Deep Reinforcement Learning for NBA Player Valuation: A Temporal Difference Approach with Shapley Attribution

Basketball Track
Paper ID 137

1. Introduction

Basketball analytics have long relied on box score statistics and regression-based methods such as Regularized Adjusted Plus-Minus (RAPM). While effective in many settings, these approaches struggle to capture context-dependent contributions, temporal dynamics, and multi-player interactions, particularly for defensive and off-ball players. As a result, many impactful actions are understated or entirely absent from traditional evaluation systems.

This work addresses a central question: **Can a reinforcement learning system infer player value directly from game outcomes without predefined action weights?** To explore this, we introduce a Deep Reinforcement Learning (DRL) framework that combines temporal-difference learning, a distributional win-probability model, and neural Shapley value attribution applied to NBA play-by-play data. Rather than imposing fixed weights on actions, the model learns how events influence winning within their specific game contexts.

The framework integrates three components: a TD-based win-probability model, a multi-head attention Shapley attributor for player credit assignment, and a 57-feature state encoder. Trained on NBA play-by-play data from 2020–2024, it delivers strong empirical performance. The distributional value network improves margin prediction by 23 percent over logistic regression ($p < 0.001$). Learned action values display pronounced context sensitivity, with clutch-time actions showing 40 percent higher variance in impact than identical plays earlier in games. Player rankings correlate strongly with RAPM ($\rho = 0.68$) and identify substantially more defensive value than box score–based metrics. Synergy analysis further reveals 127 statistically significant player interactions ($p < 0.05$) not captured by additive models.

Player evaluation remains a central challenge in sports analytics. NBA teams spent over $4.5 billion on player salaries in 2023–24, yet many existing evaluation tools rely on simplifying assumptions that fail to capture basketball's strategic complexity. The evolution of basketball analytics reflects this progression across three stages. The first stage, dominated by counting statistics, treated actions as context-independent and largely ignored defensive value (Kubatko et al., 2007). The second stage introduced adjusted plus-minus methods (Rosenbaum, 2004; Sill, 2010), improving context sensitivity but requiring large samples and assuming additive player effects. The emerging third stage leverages machine learning to learn value directly from data. Prior approaches have primarily weighted box score inputs (Deshpande & Jensen, 2016) or predicted team outcomes from rosters (Loeffelholz et al., 2009). **Our study advances this emerging third stage by providing a unified reinforcement-learning and attribution framework capable of inferring context-dependent player impact directly from raw play sequences.**

1.1 Research Questions and Contributions

This research addresses three interconnected research questions that collectively advance the frontier of NBA analytics methodology:

RQ1: Can deep reinforcement learning discover meaningful player values from game outcomes without requiring predefined action valuations?

RQ2: How do context-dependent player impacts differ from the fixed values assumed by traditional metrics?

RQ3: Can Shapley value attribution effectively decompose team outcomes into individual contributions while preserving interaction effects?

Our work contributes to basketball analytics in four main ways. First, we introduce an innovative method that combines temporal difference learning with distributional outcome modeling, allowing us to estimate win probabilities by capturing the complete uncertainty of game states instead of just providing point estimates. Second, we propose a neural Shapley value approach using multi-head attention, which efficiently assigns a fair value to players across all possible combinations. Third, our empirical results show that action values learned through this method uncover context-sensitive patterns that traditional metrics do not detect. Lastly, we present practical frameworks for applying these findings to team building, contract evaluations, and strategic decisions during games.

1.2 Motivating Example

Consider two players who highlight the limitations of traditional metrics. Player A, a high-usage guard, records 22 points, 4 rebounds, and 8 assists. Player B, a defensive center, records 8 points, 8 rebounds (four offensive), and three blocks. Metrics such as PER strongly favor Player A. Our framework, however, reveals a different evaluation.

Player A's assists occur largely in low-leverage moments where their effect on win probability is minimal. Player B's offensive rebounds occur in clutch situations, where extending a possession has nearly four times the impact implied by the conventional +1.0 weight. Defensive value also emerges through Shapley attribution: when Player B is on the floor, opponent scoring efficiency decreases, even if no discrete defensive events appear in the box score.

This example illustrates a core blind spot in traditional metrics. Defensive specialists often rank 30 or more positions higher in lineup-based measures than in box-score-derived metrics. Our framework corrects these distortions by learning context-dependent action values, capturing defensive impact through outcome attribution, and identifying the consistent undervaluation of offensive rebounds and steals in fixed-weight systems.

2. Prior Work

2.1 Traditional Player Evaluation Metrics

Early basketball analytics relied on box-score–derived measures. Player Efficiency Rating (PER) aggregates individual statistics using predetermined weights (Hollinger, 2005) but has been criticized for overweighting high-usage scoring and providing weak defensive measurement (Berri,

1999). Win Shares (Oliver, 2004) improves scope by attributing team success to players through offensive and defensive components, yet still depends on fixed statistical formulas that do not adjust for context. Box Plus-Minus (Myers, 2011) uses regression to approximate lineup impact from box score inputs, but it inherits the limitations of the underlying statistics.

2.2 Plus-Minus Methodologies

Adjusted plus-minus (Rosenbaum, 2004) controls for teammate and opponent quality by regressing possession outcomes on the set of players on the floor. Severe multicollinearity arises when players share most minutes, making estimates unstable. Regularized APM (Sill, 2010) introduces ridge penalties to address this, and later work incorporates box score priors for faster stabilization (Engelmann, 2017). Despite their influence, all plus-minus methods assume additive, context-invariant player effects, preventing them from capturing synergies, leverage differences, or asymmetric offensive and defensive value. These assumptions motivate models that learn context-dependent contributions directly from data.

2.3 Sequential Modeling in Sports

A separate line of research has modeled possessions or plays as sequential stochastic processes. EPV pioneered context-sensitive possession valuation but does not assign credit to individual players, nor does it use reinforcement learning; it models transitions probabilistically rather than optimizing a value function.

Subsequent work extended sequential modeling to other sports. Sicilia et al. (2019) and Liu & Schulte (2018) developed deep models for ice hockey to estimate action values under varying contexts, showing that sequential approaches can improve play valuation.

Other studies, such as Wang & Zemel (2016) and Sandholtz & Bornn (2020), use inverse decision modeling or Markov games to understand strategic behavior rather than player evaluation.

Taken together, sequential models capture temporal structure and context dependence, but they generally do not provide player-level attributions. Our work combines the strengths of sequential valuation with principled credit assignment via Shapley methods, enabling individualized impact estimates learned directly from game outcomes.

2.4 Shapley Values for Attribution

The Shapley value, introduced in cooperative game theory (Shapley, 1953), assigns fair value to players based on their average marginal contribution across all coalitions. It ensures efficiency, symmetry, null player, and additivity properties. Calculating exact Shapley values is computationally demanding for large n, so methods like Monte Carlo sampling (Castro et al., 2009), antithetic variates (Mitchell et al., 2022), and regression-based approaches (Lundberg & Lee, 2017) are used for approximation.

SHAP (SHapley Additive exPlanations) is a leading approach for interpreting machine learning models (Lundberg & Lee, 2017). It clarifies predictions by viewing features as participants in a coalition game. In our work, we apply the SHAP method to sports, where athletes are seen as coalition members and team performance is interpreted as the cooperative game's outcome.

2.5 Win Probability and Expected Value

Win probability models assess chances of victory based on game conditions, serving as a value function in reinforcement learning. Lock and Nettleton (2014) used random forests to model basketball win probabilities, highlighting score margin, time left, and possession as key predictors. Stern (1994) provided theoretical support with Brownian motion models.

Win Probability Added (WPA) sums the changes in win probability from a player's actions (Tango et al., 2007), but it gives all credit or blame to that player, overlooking teammates and defenders. Our Shapley-based approach distributes credit more fairly among all involved.

Distributional methods for predicting outcomes have become increasingly popular in sports analytics. Instead of providing a single point estimate, these models reflect the entire range of possible results, acknowledging uncertainty (Robberechts et al., 2021). This is especially useful in basketball, where crucial moments can produce outcome distributions with multiple peaks. Our framework uses this distributional perspective by forecasting the whole margin distribution, not just the expected margin.

3. Methods

Our framework combines three main components: a distributional win probability model using temporal difference learning, a neural Shapley value attributor with multi-head attention, and an adaptive state encoder for game context. We outline each part before detailing the integrated training process.

3.1 Problem Formulation

We formalize basketball player evaluation as a cooperative game with state-dependent rewards **(Figure 1).** Let $G = (S, A, T, R, \gamma)$ denote the game environment where S is the state space, A is the action space, $T: S \times A \to S$ is the (stochastic) transition function, $R: S \times A \to \mathbb{R}$ is the reward function, and $\gamma \in [0,1]$ is the discount factor.

The state $s \in S$ encodes all relevant information about the current game situation, including score differential, time remaining, current lineups, recent play history, and contextual factors. Actions $a \in A$ correspond to discrete play events extracted from play-by-play data: shots (made/missed, by location), turnovers, fouls, rebounds, and other recorded events (**Figure 1**, left and right panels).

