Self-Evolving AI Framework for Evaluating Wireless Authentication Robustness in WPA2/WPA3 Networks

Dheeraj Ramasahayam
Independent Researcher
Durham, North Carolina, USA
dheeraj@example.com

***Abstract*—Wireless authentication remains one of the most operationally critical and most poorly evaluated layers of Wi-Fi security. Existing WPA2/WPA3 assessment workflows are largely static: they estimate password strength from hand-crafted heuristics, replay a fixed cracking benchmark, or report protocol-level security properties without modeling how attacker pressure evolves over time. This paper introduces a self-evolving, feedback-driven framework for evaluating wireless authentication robustness in WPA2/WPA3 personal networks. The proposed system is label-free: it does not require ground-truth attack-success labels gathered from live compromise events. Instead, it combines password-feature extraction, protocol-aware adversarial simulation, contextual attack-pressure estimation, and an adaptive risk model that continuously updates which attack families are most informative for a given candidate secret. The framework explicitly distinguishes the feasibility regimes of WPA2-PSK and WPA3-SAE, allowing the evaluator to model the different economics of offline and online guessing pressure. To provide theoretical grounding, we formalize an Adaptive Weakness Convergence Theorem showing that, under bounded exploration of attack strategies, the estimated vulnerability score converges almost surely to the true attack-susceptibility functional induced by the attacker distribution. We further specify a reproducible evaluation protocol over RockYou-derived and synthetic password corpora, together with baselines including entropy-only scoring, rule-based checkers, and zxcvbn. The manuscript is structured as a publication-ready research draft: the theoretical and methodological claims are complete, while empirical result tables are intentionally left population-ready rather than numerically fabricated.**

***Index Terms*—wireless authentication, WPA2, WPA3, password robustness, label-free evaluation, contextual bandits, adversarial simulation, security analytics**

I. Introduction

The security of Wi-Fi networks often collapses to the robustness of the credential used to authenticate clients. In WPA2-Personal deployments, a weak pre-shared key can expose the network to inexpensive offline dictionary evaluation once a valid handshake is captured. WPA3-Personal improves this situation by replacing pre-shared-key authentication with Simultaneous Authentication of Equals (SAE), a Dragonfly-based password-authenticated key exchange that raises the cost of guessing and removes the classical WPA2 handshake replay model [1], [2]. However, WPA3 does not eliminate password risk; it changes the feasible attack surface, the attack rate, and the visibility of repeated guessing attempts, and practical weaknesses in implementations and protocol choices have been demonstrated [3].

Despite this operational reality, most password-security evaluation pipelines used in wireless settings remain static. Typical practice falls into one of three categories. First, heuristic meters compute entropy or dictionary penalties for a candidate secret [4], [5]. Second, benchmark-driven studies estimate crackability against a fixed password corpus such as RockYou [6]. Third, protocol analyses reason about whether the authentication mechanism resists a given class of attack without integrating password composition, attacker adaptation, and protocol-specific feasibility into a unified evaluator. These methods are useful, but they fail to model a central property of the threat environment: adversaries adapt.

This paper addresses that gap. We propose a *self-evolving AI framework* for evaluating wireless authentication robustness in WPA2/WPA3 networks under dynamically simulated adversarial pressure. The contribution is not an attack tool and not a password cracker. Instead, it is an assessment system that estimates how vulnerable a candidate credential is under a distribution of attacker strategies that changes as the system learns. The framework is label-free: it never needs a dataset of "successful compromises" to learn vulnerability. Rather, it relies on protocol-aware simulation signals, exploration over attack families, and online updates to a risk model that tracks which strategies remain plausible for a credential and protocol context.

The motivating problem is simple: current wireless-security evaluation methods are static, dataset-bound, and misaligned with evolving attacker strategy in WPA2/WPA3 environments. A security assessment system should not assume that attacker behavior is fixed after deployment, particularly when real-world guessing mixes common-password dictionaries, leet-speak mutations, keyboard-walk patterns, affix rules, and protocol-constrained online attempts. Our framework is designed around this requirement.

