People use fast, flat goal-directed simulation to reason about novel problems

Katherine M. Collins[*1,2], Cedegao E. Zhang[*1], Lionel Wong[*1,3], Mauricio Barba da Costa[*1], Graham Todd[*4], Adrian Weller[2,5], Samuel J. Cheyette[1], Thomas L. Griffiths[6], and Joshua B. Tenenbaum[1]

[1]Massachusetts Institute of Technology
[2]University of Cambridge
[3]Stanford University
[4]New York University
[5]The Alan Turing Institute
[6]Princeton University

Abstract

Games have long been a microcosm for studying planning and reasoning in both natural and artificial intelligence (AI), especially with a focus on expert-level or even super-human play (12, 8, 21, 52, 54, 62). But real life also pushes human intelligence along a different frontier, requiring people to flexibly navigate decision-making problems that they have never thought about before. Here, we use *novice* gameplay to study how people make decisions and form judgments in new problem settings. We show that people are systematic and adaptively rational in how they play a game for the first time, or evaluate a game (e.g., how fair or how fun it is likely to be) before they have played it even once. We explain these capacities via a computational cognitive model that we call the "Intuitive Gamer". The model is based on mechanisms of *fast and flat (depth-limited) goal-directed probabilistic simulation*—analogous to those used in Monte Carlo tree-search models of expert game-play, but scaled down to use very few stochastic samples, simple goal heuristics for evaluating actions, and no deep search. In a series of large-scale behavioral studies with over 1000 participants and 121 two-player strategic board games (almost all novel to our participants), our model quantitatively captures human judgments and decisions varying the amount and kind of experience people have with a game—from no experience at all ("just thinking"), to a single round of play, to indirect experience watching another person and predicting how they should play—and does so significantly better than much more compute-intensive expert-level models. More broadly, our work offers new insights into how people rapidly evaluate, act, and make suggestions when encountering novel problems, and could inform the design of more flexible and human-like AI systems that can determine not just how to solve new tasks, but whether a task is worth thinking about at all.[†]

1 Introduction

Games have been used to study human problem solving and planning, and to model these abilities in artificial intelligence systems, since the early days of both psychology and computer science (14, 50, 42, 12, 8, 21, 38, 52, 66, 63, 62). They occupy this space for good reason: games hold a mirror to the structures, patterns and challenges that reality confronts us with, across a vast range of both familiar and merely possible experiences. Mancala lets us sow and capture

[†] * indicates equal contributions. Correspondence: {katiemc, cedzhang, zyzzyva, barba, cheyette, jbt}@mit.edu, {gdrtodd}@nyu.edu, {aw665}@cam.ac.uk, {tomg}@princeton.edu.

abstract resources. Chess lets us strategize over bloodless battlefields. Go lets us encircle and claim territories with silent stone troops. A good game lets us thrill to uncertainties and losses that we might shy from in real life. But not all games are created equal: most players outgrow childhood favorites like Tic-Tac-Toe, while games like Chess and Go have the depth to support a lifetime of play and personal improvement.

For these and other games, many people choose to devote the time and mental energy to become experts. Substantial work in cognitive science focuses on the nature and development of expertise, using games to understand how human experts recall strategies, perceive the world, make decisions efficiently and plan multiple steps ahead (55, 21, 62, 30). Human experts, correspondingly, have long served as goalposts in the quest to build intelligent machines, with a particular focus on emulating and ultimately beating humans at their own games, using more game-playing data and raw computational horsepower than any human expert could acquire in their lifetimes (8, 52, 54, 18).

But people solve problems or make plans not only in expert settings. They think productively about a wide range of new problems and systems *for the first time*—before any experience. What do people think about when they open a board game for the first time and start to play, or merely read the description on the back of the box and decide if they want to open it and try it out? Thinking about whether a system of rules and reward makes sense, whether to participate, or what actions to take first—across a wide range of novel situations—is arguably at least as important for everyday human cognition (if far less studied) than the cognitive processes that lead a very small number of humans to become expert players in any one game. Here we study the abilities of *novice* game reasoners across a range of novel games that are not entirely unlike games they have likely seen before, e.g., Tic-Tac-Toe or Connect-4, but vary in their rules and underlying dynamics (see examples in Figure 1a and Extended Data Table 1). We assess people's novel game thinking in multiple reasoning settings (Figure 1b-d): game evaluation (determining how rewarding a new game is likely to be, or how fair, challenging, or fun it might be, before having ever played); action selection (choosing their first moves the first time playing with another first-time player); and action prediction (judging the likelihood of moves that other first-time players might make, without having ever played directly themselves).