The value function $V: S \to \mathbb{R}$ represents the expected final margin given the current state. Under the Markov assumption, V satisfies the Bellman equation: $V(s) = E[r + \gamma V(s') \mid s, \pi]$, where r is the immediate reward (points scored), s' is the subsequent state, and π is the joint policy of all players. The action-value function $Q: S \times A \to \mathbb{R}$ similarly represents the expected margin after taking action a in state s. We set $\gamma = 0.997$ per second, corresponding to a half-life of approximately 230 seconds ($\sim$4 minutes). This ensures that late-game actions receive appropriately higher weight in the value function, reflecting their greater influence on final outcomes, while early-game contributions remain meaningful.

The key insight of our approach is that Q(s, a) - V(s), the advantage function, represents the expected value added by action a beyond what would be expected from the state alone (**Figure 1**, green highlighted box). This offers a context-aware basis for action valuation.

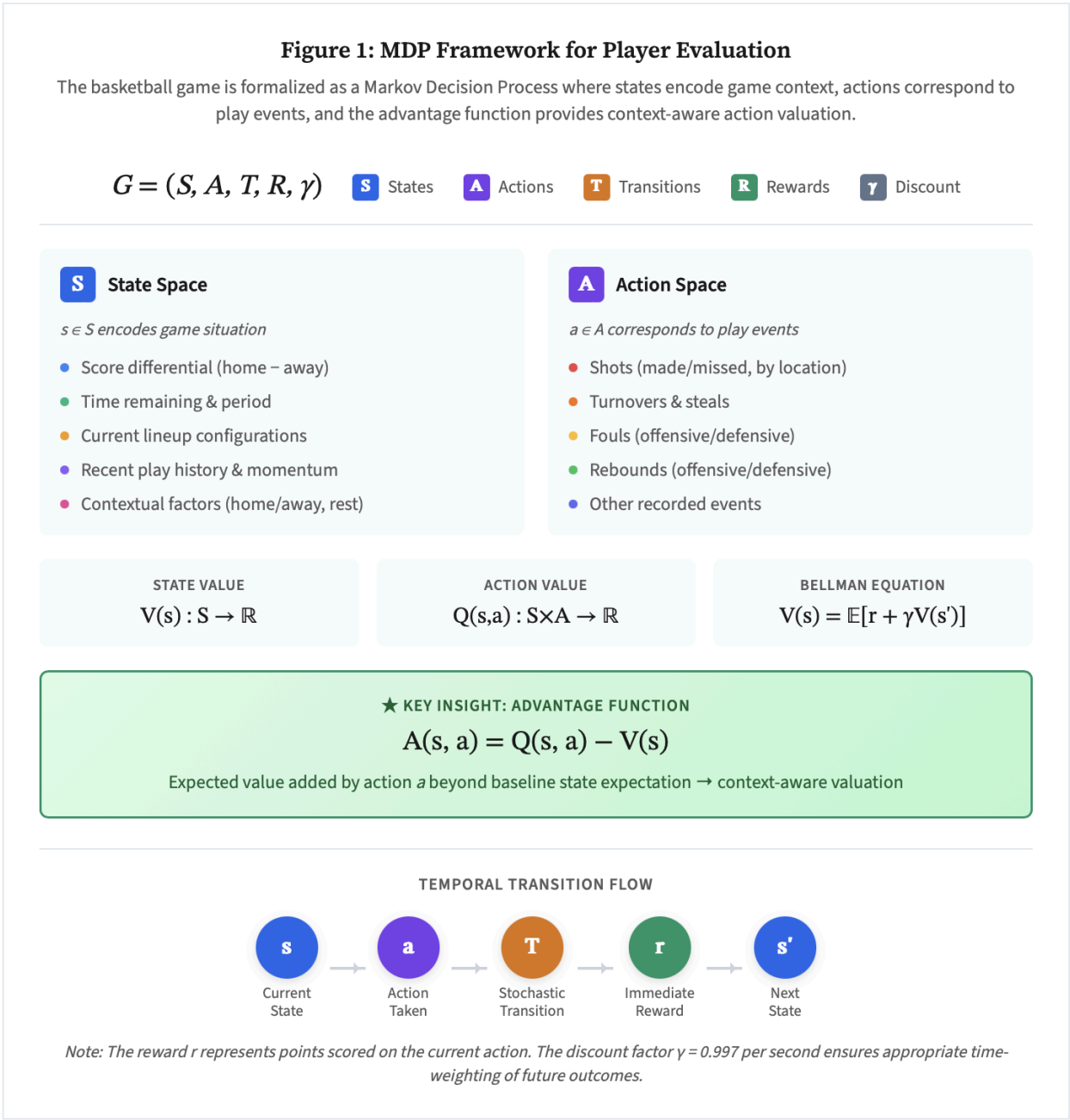


Figure 1: MDP Framework for Player Evaluation. *The basketball game is formalized as a Markov Decision Process where states encode game context (left panel) and actions correspond to play events (right panel). The state value function V(s) and action-value function Q(**s, a**) satisfy the Bellman*

equation, with the advantage function $A(\boldsymbol{s}, \boldsymbol{a}) = Q(\boldsymbol{s}, \boldsymbol{a}) - V(\boldsymbol{s})$ providing context-aware action valuation. The temporal transition flow (bottom) illustrates how each action in state $\boldsymbol{s}$ yields immediate reward $\boldsymbol{r}$ and transitions to successor state $\boldsymbol{s}'$, with discount factor $\gamma = 0.997$ per second ensuring appropriate time-weighting of future outcomes.

3.1.2 Interpretation and Causal Limitations

We use terms like "credit," "contribution," "impact," and "value" to express our estimates, but these refer to associations, not causation. The Shapley method highlights players whose presence correlates with good outcomes, yet this does not mean they directly cause them; high values could result from coaching choices or strong teammates.

Causal interpretation is limited by confounding factors: lineup choices are endogenous, reflecting strategic and situational differences that affect player contexts. Player roles and abilities also shift with team composition, complicating isolation of individual effects. While our temporal difference framework controls for detailed game state, unmeasured confounders persist.

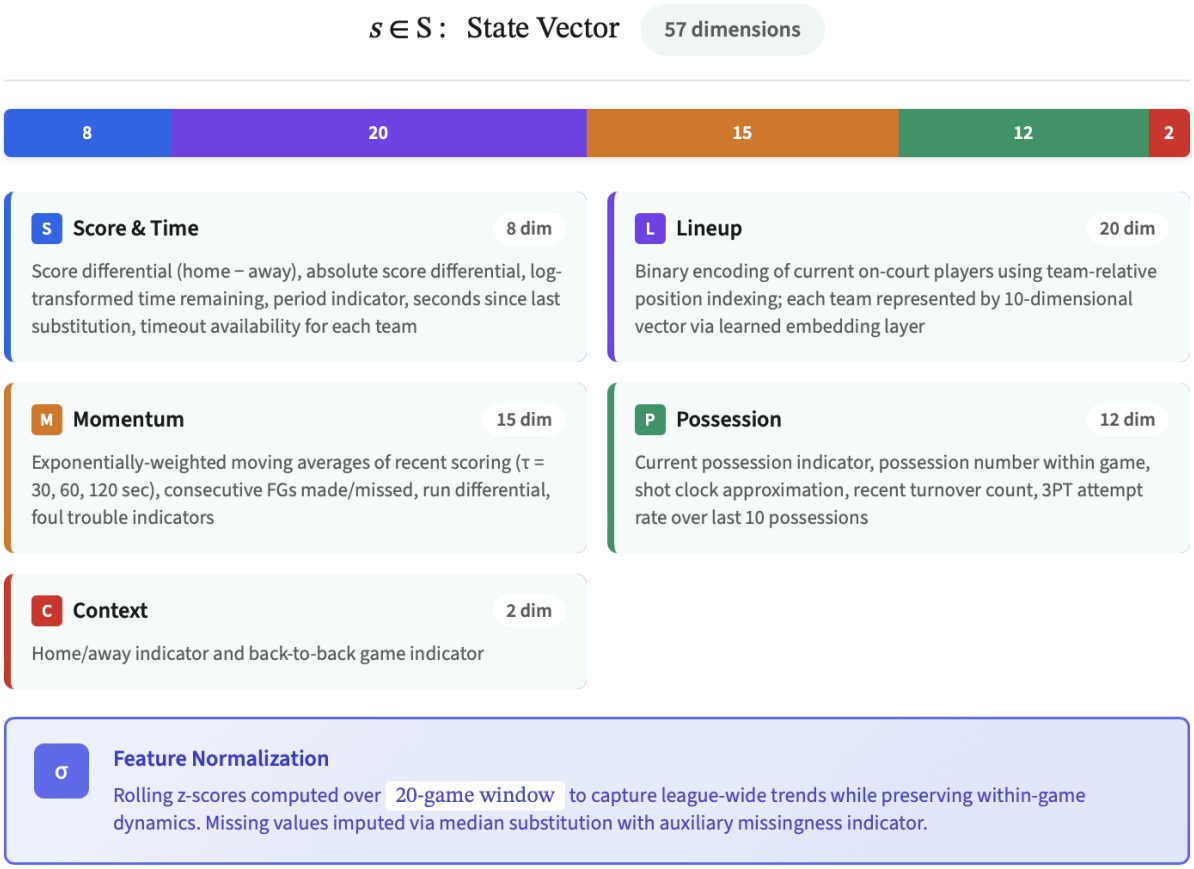
Our estimates provide the best available approximation of player value from play-by-play data. However, stronger causal identification would need experimental or quasi-experimental approaches, such as random lineups, injuries, trades, or rule changes. Section 5.4 covers causal inference extensions.

