

The Zuck-Chai Equation: A Unified Reward Function Across Reinforcement Learning, Financial Markets, and Human Behavior

David Shannon *Independent Researcher, Vancouver, Canada*

Draft — March 2026

Abstract

We identify a single mathematical structure — the risk-normalized prediction error — that appears independently across three domains traditionally studied in isolation: reinforcement learning, quantitative finance, and human behavioral neuroscience. We formalize this as the Zuck-Chai Equation:

$$R = \frac{O - E}{\rho + C}$$

where O = observed outcome, E = expected value, ρ = risk (variance/uncertainty), and C = cost of action. We demonstrate structural isomorphism between this equation and the Sharpe Ratio in finance, the normalized temporal difference error in RL, the dopaminergic reward prediction error scaled by uncertainty in neuroscience, and the attention allocation function in information theory. We argue this convergence is not coincidental but reflects a universal computation underlying decision-making and attention in any system — biological or artificial — that must allocate limited resources under uncertainty. We further demonstrate that pathological exploitation of this equation, specifically the minimization of the denominator toward zero, provides a formal account of algorithmic addiction as evidenced by recent litigation against major technology platforms.

Keywords: reward function, prediction error, attention, Sharpe ratio, reinforcement learning, algorithmic addiction, consciousness

1. Introduction

A recurring pattern emerges across disciplines that study decision-making under uncertainty. In each case, an agent — whether a neuron, a trader, or a software agent — must decide where to direct limited resources. Each domain has independently derived a solution with identical mathematical structure: compute the deviation of an observed outcome from expectation, then normalize by the cost and risk of obtaining that observation.

Despite this structural convergence, no formal unification exists in the literature. Financial economists cite the Sharpe Ratio. Reinforcement learning researchers use temporal difference error with exploration bonuses. Neuroscientists describe dopaminergic reward prediction errors modulated by uncertainty. Information theorists invoke surprise normalized by entropy. Each community treats their formulation as domain-specific.

This paper proposes that these are not analogies but instances of a single equation. We call this the Zuck-Chai Equation — named not after its discoverers, but after the technology executives whose platforms most dramatically demonstrated its power over human attention, as evidenced by multi-billion dollar litigation alleging systematic exploitation of this mechanism at global scale (Meta Platforms and Google/Alphabet addiction lawsuits, 2023-2026).

The paper proceeds as follows. Section 2 formalizes the equation and defines each term. Sections 3-5 demonstrate its instantiation in reinforcement learning, quantitative finance, and human behavioral neuroscience respectively. Section 6 presents the addiction case as empirical validation. Section 7 discusses implications for theories of attention and consciousness. Section 8 concludes.

2. The Zuck-Chai Equation

2.1 Formal Definition

For any decision-making agent operating under uncertainty, we define the reward signal as:

$$R = \frac{O - E}{\rho + C}$$

Where:

- **O (Outcome):** The observed result of an action or event. This is the realized value — what actually happened.
- **E (Expectation):** The agent's predicted outcome prior to observation. This is the baseline — what the agent believed would happen.
- **ρ (Risk):** The variance or uncertainty associated with the outcome distribution. This captures how unpredictable the environment is.
- **C (Cost):** The resource expenditure required to obtain the observation. This includes time, energy, capital, or computational resources.

2.2 Properties

Property 1: Surprise Sensitivity. The numerator (O - E) is a signed prediction error. Positive values indicate better-than-expected outcomes; negative values indicate worse-than-expected outcomes. When O = E exactly, the reward signal is zero regardless of risk or cost — no learning occurs when the world behaves as predicted.

Property 2: Risk Normalization. Dividing by ρ ensures that a given prediction error in a volatile environment produces a smaller reward signal than the same prediction error in a stable environment. This prevents overreaction to noise.

Property 3: Cost Sensitivity. Dividing by C ensures that freely obtained information produces stronger reward signals than expensive information, all else equal. This creates a natural preference for low-cost learning.

Property 4: Pathological Regime. As $(\rho + C) \rightarrow 0$, even small prediction errors produce unbounded reward signals. This is the formal condition for addiction (Section 6).

2.3 Relation to Existing Formulations

Domain	Standard Form	Zuck-Chai Mapping
Finance (Sharpe)	$(R_p - R_f) / \sigma_p$	$O = R_p, E = R_f, \rho = \sigma_p, C \approx 0$
RL (TD Error)	$r + \gamma V(s') - V(s)$	$O = r + \gamma V(s'), E = V(s), \rho + C = \text{normalization}$
Neuroscience (RPE)	$\delta = r - V(s)$	$O = r, E = V(s), \rho = \text{neural uncertainty}$
Information Theory	$-\log P(x) / H(X)$	$O - E \approx \text{surprisal}, \rho + C \approx \text{entropy}$
Signal Processing	signal / noise	$O - E = \text{signal}, \rho = \text{noise}, C = \text{measurement cost}$

3. Instantiation in Reinforcement Learning

3.1 Temporal Difference Error

The foundational update rule in RL is the TD error (Sutton & Barto, 1998):

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

This is precisely the numerator of the Zuck-Chai Equation: outcome minus expectation. The observed outcome is the immediate reward plus discounted future value. The expectation is the

current value estimate.

