The Harmony Index: Evaluating, Predicting, and Visualizing Effectiveness in Multi-Agent Team Dynamics

Darryl Roman, Noah Ari, Johnathan Mell
University of Central Florida
darryl.roman@ucf.edu, noah.ari@ucf.edu, johnathan.mell@ucf.edu

Abstract

Team-based games are a keystone pillar of the gaming industry. Sadly, the understanding of team dynamics—and the recommendations for both human and AI-based teammates—are based on a rudimentary understanding of human-AI teaming. We propose a superior metric, which provides information about team effectiveness in an efficient and easily-replicable manner. Without an accurate and effective metric for team evaluation, it is nigh-impossible to provide feedback to players and game designers to improve team balance. We provide such a metric. The Harmony Index, a novel algorithm using real-world data, provides simpler and more accurate actionable directives to improve game design across MOBAs and other game genres. We prove its predictive power in a separate analysis and make recommendations for its use in assessing team effectiveness as well as its future use in additional domains.

Introduction

Team effectiveness is a complex interplay of relationships and interactions between its members, making it challenging to assess the institutional and norm-based complexities of team coordination. To overcome this challenge, we propose the Harmony Index, a simple and accurate measure of team effectiveness and synergy that can be used as a gold standard for comparisons across teams in various domains.

For teams consisting of human players playing with and against digital characters, accurate assessment and proficiency to predict their success and improvement is a complex task which requires metrics that are easily understandable, comparable, and objectively derived.

The Harmony Index is based upon the geometric mean—which is used heavily in the statistical analysis toolbox. However, our proposal goes beyond prior attempts by making a metric which is simultaneously accurate, easier to calculate, and provides a series of easy to visualize connections which game designers and player can use to increase the impact factor of their analyses. Today the geometric mean is heavily leaned upon with variations used in domains such as Biological Fitness, Finance, Group Decision Making, and Risk Evaluation (Buckland et al. 2011; Estrada 2010; Escobar, Aguarón, and Moreno-Jiménez 2004; Wang et al. 2009).

Our novel formula bolsters the fidelity of existing methods by integrating both inclusive and exclusive success rates of each agent with the other members of the prospective team. Our methodology utilizes a bottom-up utilitarian strategy to organize virtual agents into a single Harmony Index, regardless of whether the team composition had performed together before. Furthermore, we provide predictive impact, displaying information about novel team compositions which have not yet been brought together, demonstrating superior effectiveness to current methods.

We achieve this by analyzing the pairwise effect of each pair or dyad on team success rates and categorizing each agent dyad into four harmonic classes, based on the degree of effect on each agent; Harmony, Uplift, Depress, or Discord. These classes demonstrate the benefit to the overall success rate, as well as the benefit to each member of the dyad. This approach can be applied to a team of arbitrary size to measure the benefit to the team-wide organization.

Our work builds on previous multilevel models that calculate individual agents and team metrics to determine team performance, with a focus on the top-down effect of agent behavior on the success of the team as a whole (DeShon et al. 2004). In contrast, we are fundamentally concerned with the aggregate data concerning dyadic relationships among roles. Our methodology utilizes a bottom-up approach to measure how well an arbitrary number of team members work together as opposed to apart and prioritizes the collective dynamics of roles within a team structure, offering a broader predictive framework that can be applied to diverse team compositions.

The Harmony Index neatly avoids the problem of exponential interaction terms (which are impractical to measure), and instead summarizes these interactions into an easily consumed scalar number computed using proven success rates on the intended task or set of tasks. Validated using a human-subjects data set of 1.67 million data points, we show that the Harmony Index is a useful metric which:

- Determines the team strength of a given dyad,
- Is extensible to teams of any size n,
- Measures team effectiveness more accurately than other measures,

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

- Can be used predictively to assess the strength of teams, even when those teams are unique and not contained in our data set.

Overall, our proposed Harmony Index provides a comprehensive and objective approach to measure team effectiveness, which can be applied to a broad range of domains and team compositions, including human and artificial agents. We believe that our work offers a significant contribution to the ongoing scientific study of measuring team effectiveness.

The present study introduces a novel quantitative framework that generates Harmony Indices and identifies four distinct classes of dyadic relationships. Additionally, we extend this framework to teams of arbitrary size, providing a comprehensive methodology for analyzing team dynamics. Our results indicate that the Harmony Indices are both statistically normal and highly descriptive, providing a robust foundation for predicting the effectiveness of previously unseen teams. Furthermore, our approach allows for near real-time adaptation to changes in team performance as additional data points become available. Overall, our framework offers a powerful tool for understanding and optimizing team dynamics, with broad applications across a range of fields.

Game Description

Our study utilizes data from the Multiplayer Online Battle Arena (MOBA) game, Heroes of the Storm, known for its competitive team-based dynamics. This game serves as a practical model for analyzing teamwork dynamics under pressure. Heroes of the Storm is used both for the comprehensiveness of its data and for the features that make it substantially more collaborative vs. individualistic compared to other MOBA games such as League of Legends.

With a roster of 90 distinct characters and over 4,000 unique interactions, the game offers diverse team compositions, including all-human, all-agent, or mixed teams. Each character starts with a neutral success rate of 50%, emphasizing the necessity of optimal team performance for success.

Characters in the game possess specialized traits and skills, categorizing them into specific roles reflecting their abilities. Teams strategically organize these roles to gain an edge over opponents.