Our core contribution is a computational account of how people reason in these novel problem settings. Expert models of gameplay typically involve deep tree search with potentially thousands of evaluations of possible problem states (Figure 2a). It seems unlikely that an everyday novice reasoner is engaging in such intensive search and evaluation before they have extensive—or any—experience with the problem. But neither are people likely to be completely random or unsystematic. Here, we hypothesize that everyday novice thinking occupies a valuable intermediate point between these extremes (Figure 2b): people are highly non-random in how they assess new problems and run computations analogous to those posited in sophisticated game reasoners—stochastic game-play simulations, like those that underlie both state-of-the-art AI game systems and cognitive models of expert human game play—but scaled down to a level that is more realistic for everyday thought.

We instantiate and test this hypothesis in an "Intuitive Gamer" model that makes all of its judgments and decisions by running mental simulations of game play that are (1) fast, (2) flat, (3) goal-directed, and (4) probabilistic. We present a novel collection of over a hundred diverse grid-based two-player strategy games and a series of large-scale behavioral studies each with hundreds of participants. Our setting is different from much prior work on games, which usually focuses on one or a small set of games and many rounds of play in that game; here, we consider many games and little (even no) experience. We show that the Intuitive Gamer model accounts very well—and significantly better than alternative models—for how people evaluate the properties of games based purely on their descriptions. Features computed from the same Intuitive Gamer model can capture people's judgments about the objective value of the game, as well as more subjective evaluations like a game's expected funness. We further show that

