

Amadeus: When Models of People Meet

Karl Hanna

Queen's University Belfast
Belfast, United Kingdom

Abstract

With the sheer constant advancements raining down in the field of Artificial Intelligence, one particular possibility that may cross our mind is whether it is possible to model agents after humans and, in turn, use these agents to carry out synthetic interactions that predict their real counterparts, or even interactions at a larger scale such as groups or societies. In this paper, we test a more controlled version of this question through chess. We use 8 elite chess players, seal their direct pairwise games, learn each player independently using different methods, and then compose the resulting models on the withheld dyads. To evaluate the generated interactions, we use two measurements: opening-family total variation distance and win-draw-loss (WDL) total variation distance. M1 reduces WDL-TV while leaving opening-family TV largely unchanged, whereas M2 substantially reduces opening-family TV while having little effect on WDL-TV. An additional post-hoc method combining components of the other two retains improvements across both measurements. These results suggest that independently learned models can recover aspects of previously unseen interactions, and that recovery across different behavioural aspects need not be mutually exclusive.

Keywords

behavioral modeling, multi-agent systems, imitation learning, personalization, agent interaction, chess

1 Introduction

With the rapid progress in machine learning and AI in recent years, it is now increasingly possible to model agents after humans. This has been explored in different applications, including modelling individual behaviour [8, 10] and identifying or reproducing a chess player's style [5, 6, 11].

One specific question remains less explored. With the premise that we have two agents modelled accurately after real humans, what happens if we have them interact together? Individual behaviour may depend on things such as the person being interacted with, relationships between individuals, and other interaction-specific effects, meaning that individual fidelity does not necessarily imply interaction fidelity. However, if the composition between individual models holds, then this could enable simulations involving interacting models of real individuals, small teams, and even larger social systems.

In this paper, we test this question. For a cohort of target players, we learn each player's behaviour using only games against players outside the target cohort. All interactions between target players are withheld. We then compose pairs of learned player models and compare their synthetic interactions with the corresponding sealed real interactions.

We use chess as the domain for several reasons. The interaction between two players is pairwise and explicit, their actions are

discrete and completely observable, and the environment and rules are fixed. Chess also provides repeated observations of each player and repeated interactions between the same pairs, while allowing all interactions between a target pair to be cleanly withheld to prevent leakage. Outcomes and behavioural properties, such as the openings played, can also be quantitatively measured. Finally, unlike something more open-ended such as conversation, there is little ambiguity about what the interaction trajectory actually was.

We use Amadeus to test this. Amadeus uses a ChessFormer-based model trained from scratch, which allows us to control the data seen by the base model, rather than using an already-trained model and risk leakage from games between the eight elite target players. We then use two different methods to model specific players while keeping all interactions between the target players sealed from training: Policy-Level Personalization (M1) and Candidate-Level Personalization (M2). These player models are composed in synthetic games and compared with the real interactions that were withheld. After evaluating M1 and M2, we also explore a hybrid (M3) that combines their already-trained mechanisms without additional training.

Our main contributions are as follows:

- We study whether independently learned models of specific individuals can reproduce their real interactions when those interactions are completely withheld during training.
- We introduce Amadeus, a controlled chess-based setup using eight elite players and strict pairwise sealing to test this question.
- We evaluate two different methods for modelling player-specific behaviour and compare their synthetic interactions against withheld real interactions.

2 Related Work

Chess has increasingly been used to model human rather than optimal decision-making. Maia trains policies on human games to predict moves at different skill levels [4], while Maia-2 unifies behavior across Elo levels within a single skill-conditioned model [12]. Subsequent work has moved from population-level behavior toward specific individuals. Per-player fine-tuning improves prediction of individual players [5], while behavioral stylometry shows that player identity can be recovered from decision traces alone [6]. More recent approaches improve the efficiency or representation of personalization: Maia4All uses prototype-based initialization to learn individual behavior from substantially fewer games [13], while champion-specific embeddings have been combined with distributional style evaluation based on Jensen-Shannon divergence [11]. Matilda instead combines a frozen human policy with independent search evidence and lightweight player-style embeddings [1]. Our base architecture, ChessFormer, provides the Maia-3 human-move models and reports 57.1% move-matching accuracy [7].

Individual-level behavioral modeling also extends beyond chess. Omi et al. formulate individual behavior generation as a multi-task learning problem, learning compact person-specific style vectors within a shared model at scale across chess and Rocket League [8]. In a substantially different setting, Park et al. construct LLM agents representing 1,052 real individuals from interview and survey self-reports and evaluate them on held-out attitudes, personality measures, economic-game behavior, and experimental responses [10]. Their combined interview-and-survey agents reach performance equivalent to 86% of participants' own two-week test-retest consistency on held-out survey questions. Together, these results illustrate that models can capture aspects of specific individuals across different domains and forms of behavioral evidence.