The paper makes four concrete contributions:

1) It formalizes *label-free adaptive security evaluation* for wireless passwords as an online estimation problem over protocol-conditioned attacker distributions.
2) It presents a four-module architecture comprising feature extraction, a protocol-aware adversarial simulation

engine, a self-adaptive controller based on bandit-style exploration, and a risk scoring engine that outputs compromise likelihood and simulated time-to-crack.
3) It proves an *Adaptive Weakness Convergence Theorem* showing that the vulnerability estimate converges to the true attack-susceptibility functional under bounded exploration and consistent reward estimation.
4) It defines a reproducible experimental protocol using RockYou-derived and synthetic password datasets, together with evaluation metrics and reporting templates that allow rigorous comparison against entropy-only and rule-based baselines.

The remainder of the paper is organized as follows. Section II reviews related work. Section III formulates the problem. Section IV presents the proposed framework. Section V provides theoretical analysis. Section VI defines the experimental design and reporting structure. Section VII discusses expected outcomes, limitations, and ethical boundaries. Section VIII concludes.

II. Related Work

A. Wireless Authentication and Password Exposure

WPA2 introduced robust link-layer protections for IEEE 802.11 networks, but WPA2-Personal still depends on a password-derived credential shared by all clients on the network [1]. This design makes password quality operationally decisive. Captured handshakes can be used to validate password guesses offline, so weak or patterned passphrases remain a common source of compromise risk.

WPA3-Personal replaces the classical WPA2 password exchange path with SAE, a Dragonfly-based password-authenticated key exchange intended to reduce exposure to offline guessing and forward-secrecy failures [2]. However, WPA3 changes the economics of attack rather than removing the need for credential evaluation. Empirical work on Dragonblood showed that implementation choices and protocol negotiation details still matter, and that WPA3 deployments may remain vulnerable to side-channel, downgrade, or misconfiguration risks [3]. This motivates a protocol-aware evaluator that treats WPA2 and WPA3 as different but related feasibility regimes.

B. Password Strength Estimation

Password-strength estimation has been studied from both heuristic and data-driven perspectives. Shannon entropy remains a canonical measure of uncertainty, but entropy alone is known to miss structured weakness because human-selected passwords are far from uniformly random [4]. Large-scale password leaks such as RockYou enabled empirical studies of real password distributions and the limits of policy-driven metrics [6]. Modern password meters such as zxcvbn incorporate dictionaries, pattern matching, and human-memorable structures to estimate guessability more realistically than entropy-only measures [5]. Still, such meters are static estimators: they do not adapt their internal picture of attacker pressure during evaluation, and they are typically protocol-agnostic.

C. Online Learning and Adaptive Estimation

Multi-armed and contextual bandit methods offer a principled way to balance exploration and exploitation when the quality of actions is initially unknown [7], [8]. They are particularly suitable when the evaluator must learn which attack families produce the most informative pressure signals for a password class without exhaustively simulating every strategy at every step. Related ideas also appear in adaptive security analytics, continual learning, and test-time adaptation, where systems update themselves under distribution shift without dense supervision. Our setting differs in two ways: the objective is not classification accuracy on labeled attack data, and the environment is explicitly shaped by protocol-specific feasibility constraints.

D. Gap in the Literature

The literature provides strong components but no unified framework for the problem studied here. Wireless-authentication research gives protocol guarantees and attack case studies. Password-meter research gives static guessability estimates. Online learning gives adaptive optimization tools. What is missing is a *label-free, protocol-aware, self-evolving evaluator* that treats vulnerability estimation as an online inference problem over attacker behavior. This paper fills that gap.

III. Problem Formulation

A. Setting

Let p denote a candidate password or passphrase, and let $c \in \mathcal{C}$ denote the wireless context, including protocol type (WPA2-Personal or WPA3-Personal), authentication policy constraints, observed rate-limiting assumptions, and device-specific deployment metadata. The evaluation system does not interact with a live target. Instead, it computes a risk estimate from password features and protocol-aware adversarial simulation.

Let $\mathcal{A} = \{a_1, \ldots, a_K\}$ be a finite family of attack strategies. Example families include ranked dictionary guesses, mangling rules, keyboard-walk mutations, year-affix expansions, and protocol-constrained online attempt schedules. Not every strategy is feasible in every context: WPA2 permits offline verification models that are not available in the same form for WPA3-SAE, while WPA3 makes online attempt costs and rate constraints more important. We encode this with a feasibility mask $m(a_k, c) \in \{0, 1\}$.