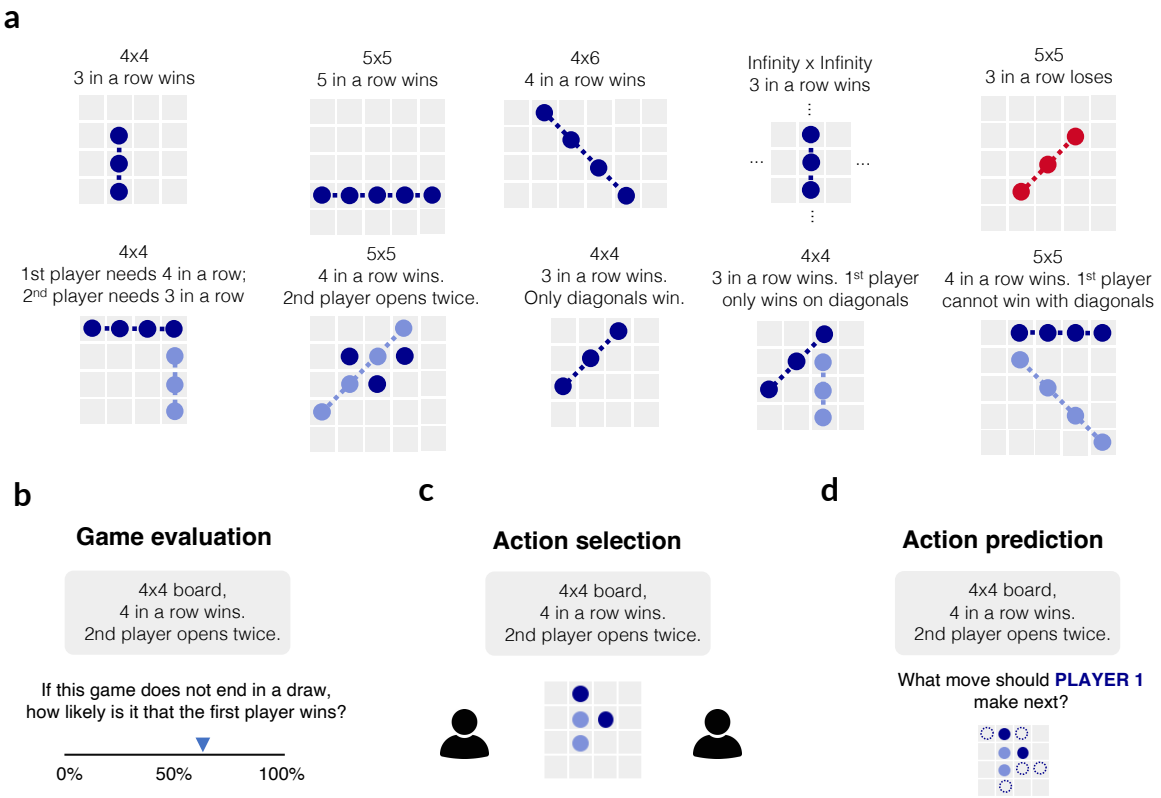


Figure 1: Our novel game dataset and suite of game tasks. **a,** Ten example games from our 121 game dataset. Games vary in board sizes and rules, such as what it takes to win and how many pieces any player can make on their opening move. **b-d,** We assess three people's reasoning about novel game through three behavioral studies designed to test how people **b,** reason about games before they even play a single game; **c,** how people decide what actions to make in their first instance of play; and **d,** how people predict others should play when watching them play.

action choices simulated by this same model capture how novices actually play in games with which they have no experience and how they make predictions about unfolding games that they have never played themselves.

The Intuitive Gamer model

The capacity to evaluate properties of a new game and decide whether the game is worth engaging with might intuitively feel like it comes prior to developing a policy for actually playing the game. However, our intuition as modelers is the opposite: game properties such as expected values depend on both the rule specifications of a game and the policies adopted by the players. We hypothesize this could be true for novice human game reasoners—before taking a single action in the game—if their policies are extremely simple to construct, fast to execute, and support probabilistic sampling of reasonable action choices. If so, a policy can be queried just a small number of times to simulate gameplay, from which the game reasoner can draw reasonable inferences about the game properties over these simulated traces. We instantiate this hypothesis with our Intuitive Gamer model. The model consists of two modules: a"flat" game-playing agent, which runs an extremely depth-limited search to selects actions probabilistically based on abstract yet locally evaluable "goal-directed" heuristic functions (Figure 2c); and a "fast" game reasoner that makes probabilistic inferences about game propositions and their

expected values using a small number of simulated games run with the agent model simulating both players (Figure 2d).

Formally, one can think of a system, problem, or game G in terms of a space of feasible states $\mathcal{S}$; possible actions $\mathcal{A}$; rules $\mathcal{T}$ specifying valid actions and state transitions, and governing the overall dynamics; and goal functions mapping from states $\mathcal{S}$ to possible rewards $\mathcal{R}$. We are interested in how a reasoner infers properties of a new game $\psi(G)$ (e.g. whether a game is likely to end in a draw or have a bias towards a particular player; whether the game is likely to be engaging and fun) as well as a policy $\pi_G(a_t \mid s_t)$ for how to play (choosing actions given the current state at time t, to achieve their goal, e.g., to win), without experience of actual traces of gameplay and relying instead on *simulated* or imagined traces. Our aim is to model how people *approximate* $\psi(G)$ and π_G in a way that can (1) take in any G as input, and (2) do so with a limited compute budget and no direct experience with game G. We next describe how each of the two modules of the Intuitive Gamer model implement on such algorithm which can compute a range of properties and policies for the subspace of novel two-player competitive games that we study empirically.