3.2 Exploration Bonuses and Curiosity-Driven RL

Modern RL introduces intrinsic motivation signals that reward agents for visiting novel states (Pathak et al., 2017; Burda et al., 2018). The curiosity reward is typically formulated as prediction error of a forward dynamics model — how surprised was the agent by what happened?

The exploration-exploitation tradeoff is governed by weighing this surprise signal against the cost of exploration (potential suboptimal actions) and the risk of the environment (stochastic transitions). This maps directly to the Zuck-Chai structure:

$$R_{intrinsic} = \frac{\text{prediction error of forward model}}{\text{exploration cost} + \text{environment stochasticity}}$$

3.3 Risk-Sensitive RL

Recent work in risk-sensitive RL (Tamar et al., 2015) explicitly incorporates variance into the objective function. Rather than maximizing expected return, agents maximize a risk-adjusted return — the Sharpe Ratio of cumulative reward. This is the Zuck-Chai Equation applied as the objective function itself, not merely the reward signal.

4. Instantiation in Quantitative Finance

4.1 The Sharpe Ratio

The Sharpe Ratio (Sharpe, 1966) is the standard measure of risk-adjusted return:

$$S = \frac{R_p - R_f}{\sigma_p}$$

Where R_p is portfolio return, R_f is the risk-free rate, and σ_p is portfolio standard deviation.

The mapping is immediate:

- $O = R_p$ (what the portfolio actually returned)
- $E = R_f$ (what you would have earned doing nothing — the baseline expectation)
- $\rho = \sigma_p$ (how volatile the returns were)
- C is implicitly zero or absorbed into R_f (transaction costs, opportunity costs)

4.2 The Information Ratio

The Information Ratio generalizes the Sharpe Ratio for active management:

$$IR = \frac{R_p - R_b}{\sigma_{p-b}}$$

Where R_b is the benchmark return and $\sigma(p-b)$ is tracking error. This is the same equation with a different baseline — the expectation is now "what would a passive strategy have returned?" rather than the risk-free rate.

4.3 Kelly Criterion Connection

The Kelly Criterion for optimal bet sizing:

$$f^* = \frac{bp - q}{b}$$

Where b = odds, p = probability of winning, q = probability of losing. This can be rewritten as:

$$f^* = \frac{E[\text{gain}] - E[\text{loss}]}{\text{maximum downside}}$$

Again: (outcome - expectation) / risk.

5. Instantiation in Human Behavioral Neuroscience

5.1 Dopaminergic Reward Prediction Error

Schultz et al. (1997) demonstrated that midbrain dopamine neurons encode reward prediction errors — the difference between received and expected reward. This is the numerator of the Zuck-Chai Equation implemented in biological hardware.

Crucially, subsequent research (Fiorillo et al., 2003) showed that dopamine neurons also encode uncertainty about predictions. Neurons showed sustained activation proportional to reward uncertainty during the delay period between cue and outcome. This suggests the brain computes both components: the prediction error AND the risk normalization.

5.2 Precision-Weighted Prediction Errors

The Free Energy framework (Friston, 2009) formalizes perception and action as minimization of prediction error weighted by precision (inverse variance). The update rule is:

$$\Delta\mu = \frac{\text{prediction error}}{\text{precision}^{-1}} = \frac{O - E}{\sigma^2}$$

This is the Zuck-Chai Equation where $\rho = \sigma^2$ (variance) and C is the metabolic cost of neural computation. The brain literally implements risk-normalized surprise as its fundamental update rule.

5.3 Attention as Resource Allocation

Attention in cognitive neuroscience is understood as competitive resource allocation (Desimone & Duncan, 1995). Stimuli compete for processing based on their salience — defined as how much they deviate from background expectation, normalized by noise.

The Zuck-Chai Equation formalizes this: attention flows to the stimulus that maximizes R . The numerator captures salience (how surprising is this stimulus?). The denominator captures cost (how expensive is it to process?). Attention is the solution to the optimization problem: allocate resources to maximize total R across all available stimuli.