Data collection was performed by the statistical analysis website HotSLogs.com via a client-side application that players voluntarily downloaded and installed on their local machines. This application enabled the upload of full game log files to a remote server, where analysis was performed and game data was presented back to the players. These data were matched to player information obtained via an API provided by Blizzard Entertainment. For our analysis, the data underwent a rigorous cleaning process to remove any features that could be traced back to identifiable players, ensuring anonymity. These data were shared freely, with permission from the owners, with our research team[1].

[1]The datasets used and/or analyzed during the current study are available from the corresponding author upon request.

Our data collection involved representing team members as in-game characters, each with their unique attributes. Teams were matched based on collective skill levels, enabling us to observe both individual character strengths and team performance. Our study scrutinized a substantial dataset comprising 1.67 million character-specific data points, representing one of the most comprehensive sources of team dynamics that is available for analysis—we provide detailed information about cooperation, competition, and the diversity of team analytics available among a deeply influential genre in competitive games.

Our method employs mathematical generalization, utilizing real performance data to analyze various factors affecting team dynamics, such as individual character features and interpersonal dynamics. Our findings offer insights applicable to assembling effective cross-functional teams in any domain requiring effective teamwork for success.

Related Work

In previous research, a team has been characterized as a collection of agents united by shared goals, but with distinct roles and responsibilities to accomplish tasks as an organized entity (Kozlowski and Ilgen 2006). However, an open question remains: is team effectiveness merely the sum of each member's measures of individual effectiveness, or does it instead arise from a diverse set of specializations that contribute to achieving the overarching objective (Kozlowski and Klein 2000). While some investigation has explored the psychological factors that impact team performances (Hastings et al. 2018), other research has proposed that teams could be considered as an organizational framework factoring for the members' interdependence (Walliser et al. 2019). Our study offers a unique methodology of examining the quantitative measures of team performance and the resulting inter-agent relationships between team members. By focusing on this aspect, we aim to provide novel insights into the underlying mechanisms that drive successful teams.

MOBAs have been a prominent focus of research on team performance, with numerous studies investigating this fascinating space. For example, one study delved deeply into the impact of individual player performance over time, as well as the effects of training on their improvement (Sapienza et al. 2018). Moreover, past literature has demonstrated that MOBAs are an ideal setting for exploring the intricacies of team dynamics. A particularly intriguing study by J. Kim, Keegan, Park, and Oh (2016) examined the various methods that players employ to construct a cohesive team that can successfully complete sub-tasks and achieve the main objectives before their opponents do (Kim et al. 2016). These works provide compelling insights into the nature of teamwork in MOBAs, shedding light on the factors that contribute to success in this challenging and competitive domain. MOBAs represent one of the most optimal environments for studying machine-human collaboration and competition.

Their paper investigates the impact of expertise, self-efficacy, and cohesion on team performance in the context of MOBAs (Kim et al. 2016). In 2017, Y. Kim et al.

were able to demonstrate that teaming in MOBAs is affected by many of the same factors that impact teaming in other domains (Kim et al. 2017). it's worth noting that the study didn't attempt to assess or predict individual skill levels, which may require more in-depth mathematical modeling of each agent's unique characteristics and features. This simplicity and generality is key to the Harmony Index, which enables effective predictions without explicitly modeling every individual feature. While these features are accounted for, they are not expressly formalized. Instead, significant individual traits and their interactions are observed only through their dyadic relationships. By abstracting general properties into relationship information we can recombine these pairwise relationships, sidestepping the exponential complexity of explicit coding. Leveraging established data on these relationships, the efficacy of larger teams is assessed, yielding valuable insights into the dynamics of MOBA teamwork.

Building on prior research, our study explores the notion that metrics for team effectiveness are most accurately measured at the level of the entire system, rather than on an individual level (Gorman, Grimm, and Dunbar 2018). This approach is particularly relevant to the study of Human-Agent teaming, where the ultimate goal is to construct teams that can consistently and effectively achieve desired outcomes (Damacharla et al. 2018; Berretta et al. 2023), as well as to the work of optimizing trust and collaboration between human and digital agents (Seeber et al. 2020; Dennis, Lakhiwal, and Sachdeva 2023). By adopting this broader perspective, our study seeks to shed light on the mechanisms that enable successful team performance in complex systems, providing insights that can inform the development of more effective Human-Agent teams.

Hill et al. conducted a thought-provoking study that examined the impact of systemic complexities and resilience on team performance, using dyadic pairs as a lens (Hill et al. 2020). Their hypothesis was that reductions in resilience and adaptability within the team environment would result in a reactive reduction in dyadic performance. Inspired by this work, our study proposes a novel methodology for measuring dyads and generating a scalar representation of their resilience, which can be extended to assess the resiliency of teams of any size. This approach promises to offer a more comprehensive understanding of how teams function and how their resilience impacts their ability to succeed in complex environments, ultimately contributing to the development of more effective team-building strategies.

Drawing from the dynamic and complex domain of MOBAs, our study aims to shed light on effective methods for measuring team success rates, with the goal of building institutional teams that combine diverse skill sets, knowledge, and effective cognitive and behavioral processes. Building on prior work by (Hall et al. 2018) and (Cheng et al. 2019), we recognize that team composition has a significant impact on team effectiveness and on the balance of the match as a whole. To this end, we present a novel method for maximizing the selection of criteria optimized for a specific objective and offering insights into building more effective and well-rounded teams able to tackle complex tasks with greater efficiency and success. Our metric is complementary to existing systems like MMR, but differs insofar as it measures team performance as individual components, rather than directly measuring individual or team performance as an atomic entity. By understanding team dynamics using a straightforward metric, we can provide actionable insight to players and designers.