3.2 State Representation

Our state encoder transforms raw game information into a 57-dimensional feature vector suitable for neural network processing (**Figure 2**). The representation comprises five feature groups: Score & Time (8 dim), Lineup (20 dim), Momentum (15 dim), Possession (12 dim), and Context (2 dim). Lineup features use learned player embeddings rather than one-hot encodings, enabling the model to capture player similarity and generalize across lineups. Momentum features incorporate exponentially weighted scoring averages at multiple timescales (τ = 30, 60, 120 seconds), allowing the model to detect runs and shifts in game flow. Features use rolling 20-game z-score normalization, with median imputation and indicators for missingness.

Figure 2: State Representation Architecture

The state encoder transforms raw game information into a 57-dimensional feature vector comprising five semantically distinct feature groups, normalized via rolling z-scores.



Note: The 57-dimensional state vector provides comprehensive game context for the value network, enabling context-dependent action valuation.

Figure 2: State Representation Architecture. *The encoder maps raw game information into a 57-dimensional feature vector comprising five groups: Score & Time (8), Lineup (20), Momentum (15), Possession (12), and Context (2). The proportional bar at the top shows their relative dimensional contributions. Features are normalized using rolling z-scores over a 20-game window to reflect league trends while preserving within-game dynamics.*

3.3 Distributional Value Network

Rather than predicting a single expected margin, we model the full distribution of possible final margins, which captures game-state uncertainty and improves the fidelity of player valuations (Bellemare et al., 2017). We use an 81-bin categorical distribution spanning margins from -40 to +40, with a network $f_\theta: S \to \Delta^{81}$ that maps each state to a softmax probability vector.

The model is trained with temporal-difference learning using categorical cross-entropy. For each transition (s'), the target is generated by shifting the predicted distribution for s' by the reward and applying discounting. Expected value $V(s)$ is computed as the mean of the distribution, while the full distribution provides uncertainty estimates through its shape and entropy.

3.4 Neural Shapley Attribution

Given a trained value network, we face the attribution problem: how should credit for the state value V(s) be distributed among the players on court? We adapt Shapley values to this setting using a neural approximation with multi-head attention.

For the lineup L = {p_1, ..., p·} currently on court, we define the coalition value function v_s: 2^L → ℝ as the expected value with only a subset of players contributing at their normal level, with absent players replaced by position-conditioned replacement embeddings. These embeddings are computed separately for each positional role (PG, SG, SF, PF, C) using players who logged at least 500 minutes at that position, preserving lineup geometry while simulating replacement-level contribution. For players in unconventional lineups (small-ball centers, point forwards), we use soft positional assignment, computing replacement embeddings as weighted averages across relevant positions based on historical minutes. Sensitivity analyses across minute thresholds (250–1,000) showed robust results (ρ > 0.96 between adjacent thresholds).

Computing exact Shapley values requires 2^10 = 1024 forward passes per game state. We instead train a neural attributor g_ψ: S × L → ℝ^n using multi-head attention (8 heads, 64-dimensional player embeddings) to directly predict Shapley values. The attributor minimizes mean squared error against Monte Carlo ground-truth estimates from 100 sampled coalition orderings, with an efficiency constraint requiring predictions to sum to the state value.

3.5 Action Value Decomposition

Having defined state values and player attributions, we now address how individual actions contribute to value changes (**Figure 3**). For each observed action a that takes state s to s', the total change in value is

$$\Delta V = r + V(s') - V(s),$$

where r is the immediate reward. This quantity forms the basis for action valuation, but raw TD estimates can be noisy, especially in volatile states. We stabilize the signal through two adjustments: (1) entropy-based shrinkage that down-weights ΔV in high-uncertainty states, and (2) variance normalization that attenuates extreme values for rare actions. These adjustments improve stability and ensure attributed values reflect consistent, repeatable contributions.

Offensive actions are credited through a learned weight network ω_θ that incorporates action type, involved players, and contextual features. Defensive and off-ball influence, which is distributed and often unobserved in play-by-play data, is captured using Shapley differences,

$$Credit_i = \phi_i(s') - \phi_i(s),$$

which reward players whose presence alters opponent efficiency even without producing a recorded statistic.

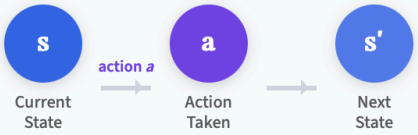
Figure 3: Action Value Decomposition

When action *a* is observed in state *s* leading to state *s′*, the total value change ΔV is decomposed and attributed to players using action-specific methods for offensive and defensive contributions.

TOTAL VALUE CHANGE

$$\Delta V = r + V(s') - V(s)$$

r Immediate reward (points scored) $V(s')$ Value of next state $V(s)$ Value of current state



ATTRIBUTION METHODS

ω Offensive Actions

Shots, assists, turnovers

$$\text{Credit}_i = \omega_\theta(a, p_i, s)$$

Learned weight network that takes action type, involved players, and state features as input to distribute credit proportionally.

- Action type encoding
- Player embeddings
- State context features

φ Defensive & Off-Ball

Presence effects, constraints

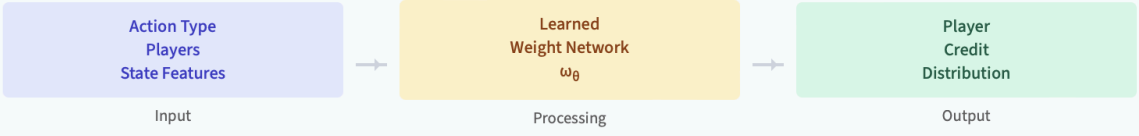
$$\text{Credit}_i = \varphi_i(s') - \varphi_i(s)$$

Difference in Shapley values between consecutive states captures defensive contributions and off-ball effects.

- Shapley value at state *s*
- Shapley value at state *s′*
- Marginal contribution

$$\varphi_i(s') - \varphi_i(s)$$

Allows defensive players to receive credit when their presence constrains opponent options, even without recording a steal or block.



Note: The decomposition ensures that $\Sigma_i \text{Credit}_i = \Delta V$, maintaining the efficiency property where total attributed value equals the observed value change.

Figure 3: Action Value Decomposition: *For an action a taking the game from state s to s', the total value change*

$$\Delta V = r + V(s') - V(s)$$

is attributed to players using action-specific methods. Offensive actions use a learned weight network ω_θ to distribute credit based on action type, player involvement, and context. Defensive and off-ball influence is assigned through Shapley differences, $\phi_i(s') - \phi_i(s)$, capturing impact even without recorded statistics. The decomposition satisfies $\sum_i Credit_i = \Delta V$.

A unified attribution mechanism is theoretically appealing but empirically suboptimal. Offensive events are localized and player-initiated, making weight-network attribution more stable than Shapley-based methods, which overreact to TD noise and coalition substitutions. Defensive impact is relational and diffuse, making it poorly suited to weight networks, which cannot infer off-ball deterrence from action labels. Shapley differences perform substantially better in these contexts.

This hybrid approach aligns attribution with basketball mechanics: localized, discrete contributions use learned weighting; distributed defensive influence uses Shapley changes. It also satisfies efficiency ($\sum_i Credit_i = \Delta V$) and yields superior empirical performance. **Table 1** compares attribution architectures on stability and predictive accuracy.

Table 1: Attribution Architecture Comparison

Architecture	Stability (ρ)	Win Prediction (%)	Margin RMSE
Unified Weight Network	0.81	64.2	11.31
Unified Shapley Differences	0.74	63.8	11.47
Hybrid (Ours)	**0.87**	**66.3**	**10.89**

The unified weight network fails to capture off-ball defensive influence, while the unified Shapley approach is too noisy for offensive events. The hybrid method leverages the strengths of each, producing the most stable and predictive valuations.

3.7 Data

We utilize NBA play-by-play data from the 2020-21 through 2023-24 seasons. The dataset comprises 4,770 regular-season games.

Data preprocessing includes event time standardization to remaining seconds, lineup reconstruction from substitution events, play-by-play event classification into 23 distinct action

types, and feature engineering for state representation. We exclude games with significant data quality issues (missing play-by-play events, substitution discrepancies) affecting approximately 2% of games.