For a password-context pair (p, c), define the latent attack susceptibility of strategy a_k as

$$s_k(p, c) \in [0, 1], \tag{1}$$

which measures the probability that the password is exposed under the modeled attack family. The full susceptibility distribution is the vector

$$\mathbf{s}(p, c) = (s_1(p, c), \ldots, s_K(p, c)). \tag{2}$$

The evaluator does not observe $\mathbf{s}(p, c)$ directly. Instead, at round t it selects a feasible strategy A_t, simulates pressure,

and obtains a bounded feedback signal $R_t \in [0, 1]$. This signal is not a real compromise label; it is a surrogate derived from protocol-aware simulation, for example rank proximity, mutation coverage, or estimated online feasibility under a given attempt budget.

B. Goal

The objective is to estimate a scalar vulnerability score

$$V^\star(p, c) = \sum_{k=1}^{K} q_k(p, c)\, s_k(p, c), \tag{3}$$

where $q_k(p, c)$ is the probability mass assigned to attack family a_k by the modeled adversary distribution for (p, c). Equation (3) can be interpreted as the expected attack susceptibility of the password under the evaluator's attacker model.

The system must estimate $V^\star(p, c)$ without labeled compromise data and while continuously updating its estimate of which attack families matter for the observed password class. This makes the task an online adaptive estimation problem rather than supervised classification.

C. Design Requirements

The framework must satisfy four requirements:

1) **Label-free operation**: no reliance on ground-truth success labels from live compromise events.
2) **Protocol awareness**: explicit separation of WPA2 and WPA3 feasibility regimes.
3) **Adaptive pressure modeling**: the evaluator must revise its attacker model as simulation reveals more informative strategies.
4) **Actionable outputs**: the system must return a calibrated risk score and a simulated time-to-crack estimate under the modeled attacker budget.

IV. Self-Evolving Evaluation Framework

A. Overview

The proposed architecture consists of four modules shown conceptually in Table I. Given a password p and context c, the system extracts descriptive features, runs protocol-aware adversarial simulations, updates a self-adaptive controller that allocates simulation effort across attack families, and produces risk estimates. The framework is "self-evolving" because the controller modifies future attack allocation using prior feedback, thereby altering the evidence used to score later passwords.

B. Module 1: Feature Extraction

Each password is mapped to a feature vector $\phi(p) \in \mathbb{R}^d$. The proposed feature set contains four groups.

Entropy features. We compute Shannon-style character entropy estimates, normalized entropy per character, and entropy-gap indicators that compare the password to expected human-generated distributions [4].

Character diversity features. These include password length, counts of lowercase, uppercase, digit, and symbol classes, transition counts between classes, and repeated-token statistics.

TABLE I
Core modules of the proposed framework.

Module	Function
Feature extraction	Computes entropy, character diversity, length statistics, pattern signatures, and similarity to known weak-password clusters.
Adversarial simulation	Generates protocol-aware attack pressure from dictionary, mutation, affix, and structural guessing families under WPA2/WPA3 feasibility constraints.
Self-adaptive model	Uses a bandit-style controller to update which attack families are most informative for a password class without labeled success data.
Risk scoring engine	Aggregates attack-pressure evidence into vulnerability score, protocol-conditioned compromise likelihood, and simulated time-to-crack estimate.

Structural-pattern features. We compute indicators for keyboard walks, common year suffixes, l33t substitutions, affix rules, repeated words, mirrored substrings, and popular token stems derived from the password corpus.

Similarity features. We estimate similarity to known weak-password neighborhoods using token overlap, edit distance to common patterns, and rank-derived corpus priors. These features allow the evaluator to detect non-obvious weakness that pure entropy misses, such as high-length but highly patterned strings.

C. Module 2: Protocol-Aware Adversarial Simulation Engine

The simulation engine applies attack families from $\mathcal{A}$ to the feature representation and password candidate space. It does not attempt unauthorized access to real networks. Instead, it produces bounded *attack-pressure signals* that estimate how exposed the password would be to each modeled strategy.

For WPA2 contexts, the engine can model offline dictionary economics because captured handshakes admit large-volume guess validation. For WPA3 contexts, the engine suppresses those offline channels and instead emphasizes online, rate-constrained, and policy-visible pressure models consistent with SAE-style deployment assumptions [2], [3]. This distinction is encoded through context-dependent feasibility masks and attack budgets.