Component	*Representing*
$P(A)$	Win % given that Agent A is Team T
$P(B)$	Win % given that Agent B is Team T
$P(A \cap B)$	Win % given that Agents A & B are on Team T
$P(A \cap \bar{B})$	Win % given that Agent A is on Team T & Agent B is not.
$P(B \cap \bar{A})$	Win % given that Agent B is on Team T & Agent A is not.
n	Number of team members whose relationship we are testing

Table 1: Components of the Harmony Index

Furthermore, in a recent study, researchers proposed to use the harmonic mean, a classic Pythagorean mean, to approach team assembly specifically for Soccer (Afshar et al. 2018). Their approach involved characterizing the skill sets of potential team members and deriving a harmonic mean of those roles. In contrast, our study introduces a novel methodology for calculating a unified index metric that takes into account individual character proficiencies, skills, and effects to determine the neighbor overlap, the performance effect of one agent upon another agent, for any given agent pairing. By doing so, we aim to provide a more comprehensive and nuanced understanding of the factors contributing to team performance. Our approach provides valuable insights into team dynamics that can be applied across a range of team-based contexts, beyond the domain of Soccer and MOBAs.

Mathematical Formulation

Harmony Index

We devised a streamlined method to quantify team strength, leveraging a generalized formula yielding a singular scalar value. This formula, outlined in Table 1, begins by employing the geometric mean by multiplying individual members' success rates and raising the result to the reciprocal of the team size. However, if stopped here, this approach overlooks nuanced factors impacting team dynamics.

Mathematically expressing factors influencing team effectiveness is traditionally labor-intensive, involving exhaustive testing and manual formulation. Especially in intricate scenarios, the computational overhead of comparing numerous distinct features across agents is substantial.

To mitigate this challenge, we introduce the Harmony Index—a lightweight metric emphasizing historical interactions among agents. By focusing on dyadic relationships, this metric facilitates the construction of effective teams. Dyads, the most fundamental level of relationship data given that any team of size n can be composed of n(n-1)/2 unique dyads, serve as the elemental building blocks for for reasoning and extrapolating about teams of size n.

In a team comprising n members, each of the $n(n$-1)/2 unique dyads encapsulates bidirectional relationship insights. To extract this information, we decompose each unique dyad into its constituent unidirectional components, generating new dyads from each agent's standpoint. For instance, in a three-member team with agents A, B, and C, the unique dyads are (A,B), (A,C), and (B,C), which decompose into (A,B), (B,A), (A,C), (C,A), (B,C), and (C,B).

Subsequently, these unidirectional dyads represent the first agent's perspective on their relationship strength with the second agent, facilitating team construction. Finalizing the calculation necessitates expressing this perspective mathematically.

Procedure

When assessing relationship efficacy from an agent's viewpoint, relying solely on dyadic success metrics overlooks the unique impact of the specific agent pairing on individual performance. To address this, we contrast an agent's success rate when teamed with another agent ($P(A \cap B)$) against their success rate without that agent ($P(A \cap \bar{B})$). This yields $\frac{P(A \cap B)}{P(A \cap \bar{B})}$, representing agent A's perspective on the relationship strength. However, this formulation lacks consideration for agent B's perspective, which is then captured by using $\frac{P(B \cap A)}{P(B \cap \bar{A})}$ as detailed in Equation 1.

$$\frac{P(A \cap B)}{P(A \cap \bar{B})} * \frac{P(A \cap B)}{P(B \cap \bar{A})} \quad (1)$$

We then incorporate both agents' perspectives on dyadic effectiveness into the previous geometric mean, as outlined in Equation 2.

$$\sqrt[2]{\frac{P(A \cap B)}{P(A \cap \bar{B})} * \frac{P(A \cap B)}{P(B \cap \bar{A})}} \quad (2)$$

After calculating the intersections and inclusions as prescribed by equations 1 and 2, we can combine them using a new geometric mean to measure the relationship between these two specific team members. This provides a strong indicator of how each agent was either uplifted or depressed by their teammate.

For larger teams, we consider each agent's perspective on its relationship with every team member (yielding $n^2 - n$ unidirectional dyads), and then apply the geometric mean across these dyadic values using the generalized formula in Equation 3. The resulting scalar value, termed the Harmony Index, quantifies team effectiveness based on historical member interactions.

$$\sqrt[n(n-1)]{\prod_{A \in T} \prod_{B \in T} \frac{P(A \cap B)}{P(A \cap \bar{B})}} \quad (3)$$

Although our dataset comprised teams of size n=5, we analyzed it at the dyadic level rather than the full team level to compute the Harmony Index. Indeed, this dyadic approach is key to our contribution. Additionally, for evaluating the Harmony Index of a specific n=5 team, we excluded all data related to that team's actual performance from the dataset, reserving it as a ground truth for later assessment.

Thus, even if data for a particular n-sized team is incomplete, the Harmony Index remains computable through fundamental dyads. By anchoring our metric in dyadic data, we ensure its broad applicability. This process allows any team size k to be deconstructed into dyads and reconstructed into teams of size n, yielding a metric that accurately reflects team strength in terms of coordination. Our contribution lies in extending known dyadic relationships to untested teams of arbitrary size. We predict these teams' efficacy by expanding and amalgamating inter-agent relationships, validating our predictions against existing data on larger n-sized teams.

To validate our predictions on untested teams, we isolated all data for each specific team composition and recalculated to generate an Excised Harmony Index (Excised HI). Subsequent analysis in Results will contrast these values with the standard Harmony Index and geometric mean, demonstrating predictive capability for unseen n-sized teams.