Player module. A long tradition of modeling human behavior in games has used some combination of game-state evaluation functions and search over the decision tree of game states (41, 11), as depicted in Figure 2b, to construct a policy $\pi_G(a_t \mid s_t)$. Our approach broadly falls into this tradition. However, traditional accounts typically assume highly tuned or expert-designed value functions and deep simulation (8, 16, 28, 53, 62), which by definition cannot be expected of a novice. Rather, we propose a model that combines a general-purpose abstract value function—which could be easily constructed from understanding the rules of the game—along with depth-limited search. Instead of simulating the downstream effects of a move via extensive forward search, the Intuitive Gamer player module uses only a single step of "look-ahead" and selects an action probabilistically by evaluating the resulting board states according to a small collection of general heuristics (see Figure 2c). These heuristics are built on the assumption that players understand and pursue their goals (as defined by the game rules) and attempt to prevent their opponents from doing the same, subject to constraints on computational resources. The Intuitive Gamer player module then is a more general, compute-bounded version of prior game-based agents.

Formally, at a given board state ($s_t \in \mathcal{S}$), the Intuitive Gamer player module scores all legal next actions ($a_t \in \mathcal{A}$) according to a measure of *immediate progress* made towards a player's own goal (U_{self}) and a measure of progress *blocked* (U_{opp}) towards the opponent's goal (Figure 2c). Progress is based on the extent to which an action connects more contiguous pieces towards a winning K-in-a-row configuration (see Methods). In addition, we consider other easily-computable heuristic functions that may bear on the value of an action (U_{aux}, for "auxiliary"). In our experiments, we consider an auxiliary heuristic based on proximity to the center of the board (which allows a piece to participate in more winning terminal configurations). These heuristics are drawn from features used by prior studies in similar games (3, 17, 62), though we generalize them to our broader class of strategic games. The final heuristic value assigned to a given action on a particular board, $\tilde{\mathcal{V}}(s_t, a_t)$, is a sum of the three utility components described above:

$$\tilde{\mathcal{V}}(s_t, a_t) = U_{\text{self}}(s_t, a_t) + U_{\text{opp}}(s_t, a_t) + U_{\text{aux}}(s_t, a_t).$$

Moves are selected by the Boltzmann rule (33, 59, 19), sampling from a softmax probability distribution applied to action values. We assume that players have therefore already developed the capacity to account for multiple goals simultaneously, unlike potentially even more naive child-like game reasoners (17).

Our approach to action evaluation and choice reduces computational cost as it is strongly *local* in both space and time: it considers only the progress immediately stemming from a