At round t, the engine selects attack family A_t and produces a reward-like signal

$$R_t = g(p, c, A_t, \phi(p), B_t)\,, \tag{4}$$

where B_t is the current simulation budget and $g(\cdot)$ is a bounded scoring functional. Examples include dictionary-rank penetration, mutation coverage, pattern exposure confidence, and online-guess feasibility after rate penalties. The signal is normalized to $[0, 1]$ so that heterogeneous strategy outputs can be aggregated.

D. Module 3: Self-Adaptive Model

The self-adaptive controller maintains an estimate $\hat{\mu}_{k,t}$ of how informative attack family a_k is for passwords like p in context c. Let

$$x_t = h(\phi(p), c, \psi_t) \tag{5}$$

be the context vector combining password features, protocol metadata, and internal state ψ_t such as recent uncertainty or disagreement between attack families. A contextual bandit policy selects the next strategy according to

$$A_t \sim \pi_t(\cdot \mid x_t), \tag{6}$$

with enforced exploration:

$$\pi_t(a_k \mid x_t) \geq \varepsilon_t / K, \qquad \forall k \text{ such that } m(a_k, c) = 1. \tag{7}$$

In practice, the controller can be implemented with Thompson sampling or an upper-confidence policy over linear or generalized linear reward models [7]. The key role of the controller is not to maximize cracking success; it is to concentrate evaluation effort on strategies that reveal the most about password weakness for the given protocol setting while preserving sufficient exploration for convergence.

E. Module 4: Risk Scoring Engine

Let $\hat{s}_{k,t}(p, c)$ be the running estimate of susceptibility for attack family a_k. The overall vulnerability estimate after t rounds is

$$\hat{V}_t(p, c) = \sum_{k=1}^{K} \hat{q}_{k,t}(p, c)\, \hat{s}_{k,t}(p, c), \tag{8}$$

where $\hat{q}_{k,t}(p, c)$ is the controller's evolving estimate of the adversary-distribution mass on strategy a_k.

The risk scoring engine returns three outputs:

1) a normalized vulnerability score $\hat{V}_t(p, c) \in [0, 1]$,
2) a protocol-conditioned compromise likelihood bucket (low, moderate, high, critical),
3) a simulated time-to-crack estimate derived from the effective attack budget and the cumulative pressure assigned to feasible strategies.

The time-to-crack output is explicitly *simulated*. In WPA2 settings it reflects offline-guess economics; in WPA3 settings it reflects rate-limited online pressure and therefore typically expands sharply even for moderately weak passwords.

V. Theoretical Analysis

A. Estimator Construction

For each attack family a_k, define the empirical estimator

$$\hat{s}_{k,t}(p, c) = \frac{1}{N_{k,t}} \sum_{\tau=1}^{t} \mathbf{1}\{A_\tau = a_k\} R_\tau, \tag{9}$$

where

$$N_{k,t} = \sum_{\tau=1}^{t} \mathbf{1}\{A_\tau = a_k\}. \tag{10}$$

The attacker-distribution estimate is updated online by a stochastic approximation rule

$$\hat{q}_{t+1} = \Pi_{\Delta_K}\left(\hat{q}_t + \eta_t\, u_t\right), \tag{11}$$

where Π_{Δ_K} projects onto the probability simplex, η_t is a step size satisfying $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$, and u_t is an unbiased score estimator for the adversary mass update.

B. Adaptive Weakness Convergence Theorem

Theorem 1 (Adaptive Weakness Convergence)**.** *Fix a password-context pair (p, c) and a finite feasible attack family set $\mathcal{A}_c = \{a_k : m(a_k, c) = 1\}$. Assume:*

1) *for each $a_k \in \mathcal{A}_c$, the simulation feedback sequence $\{R_t \mid A_t = a_k\}$ is conditionally i.i.d., bounded in $[0, 1]$, and has mean $s_k(p, c)$;*
2) *the exploration policy satisfies bounded exploration, i.e., there exists $\epsilon > 0$ such that for all sufficiently large t,*

$$\pi_t(a_k \mid x_t) \geq \epsilon / |\mathcal{A}_c|, \qquad \forall a_k \in \mathcal{A}_c; \tag{12}$$