We've demonstrated the feasibility of our method by extending the generalized Harmony Index formula for n=3, as depicted in Equation 4.

$$\sqrt[6]{\begin{array}{l}\frac{P(A \cap B)}{P(A \cap \bar{B})} * \frac{P(A \cap B)}{P(B \cap \bar{A})} * \frac{P(A \cap C)}{P(A \cap \bar{C})} * \\ \frac{P(A \cap C)}{P(C \cap \bar{A})} * \frac{P(B \cap C)}{P(B \cap \bar{C})} * \frac{P(B \cap C)}{P(C \cap \bar{B})}\end{array}} \quad (4)$$

A similar expansion allowed us to conduct pairwise analysis and apply the result to the MOBA environment where n=5 in order to encompass the entire team.

Notably, while a dyadic Harmony Index of 0.0 is plausible, implying a 100% failure rate for the pair, such extreme cases were not observed in our dataset. Even the least successful dyad, with a 15.8% success rate over 19 games, yielded a Harmony Index of 0.33, indicating significant discordance. Despite the smaller sample size, this underscores the robustness of our data.

By extending beyond pairwise analysis, we attained a deeper insight into team dynamics within complex environments like MOBAs. This advancement holds key implications for team performance prediction across diverse settings.

Relationship Classes

Our analysis under Mathematical Formulation, incorporating interaction patterns and our formulated model, reveals four distinct classes of pairwise relationships: Harmony, Discord, Uplift, and Depress. The selection of these four categories is derived from the rational segmentation of the data, which clearly delineated these groups according to the relational dynamics within each dyad. Specifically the categories can be explained as so: mutually beneficial, mutually detrimental, beneficial for one and detrimental for the other (but overall beneficial) and beneficial for one and detrimental for the other (but overall detrimental). These capture the essential patterns of interaction between dyad members. We argue that these distinctions reflect recurrent differences in how agents are influenced by either uplift or depression, making these categories representative of the possible dyadic relationships. Leveraging these classifications provide a comprehensive framework for analyzing team dynamics in dynamic environments like MOBAs.

Our findings provide important insights into the dynamics of team coordination and can be used to improve team performance in various settings or to enhance efficiency and speed of matchmaking. We believe that our approach offers a valuable contribution to the field of coordination research and look forward to feedback from the scientific community.

Harmony, Discord, Uplift, & Depress

Our analysis revealed two classes of interaction where both agents receive the same effect from their collaboration. The first class, Harmony, is characterized by an increase in task success rate for both agents relative to their individual success rates. In the vast majority of cases, Harmonious pairings resulted in a success rate above the target rate of $> 50\%$. For instance, we observed the highest Harmony Index in our data set between agent 40 ($P(A)$ = 46.44%) and agent 66 ($P(B)$ = 48.67%), who had a Harmony Index of 1.28 and a joint task success rate of 60.81% — an average increase of 13.2%.

On the other hand, we defined Discordant pairings as those where the joint task success rate is lower than the success rate of each individual agent. This class of pairings indicates that the agents are not effective in improving their effectiveness and, in fact, they may even interfere with each other. The lowest Harmony Index in our data set was between agent 53 ($P(A)$ = 41.55%) and agent 71 ($P(B)$ = 42.70%), with a Harmony Index of 0.72 and a paired success rate of 30.49% — an average decrease of 11.63%.

The Uplift and Depress classes reveal nuanced outcomes of pairwise interactions. In the Uplift class, one agent experiences an increase in effectiveness while the other experiences a decrease, resulting in an overall increase in pairwise task success rate. For instance, agents 7 and 39 achieved a pairwise success rate of 57.13%, even though agent 39's effectiveness decreased by 1.48%, while simultaneously uplifting agent 7 by 10.84%. In contrast, the Depress class presents potentially divisive pairings where the average success rate of the agents decreases, but one agent still benefits. For example, agents 53 and 37 had a pairwise success rate of 42.56%, with agent 53 experiencing a 1.01% increase in effectiveness and agent 37 suffering a 10.37% decrease. This creates a Harmony Index of 0.90, indicating an overall negative impact. However, specialized tasks or multiple vectors for success could still make this trade-off effective in certain scenarios.

These findings offer valuable insights into the dynamics of pair-wise coordination in MOBAs and provide a framework for understanding the effects of collaboration on task success rates. By categorizing pair-wise interactions into distinct classes, we are able to identify which pairs are most effective at achieving task success and which pairs may need additional support or intervention. We believe that these findings have important implications for improving team coordination in dynamic environments and can be applied to various domains beyond MOBAs.

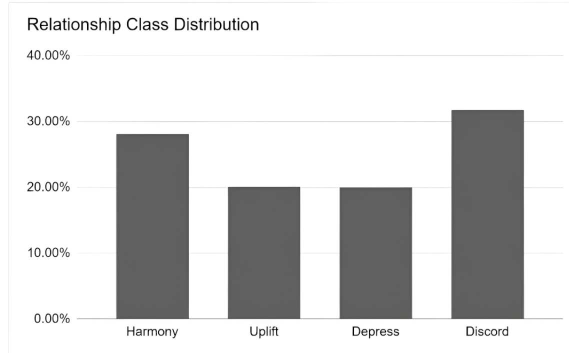

Figure 1: Harmony Index, Class Distribution

Class Distribution

Through analysis of the 1.67 million data points across 90 agents in the n=2 scenario, we identified four distinct classes of pairwise relationships: Harmonious, Uplifting, Depressed, and Discordant. Cleaned results, depicted in Figure 1, unveiled 812 purely Harmonious interactions, 581 Uplifting interactions, 579 Depressed interactions, and 918 purely Discordant interactions.