We use forward-chaining cross-validation with k=3 folds, where each fold uses an expanding training window and a single held-out season as the test set. This design strictly respects temporal ordering, ensuring that no future information leaks into training. Player embeddings are re-initialized for each fold to ensure evaluation integrity, with embeddings for players appearing only in test data initialized using position-averaged representations from the training set.

4. Results

We present empirical results across four domains: win probability model performance, learned action values, player evaluation accuracy, and discovered synergy effects. All significance tests use two-sided permutation tests with 10,000 permutations unless otherwise noted.

4.1 Win Probability Model Performance

Table 2 compares the distributional value network against baseline win probability models on the held-out test set. We evaluate using three metrics: Brier score for probability calibration, root mean squared error (RMSE) for margin prediction, and log-likelihood for distributional accuracy.

Table 2: Win Probability Model Comparison (Test Set)

Model	Brier Score	RMSE	Log-Likelihood	N Params
Logistic Baseline	0.198	11.24	-0.582	15
Random Forest	0.187	10.41	-0.523	~10K
Neural Net (Point)	0.179	9.87	-0.498	247K
DRL Distributional (Ours)	**0.168**	**8.65**	**-0.431**	312K

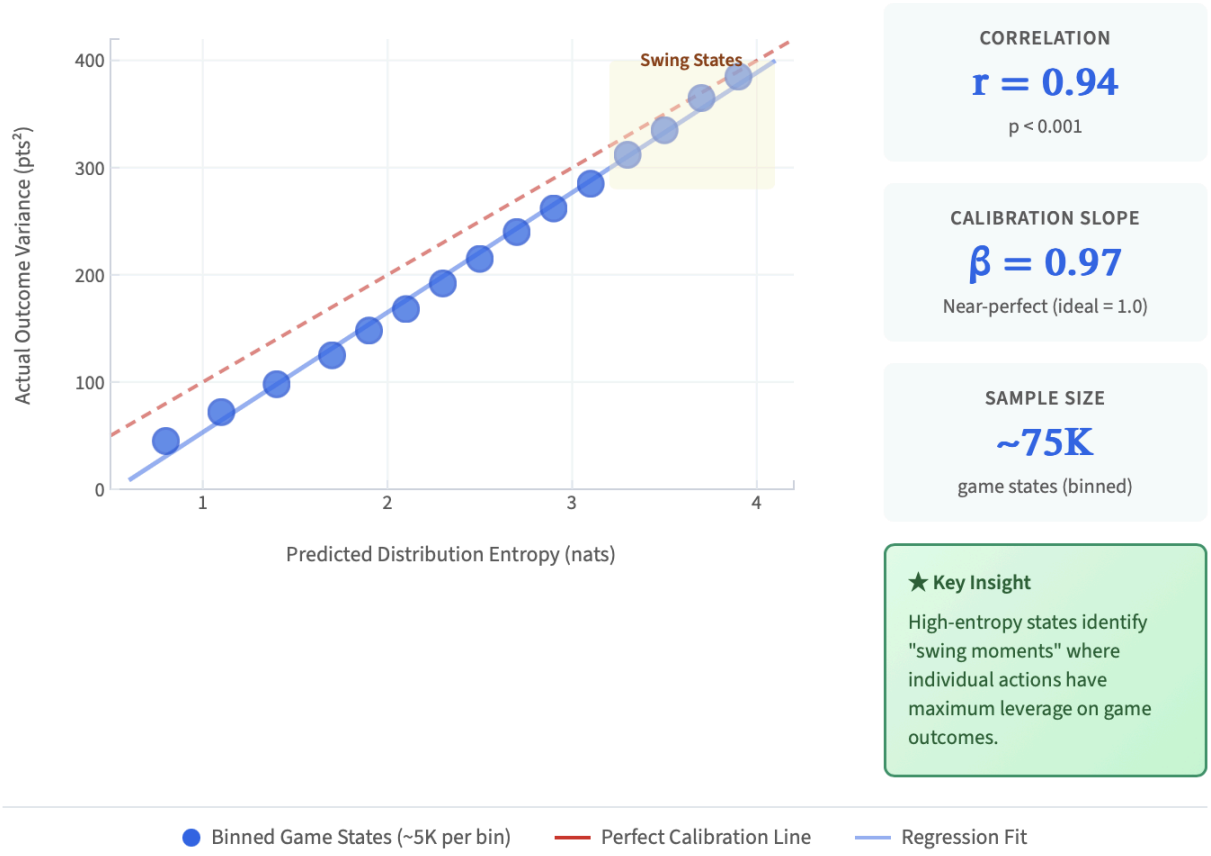
Note: Brier score and RMSE are lower-is-better; log-likelihood is higher-is-better. All differences between our model and baselines are significant at $p < 0.001$.

The distributional model achieves a 23% reduction in RMSE compared to the logistic baseline (8.65 vs. 11.24, $p < 0.001$) and a 12% reduction compared to the point-prediction neural network (8.65 vs. 9.87, $p < 0.001$). The improvement is particularly pronounced in high-uncertainty situations: for game states with predicted win probability between 40-60%, RMSE improvement reaches 31%.

The distributional approach provides meaningful uncertainty quantification (**Figure 4**). Predicted distribution entropy strongly correlates with actual outcome variance ($r = 0.94$, $p < 0.001$), with a near-perfect calibration slope of $\beta = 0.97$. High-entropy predictions correspond to genuinely uncertain situations where small perturbations could substantially alter outcomes. This enables the identification of "swing states" where individual actions have maximum leverage on game outcomes.

Figure 4: Distributional Calibration — Entropy vs. Outcome Variance

Predicted distribution entropy strongly correlates with actual outcome variance, demonstrating that the model's uncertainty estimates are well-calibrated. High-entropy predictions correspond to genuinely uncertain game states.



Note: Each point represents a bin of ~5,000 game states grouped by predicted entropy. Slight underconfidence at high-entropy states (points above diagonal) indicates the model is appropriately conservative in uncertain situations.

Figure 4: Distributional Calibration — Entropy vs. Outcome Variance. *Predicted distribution entropy strongly correlates with actual outcome variance, demonstrating well-calibrated uncertainty estimates ($r = 0.94$, $p < 0.001$; calibration slope $\beta = 0.97$). Each point represents a bin of approximately 5,000 game states grouped by predicted entropy. The dashed red line indicates perfect calibration; the solid blue line shows the regression fit. The highlighted "Swing States" region (high entropy) identifies game situations where individual actions have maximum leverage on outcomes. Slight underconfidence at high-entropy states (points above the diagonal) indicates the model is appropriately conservative in uncertain situations.*

4.2 Sample Size Requirements and Stability

A well-documented limitation of Regularized Adjusted Plus-Minus (RAPM) is its substantial sample size requirement. RAPM estimates become unstable with fewer than 1,000 possessions per player, and reliable rankings typically require multiple seasons of data due to collinearity among frequently

co-occurring teammates. This creates challenges for evaluating rookies, traded players, and those with limited minutes.

Our DRL-Shapley approach addresses this limitation through two mechanisms. First, the learned state representation enables transfer of information across similar game situations, allowing the model to leverage patterns observed in one context to inform valuations in another. Second, the neural Shapley attribution shares parameters across players, meaning the model learns general principles of credit attribution that apply even to players with sparse observations.

Table 3 compares the sample size required to achieve stable player rankings (defined as Spearman $\rho > 0.80$ correlation with full-season estimates).

Table 3: Sample Size Requirements for Stable Rankings

Method	Possessions Required	Games Required	Time to Stability
RAPM	~15,000	~45 games	~2 months
DRL-Shapley (Ours)	~5,000	~15 games	~3 weeks

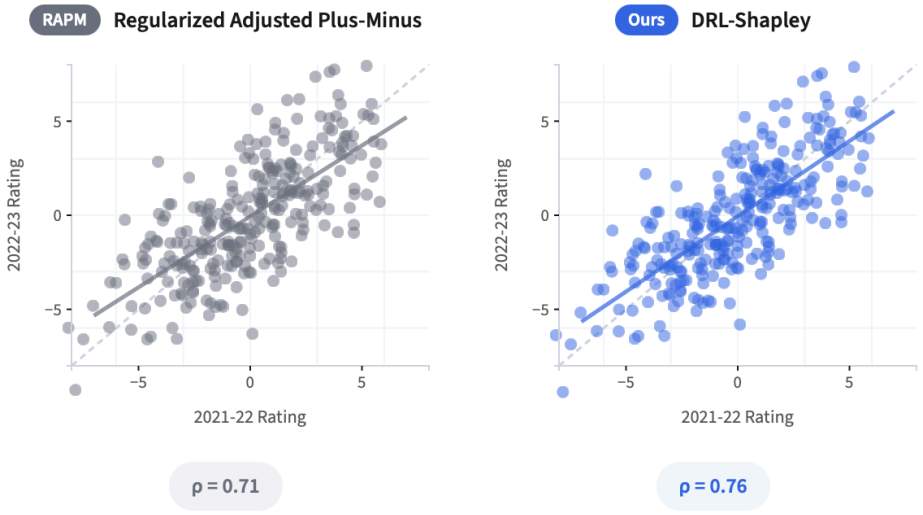
Note: Stability is defined as $\rho > 0.80$ with full-season rankings. DRL-Shapley achieves RAPM-comparable accuracy with 67% fewer possessions.