A parallel line of work studies interactions among generative agents. Generative Agents places multiple language-model agents in a shared environment and demonstrates emergent behaviors including information diffusion, relationship formation, and coordination [9]. SOTOPIA provides an environment for evaluating agents across social interactions such as cooperation, negotiation, and competition [15], while Concordia provides infrastructure for constructing grounded generative agent-based simulations [14]. These works study believable, socially capable, or emergent interaction among agents, rather than whether interactions between models of specific real individuals reproduce those individuals' real interactions. In contrast, Amadeus withholds all interactions among a target cohort during learning, composes the resulting individual models afterward, and evaluates the synthetic interactions against the corresponding withheld real interactions.

3 Methodology

3.1 Problem Formulation

For target players A and B , let π_A and π_B denote the learned behavioral policies for the two players, trained without access to direct A - B interactions. Composing these policies in the shared chess environment induces a synthetic interaction distribution,

$$\hat{P}_{AB} = C(\pi_A, \pi_B), \quad (1)$$

where C denotes the interaction process between the two policies. Let P_{AB}^{real} denote the empirical distribution of the sealed real interactions between A and B . Our goal is to test whether the synthetic interaction distribution $\hat{P}_{AB}$ resembles P_{AB}^{real} .

3.2 Base Chess Model

Both personalization methods use a base model based on ChessFormer [7].

The board is first canonicalized and encoded as a tensor of shape $m \times 64 \times 96$, where m is the batch size. Each square contains eight groups of 12 one-hot features, representing the current board and the previous seven board states.

The model also receives two 128-dimensional Elo embeddings, one for the mover and one for the opponent. Each embedding is obtained by interpolating between learned low- and high-Elo vectors. These embeddings are broadcast across the board and concatenated with each square's piece and history features, giving a 352-dimensional input per square.

Each square's input is projected to $d_{\text{model}} = 1024$ and passed through eight ChessFormer blocks. Each block contains GAB augmented multi-head self-attention followed by a feed-forward network.

To describe the GAB computation, let N denote the number of board squares, h the number of attention heads, and d_1 , d_2 , and d_3 the GAB projection dimensions. The block first projects each square representation from d_{model} to d_1 . The resulting $N \times d_1$ representation is flattened into a vector of dimension Nd_1 , then projected to d_2 , followed by GELU and LayerNorm. A further projection maps this vector to hd_3 , again followed by GELU and LayerNorm. The result is reshaped into $h \times d_3$, giving one latent vector per attention head.

The same linear map is applied to each head's vector, projecting it from d_3 to N^2 . Reshaping these outputs produces one $N \times N$ GAB bias matrix per head. Each bias matrix is added to the corresponding standard attention logits before softmax. The resulting attention weights are then multiplied by the value representations, as in standard multi-head self-attention.

We use $N = 64$, $d_{\text{model}} = 1024$, $h = 32$, $d_1 = 32$, and $d_2 = d_3 = 128$.

After the final block, the square representations are passed to policy and value heads. The policy head uses separate projections to represent each square as a possible source (FROM) and destination (TO). Scaled pairwise dot products between these representations produce $64 \times 64 = 4096$ source-destination logits. Promotions are handled separately because the same source and destination can correspond to different promotion pieces, adding another 256 logits. The full policy output therefore contains 4352 logits.

The value head mean-pools the square representations and passes the resulting vector through an MLP with dimensions $1024 \rightarrow 1024 \rightarrow 3$, producing three outcome logits.

We train the base model from scratch because the publicly released weights may have been trained on games involving our target players.

Let x_t denote the encoded board history immediately before ply t , and let $e(r) \in \mathbb{R}^{128}$ denote the base model's Elo embedding for rating r . We write the base policy logits as

$$\ell_t = f_{\theta}(x_t, e(r_t), e(r_t^{\text{opp}})) \in \mathbb{R}^{4352}, \quad (2)$$

where r_t and r_t^{opp} are the mover's and opponent's ratings at ply t , respectively.

3.3 Target Cohort and Interaction Sealing

Our target cohort contains eight players: Magnus Carlsen, Hikaru Nakamura, Ian Nepomniachtchi, Wesley So, Fabiano Caruana, Maxime Vachier-Lagrave, Levon Aronian, and Alireza Firouzja. These players were selected to provide dense mutual-game coverage across all 28 target-player pairs.