To ensure statistical robustness, we excluded agent pairings with fewer than 1,000 intersecting data points.

Results

Comparison to Existing Measures

In an ideal scenario with complete data on all possible teams, a simple geometric mean would allow direct computation of success rate. However, our focus lies in scenarios where data are incomplete. When the examined team composition is unprecedented, traditional methods fail to determine its effectiveness. Yet, by constructing a matrix of pairwise relationships, we leverage past team compositions to assess the new team's efficacy. We thus contribute to the existing methods by providing an improved metric.

Moreover, while a geometric mean is conventional for combining probabilistic values like win rates, it overlooks crucial interaction terms among individual team members. These interactions, which could positively or negatively influence effectiveness, multiply exponentially with team size. Consequently, employing a simple regression and geometric mean becomes impractical due to the exponential growth in interaction terms. Utilizing the Harmony Index of dyads enables the incorporation of each member's impact on their peers into the calculation.

These interaction possibilities are exponential as team size increases. Therefore, calculating these interactions using a simple regression and geometric mean is impractical as the number of interaction terms grows exponentially. Using the Harmony Index of dyads allows the impact of each prospective member on its peers to be factored into the calculation.

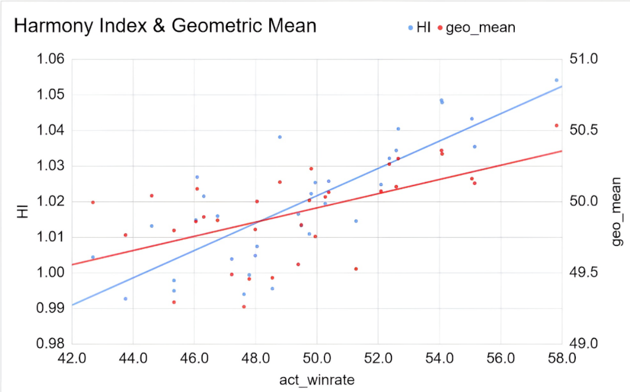

Figure 2: Harmony Index and Geometric Mean

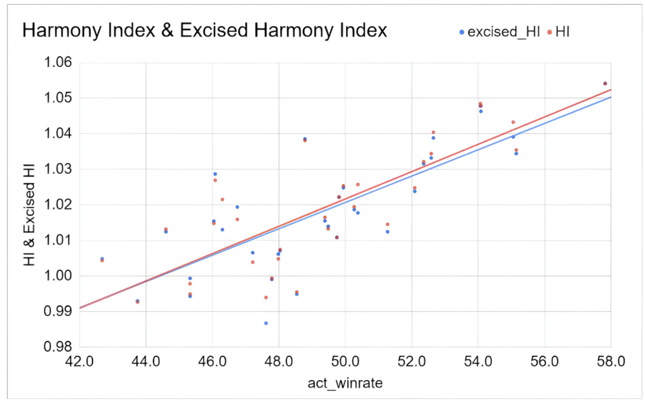

Figure 3: Harmony Index and Excised Harmony Index

	HI	Geo Mean
Pearson's Corr.	.799**	.563**
Sig (2-tailed)	$< .001$	$< .001$
n	33	33

Table 2: Pearson's Correlations

Our analysis reveals a robust correlation between the Harmony Index (HI) and the actual team win rate. Illustrated in Figure 2 and summarized in Table 2, this correlation yields a Pearson's coefficient of 0.799, with a significance level below 0.001. Notably, this correlation is stronger than that observed between the geometric mean of standard individual win rates and the actual win rate (0.563). Hence, the Harmony Index emerges as a dependable predictor of a team's actual win rate.

Furthermore, our results remain robust even for teams with a smaller sample size of 50 games. In this instance, the Pearson's correlation coefficient is 0.522, with a significance level of 0.01 for a sample size of 58 teams. These findings carry significant implications for comprehending and predicting team performance across diverse contexts.

Predictive Analysis

The ability to predict the performance of teams that have never worked together before is a crucial challenge in the field of team science. Simply assessing the past performance of superficially similar teams is not sufficient for this purpose. Our Harmony Index provides this functionality. It is crucial to emphasize that even when all possible dyadic permutations within a team of size n are considered, it does not imply that the entire team has previously performed as a cohesive unit. In fact, our analysis of 252 possible 5-member permutations of 6 roles across approximately 820,000 teams revealed 58 team permutations that were not observed performing together in our dataset. Despite this, the Harmony Index remains capable of predicting the effectiveness of these untested teams by extrapolating from the network of relationships among their atomic components.

To address this challenge, we conducted an analysis to test the predictive power of the Harmony Index. Specifically, we systematically removed all games played by a particular team composition, recalculated the team's Harmony Index, and then predicted the win rate of the team that was removed from the data set. This allowed us to simulate the effectiveness of a novel team that has never been seen before. We then compared the predicted win rate of this team to their actual performance in our data set.

We repeated this process for every unique team with more than 1,000 games played, enabling us to comprehensively evaluate the predictive strength of the Harmony Index. We believe that this approach represents a significant advancement in the ability to predict team performance and has important implications for the design and management of effective teams.

The results presented in Fig. 2 provide strong evidence that the Harmony Index continues to be a robust predictor of win rate for untested teams. The best fit slopes of the original and excised data sets clearly demonstrate this, underscoring the reliability and generalizability of the Harmony Index. These findings have important implications for researchers and practitioners interested in team performance and highlight the importance of considering the Harmony Index as a key factor in predicting team success.