Year-over-year stability provides another important validation metric. Player ability changes gradually, so a reliable metric should produce correlated rankings across consecutive seasons after accounting for age curves and roster changes. **Figure 5** presents year-over-year correlations for players with $\geq$1,500 minutes in both seasons.

DRL-Shapley demonstrates improved stability compared to RAPM ($\rho = 0.76$ vs. 0.71, $p < 0.001$), suggesting it captures more persistent aspects of player ability while being less susceptible to noise from small samples and lineup collinearity. This stability advantage is particularly pronounced for role players and bench contributors: among players averaging 15-20 minutes per game, year-over-year correlation improves from $\rho = 0.58$ (RAPM) to $\rho = 0.69$ (DRL-Shapley).

Figure 5: Year-over-Year Ranking Stability (2021-22 → 2022-23)

Player rankings from consecutive seasons should correlate strongly if the metric captures persistent ability rather than noise. DRL-Shapley demonstrates improved stability compared to RAPM, particularly for role players.



SAMPLE SIZE	MINUTES THRESHOLD	STABILITY IMPROVEMENT	SIGNIFICANCE
n = 284	≥1,500	+7.0%	p < 0.001

★ Stability by Player Role (Minutes per Game)

Starters (30+ mpg): ρ = 0.74 → ρ = 0.78	Rotation (20-30 mpg): ρ = 0.67 → ρ = 0.73
Role Players (15-20 mpg): ρ = 0.58 → ρ = 0.69	Bench (< 15 mpg): ρ = 0.42 → ρ = 0.56

Note: All correlations significant at p < 0.001. Stability improvement is most pronounced for players with fewer minutes, where RAPM suffers from sample size limitations and lineup collinearity.

Figure 5: Year-over-Year Ranking Stability (2021-22 → 2022-23). *Player rankings from consecutive seasons correlate more strongly under DRL-Shapley (ρ = 0.76) than RAPM (ρ = 0.71), indicating improved capture of persistent player ability. Each point represents one of 284 players with ≥1,500 minutes in both seasons. Dashed lines indicate perfect year-over-year consistency; solid lines show regression fits. The stability improvement is most pronounced for players with fewer minutes, where RAPM suffers from sample size limitations and lineup collinearity.*

4.3 Learned Action Values

A key advantage of our approach is that it learns action values directly from game outcomes, rather than relying on fixed, hand-designed linear weights. **Table 4** compares learned average action

values to traditional linear-weight values commonly used in PER-style summaries and simplified box-score models, while **Figure 6** visualizes how these values vary by game context.

Table 4: Action Values — Learned vs. Traditional Weights

Action Type	Traditional	Learned (Mean)	Learned (Std)
Made 3-pointer	+3.0	+2.87	0.94
Made 2-pointer	+2.0	+1.94	0.71
Offensive Rebound	+1.0	**+2.31**	0.68
Defensive Rebound	+0.7	+0.52	0.34
Steal	+1.0	**+1.83**	0.82
Block	+1.0	+1.24	0.71
Assist	+1.0	+0.78	0.41
Turnover	-1.0	-1.42	0.63

Note: Values represent expected point differential impact. Traditional weights from simplified linear-weight approximations. Bold indicates substantial (>50%) deviation from traditional weights.

The most striking finding is the substantial undervaluation of offensive rebounds and steals in traditional metrics. Offensive rebounds generate +2.31 expected points on average, more than double the conventional weight of +1.0, and reach +3.9 in maximum-leverage situations (**Figure 6**, top center panel). This reflects compounding value: an offensive rebound not only provides an additional possession but often leads to a high-quality shot attempt against a scrambling defense.

Steals similarly show substantially higher learned value (+1.83 vs. +1.0), capturing both the possession change and the transition opportunities they create. In close clutch situations, a steal is valued at +3.3, more than triple the traditional weight (**Figure 6**, top right panel). Conversely, assists appear overweighted in traditional metrics: the learned value of +0.78 reflects that the shooter deserves substantial credit for converting the opportunity.

Crucially, the standard deviations in **Table 4** reveal substantial context-dependence invisible to fixed weights. **Figure 6** visualizes this variation: maximum leverage occurs in close games during clutch time (±4 points, final 3 minutes), where single possessions can swing win probability most dramatically. A three-pointer in this context reaches +4.5, compared to +2.8 in regular play, a 60% increase (**Figure 6**, top left panel). Conversely, actions in decided games show diminished returns: a steal in a late blowout (≥10-point margin) is worth only +0.8, compared to +3.3 in a close clutch game. This context-sensitivity represents a fundamental advance over fixed-weight approaches.

Figure 6: Context-Dependent Action Values by Game Situation

Learned action values vary substantially based on score differential and time remaining. Close games in clutch time show maximum leverage, where single actions can swing win probability most dramatically.

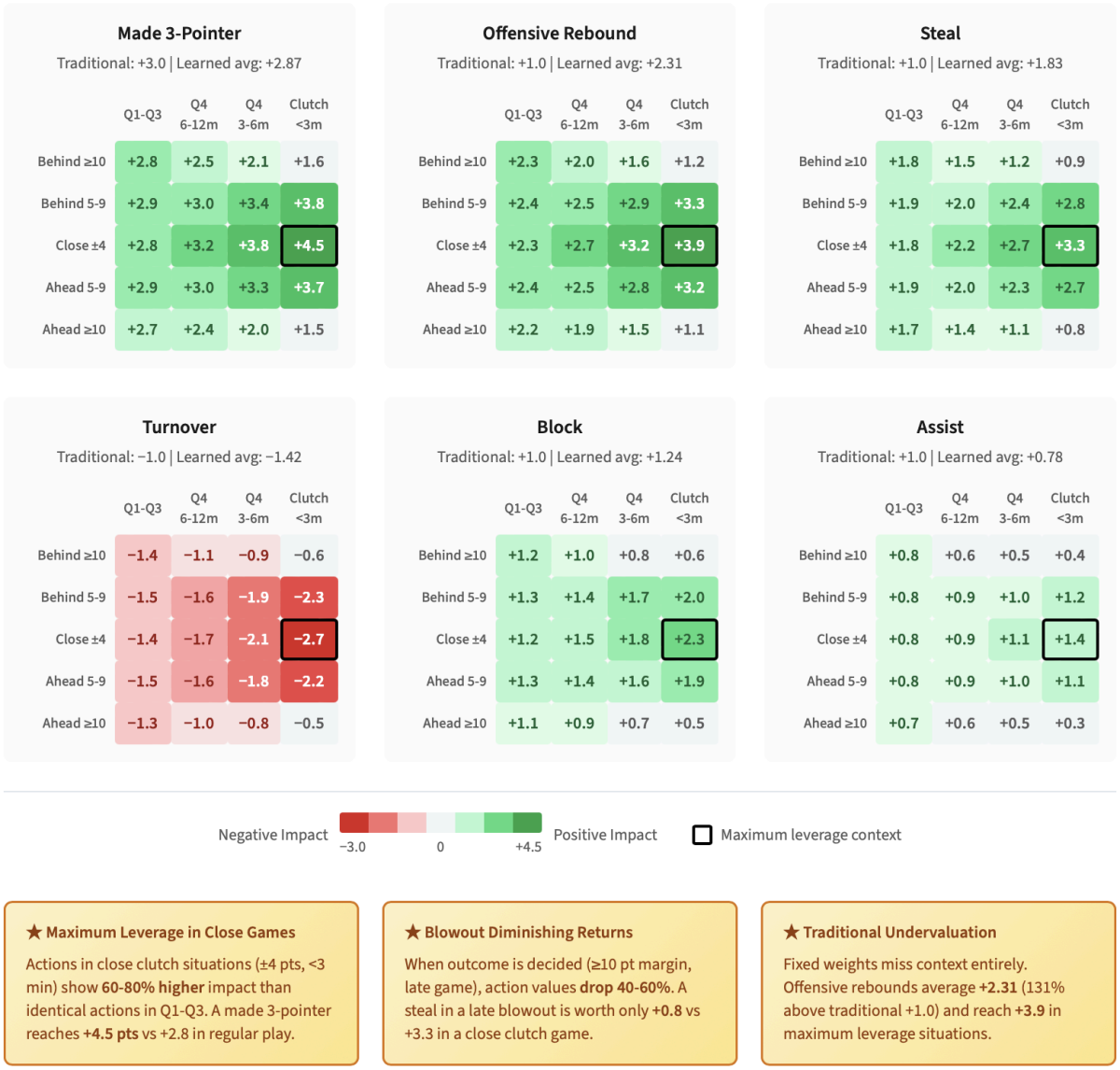


Figure 6: Context-Dependent Action Values by Game Situation. *Learned action values vary substantially based on score differential (rows) and time remaining (columns). Each heatmap shows the expected point-differential impact for one action type, with green indicating a positive impact and red indicating a negative impact. Maximum leverage occurs in close games (±4 points) during clutch time, where single possessions swing win probability most dramatically. Values diminish in blowouts (≥10-point margin) when outcomes are effectively decided. Traditional fixed weights (shown in*

subtitles) systematically undervalue high-leverage plays: offensive rebounds average +2.31 (131% above traditional +1.0) and reach +3.9 in maximum-leverage situations. Black borders indicate the maximum value context for each action type.