Base-model training consists of two stages. In the first stage, we train on public Lichess games from June of each year from 2017 to 2021. Using the same calendar month across years is intended to limit seasonal variation in the training data.

We exclude games involving any identified Lichess account associated with a target player, including accounts with low-confidence attribution, to reduce the risk of leakage. The full list of excluded usernames is provided in the appendix.

Games are processed in chunks of 20,000 and assigned to Elo bins using the mean of White's and Black's ratings. We use 22 bins: one below 600, twenty 100-point bins covering [600, 2600), and one for ratings of 2600 and above. Within each chunk, we retain at most 10 games per bin. The counts reset for the next chunk, and processing of a chunk can stop early once all bins are full. This sampling is intended to balance representation across playing strengths rather than concentrate training on the middle of the Elo distribution.

In the second stage, we fine-tune the base model for an additional 200,000 steps on the Lichess Broadcast corpus [2], all games involving any of the eight target players.

For both personalization methods, each target player's eligible data consists only of games against players outside the target cohort. These games are split at the game level for training and individual behavioral validation. All games between target players remain sealed and are excluded from training, tuning, and calibration.

3.4 Individual Behavioral Modeling

We use two personalization methods. Policy-Level Personalization changes the player conditioning supplied to the base model, while Candidate-Level Personalization adds player-specific adjustments to candidate move logits.

3.4.1 Policy-Level Personalization (M1). M1 adapts the player personalization approach of Sogliuzzo et al [11], which learns a player-specific embedding while keeping the Maia-2 backbone frozen, to ChessFormer's continuous Elo conditioning.

For each target player p , M1 introduces a trainable player vector $z_p \in \mathbb{R}^{128}$, initialized from the base model's Elo embedding at that player's representative Elo,

$$z_p^{(0)} = e(r_p). \quad (3)$$

The base-model parameters θ remain frozen. When player p is the mover at ply t , the ordinary mover-Elo embedding is replaced by z_p , giving personalized policy logits

$$\ell_t^{(p)} = f_\theta(x_t, z_p, e(r_t^{\text{opp}})). \quad (4)$$

The opponent's Elo conditioning remains unchanged, and z_p is the only parameter optimized for that player.

During generation, illegal moves are masked before softmax, giving a probability distribution over legal moves. Because the base model was initially trained on a broad range of playing strengths and is used here to model elite grandmasters, we apply a Stockfish based sampling adjustment.

Production sampling. For generation, illegal moves are first masked and the calling behavioral policy is restricted to its top- K legal candidates, $C_t = \{a_{t,1}, \dots, a_{t,K}\}$, with $K = 5$. Let $p_{t,i}$ denote the behavioral-policy probability of candidate centipawn loss relative the best candidate in C_t , estimated by Stockfish 19 at depth 8. The final sampling distribution is

$$q_{t,i} = \text{softmax}_i \left(\log p_{t,i} - \lambda \frac{c_{t,i}}{100} \right), \quad \lambda = 2. \quad (5)$$

The coefficient λ is frozen using nonsealed data and is chosen to limit intervention in the behavioral distribution while discouraging poor moves. The same sampling rule is used for the generic model, M1, M2, and the exploratory M3 hybrid.

3.4.2 Candidate-Level Personalization (M2). M2 begins from the top- K legal candidates of the frozen base policy, with $K = 5$. Let $C_t = \{a_{t,1}, \dots, a_{t,K}\}$ enote this candidate set at ply t , and let $\ell_{t,i}$ denote the frozen base-policy logit of candidate $a_{t,i}$.

A shared move-style network g_ψ encodes the current board and candidate move as a d -dimensional feature vector,

$$\phi_{t,i} = g_\psi(x_t, a_{t,i}) \in \mathbb{R}^d, \quad d = 32. \quad (6)$$

Each target player p has a learned style vector $u_p \in \mathbb{R}^d$. The candidate and player representations are L2-normalized,

$$\hat{\phi}_{t,i} = \frac{\phi_{t,i}}{\|\phi_{t,i}\|_2}, \quad \hat{u}_p = \frac{u_p}{\|u_p\|_2}. \quad (7)$$

M2 adds a player-specific residual to the frozen base-policy logit,

$$\tilde{\ell}_{t,i}^{(p)} = \ell_{t,i} + \sigma \sqrt{d} \hat{\phi}_{t,i}^\top \hat{u}_p, \quad (8)$$

where σ is a learned scalar shared across players and candidates. The shared move-style parameters ψ , player vectors u_p , and σ are trained jointly across all eight target players, while the base-model parameters θ remain frozen.