The results presented in this study provide compelling evidence that the Harmony Index can predict the win rates of unseen team compositions. Our analysis of the original and excised data sets reveals that the Harmony Index consistently outperforms other metrics in predicting team success. Specifically, as illustrated in Fig. 3, the best fit slopes for both data sets confirm the strong predictive power of the Harmony Index.

Moreover, to assess the strength of the correlation between the excised Harmony Index and actual win rates, we conducted a Pearson's correlation analysis. The results, reported in Table 3 & Fig. 3, further underscore the robustness of the Harmony Index, revealing a high correlation coefficient of .777 and a significance level of $< .001$. These findings suggest that the Harmony Index is a valuable tool for predicting team success and should be considered a reliable metric in future studies of team dynamics.

Our analysis reveals that both the Harmony Index and the

	Excised HI	Geo Mean
Pearson's Corr.	.777**	.563**
Sig (2-tailed)	$< .001$	$< .001$
n	33	33

Table 3: Pearson's Correlations

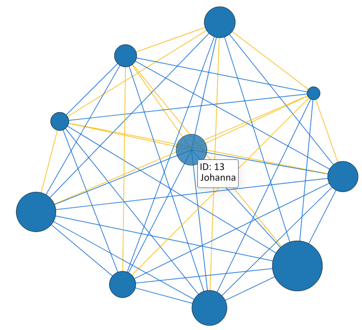

Figure 4: Force-Directed Weighted Graph, n=10

excised Harmony Index demonstrate a remarkable correlation to actual win rates. Notably, the strength of these correlations surpasses the Pearson's correlation coefficient of .563 for the geometric mean of individual win rates. This finding strongly supports the validity and predictive power of the Harmony Index as a metric for assessing team effectiveness.

Furthermore, analysis indicates that the Harmony Index calculated using our simulation method is significantly correlated with both the actual win rate and the Harmony Index in the excised ground truth data for untested teams.

These findings illustrate that the Harmony Index remains significantly correlated with the actual win rate, even when a team's past performance data has been removed from the dataset, indicating that the team was not observed performing together in our excised data, and providing valuable insights for team building and optimization.

Taken together, our findings highlight the significance of the Harmony Index as a reliable and robust metric for assessing team effectiveness, particularly in novel team compositions. These insights have important implications for team management and organizational strategy, emphasizing the importance of considering team dynamics and composition in optimizing team performance.

Visualization

The Harmony Index provides insights into these dyadic and team-wide relationships in raw form. While tables may efficiently present quantitative data, for human planners they often fall short in capturing the nuanced and complex nature of interconnections within those organizations.

In 2010, Xiang, Nevilla, and Rogati proposed modelling the strength of relationships as a linked graph (Xiang, Neville, and Rogati 2010). Lin, Shang, and Liu, in 2014, extended that proposal as they implemented a weighted network graph which indicated the weights of the relationships.

In our work we provide a visualization that generates a force-directed weighted network graph in order to construct a dynamic and visually compelling alternative to represent the dyadic and organizational relationships, facilitate an intuitive understanding of the intricate network structures inherent providing valuable insights into the dynamics and strengths of the connections within the team, and empower researchers and practitioners to make informed decisions that contribute to the overall effectiveness of the team[2].

The first aspect of the visualization focuses on the count of Harmonic relationships demonstrated by each agent in its neighborhood. Each agent is depicted as a node in the graph, with the size of the node determined by the number of its dyads resulting in a Harmonic pairing. Node size follows a linear scale.

The second aspect involves assessing the strength of the relationship between each set of dyads. A link is established between each dyad, connecting the agent to its dyadic counterpart. This link is assigned the value of its Harmony Index to select and apply an appropriate color to the link, which indicates the strength and valence of the relationship of this specific dyad.

The concluding element employs the Harmony Index as a weight to simulate the strength of the relationship, representing it as a force between the two nodes. This application is extended across the entire array of potential dyadic permutations, resulting in a network of nodes that accentuates connection density and aids in identifying central nodes and subgroups within the network.

While it is possible to illustrate the comprehensive set of dyads derived from the 90 agents within our experimental domain, we recommend that planners numerically refine their selections to generate a targeted pool of agents and enhance visibility of the most suitable team members.

In Fig. 4, the visualization presents a subset of 10 potential agents, allowing for a more detailed exploration of specific relationships within and across dyads. This view assists planners in identifying subsets of agents with the maximum number of Harmonic relationships, streamlining the selection process. Furthermore, a notable feature of employing the d3.js library is the capability to drag and drop nodes. This interactive functionality aides in the isolation of a set of 6 nodes characterized by positive relationships among the 15 possible dyads within the chosen subset. This process highlights 6 highly Harmonic teams n=5 teams from the initial pool of 252.

It is possible there is a stronger set of agents that include Depress class dyads. To assist with analysis of that scenario, In Fig. 5 we have also applied the Harmony Index values as a locally-relevant gradient of color rather than a simple binary valence. This example applied a 5-color gradient to expand upon the positive and negative valences in an IBM Color Blindness Palette.