3) *the update for $\hat{q}_t$ is a consistent stochastic approximation to the true adversary distribution $q(p, c)$ on $\mathcal{A}_c$.*

Then the vulnerability estimator in (8) *converges almost surely to the true attack-susceptibility functional in* (3)*:*

$$\hat{V}_t(p, c) \xrightarrow[t\to\infty]{a.s.} V^\star(p, c). \tag{13}$$

Proof. By Assumption 2, every feasible attack family is selected infinitely often almost surely, so $N_{k,t} \to \infty$ for each $a_k \in \mathcal{A}_c$. By Assumption 1 and the strong law of large numbers, the empirical estimator in (9) converges almost surely:

$$\hat{s}_{k,t}(p, c) \xrightarrow[t\to\infty]{a.s.} s_k(p, c). \tag{14}$$

By Assumption 3, the adversary-distribution estimate also converges almost surely:

$$\hat{q}_{k,t}(p, c) \xrightarrow[t\to\infty]{a.s.} q_k(p, c), \qquad \forall a_k \in \mathcal{A}_c. \tag{15}$$

Because the attack family set is finite and both sequences are bounded, the product sequence is almost surely convergent termwise. Therefore,

$$\sum_{k=1}^{K} \hat{q}_{k,t}(p, c)\hat{s}_{k,t}(p, c) \xrightarrow[t\to\infty]{a.s.} \sum_{k=1}^{K} q_k(p, c) s_k(p, c) = V^\star(p, c). \tag{16}$$

Hence $\hat{V}_t(p, c)$ converges almost surely to the true vulnerability score. □

Remark 1. *The theorem is intentionally stated at the level of the modeled attacker distribution. It does not claim convergence to a universal or ground-truth compromise probability independent of modeling assumptions. Instead, it guarantees convergence to the evaluator's protocol-conditioned susceptibility functional under bounded exploration and consistent updates. This is the appropriate notion of correctness for a label-free security evaluator.*

C. Interpretation

The theorem matters because it turns the framework into more than an engineering heuristic. It establishes that continuous adversarial simulation can be made statistically meaningful: as long as the system continues to explore feasible attack families and its reward estimates remain unbiased for the modeled environment, the risk score stabilizes around the correct attack-susceptibility functional. In practical terms, the evaluator can begin with uncertainty, adapt to newly observed weak patterns, and still retain a principled convergence guarantee.

VI. Experimental Design and Reporting Template

A. Datasets

The proposed evaluation uses three complementary password sources.

RockYou-derived corpus. A cleaned and anonymized subset of RockYou-style passwords provides realistic weak and human-generated patterns [6]. The subset should be deduplicated, length-filtered to WPA-compliant ranges, and stripped of user-identifying metadata.

Synthetic WPA password set. A generator produces passwords that match WPA deployment constraints while spanning structured weakness modes including years, dictionary concatenations, affix patterns, keyboard walks, and symbol substitutions.

High-entropy reference set. Uniformly sampled high-entropy strings provide a hard negative set representing passwords that should remain difficult under both WPA2 and WPA3 pressure models.

B. Baselines

The framework should be compared against:

1) **Entropy-only scoring**: Shannon-entropy features converted into a single static robustness score.
2) **Rule-based checker**: a hand-engineered policy that penalizes short length, low diversity, and common patterns.
3) **zxcvbn**: a widely used pattern-aware password estimator [5].

C. Metrics

The following metrics should be reported.

1) **Vulnerability score accuracy**: agreement between predicted score ranking and held-out simulated susceptibility outcomes.
2) **Convergence rate**: rounds required for $\hat{V}_t$ to stabilize within a fixed tolerance of its terminal estimate.
3) **Attack simulation efficiency**: vulnerability estimation gain per unit of simulation budget.
4) **Weak-pattern detection**: recall on non-obvious but structurally weak passwords that entropy-only methods overrate.
5) **Protocol sensitivity**: change in risk ranking when moving the same password from WPA2 to WPA3 feasibility conditions.

TABLE II
Primary result template for empirical population after experiment execution.