During training, if the observed human move is absent from the raw top- K candidate set, it is appended so that the objective is defined. At inference, the candidate set remains the raw top- K . Softmax over the adjusted logits gives the M2 behavioral probabilities, which are passed through the same production sampling rule defined for M1.

3.5 Individual Behavioral Validation

Before evaluating composed interactions, we test whether each of the two personalization methods learns player-specific behavioral information. This validation uses only held-out target-vs-outsider games, with game-level train-validation splits. No sealed target-target interactions are used.

For each target player, we evaluate their held-out moves using the correct learned player representation, the generic model, and the representations of the other seven target players. We compare the negative log likelihood assigned to the observed human move. A lower negative log likelihood means that the model assigns greater probability to the move the player actually made.

We also rank all eight learned player representations by their likelihood on each player's held-out data. This tests whether the correct representation is more consistent with that player's behavior than the representations learned for other players.

These likelihoods are measured directly from the behavioral models, before and independently of the Stockfish sampling adjustment. For M1, validation is performed over the full legal action space. For M2, the generic, correct-player, and wrong-player comparisons use the same candidate set: the raw top-five legal moves from the frozen base policy, with the observed human move appended if it is absent.

We separately measure raw top-five coverage before appending the observed move. This quantifies how often the human move is already available to the candidate-level personalization mechanism.

3.6 Exploratory Hybrid Composition (M3)

After inspecting the M1/M2 interaction results, we introduce M3 as a post-hoc exploratory hybrid of the two personalization mechanisms. M3 uses the already-trained M1 and M2 components without additional training or optimization.

The M1-personalized policy produces the top- K legal candidate moves. The already-trained M2 candidate-level mechanism then reranks these candidates by adding its player-specific residual to their M1-personalized policy logits. The resulting candidates are passed to the same production sampler used by M1 and M2. All learned components remain frozen, and no parameters or sampling settings are tuned on sealed interactions.

3.7 Compositional Interaction Protocol

The primary sealed-dyad evaluation compares M1 and M2. We also apply the same protocol to M3 as a post-hoc exploratory evaluation.

The eight target players form 28 unordered dyads. For a target player p , let π_p^G denote the generic generation policy and let π_p^M denote the personalized generation policy under method M , where

$$M \in \{M1, M2, M3\}.$$

For each dyad (A, B) , the four experimental conditions produce the following synthetic interaction distributions:

$$\hat{P}_{AB}^{GG,M} = C(\pi_A^G, \pi_B^G), \quad (9)$$

$$\hat{P}_{AB}^{AG,M} = C(\pi_A^M, \pi_B^G), \quad (10)$$

$$\hat{P}_{AB}^{GB,M} = C(\pi_A^G, \pi_B^M), \quad (11)$$

$$\hat{P}_{AB}^{AB,M} = C(\pi_A^M, \pi_B^M). \quad (12)$$

Thus, **GG** uses generic models for both players, **AG** personalizes only A , **GB** personalizes only B , and **AB** personalizes both players. The labels identify the modeled players rather than their colors. Each condition is compared with the same sealed empirical distribution P_{AB}^{real} .

For each dyad and condition, we generate 10,000 games: 5,000 with the A side playing White and the B side playing Black, and 5,000 with the colors reversed. Across 28 dyads and four conditions, this gives 1.12 million games per method.

Games use normal rule-based termination, including claimable draws by threefold repetition and the fifty-move rule, implemented through `board.outcome(claim_draw=True)`. We do not model voluntary human actions such as resignation or agreed draws. An additional emergency limit is set at 500 plies. Games reaching this limit without a rule-based outcome are censored rather than counted as draws.

3.8 Evaluation Metrics

We use two main metrics to compare synthetic interactions with the sealed real games: total variation distance between win-draw-loss (WDL) distributions and total variation distance between opening-family distributions. For both metrics, we use total variation distance (TV) to be able to compare across different dyads and generation methods.

Table 1: Individual behavioural validation on held-out non-sealed data. ΔNLL denotes generic-minus-correct NLL. M2 uses candidate-matched NLL and is therefore not directly comparable in scale with M1.

Method	Rank 1	Mean ΔNLL	Wrong-correct	Positions
M1	8/8	0.006802	0.004365	47,136
M2	8/8	0.009278	0.020449	46,950

$$\text{TV}(P, Q) = \frac{1}{2} \sum_x |P(x) - Q(x)|. \quad (13)$$

For WDL evaluation, outcomes are defined by player identity: an A win, a draw, or a B win. The two color orientations are evaluated separately, and their WDL total variation distances are combined with equal 50/50 weight. Censored 500-ply games do not enter the WDL denominator and are reported separately rather than being treated as draws.