Discussion & Future Work

The Harmony Index revealed several undiscovered areas where it exceeded a value of 1—areas where we believe

[2]JavaScript library, D3.js, provided by ObservableHQ

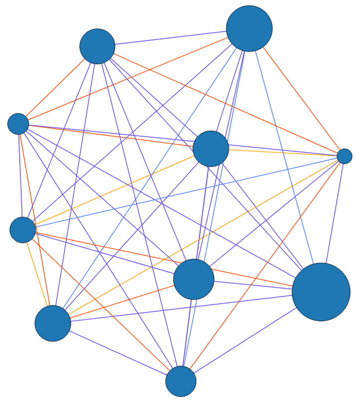

Figure 5: Force-Directed Weighted Graph in Color Gradient, n=10

designers and players each may be missing future analysis. Within the Harmonic Classes, we identified 17 agent pairings where the Harmony Index exceeded 1, indicating mutual benefit, yet their joint success rate fell below the minimum target of 50%. For instance, Agent 7 and Agent 68 formed a Harmonious pair with a Harmony Index of 1.05, but their joint success rate remained well below 50% (47.23%). This subset highlights how harmoniousness doesn't always ensure success, emphasizing the nuanced complexities of team composition.

Similarly, we found Discordant pairings with success rates surpassing the 50% target. For example, Agent 2 and Agent 83, with a Discordant Harmony Index of 0.91, achieved a joint success rate above 50% (50.38%). These findings underscore the need to consider the broader team context when evaluating pairings.

Exploring human intuition in interacting with such teams reveals conflicting priorities. While some research suggests a focus on team performance over individual success, other studies indicate instances where agents prioritize personal benefits or even hinder team effectiveness for personal gain (Velez 2015). This highlights the multifaceted dynamics of team interactions and the diverse ways individual agents can influence overall team effectiveness.

Abuse of the Depress Class

We believe the depress class is singularly at risk for being under-analyzed by human participants. One potential unintended consequence of using the Harmony Index (HI) to optimize team performance is the possibility of egoist behavior in certain agent pairings. These edge cases occur when one agent benefits significantly from the pairing at the cost of the success rate of the other agent falling below the Target Success Rate. This type of pairing may be susceptible to abuse by computer, human, or robotic agents that prioritize their own welfare while reducing the welfare of their teammate or team.

Given that not all parties in a team may be utilitarian in nature, it is crucial to consider the potential for egoist behavior in team optimization. Previous studies in the sports domain have demonstrated the occurrence and benefit of such behavior to the individual over the team (Uhlmann and Barnes 2014), while strategic games provide a framework for incorporating economic concepts into effectiveness calculations (Apt and Schafer 2014). To address this issue, we propose further research on the integration of philosophical concepts such as the Prisoner's Dilemma, Nash Equilibrium, Pareto Optimization, and the psychology of intention to enable agents to account for egoist behaviors in prospective teammates and teams. It is also important to explore whether such behavior can be predicted from partial data points to prevent abuse by egoist agents. By doing so, we create more effective and sustainable team structures to account for the reality of human behavior.

Future Opportunities

The Index currently does not fully account for the impact of skill specializations on neighbor overlap. Further research could enhance the Harmony Index algorithm by incorporating subject matter experts' breakdown of skills into many-to-one ratios to refine team compositions beyond traditional roles and better aligning with overlapping skill sets in diverse team structures. Diverging across specialized skills to generate multiple more specific Harmony Indexes could further elucidate enhance matchmaking by elucidating neighbor overlaps relevant to localized task types or maps.

Comparing general Harmony Index values across characters versus specific HI values for players within characters could provide a novel assessment of a player's skill proficiency in particular roles rather than specific skill sets. In Draft mode, this could offer valuable insights for adjustments to each player's MMR post-match.

Expanding Harmonic classes to n=5 and analyzing highly effective team compositions, such as the 4:1 "Carry," would illuminate the resonance of known team configurations, as observed in similar game environments (Eaton et al. 2017).

Exploring player perception of Harmonic classes in relation to success rates could unveil links between team performance and individual perceptions, offering insights into team dynamics and pathways to enhance perception of fairness. Additionally, investigation perceptions of balance between assisting teammates and personal benefit could shed light on toxic team dynamics.

Ultimately, the application of the Harmony Index extends beyond the experimental domain, with practical implications across sectors leveraging team dynamics, such as military, academia, and healthcare. Incremental enhancement in predicting team effectiveness would greatly benefit industries increasingly integrating artificial agents and human-agent teaming.

Conclusion

Effective team optimization demands a reliable and objective metric for measuring performance and effectiveness. Our study demonstrates that the Harmony Index (HI) serves as a crucial tool in this regard, surpassing previous measures in predicting team success and coordinating collaborative efforts. The HI is a novel tool to supplement player ratings and enhance accuracy of Matchmaking by allowing a lesser rated team to be formed with agents showing higher team-wide HIs, to feed back into the MMR impact assessment by factoring the Harmony Index of a given team against the opposing team's HI to re-score the players involved in the match, or by merging with results of clustered character build analysis to reveal possible imbalances (Renaudie et al. 2020). There is value in the difference between winning with a highly harmonic composition versus a deeply discordant one. The adaptability of the HI to new datasets further enhances its value for ongoing analysis and optimization.

In conclusion, while the HI offers a solid foundation for team optimization, ongoing research is vital to fully grasp the complexities of team dynamics. The HI is a single piece of the larger puzzle of team optimization. Further research is essential to unravel the specific factors contributing to team effectiveness and to understand potential unintended consequences. By embracing a holistic approach that considers the nuanced details of team composition and collaboration, we can leverage the HI and other metrics to drive continuous improvements in team building and matchmaking effectiveness.

References

Afshar, J.; Roudsari, H.; A., L.; C. C., E.; C. S. H., L. W.; and N., A. 2018. Harmonic mean based soccer team formation problem. In *In Proceedings of the 7th International Conference on Emerging Databases: Technologies, Applications, and Theory*, 240–246. Springer Singapore.