6. The Addiction Case: Empirical Validation at Scale

6.1 The Pathological Regime

Recall Property 4: as $(\rho + C) \rightarrow 0$, even small prediction errors produce unbounded reward signals. Social media platforms create exactly this condition:

- **Cost $\rightarrow 0$:** Infinite scroll eliminates action cost. The next stimulus requires only a thumb movement. There is no paywall, no time gate, no friction.
- **Risk $\rightarrow 0$:** Algorithmic curation eliminates uncertainty. The feed is personalized to reliably deliver content that triggers prediction errors for each individual user. The variance of "will this be interesting?" approaches zero because the algorithm already knows it will be.
- **Numerator stays positive:** Content is selected to maximize $(O - E)$ — each item is slightly more novel, outrageous, or emotionally activating than the user's adapted baseline.

The result: $R \rightarrow \infty$. The attention system cannot disengage because no alternative stimulus offers a comparable ratio.

6.2 Contrast with Non-Addictive Systems

Traditional media has natural denominators:

- **Books:** High cost (sustained attention, cognitive effort). The Zuck-Chai ratio self-limits.

- **Television:** Moderate cost (must wait for scheduled programming, watch advertisements). Denominator stays nonzero.
- **Gambling:** High financial risk. The denominator is large, which is why gambling addiction requires a specific predisposition — most people's risk sensitivity keeps R bounded.
- **Social media:** Near-zero cost, near-zero risk, algorithmically maximized numerator. The ratio explodes for everyone, not just predisposed individuals.

6.3 Legal Evidence

The addiction lawsuits against Meta Platforms (2023-ongoing) and Google/Alphabet (2023-ongoing) allege that these companies deliberately engineered their platforms to maximize engagement through mechanisms including infinite scroll, autoplay, intermittent variable reward schedules, and personalized content ranking. Each of these mechanisms maps to a specific term in the Zuck-Chai Equation:

Platform Feature	Zuck-Chai Effect
Infinite scroll	$C \rightarrow 0$ (eliminates action cost)
Algorithmic feed	Maximizes $(O - E)$ per user
Autoplay	$C \rightarrow 0$ (removes choice cost)
Variable reward schedule	Maintains $\rho > 0$ just enough to sustain novelty
Push notifications	Injects external $(O - E)$ signals at zero cost

7. Implications

7.1 A Universal Decision-Making Primitive

The convergence of identical mathematical structure across RL, finance, neuroscience, and information theory suggests this is not a coincidence of formalism but a reflection of a universal computational constraint. Any system that must allocate finite resources to maximize information acquisition under uncertainty will converge on this ratio.

7.2 Connection to Attention and Consciousness

If the Zuck-Chai Equation is the fundamental computation of attention allocation, and if attention is constitutive of (rather than merely correlated with) conscious experience, then this equation describes the mechanism by which consciousness selects its contents. This connects to

the ACU (Attention Computing Unit) framework (Shannon, 2026), which proposes the identity $A \equiv C \equiv U$: Attention equals Consciousness equals Universe. The Zuck-Chai Equation would then be the operational definition of the ACU — the specific computation that each unit of attention performs.

7.3 Testable Predictions

1. **RL agents** using Zuck-Chai normalized rewards should converge faster than agents using raw TD error in high-variance environments.
 2. **Trading strategies** that explicitly optimize the Zuck-Chai Ratio (generalizing the Sharpe Ratio to include transaction costs in the denominator) should outperform standard Sharpe-optimized strategies.
 3. **Neural recordings** should show that the magnitude of dopaminergic response to a given prediction error is inversely proportional to both environmental uncertainty AND metabolic cost of processing.
 4. **Platform design** interventions that increase C (adding friction) or ρ (reducing algorithmic precision) should measurably reduce addictive engagement, as predicted by the equation.
-

8. Cross-Domain Transferability: If $A = B = C$, Then Engineering in A Implies Engineering in B and C

8.1 The Transferability Thesis

If the Zuck-Chai Equation is not merely analogous but structurally identical across domains, a strong implication follows: optimization techniques developed in one domain should be directly transferable to any other domain governed by the same equation.

Formally: if domains A (human attention), B (reinforcement learning), and C (financial markets) all instantiate $R = (O - E) / (\rho + C)$, then engineering solutions that successfully optimize R in domain A should, with appropriate translation of terms, optimize R in domains B and C.

This is not a loose metaphor. It is a logical consequence of the claimed isomorphism. If the isomorphism holds, then the transferability is guaranteed by the structure of the mathematics itself.

8.2 Case Study: From Attention Engineering to Quantitative Trading

Meta Platforms and Google did not merely discover the Zuck-Chai Equation in human attention. They built industrial-scale infrastructure to optimize each of its terms in real time:

- **Baseline learning:** Algorithms that continuously model each user's expectation E and adapt content delivery to maximize the numerator $(O - E)$ relative to that shifting baseline.
- **Cost minimization:** Interface design (infinite scroll, autoplay, one-tap interactions) that drives $C \rightarrow 0$, ensuring the denominator remains small.
- **Risk calibration:** Personalization engines that control ρ precisely — maintaining enough variance to sustain novelty while reducing uncertainty sufficiently that users remain engaged rather than overwhelmed.