For opening-family evaluation, games are classified using a pinned revision of the `lichess-org/chess-openings` database [3]. Each game is replayed, and its deepest recognized opening position is used to assign an opening name. The full name is then collapsed to its top-level family, generally using the part before the first colon. For example, a named Sicilian variation is assigned to the *Sicilian Defense* family. Games with no recognized opening are assigned an explicit Unknown family rather than discarded.

We compare opening-family frequencies in the generated games with those in the sealed real games using total variation distance. For both metrics, lower distance indicates closer agreement with the sealed real distribution for the property being measured.

4 Results

4.1 Individual Behavioural Validation

Before evaluating sealed interactions, we first perform a sanity check that the learned representations capture player-specific behaviour. For each of the eight target players, we evaluate held-out nonsealed positions using the correct player representation, the generic representation, and the representations of the seven other target players.

For M1, we evaluate the legal-masked full-action policy negative log-likelihood (NLL). For M2, whose personalized scorer operates over a candidate set, we instead use candidate-matched NLL over the raw top-5 candidates, appending the observed move when it is absent. The NLL values of M1 and M2 therefore have different definitions and should not be directly compared across methods.

For both M1 and M2, the correct player representation achieves the lowest NLL of all eight player representations for every target player. The correct representation ranks first for 8/8 players under both methods. Relative to the generic representation, the mean NLL decreases by 0.006802 for M1 and 0.009278 for M2. Relative to the mean of the incorrect player representations, the corresponding gaps are 0.004365 and 0.020449, respectively. M2's raw top-5 candidate set contains the observed move for 94.98% of positions on average.

4.2 Sealed Pairwise Interaction Fidelity

We next evaluate the central question of the paper: whether the independently learned player models reproduce real interactions that were completely sealed from personalization. We compare the four conditions GG, AG, GB, and AB over all 28 target-player dyads. Results are averaged equally across dyads and across both color orientations. For both WDL-TV and opening-family TV, lower values indicate that the synthetic interaction distribution is closer to the withheld real interaction distribution.

The two personalization mechanisms produce notably different patterns. For M1, personalization progressively reduces WDL-TV from 0.164171 under GG to 0.150148 under AB, a reduction of 0.014024. Personalizing only one side also reduces the distance, with AG and GB reaching 0.157753 and 0.156141, respectively. M1 also improves opening-family fidelity, although by a smaller amount: opening-family TV decreases from 0.554433 under GG to 0.536973 under AB.

M2 shows almost the opposite pattern across the two behavioural measures. WDL-TV remains effectively unchanged, moving from 0.163166 under GG to 0.163408 under AB. In contrast, opening-family TV decreases substantially. Personalizing a single player reduces it to 0.501370 under AG and 0.488343 under GB, while composing both personalized players reduces it further to 0.437202. This corresponds to an $AB - GG$ change of -0.117509.

These results suggest that the two personalization mechanisms transfer different aspects of player-specific behaviour into unseen interactions. M1 more clearly changes the distribution of game outcomes, whereas M2 produces a much larger change in opening-family behaviour.

4.3 Exploratory Hybrid Model (M3)

The complementary behaviour of M1 and M2 motivates an exploratory hybrid, M3, which combines the already-trained M1 player vectors with the already-trained M2 candidate reranker. M3 introduces no additional training or new scientific hyperparameters, and we therefore treat it as a post-hoc exploratory analysis rather than a third primary method.

M3 reduces WDL-TV from its own GG value of 0.163073 to 0.152036 under AB, for a change of -0.011037. Its AB WDL-TV is therefore close to the M1 value of 0.150148 and substantially lower than the M2 value of 0.163408.

For opening behaviour, M3 reduces opening-family TV from 0.554395 under GG to 0.444428 under AB, for a change of -0.109967. This is close to the M2 AB value of 0.437202 and substantially below the M1 value of 0.536973. Thus, descriptively, the hybrid retains much of the M1-like improvement in outcome fidelity while also retaining much of the M2-like improvement in opening-family fidelity.

4.4 Dyad-Level Variation

The aggregate results can obscure whether the observed changes are consistent across player pairs. We therefore examine $AB - GG$ separately for each of the 28 dyads. Negative values indicate that personalization brings the synthetic interaction closer to the real interaction.

For M1, the aggregate improvements are broadly reflected at the dyad level. WDL-TV decreases for 22 of 28 dyads (78.6%), with a median $AB - GG$ change of -0.010900. Opening-family TV also decreases for 22 of 28 dyads, with a median change of -0.018883. The effect is nevertheless heterogeneous: the WDL-TV changes range from -0.088700 to 0.046300, while opening-family changes range from -0.058385 to 0.027417.