Apt, K. R.; and Schafer, G. 2014. Selfishness level of strategic games. *Journal of Artificial Intelligence Research*, 49: 207–240.

Berretta, S.; Tausch, A.; Ontrup, G.; Gilles, B.; Peifer, C.; and Kluge, A. 2023. Defining human-AI teaming the human-centered way: a scoping review and network analysis. *Frontiers in Artificial Intelligence*, 6.

Buckland, S. T.; Studeny, A. C.; Magurran, A. E.; Illian, J. B.; and Newson, S. E. 2011. The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere*, 2(9): 1–15.

Cheng, Z.; Yang, Y.; Tan, C.; Cheng, D.; Cheng, A.; and Zhuang, Y. 2019. What makes a good team? a large-scale study on the effect of team composition in honor of kings. *In The World Wide Web Conference*, 2666–2672.

Damacharla, P.; Javaid, A. Y.; Gallimore, J. J.; and Devabhaktuni, V. K. 2018. Common metrics to benchmark human-machine teams (HMT): A review. *IEEE Access*, 6: 38637–38655.

Dennis, A. R.; Lakhiwal, A.; and Sachdeva, A. 2023. AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With. *Journal of Management Information Systems*, 40(2): 307–337.

DeShon, R. P.; Kozlowski, S. W.; Schmidt, A. M.; Milner, K. R.; and Wiechmann, D. 2004. A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of applied psychology*, 89(6): p.1035.

Eaton, J. A.; Sangster, M. D. D.; Renaud, M.; Mendonca, D. J.; and Gray, W. D. 2017. Carrying the team: the importance of one player's survival for team success in league of legends. *In Proceedings of the human factors and ergonomics society annual meeting.*, 61(1): 272–276.

Escobar, M. T.; Aguarón, J.; and Moreno-Jiménez, J. M. 2004. A note on AHP group consistency for the row geometric mean priorization procedure. *European Journal of Operational Research*, 153(2): 318–322.

Estrada, J. 2010. Geometric mean maximization: an overlooked portfolio approach? *The Journal of Investing*, 19(4): 134–147.

Gorman, J. C.; Grimm, D. A.; and Dunbar, T. A. 2018. Defining and measuring team effectiveness in dynamic environments and implications for team ITS. In *Building Intelligent Tutoring Systems for Teams: What Matters*, 55–74. Emerald Publishing Limited.

Hall, K. L.; Vogel, A. L.; Huang, G. C.; Serrano, K. J.; Rice, E. L.; Tsakraklides, S. P.; and Fiore, S. M. 2018. The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American psychologist*, 73(4): p.532.

Hastings, E. M.; Jahanbakhsh, F.; Karahalios, K.; Marinov, D.; and Bailey, B. P. 2018. Structure or nurture? the effects of team-building activities and team composition on team outcomes. In *Proceedings of the ACM on Human-Computer Interaction 2*, 1–21. CSCW.

Hill, Y.; Den Hartigh, R.; Cox, R.; De Jonge, P.; and Van Yperen, N. W. 2020. Predicting resilience losses in dyadic team performance. *Nonlinear dynamics, psychology, and life sciences*, 24(3): 327–351.

Kim, J.; Keegan, B. C.; Park, S.; and Oh, A. 2016. May. In *The proficiency-congruency dilemma: Virtual team design and performance in multiplayer online games*, 4351–4365. In Proceedings of the 2016 CHI conference on human factors in computing systems.

Kim, Y. J.; Engel, D.; Woolley, A. W.; Lin, J. Y.-T.; McArthur, N.; and Malone, T. W. 2017. What makes a strong team? Using collective intelligence to predict team performance in League of Legends. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 2316–2329.

Kozlowski, S. W.; and Ilgen, D. R. 2006. Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3): 77–124.

Kozlowski, S. W.; and Klein, K. J. 2000. *A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes*. Jossey-Bass.

Renaudie, D.; Lizatovic, R.; Azadvar, A.; Elmqvist, R.; Hofmeister, K.; and Kristmannsson, B. 2020. Categorical Clustering Applied to the Discovery of Character Builds in TCTD2: the BaT Approach. In *2020 IEEE Conference on Games (CoG)*, 152–159.

Sapienza, A.; Zeng, Y.; Bessi, A.; Lerman, K.; and Ferrara, E. 2018. Individual performance in team-based online games. *Royal Society open science*, 5(6): p.180329.

Seeber, I.; Bittner, E.; Briggs, R. O.; De Vreede, T.; De Vreede, G. J.; Elkins, A.; Maier, R.; Merz, A. B.; Oeste-Reiß, S.; Randrup, N.; and Schwabe, G. 2020. Machines as teammates: A research agenda on AI in team collaboration. *Information & management*, 57(2): p.103174.

Uhlmann, E. L.; and Barnes, C. M. 2014. Selfish play increases during high-stakes NBA games and is rewarded with more lucrative contracts. *PloS one*, 9(4): p.e95745.

Velez, J. A. 2015. Extending the theory of Bounded Generalized Reciprocity: An explanation of the social benefits of cooperative video game play. *Computers in Human Behavior*, 48: 481–491.

Walliser, J. C.; de Visser, E. J.; Wiese, E.; and Shaw, T. H. 2019. Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, 13(4): 258–278.

Wang, Y.; Chin, K.; Poon, G.; and Yang, J. 2009. Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean. *Expert systems with applications*, 36(2): 1195–1207.

Xiang, R.; Neville, J.; and Rogati, M. 2010. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, 981–990.