These are not ad hoc design choices. They constitute a systematic, real-time optimization framework for the Zuck-Chai Equation.

Now consider the identical framework applied to quantitative finance:

Attention Engineering (Meta/Google)	Quantitative Trading (Transposed)
Model each user's expectation baseline E	Model the market's expected price behavior E
Serve content maximizing $O - E$ (surprise)	Identify trades where $O - E$ is maximized (alpha generation)
Minimize interaction cost $C \rightarrow 0$	Minimize transaction costs, slippage, latency
Calibrate ρ to sustain engagement	Model volatility ρ , trade only when ratio is favorable
Real-time feedback loop updating E	Real-time market data updating E
Personalization per user	Strategy adaptation per asset/regime

The architectural parallels are exact. The personalization engine that selects content for a user is structurally identical to a trading agent that selects positions in a market. Both are solving the same optimization problem: maximize $R = (O - E) / (\rho + C)$ over a sequence of decisions with a continuously updating baseline.

8.3 Implications for RL Agent Design

This transferability has immediate engineering implications. Reinforcement learning agents designed for financial markets can borrow directly from the engagement optimization literature:

1. **Adaptive baseline estimation:** Social media algorithms excel at tracking non-stationary user preferences. The same techniques (online learning, contextual bandits) apply to tracking non-stationary market regimes.
2. **Cost-aware action selection:** Platform design obsessively minimizes user friction. Trading agents should similarly optimize not just for expected return but for the full Zuck-Chai

Ratio, incorporating transaction costs and market impact directly into the reward signal.

3. **Uncertainty-calibrated exploration:** Social media platforms modulate content novelty to maintain engagement without overwhelming users. Trading agents should similarly calibrate exploration intensity to market volatility — exploring more in stable regimes (where ρ is low and the denominator is small, making exploration cheap) and exploiting more in volatile regimes (where ρ is high and only large prediction errors justify action).

8.4 Generalization: A Universal Optimization Architecture

The broader claim is that any system optimizing the Zuck-Chai Equation — regardless of domain — requires the same four engineering components:

1. A **baseline model** that continuously estimates E from sequential observations.
2. A **surprise detector** that computes $O - E$ in real time.
3. A **cost minimizer** that reduces C for each decision cycle.
4. A **risk estimator** that tracks ρ and normalizes the surprise signal accordingly.

These four components constitute a universal optimization architecture. Meta built it for attention. A quant fund builds it for markets. A defense AI system builds it for threat detection — where O is observed sensor data, E is predicted threat behavior, ρ is battlefield uncertainty, and C is the cost of deploying countermeasures.

The equation does not change. Only the domain-specific definitions of its terms change.

9. Conclusion

We have identified a single equation — $R = (O - E) / (\rho + C)$ — that unifies reward computation across reinforcement learning, quantitative finance, and human behavioral neuroscience. The structural isomorphism is exact, not analogical. Each domain independently derived the same solution to the same problem: how to value new information when resources are finite and outcomes are uncertain.

The equation's pathological regime, where the denominator approaches zero, provides the first formal mathematical account of algorithmic addiction — explaining why social media platforms are uniquely addictive compared to prior media technologies, and why this addiction is population-wide rather than limited to predisposed individuals.

Crucially, if the isomorphism holds, then engineering techniques transfer across domains. The real-time optimization infrastructure built by social media platforms to exploit human attention is architecturally identical to what is needed for autonomous trading agents and defense AI systems. This is not an analogy — it is a mathematical consequence of the unified equation.

We propose that this convergence reflects a universal computation underlying attention allocation in any decision-making system, and that the four-component optimization architecture (baseline model, surprise detector, cost minimizer, risk estimator) constitutes the minimal engineering specification for any system that allocates finite resources under uncertainty.

The equation is named after Mark Zuckerberg and Sundar Pichai — not as its discoverers, but as the executives whose platforms provided the largest-scale empirical demonstration of its power and its pathology.

References

Burda, Y., Edwards, H., Storkey, A., & Klimov, O. (2018). Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1), 193-222.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614), 1898-1902.

Friston, K. (2009). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *ICML*.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.

Shannon, D. (2026). The David-Shannon Identity: $A \equiv C \equiv U$. *Manuscript in preparation*.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 39(1), 119-138.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Tamar, A., Di Castro, D., & Mannor, S. (2015). Policy gradients with variance related risk criteria. *ICML*.

Correspondence: [email to be added] The author declares no conflicts of interest. No funding was received for this research.