The dyad-level results for M2 further separate its two behaviours. WDL-TV improves for only 13 of 28 dyads (46.4%), with a median change of 0.001650, consistent with the near-zero aggregate effect. In contrast, opening-family TV improves for 27 of 28 dyads (96.4%), with a median change of -0.105987. The opening effect is therefore not explained solely by a small number of player pairs, although its magnitude varies considerably across dyads.

M3 shows an intermediate WDL pattern, improving 19 of 28 dyads (67.9%) with a median change of -0.010500. Its opening-family result is again highly consistent across pairs, improving 27 of 28 dyads with a median change of -0.090846.

5 Discussion

Looking closely at the results from the previous section, we see that the different methods recover some aspects of interaction fidelity. M1 achieves better WDL fidelity than the compositions involving the general model, while M2 results in a substantial improvement in opening-family fidelity. M3 then retains some of both. Together, these results suggest that independently learned models can recover aspects of previously unseen interactions, but that this recovery is not uniform across methods or behavioural dimensions.

The difference in behaviour between the methods also suggests that interaction fidelity is not a single all-or-nothing property. One personalization mechanism preserves outcome tendencies better, another preserves opening behaviour much better, while M3 retains some of both. This suggests that different modelling methods can recover different aspects of an interaction, while the M3 result indicates that these gains need not be mutually exclusive.

We can therefore observe partial recovery of interaction fidelity, but we cannot determine how much of the remaining gap comes from imperfect individual modelling versus properties that are genuinely specific to the interaction itself. Using a much stronger individual modelling method would make this distinction clearer. If stronger individual models lead to substantially better interaction fidelity, this would suggest that much of the current gap comes from limitations in the individual models. If a substantial gap remains, however, this would suggest that some parts of an interaction may be irreducibly specific to the pair itself.

The scope of the current study is also limited to eight elite chess players, chess as a single interaction domain, WDL and opening-family distributions as the main behavioural measurements, and an exploratory post-hoc hybrid in M3. Natural future directions therefore include stronger individual modelling methods, a richer set of behavioural metrics that capture more aspects of an interaction, and evaluation in other domains with repeated pairwise interactions. These extensions would help clarify both how much of an interaction can be recovered from independently learned individuals and where the limits of that recovery lie.

Table 2: Sealed target–target interaction fidelity for M1 and M2. Lower TV is better. The final column reports $AB - GG$, so negative values indicate that composing both personalized players reduces the distance from the corresponding real interaction.

Method	Metric	GG	AG	GB	AB	$AB - GG$
M1	WDL-TV	0.164171	0.157753	0.156141	0.150148	-0.014024
M1	Opening-family TV	0.554433	0.547406	0.544765	0.536973	-0.017461
M2	WDL-TV	0.163166	0.162582	0.162589	0.163408	0.000241
M2	Opening-family TV	0.554711	0.501370	0.488343	0.437202	-0.117509