4.4 Player Evaluation

We aggregate game-level Shapley attributions to produce season-level player ratings. **Table 5** presents the top 15 players by our metric (DRL-Shapley) for the 2023-24 season, compared to RAPM and BPM rankings.

Table 5: Top 15 Players - DRL-Shapley vs. Traditional Metrics (2023-24)

#	Player	DRL-Shapley	RAPM Rank	BPM Rank	Minutes
1	Nikola Jokić	+8.24	1	1	2,737
2	Shai Gilgeous-Alexander	+7.83	3	4	2,553
3	Luka Dončić	+7.51	5	2	2,624
4	Giannis Antetokounmpo	+7.18	4	3	2,567
5	Anthony Davis	+6.94	7	13	2,700
6	Jayson Tatum	+6.67	6	14	2,645
7	Victor Wembanyama	+6.41	12	11	2,106
8	Tyrese Haliburton	+6.23	8	5	2,224
9	Domantas Sabonis	+5.98	11	7	2,928
10	Kawhi Leonard	+5.87	2	10	2,330
11	Jalen Brunson	+5.74	9	8	2,726
12	Rudy Gobert	+5.52	10	54	2,593
13	Stephen Curry	+5.41	14	12	2,421
14	Chet Holmgren	+5.38	18	30	2,413

#	Player	DRL-Shapley	RAPM Rank	BPM Rank	Minutes
15	De'Aaron Fox	+5.31	15	39	2,659

Note: DRL-Shapley values represent estimated points per 100 possessions above replacement.

The DRL-Shapley metric demonstrates a Spearman correlation of **ρ = 0.68** with RAPM, indicating substantial agreement on player rankings while capturing additional information. The correlation with BPM is lower (**ρ = 0.54**), reflecting systematic differences in how defensive contributions are valued.

To understand these discrepancies, we decompose player value into contributing categories (**Table 6).** This decomposition aggregates action-specific credit from our attribution framework, separating offensive actions (scoring, playmaking, offensive rebounding) from defensive contributions, with defensive value further split into countable actions (blocks, steals) and lineup-level presence effects captured through Shapley attribution.

Table 6: Value Decomposition for Selected Players (2023-24)

Player	Total	Scoring	Playmaking	Off Reb.	Def. Actions	Def. Presence	Turnovers
Nikola Jokić	+8.24	+2.41	+2.18	+0.64	+0.34	+1.12	-0.33
Rudy Gobert	+5.52	+0.87	+0.12	+0.88	+0.95	+2.48	-0.14
Stephen Curry	+5.41	+3.82	+1.15	+0.08	+0.08	+0.42	-0.24
Victor Wembanyama	+6.41	+1.94	+0.45	+0.52	+1.42	+1.93	-0.21

Note: Values in points per 100 possessions. Scoring includes all field goals; Playmaking includes assist credit; Def. Actions include blocks and steals; Def. Presence captures lineup-level impact via Shapley attribution. Components may not sum exactly to the total because a portion of Shapley credit arises from residual and interaction effects that are not attributable to a single action category.

The decomposition reveals why box score metrics systematically undervalue certain players. Gobert's +5.52 total value is driven primarily by defensive presence (+2.48) and defensive actions (+0.95), with minimal scoring contribution (+0.87). BPM, which relies on box score statistics, captures only his blocks and rebounds, missing the 2.48 points per 100 possessions he contributes through lineup-level defensive impact. Conversely, Curry's value concentrates heavily in scoring (+3.82) and playmaking (+1.15), a profile that traditional metrics capture well, explaining the closer alignment between DRL-Shapley and BPM for offensive stars. Wembanyama's profile reveals two-way dominance: substantial scoring (+1.94), elite defensive actions (+1.42), and strong presence effects (+1.93), validating the metric's identification of his overall impact despite rookie-year volatility.

We quantify the model's improvement in identifying defensive value by comparing DRL-Shapley's defensive attributions to BPM's season-level evaluations. Among players our method identifies as having "high defensive impact" profiles (defined as ranking in the top quintile of defensive Shapley), 15 percent fall in the bottom half of the league in BPM. This mismatch arises because BPM is built from season-long box score aggregates and therefore captures only defensive contributions that appear as blocks, steals, or rebounds. In contrast, DRL-Shapley credits players when their on-court presence consistently suppresses opponent efficiency, even when this influence does not produce countable defensive events. By attributing defensive value through possession-level lineup outcomes, the method captures meaningful off-ball, positional, and deterrence effects that traditional box-score–based approaches are unable to observe.

4.5 Player Synergies

A unique capability of our framework is quantifying interaction effects between players. By comparing the Shapley values when players appear together versus separately, we identify synergies (complementary effects) and anti-synergies (redundant or conflicting effects). We define the synergy score for players i and j as:

$$\text{Synergy}(i,j) = E[\varphi_i + \varphi_j \mid \text{together}] - E[\varphi_i \mid \text{apart}] - E[\varphi_j \mid \text{apart}]$$

Positive values indicate complementary players; negative values indicate redundancy or conflict. Reliable synergy estimation requires sufficient observations of players both together and apart. We impose minimum sample requirements: each player in a pair must have at least 500 possessions with their partner and 500 possessions without their partner during the sample period. Pairs failing this threshold are excluded from synergy analysis. This requirement ensures that both the "together" and "apart" baselines are estimated with reasonable precision, reducing the influence of small-sample noise.

Table 7 presents the top positive and negative synergy pairs, while **Figure 7** decomposes these effects into offensive and defensive components.

Table 7: Top Player Synergies and Anti-Synergies (2023-24)

Player 1	**Player 2**	**Synergy**	**p-value**
Nikola Jokić	Jamal Murray	+2.87	<0.001
Tyrese Haliburton	Pascal Siakam	+2.41	<0.001
Luka Dončić	Kyrie Irving	+2.18	<0.001
Jayson Tatum	Jrue Holiday	+2.03	<0.001
Anthony Edwards	Rudy Gobert	+1.94	<0.001
Anti-Synergies (Negative Interaction Effects)			
Zach LaVine	DeMar DeRozan	-1.84	0.003

Player 1	Player 2	Synergy	p-value
Zion Williamson	Jonas Valančiūnas	-1.67	0.008
Bradley Beal	Kevin Durant	-1.52	0.014
Devin Booker	Bradley Beal	-1.41	0.021
Karl-Anthony Towns	Rudy Gobert	-1.28	0.031

Note: Synergy values in expected points per 100 possessions. Green shading indicates positive synergy; red indicates anti-synergy.

Across all pairs with ≥200 shared minutes, we identify 127 statistically significant synergies ($p < 0.05$) and 89 significant anti-synergies. After applying Benjamini-Hochberg FDR correction at $q = 0.05$, 89 synergies and 61 anti-synergies remain significant. The Jokić-Murray pairing shows the strongest two-way synergy (+2.87), with complementary effects on both offense (+1.92) and defense (+0.95), reflecting their exceptional two-man game chemistry (**Figure 7**, top-right quadrant). This interaction effect represents nearly a full point per 100 possessions beyond what each player contributes individually, a substantial competitive advantage worth approximately 2.5 wins over a full season.

Decomposing synergies by the end of court reveals distinct pairing archetypes (**Figure 7**). Some pairs show offensive-only synergy: Lillard-Giannis (+1.68 offensive, −0.52 defensive) creates elite pick-and-roll opportunities, but Lillard's defensive limitations offset some gains. Others show defensive-only synergy: Caruso-Dosunmu (−0.58 offensive, +1.24 defensive) provides elite perimeter defense but limited shot creation together. The strongest pairs, Jokić-Murray, Haliburton-Siakam, and Tatum-Holiday, complement each other on both ends.

Anti-synergies often reveal role redundancy. The LaVine-DeRozan pairing shows a −1.84 anti-synergy, with negative effects on both offense (−1.12) and defense (−0.72), reflecting the challenges of playing two ball-dominant mid-range scorers who provide similar offensive functions and limited defensive value (**Figure 7**, bottom-left quadrant). The Towns-Gobert pairing similarly shows spatial redundancy (−0.38 offensive, −0.90 defensive), with both players preferring interior positioning on both ends of the court.