Method	Score accuracy	Convergence rate	Efficiency	Weak-pattern recall
Entropy-only	To populate	N/A	To populate	To populate
Rule-based checker	To populate	N/A	To populate	To populate
zxcvbn	To populate	N/A	To populate	To populate
Proposed framework	To populate	To populate	To populate	To populate

D. Reporting Structure

To avoid overstating evidence, this manuscript intentionally leaves the result tables population-ready. Table II defines the primary reporting structure to be filled after experiment execution. A companion implementation should export the final values from the evaluation pipeline rather than inserting hand-written numbers.

E. Expected Outcome Pattern

The framework is designed to support three falsifiable hypotheses:

1) adaptive simulation will surface weak patterns faster than static estimators because it reallocates effort toward the most informative attack families;
2) protocol awareness will prevent over-penalization of passwords in WPA3 settings where offline attack feasibility is materially reduced;
3) the combined feature-plus-simulation model will identify high-length but pattern-heavy passwords that entropy-only scoring incorrectly treats as robust.

If these hypotheses are validated, the paper can make the qualitative claim that the model adapts faster to weak patterns, detects non-obvious weak passwords, and outperforms static estimators. Numerical claims, however, should only be inserted after actual runs.

VII. Discussion

A. Why This Framing Is Strong

The strength of the paper lies in its reframing of password evaluation as a self-adaptive estimation problem. Most related work either predicts whether a password is weak or analyzes a protocol in isolation. The proposed framework instead models a moving adversary. This makes the contribution more general than a new password meter and more practical than a purely formal protocol analysis. It also aligns with how security teams reason about risk in deployed wireless environments: not "is this password theoretically strong," but "how does its exposure

change under realistic attacker pressure for this protocol and deployment context?"

B. Limitations

Three limitations remain. First, the framework converges to a modeled attacker distribution, not to a universal notion of password compromise risk. Second, simulation quality depends on the realism of the attack family library and feasibility assumptions. Third, WPA2/WPA3 evaluation in enterprise environments with 802.1X, device certificates, or mixed transition modes would require an expanded context model beyond the personal-network focus of this paper.

C. Ethical Boundary

The system is intended for defensive evaluation, password policy testing, and offline robustness analysis of credentials under authorized conditions. It should not be connected to real wireless targets or used to conduct guessing attempts against networks that the operator does not own or control. The simulation engine is deliberately framed as an estimator rather than an exploit workflow.

VIII. Conclusion

This paper introduced a self-evolving AI framework for evaluating wireless authentication robustness in WPA2/WPA3 networks. The core idea is label-free adaptive security evaluation: instead of learning from labeled compromise events, the system models password weakness through protocol-aware adversarial simulation, bounded exploration, and online refinement of attack-pressure estimates. The resulting architecture integrates password features, adaptive attack-family selection, and actionable risk scoring in a single framework. The proposed Adaptive Weakness Convergence Theorem gives the method theoretical grounding by proving convergence of the vulnerability estimate under bounded exploration. Finally, the paper defines a reproducible empirical protocol over RockYou-derived and synthetic datasets with clear baselines and metrics. As a result, the work offers a defensible path from static password scoring toward a more realistic, feedback-driven evaluation paradigm for wireless authentication security.

References

[1] *IEEE Standard for Information Technology—Telecommunications and Information Exchange between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 6: Medium Access Control (MAC) Security Enhancements*, IEEE Std. IEEE Std 802.11i-2004, 2004.

[2] D. Harkins, "Dragonfly key exchange," RFC Editor, Tech. Rep. RFC 7664, 2015.

[3] M. Vanhoef and E. Ronen, "Dragonblood: Analyzing the dragonfly handshake of wpa3 and eap-pwd," in *2020 IEEE Symposium on Security and Privacy (SP)*, 2020, pp. 517–533.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[5] Dropbox, "zxcvbn: Low-budget password strength estimation," https://github.com/dropbox/zxcvbn, accessed April 17, 2026.

[6] M. Weir, S. Aggarwal, M. Collins, and H. Stern, "Testing metrics for password creation policies by attacking large sets of revealed passwords," in *Proceedings of the 17th ACM Conference on Computer and Communications Security*, 2010, pp. 162–175.

[7] S. Agrawal and N. Goyal, "Thompson sampling for contextual bandits with linear payoffs," in *Proceedings of the 30th International Conference on Machine Learning*, 2013, pp. 127–135.

[8] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The non-stochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002.