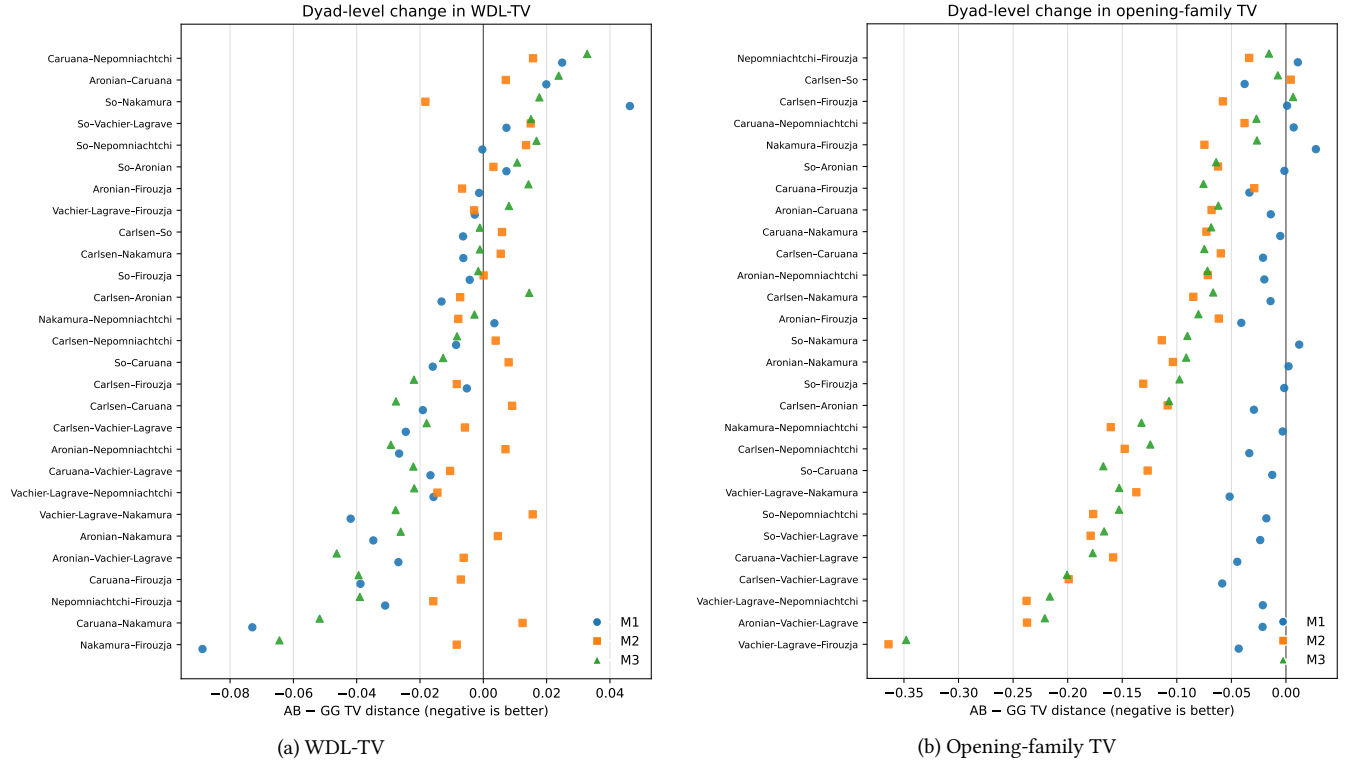


Figure 1: Dyad-level change from GG to AB for each method. Negative values indicate that the fully personalized interaction is closer to the withheld real interaction. M1 shows a broader reduction in WDL-TV, whereas the opening-family reductions for M2 and M3 are substantially larger and occur for nearly all dyads.

Table 3: Exploratory M3 results. The M1 and M2 columns show the corresponding fully personalized AB conditions for reference. Lower TV is better.

Metric	M1 AB	M2 AB	M3 AB
WDL-TV	0.150148	0.163408	0.152036
Opening-family TV	0.536973	0.437202	0.444428

6 Conclusion

This paper studies whether independently trained models of individuals can recover aspects of previously unseen pairwise interactions. We observe partial recovery of interaction fidelity across different

personalization methods: M1 improves WDL fidelity over compositions involving the general model, M2 substantially improves opening-family fidelity, and M3 retains some of both.

These results suggest that interaction recovery is possible in part, but depends strongly on both the individual modelling method and the aspect of the interaction being measured. This raises a broader question: how much of an interaction can ultimately be recovered from strong independently modelled individuals, and how much reflects genuinely interaction-specific structure that cannot be recovered from the participants alone?

7 Limitations

The present study is limited in several ways. First, the experiments consider only eight elite chess players within a single interaction domain. It is therefore unclear whether the same patterns would

Table 4: Variation in AB – GG across the 28 target-player dyads. “Improved” reports the number of dyads for which AB has lower TV than GG.

Method	Metric	Improved	Mean	Median	Min	Max
M1	WDL-TV	22/28	-0.014024	-0.010900	-0.088700	0.046300
M1	Opening-TV	22/28	-0.017461	-0.018883	-0.058385	0.027417
M2	WDL-TV	13/28	0.000241	0.001650	-0.018300	0.015700
M2	Opening-TV	27/28	-0.117509	-0.105987	-0.364176	0.004373
M3	WDL-TV	19/28	-0.011037	-0.010500	-0.064400	0.032800
M3	Opening-TV	27/28	-0.109967	-0.090846	-0.347976	0.006572

hold for weaker players, more volatile populations, or interactions outside chess. The evaluation is also restricted to WDL and opening-family distributions, which capture only selected aspects of an interaction rather than the interaction in its entirety. Other behavioural dimensions may exhibit different levels of recoverability.

Second, the individual models themselves do not perfectly reproduce the behaviour of the target players. As a result, the remaining interaction gap cannot be cleanly attributed to limitations in individual modelling versus behaviour that is genuinely specific to the interaction between a particular pair. Stronger individual models would be required to better separate these two sources of error. One additional limitation is that the models currently contain things such as board-defined termination, but not behavioral actions such as resignation or agreed draws. Such decisions may carry player or dyad signal and may affect both game length, and observed outcomes.