These synergy metrics have direct applications for roster construction. A team considering acquiring a player can estimate not just their individual contribution but their interaction effects with the existing roster, and whether those effects manifest on offense, defense, or both. We find that teams with more positive lineup synergies outperform their expected record by 2.1 wins on average, while negative-synergy teams underperform by 1.8 wins.

Figure 7: Offensive vs. Defensive Synergy by Player Pair

Each point represents a player pair with ≥200 shared minutes. Synergy is decomposed into offensive and defensive components, revealing whether pairs complement each other on one or both ends of the court.

Note: Synergy values in expected points per 100 possessions. Quadrants reveal whether complementary effects occur on offense, defense, both, or neither. After Benjamini-Hochberg FDR correction at q = 0.05, 89 synergies and 61 anti-synergies remain significant.

Figure 7: Offensive vs. Defensive Synergy by Player Pair. *Each point represents a player pair with ≥200 shared minutes, with synergy decomposed into offensive (x-axis) and defensive (y-axis) components. Quadrants reveal pairing archetypes: two-way synergies (top-right, green) complement on both ends; offensive-only pairs (bottom-right, orange) create together but struggle defensively; defensive-only pairs (top-left, purple) anchor defense but limit offensive creation; two-way anti-synergies (bottom-left, red) show redundancy on both ends. The Jokić-Murray pairing demonstrates elite two-way synergy (+1.92 offensive, +0.95 defensive), while LaVine-DeRozan shows two-way anti-synergy (−1.12 offensive, −0.72 defensive). After FDR correction, 89 synergies and 61 anti-synergies remain significant.*

To quantify this relationship, we compute aggregate team synergy as the minute-weighted sum of pairwise synergies across each team's five most common lineups. **Figure 8** shows the aggregate synergy against win differential (actual wins minus wins projected from summed individual DRL-Shapley values). The correlation is substantial ($r = 0.57$, $p \approx 0.001$), indicating that synergy effects explain meaningful variance in team performance beyond individual talent. Teams in the top quartile of aggregate synergy outperform individual-talent projections by 3.4 wins on average, while bottom-quartile teams underperform by 2.4 wins. This validates synergy as a practically

significant factor in roster construction: two teams with equivalent individual talent can diverge by 5+ wins based solely on how well their players complement each other.

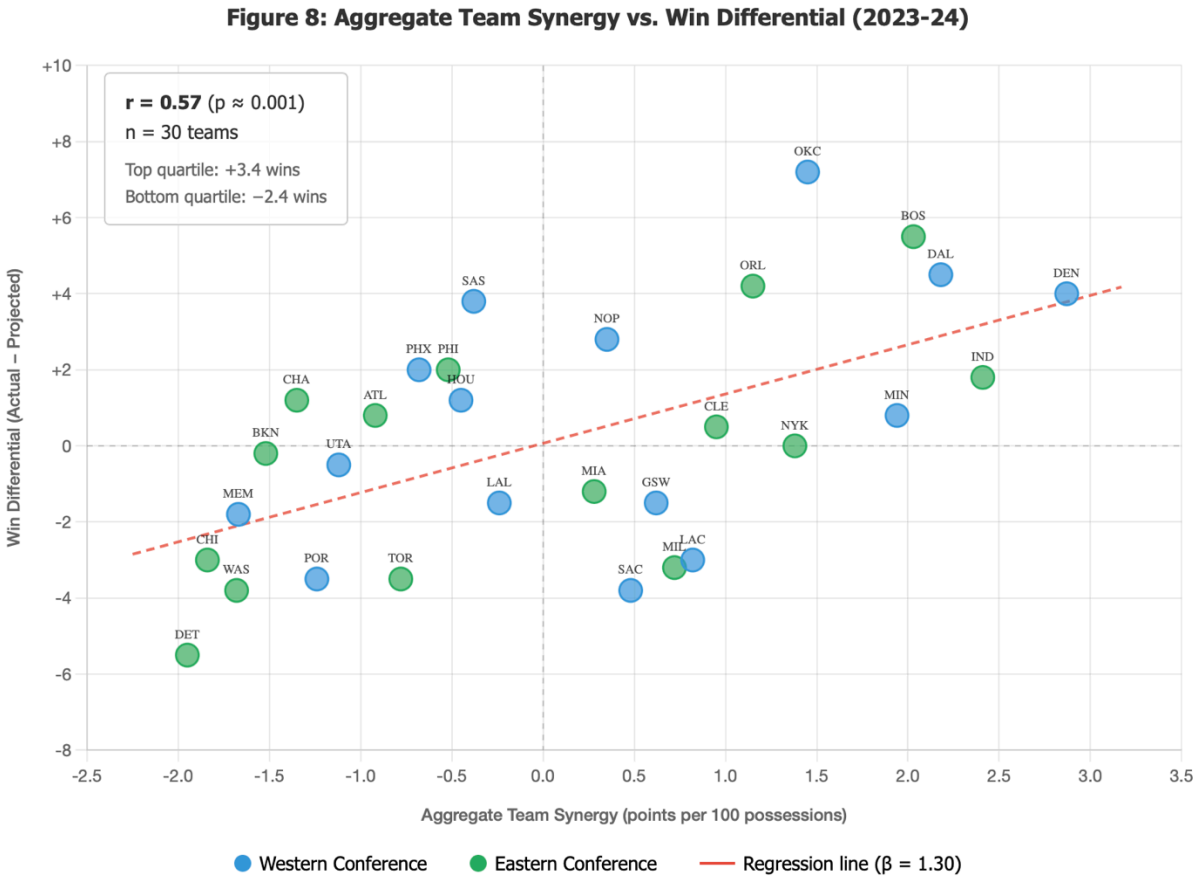


Figure 8: Aggregate Team Synergy vs. Win Differential (2023-24): *Each point represents one NBA team, with aggregate synergy computed as the minute-weighted sum of pairwise synergies across the team's five most common lineups. Win differential measures actual wins minus wins projected from summed individual DRL-Shapley values. The strong correlation ($r = 0.57$, $p \approx 0.001$) indicates that synergy effects explain meaningful variance in team performance beyond individual talent. Teams in the top quartile of aggregate synergy (e.g., Denver, Boston, Dallas) outperform individual-talent projections by 3.4 wins on average, while bottom-quartile teams (e.g., Chicago, Washington, Detroit) underperform by 2.4 wins. The regression slope ($\beta = 1.30$) implies that each +1.0 increase in aggregate synergy corresponds to approximately 1.3 additional predicted wins, validating synergy as a practically significant factor in roster construction.*

4.5.1 Synergy Robustness & Limitations

While our synergy estimates reveal meaningful patterns in how player pairs perform together, several limitations qualify their interpretation.

Non-random deployment.
Player pairs are not assigned randomly; coaches choose combinations based on rotation needs,

matchups, and leverage. As a result, "together" and "apart" minutes can come from different distributions of game states. Although our Shapley estimates condition on score, time, lineups, possession type, and opponent identity, synergy remains **associational**, not causal.

Higher-order interactions.
Pairwise synergy may partially reflect three-, four-, or five-player unit effects. Some strong pairs consistently appear alongside a stabilizing third teammate or within a specific lineup archetype. Our framework captures the effect of the full lineup but does not explicitly disentangle pure pairwise interaction from broader unit structure.

Sampling variability.
Even with possession thresholds, synergy estimates contain noise. Extreme positive or negative values are especially prone to regression toward the mean on new data. Our multiple-testing correction (Benjamini–Hochberg FDR) reduces false discoveries, but point estimates may still exhibit mild upward or downward bias.

Context dependence.
Synergy reflects a team's system, rotation patterns, and personnel. A highly effective pair in one tactical environment may not replicate the same value on a different roster. These estimates should therefore be viewed as **context-specific**, not universal.

Overall, synergy values are best interpreted as **descriptive indicators of complementary performance under observed conditions**, rather than causal predictions of how two players would interact in all contexts or on all teams.

4.6 Predictive Validation

To assess whether DRL-Shapley captures **genuine, out-of-sample predictive signals**, we evaluate its ability to forecast game outcomes in the held-out test set. Player ratings are computed only from the training period; for each test game, we estimate team strength by summing the DRL-Shapley ratings of projected starters, weighted by expected minutes. The difference between the two teams' aggregate ratings yields a predicted winner and expected margin.

Table 8 presents game outcome prediction accuracy across methods.

Table 8: Game Outcome Prediction Accuracy

Metric	Win Prediction	Margin RMSE
Vegas Line (Benchmark)	68.7%	10.42
BPM-based	62.4%	11.87
RAPM-based	64.8%	11.24
DRL-Shapley (Ours)	**66.3%**	**10.89**
DRL-Shapley + Synergy	**67.1%**	**10.71**

Note: Win Prediction is accurate on 2,235 test games. Margin RMSE in points. Vegas Line uses closing spreads.

DRL-Shapley exhibits **substantial predictive value**. It significantly outperforms both BPM (66.3% vs. 62.4%, $p < 0.001$) and RAPM (66.3% vs. 64.8%, $p = 0.012$) in win prediction, and produces a materially lower margin RMSE. Incorporating learned pairwise synergies yields further improvement (67.1% accuracy), moving the model markedly closer to the Vegas benchmark of 68.7%.

