311–328.
- Leshner, A.I. (1997) 'Addiction is a brain disease, and it matters', *Science*, 278(5335), pp. 45–47.
- Levy, N. (ed.) (2013) *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience*. New York: Oxford University Press.
- Li, W., Mills, D. and Nower, L. (2019) 'The relationship of loot box purchases to problem video gaming and problem gambling', *Addictive Behaviors*, 97, pp. 27–34.
- Maani, N., Petticrew, M. and Galea, S. (eds) (2023) *The Commercial Determinants of Health*. New York: Oxford University Press.

- Macey, J. and Hamari, J. (2018) 'Investigating relationships between video gaming, spectating esports, and gambling', *Computers in Human Behavior*, 80, pp. 344–353.
- Macey, J., Hamari, J. and Adam, M. (2024) 'A conceptual framework for understanding and identifying gamblified experiences', *Computers in Human Behavior*, 152, 108087.
- Matthias, A. (2004) 'The responsibility gap: ascribing responsibility for the actions of learning automata', *Ethics and Information Technology*, 6(3), pp. 175–183.
- McGarrigle, J., Smith, J., Griffiths, J., Torrance, J., Quigley, M. and Dymond, S. (2026) 'Dark patterns in online gambling: a scoping review and classification of deceptive design practices', *Journal of Behavioral Addictions*, 15(1), pp. 99–114.
- Nass, C. and Moon, Y. (2000) 'Machines and mindlessness: social responses to computers', *Journal of Social Issues*, 56(1), pp. 81–103.
- Ofcom (2024) *Children and Parents: Media Use and Attitudes*. London: Ofcom.
- Parke, J. and Griffiths, M.D. (2004) 'Gambling addiction and the evolution of the “near miss”', *Addiction Research & Theory*, 12(5), pp. 407–411.
- Pickard, H. (2017) 'Responsibility without blame for addiction', *Neuroethics*, 10(1), pp. 169–180.
- Pickard, H. and Pearce, S. (2013) 'Addiction in context: philosophical lessons from a personality disorder clinic', in Levy, N. (ed.) *Addiction and Self-Control*. New York: Oxford University Press, pp. 165–189.
- Reeves, B. and Nass, C. (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge: Cambridge University Press.
- Reith, G. (1999) *The Age of Chance: Gambling in Western Culture*. London: Routledge.
- Reith, G. (2007) 'Gambling and the contradictions of consumption: a genealogy of the “pathological” subject', *American Behavioral Scientist*, 51(1), pp. 33–55.
- Reith, G. (2018) *Addictive Consumption: Capitalism, Modernity and Excess*. London: Routledge.
- Reith, G. and Wardle, H. (2022) 'The framing of gambling and the commercial determinants of harm: challenges for regulation in the UK', in Nikkinen, J., Marionneau, V. and Egerer, M. (eds) *The Global Gambling Industry: Structures, Tactics and Networks of Impact*. Wiesbaden: Springer Gabler, pp. 71–86.
- Ross, D. (2013) 'The picoeconomics of gambling addiction and supporting neural mechanisms', in Levy, N. (ed.) *Addiction and Self-Control*. New York: Oxford University Press, pp. 38–47.
- Schüll, N.D. (2012) *Addiction by Design: Machine Gambling in Las Vegas*. Princeton, NJ: Princeton University Press.

- Shibuya, A., Teramoto, M., Shoun, A. and Akiyama, K. (2019) 'Long-term effects of in-game purchases and event game mechanics on young mobile social game players in Japan', *Simulation & Gaming*, 50(1), pp. 76–92.
- Sinnott-Armstrong, W. (2013) 'Are addicts responsible?', in Levy, N. (ed.) *Addiction and Self-Control*. New York: Oxford University Press, pp. 122–143.
- Spicer, S.G., Nicklin, L.L., Uther, M., Lloyd, J., Lloyd, H. and Close, J. (2022) 'Loot boxes, problem gambling and problem video gaming: a systematic review and meta-synthesis', *New Media & Society*, 24(4), pp. 1001–1022.
- Susser, D., Roessler, B. and Nissenbaum, H. (2019) 'Online manipulation: hidden influences in a digital world', *Internet Policy Review*, 8(2).
- Thompson, D.F. (1980) 'Moral responsibility of public officials: the problem of many hands', *American Political Science Review*, 74(4), pp. 905–916.
- van de Poel, I., Royakkers, L. and Zwart, S.D. (2015) *Moral Responsibility and the Problem of Many Hands*. New York: Routledge.
- van Schalkwyk, M.C.I., Hawkins, B. and Petticrew, M. (2022) 'The politics and fantasy of the gambling education discourse: an analysis of gambling industry-funded youth education programmes in the United Kingdom', *SSM - Population Health*, 18, 101122.
- van Schalkwyk, M.C.I., Petticrew, M., Cassidy, R., Adams, P., McKee, M., Reynolds, J. and Orford, J. (2021) 'A public health approach to gambling regulation: countering powerful influences', *The Lancet Public Health*, 6(8), e614–e619.
- Wardle, H., Reith, G., Langham, E. and Rogers, R.D. (2019) 'Gambling and the health of the public: policy and action for harm prevention', *BMJ*, 365, 11807.
- World Health Organization (2019) *International Classification of Diseases, 11th Revision (ICD-11)*. Geneva: World Health Organization.
- Xiao, L.Y. (2024) 'Loot box state of play 2023: law, regulation, policy, and enforcement around the world', *Gaming Law Review*. doi:10.1089/glr2.2024.0006.
- Xiao, L.Y. and Lund, M.L. (2025) 'Non-compliance with and non-enforcement of UK loot box industry self-regulation on the Apple App Store: a longitudinal study on poor implementation', *Royal Society Open Science*, 12(5), 250704.
- Xue, S., Wu, M., Kolen, J., Aghdaie, N. and Zaman, K.A. (2017) 'Dynamic difficulty adjustment for maximized engagement in digital games', in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 465–471.
- Yeung, K. (2017) "'Hypernudge': Big Data as a mode of regulation by design", *Information, Communication & Society*, 20(1), pp. 118–136.
- Zendle, D. (2020) 'Beyond loot boxes: a variety of gambling-like practices in video games are linked to both problem gambling and disordered gaming', *PeerJ*, 8, e9466.