M3 should also be interpreted cautiously, as it is an exploratory post-hoc hybrid constructed from already trained M1 and M2 components rather than a separately trained personalization method. In addition, the study considers only pairwise interactions between two independently learned models and does not test whether similar behaviour would hold in interactions involving larger groups of models. Finally, the reported interaction distributions are estimated from a finite number of generated games and therefore remain subject to sampling variation.

References

- [1] Jason Carlson. 2026. Matilda: Engine-Agnostic Search with Human Policy Guidance. arXiv:2606.25176 [cs.AI] <https://arxiv.org/abs/2606.25176>
- [2] Lichess. [n.d.]. Lichess Broadcast Database. <https://database.lichess.org/broadcast/>. Monthly compressed PGN archives of Lichess-hosted tournament and broadcast games. CC0 licensed via the Lichess Open Database. Accessed 2026-09-24..
- [3] Lichess. 2026. chess-openings: An Aggregated Data Set of Chess Opening Names. <https://github.com/lichess-org/chess-openings/commit/4b8622759e7ae6f93f011cc6c83a3823401ab45e>. Pinned to commit 4b8622759e7ae6f93f011cc6c83a3823401ab45e (2026-08-04), CC0-1.0 licensed..
- [4] Reid McIlroy-Young, Siddhartha Sen, Jon M. Kleinberg, and Ashton Anderson. 2020. Aligning Superhuman AI with Human Behavior: Chess as a Model System. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1677–1687. doi:10.1145/3394486.3403219
- [5] Reid McIlroy-Young, Russell Wang, Siddhartha Sen, Jon M. Kleinberg, and Ashton Anderson. 2022. Learning Models of Individual Behavior in Chess. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 1253–1263. doi:10.1145/3534678.3539367
- [6] Reid McIlroy-Young, Yu Wang, Siddhartha Sen, Jon Kleinberg, and Ashton Anderson. 2021. Detecting Individual Decision-Making Style: Exploring Behavioral Stylometry in Chess. *Advances in Neural Information Processing Systems* 34 (2021), 24482–24497. https://proceedings.neurips.cc/paper_files/paper/2021/hash/cdf8111910291ba472b385e9c5f59099-Abstract.html
- [7] Daniel Monroe, George Eilender, Philip Chalmers, Zhenwei Tang, and Ashton Anderson. 2026. Chessformer: A Unified Architecture for Chess Modeling. In *The Fourteenth International Conference on Learning Representations*. <https://openreview.net/forum?id=2ltBrzEHYd>
- [8] Nabil Omi, Lucas Caccia, Anurag Sarkar, Jordan T. Ash, and Siddhartha Sen. 2025. Generative Modeling of Individual Behavior at Scale. arXiv:2502.14998 [cs.LG] <https://arxiv.org/abs/2502.14998>
- [9] Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, Article 2, 22 pages. doi:10.1145/3586183.3606763
- [10] Joon Sung Park, Carolyn Q. Zou, Jonne Kamphorst, Niles Egan, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Percy Liang, Robb Willer, and Michael S. Bernstein. 2026. LLM Agents Grounded in Self-Reports Enable General-Purpose Simulation of Individuals. arXiv:2411.10109 [cs.AI] <https://arxiv.org/abs/2411.10109>
- [11] Loris Sogliuzzo, Alois Rautureau, and Éric Piette. 2026. Toward Modeling Player-Specific Chess Behaviors. arXiv:2605.11893 [cs.AI] <https://arxiv.org/abs/2605.11893>
- [12] Zhenwei Tang, Difan Jiao, Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, and Ashton Anderson. 2024. Maia-2: A Unified Model for Human-AI Alignment in Chess. *Advances in Neural Information Processing Systems* 37 (2024), 20919–20944. doi:10.52202/079017-0659
- [13] Zhenwei Tang, Difan Jiao, Eric Xue, Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, and Ashton Anderson. 2026. Learning to Imitate with Less: Efficient Individual Behavior Modeling in Chess. *Transactions on Machine Learning Research* (2026). <https://openreview.net/forum?id=iw4kjcw319>
- [14] Alexander Sasha Vezhnevets, John P. Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A. Duéñez-Guzmán, William A. Cunningham, Simon Osindero, Danny Karmon, and Joel Z. Leibo. 2023. Generative Agent-Based Modeling with Actions Grounded in Physical, Social, or Digital Space Using Concordia. arXiv:2312.03664 <https://arxiv.org/abs/2312.03664>
- [15] Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and Maarten Sap. 2024. SOTOPIA: Interactive Evaluation for Social Intelligence in Language Agents. In *International Conference on Learning Representations*. https://proceedings.iclr.cc/paper_files/paper/2024/hash/b3075b88e583a0e98d8b24338a613060-Abstract-Conference.html