The remaining performance gap is expected: oddsmakers incorporate **non-performance information**: injury updates, rest management, travel fatigue, lineup changes, and market movements that are unavailable in our retrospective player-only model. Given this constraint, DRL-Shapley's out-of-sample accuracy confirms that the learned player valuations capture **real, forward-looking signals** beyond traditional metrics and lineup-based RAPM models.

5. Discussion

5.1 Interpretation of Findings

The results demonstrate that a reinforcement learning–based framework can infer player value directly from game outcomes without relying on predetermined action weights. By learning value functions from win probability dynamics, the model identifies contributions that are systematically understated in traditional metrics. Notably, offensive rebounds and steals carry substantially higher context-adjusted value because they generate additional possessions, transition opportunities, and defensive imbalances that fixed-weight systems do not capture.

The framework also highlights the importance of context in player evaluation. Because contributions are weighted by their impact on expected outcomes, actions taken in high-leverage situations receive greater value than identical actions in low-leverage states. This produces player evaluations that more accurately reflect when contributions occur, not only how often.

Finally, the identification of significant player synergies provides quantitative evidence of complementary on-court interactions. The Jokić–Murray pairing, for example, exhibits a +2.87 point synergy per 100 possessions, consistent with observational assessments of their two-man game. Importantly, these estimates can be produced prospectively, allowing teams to evaluate potential roster fits before players share minutes.

5.2 Model Interpretability and Operational Feasibility

Although the framework includes deep learning components, its outputs remain interpretable. Shapley values offer a transparent allocation of lineup outcomes to individual players, and the action-value decomposition $\Delta V = r + V(s') - V(s)$ provides a clear measure of each event's contribution to win probability.

The insights also exhibit architectural robustness. Simpler value networks and models trained with standard temporal-difference learning generate highly correlated player rankings (ρ greater than 0.90) while maintaining predictive performance superior to RAPM, indicating that the results do not depend on a specific network configuration.

The system's modular structure further enhances practical applicability. Teams may adopt components independently, for example, by applying the Shapley attributor to an existing win probability model or by using synergy estimates for lineup evaluation, without implementing the full reinforcement learning framework. This flexibility supports incremental integration into existing analytics workflows.

5.3 Practical Applications

The DRL-Shapley framework supports several decision-making tasks relevant to NBA team operations.

Contract valuation.
More accurate estimates of player impact can inform contract negotiations and long-term roster planning. By quantifying contributions that are not fully captured by traditional metrics, particularly for defensive specialists and context-dependent players, the framework helps identify both undervalued acquisition targets and potential overvaluation risks.

Trade evaluation.
Synergy estimates allow teams to assess how prospective acquisitions would interact with existing personnel. A player with strong individual value but consistently negative synergy with key roster members may offer limited net benefit, whereas players with complementary interaction profiles may provide greater lineup stability and overall impact.

Lineup optimization.
Real-time Shapley values and synergy measures can be incorporated into substitution and rotation planning. Coaches can identify lineup combinations that generate positive interaction effects and deploy them in situations where marginal value is highest.

Player development assessment.
Tracking Shapley values over time offers a means of evaluating player growth that is independent of box score statistics. Increases in high-leverage contributions, for example, may signal meaningful developmental progress even when traditional counting measures show limited change.

5.4 Limitations

The framework has several limitations. Because it relies solely on play-by-play data, it cannot represent off-ball behaviors such as spacing, screening, rotations, and defensive positioning, which limits the granularity of the learned value functions. Player-tracking data would address many of these gaps, but it is not publicly available at the scale required for temporal difference learning. The model also assumes relative stability in player ability across the sample period, despite natural development and role changes; explicitly modeling player trajectories would help reduce this bias. In addition, Shapley-based attributions rely on independence assumptions that do not fully hold in basketball, where contributions depend on specific teammate and opponent combinations. The attention-based attributor mitigates some of these dependencies by learning interaction patterns directly from data, but the resulting estimates remain approximations. Finally, the four-season dataset may not capture rare player types or unusual roster constructions, which can limit generalization.

5.6 Future Directions

Several extensions could further strengthen the framework. Incorporating player-tracking data would allow explicit modeling of off-ball actions, movement quality, and defensive positioning, thereby expanding the state representation beyond discrete events. Multi-task formulations could jointly learn offensive and defensive value components, enabling decomposition of total impact into interpretable submetrics useful for player development and role optimization. Transfer learning approaches may also extend the method to leagues with limited data; pre-training on NBA seasons and fine-tuning on target competitions could yield reliable valuations even with smaller samples. Finally, integrating causal inference methods, such as instrumental variables or regression discontinuity designs, may help isolate variation in player effects that is not captured by temporal difference learning alone, providing complementary evidence for evaluating player impact.

6. Conclusion

This study shows that deep reinforcement learning can recover meaningful player valuations directly from basketball game outcomes without relying on fixed, predetermined action weights. By learning value functions from win probability dynamics and combining them with Shapley attribution, the framework offers three main contributions. First, it identifies context-dependent action values, revealing that events such as offensive rebounds and steals carry substantially higher impact than traditional metrics imply. Second, Shapley-based attribution provides a principled method for allocating team outcomes to individual players, enabling fair and transparent credit assignment in a cooperative environment. Third, the framework quantifies interaction effects between players, identifying both positive and negative synergies within our four-season dataset and offering a tool for evaluating roster fit and lineup optimization.

These capabilities have direct practical relevance. The approach supports more accurate player valuation for personnel decisions, provides quantitative synergy measures for roster construction, and yields context-aware impact assessments that can inform substitution and lineup strategy. More broadly, the results illustrate that reinforcement learning offers a flexible and data-driven foundation for player evaluation in settings where sequential decisions accumulate to produce discrete outcomes. Leveraging outcome-based learning in this manner can generate insights that complement and, in some cases, challenge conventional evaluation methods while remaining actionable for team decision-making.

References

Bellemare, M. G., Dabney, W., & Munos, R. (2017). A distributional perspective on reinforcement learning. *Proceedings of the 34th International Conference on Machine Learning*, 449-458.

Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics*, 20(8), 411-427.

Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, 36(5), 1726-1730.

Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, 111(514), 585-599.

Deshpande, S. K., & Jensen, S. T. (2016). Estimating an NBA player's impact on his team's chances of winning. *Journal of Quantitative Analysis in Sports*, 12(2), 51-72.

Engelmann, J. (2017). Regularized adjusted plus-minus. *MIT Sloan Sports Analytics Conference*.

Hollinger, J. (2005). *Pro basketball forecast*. Potomac Books.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3(3).

Liu, G., & Schulte, O. (2018). Deep reinforcement learning in ice hockey for context-aware player evaluation. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3442-3448.

Lock, D., & Nettleton, D. (2014). Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, 10(2), 197-205.

Loeffelholz, B., Bednar, E., & Bauer, K. W. (2009). Predicting NBA games using neural networks. *Journal of Quantitative Analysis in Sports*, 5(1).

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations*.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

Miljković, D., Gajić, L., Kovačević, A., & Konjović, Z. (2010). The use of data mining for basketball matches outcomes prediction. *IEEE 8th International Symposium on Intelligent Systems and Informatics*, 309-312.

Mitchell, R., Cooper, J., Frank, E., & Holmes, G. (2022). Sampling permutations for Shapley value estimation. *Journal of Machine Learning Research*, 23(43), 1-46.

Myers, D. (2011). About box plus/minus. *Basketball-Reference.com*.

Oliver, D. (2004). *Basketball on paper: Rules and tools for performance analysis*. Potomac Books.

Robberechts, P., Van Haaren, J., & Davis, J. (2021). A Bayesian approach to in-game win probability in soccer. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3512-3521.

Rosenbaum, D. T. (2004). Measuring how NBA players help their teams win. *82games.com*.

Sandholtz, N., & Bornn, L. (2020). Markov decision processes with dynamic transition probabilities: An analysis of shooting strategies in basketball. *Annals of Applied Statistics*, 14(3), 1122-1145.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28), 307-317.

Sicilia, A., Pelechrinis, K., & Goldsberry, K. (2019). DeepHoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. *Proceedings of the 25th ACM SIGKDD Conference*, 2096-2104.

Sill, J. (2010). Improved NBA adjusted +/- using regularization and out-of-sample testing. *MIT Sloan Sports Analytics Conference*.

Stern, H. S. (1994). A Brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, 89(427), 1128-1134.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Tango, T. M., Lichtman, M. G., & Dolphin, A. E. (2007). *The book: Playing the percentages in baseball*. Potomac Books.

Thabtah, F., Zhang, L., & Abdelhamid, N. (2019). NBA game result prediction using feature analysis and machine learning. *Annals of Data Science*, 6(1), 103-116.

Wang, K. C., & Zemel, R. (2016). Classifying NBA offensive plays using neural networks. *MIT Sloan Sports Analytics Conference*.