- Zendle, D. and Bowden-Jones, H. (2019) ‘Loot boxes and the convergence of video games and gambling’, *The Lancet Psychiatry*, 6(9), pp. 724–725.
- Zendle, D. and Cairns, P. (2018) ‘Video game loot boxes are linked to problem gambling: results of a large-scale survey’, *PLoS ONE*, 13(11), e0206767.
- Zendle, D., Meyer, R., Waters, S. and Cairns, P. (2020) ‘The prevalence of loot boxes in mobile and desktop games’, *Addiction*, 115(9), pp. 1768–1772.
- Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Appendix A. Audit coding protocol and sources

This appendix documents the coding protocol for the proof-of-concept audit reported in Section 6, in the interest of reproducibility. The audit is a desk-based documentary classification of public information, not primary play-testing, conducted by a single coder in May 2026; the codings are time-stamped rather than standing claims, since monetisation features change with updates.

Sampling frame. Ten titles were selected purposively to span the principal convergent categories rather than to estimate prevalence: gacha role-playing games (Genshin Impact, Honkai: Star Rail, Raid: Shadow Legends), sports loot-box modes (EA Sports FC Ultimate Team), chest-based mobile games (Clash Royale), slot-loop and social casino products (Coin Master, Slotomania), weapon-case economies (Counter-Strike 2), a user-generated-content platform (Roblox), and a deliberately included non-gambling-like control (Candy Crush Saga). All were drawn from titles consistently among the highest-grossing internationally during the audit window. The pool was defined as titles recurrently appearing in the top-grossing tiers of the major mobile storefront charts and the leading PC platforms in the months preceding the audit; the convergent categories were specified first, and for each category the most prominent exemplar by revenue and user base within that pool was selected. The selection is purposive and illustrative: no claim is made that each title is the uniquely best exemplar of its category, only that each is a leading and widely recognised instance of it.

Coding rules. Each indicator was assigned one of four codes. Present and publicly verifiable (✓) required a primary public source: an official in-game or publisher probability or pity disclosure, a store or platform listing, or a published privacy or data policy. Present but only inferable (~) was used where a feature is documented at industry level but not confirmed for the specific title. Not externally verifiable (ND) was reserved for features whose presence or absence cannot be determined from any public source, chiefly the adaptive, behaviourally-

keyed features; ND therefore records a limit of observation, not a finding of presence. Absent (–) was used where a feature could be positively established not to be present.

Sources for precondition features. Chance-based monetisation, probability or pity disclosure, and VIP or loyalty tiers were coded from official sources. HoYoverse, EA, Supercell and Valve publish drop-rate or pack-probability information for Genshin Impact and Honkai: Star Rail, EA Sports FC, Clash Royale and Counter-Strike 2 respectively, in several cases under regulatory requirement; social casino VIP schemes (Slotomania) and the slot-loop core of Coin Master are observable in-product and in published descriptions. Where probability information is not officially published (Coin Master, Slotomania, Roblox), the disclosure cell was coded ND. Collection of behavioural analytics was coded from each title’s published privacy and data policy, which uniformly disclose analytics collection. The prevalence context for chance-based monetisation draws on Zendle et al. (2020) and Xiao (2024; Xiao and Lund, 2025).

Sources for adaptive features. Personalised or loss-contingent offers and risk or churn segmentation were coded ND for every title. No publisher in the sample discloses whether, or how, behavioural models are used to personalise offers, time prompts to behavioural states, modulate friction, or segment users by risk. The existence of the underlying capability is documented at industry level (Xue et al., 2017; Auer and Griffiths, 2022), but its deployment in any specific title is not externally verifiable, which is the observability gradient reported in Section 6.2.

Verification note. Because monetisation and disclosure practices change frequently, each ✓ coding should be re-checked against current official sources before any reuse of the table; the codings are offered to demonstrate the instrument, not as a durable dataset.