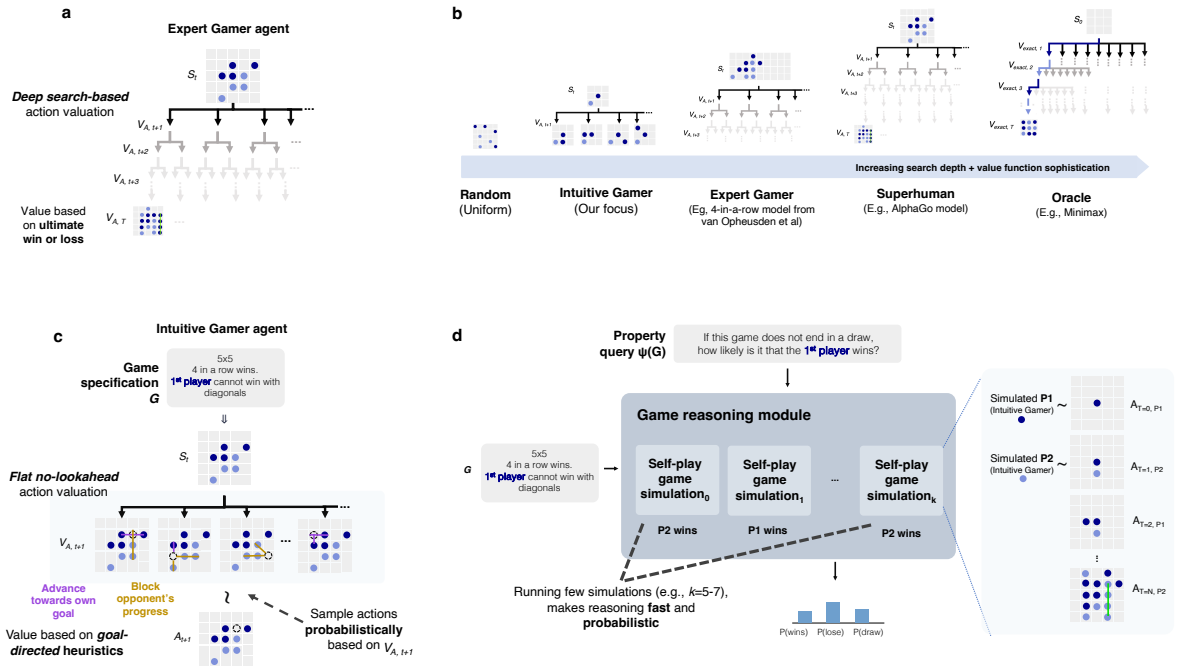


Figure 2: The Intuitive Gamer model, compared to prior models of game reasoning. a, Prior work modeling expert gameplay often involves deep tree search to determine what move to make given a board state S_t (52, 62). It is unlikely that novice human reasoners conduct such computationally expensive search and state evaluation before deciding whether to engage with the problem at all. **b,** What might novice reasoners be doing instead? Gameplay agents can differ in the amount of compute and expertise brought to bear to reason about any game (differing in search depth and value sophistication). Our proposal is that people reasoning about problems with which they have no experience, sit at the lower-end of this spectrum, but not *the* lowest. **c,** The Intuitive Gamer is conducts depth-limited search with game-general abstract goal-directed value functions that have yet to encode game-specific features. Specifically, we specify an Intuitive Gamer gameplay agent that conducts no more than a single step of lookahead when assessing deciding what action (A_{t+1}) to take from a given board state (S_t). The Intuitive Gamer is goal-directed, assessing whether any action would advance the player's own goal (purple) and how much it may block the progress of the opponent's goal (yellow). These operations are done in a "flat" fashion: thinking only one step ahead. The final action is selected probabilistically by sampling from a softmax-distribution over the estimated values per action. **d,** To reason about any new game query ψ for a given game description $\mathcal{G}$, we posit that people conduct only a few (k) self-play simulations between the gameplay agent (as depicted in panel **c**) to answer the query, which could be run to termination or probabilistically stop early. Taken together, the Intuitive Gamer model is fast (low k), flat (low-search depth), goal-directed (in the value function), and probabilistic (in action selection)—and involves mental simulations of gameplay.

specific action (hence its dependence on s_t and a_t) and does not account for either past or potential future returns. In contrast, it is common for value functions to be action-agnostic (i.e., to depend only on s_t van Opheusden et al. 62) and to explicitly capture some notion of the game's terminal rewards (e.g., via a learned value network Silver et al. 53). Intuitively, each of these alternatives represents a substantial increase in cognitive load: the former requires scanning over the entire board to evaluate any action and the latter requires mental simulation all the way to the end of the game (or access to a function that is derived from such simulations) before deciding what action to take. We provide further details on the player module and heuristic value choices in the Methods and Supplementary Information.

Reasoning module. While the Intuitive Gamer player model captures action choice during a game, the Intuitive Gamer reasoning module captures *judgments* about game properties by nesting the player module within a sample-based probabilistic inference procedure (Figure 2d).

For any game and position, the reasoning module infers answers to a query (e.g. *what is the likelihood that the game will end in a win, loss, or draw for a given player?*) by simulating k game playouts ($\{(s_0^0, s_1^0, ..., s_T^0), ..., (s_0^k, s_1^k, ..., s_{T'}^k)\}$) and iteratively calling the player module at each turn of self-play. Each simulation ends when the game reaches a win or draw state, or may terminate early with some probability distribution on stopping times. Computations over the simulated gameplay traces then inform the resulting distribution over the query values (e.g., game outcomes). We assume simulated games are independent, though this assumption could be relaxed in the future (see Discussion). In keeping with our focus on fast, resource-limited reasoning, we use only a *small number of game simulations* (which we estimate empirically, see Results and Methods) in line with prior evidence that people use a relatively small number of simulations to form beliefs and make decisions (22, 49, 64, 26, 69).

Taken together, the two modules instantiate a compute-limited, probabilistic account of both action selection and inferences about whether the game is worth playing (before any actual play) in novices. This account formalizes the intuition that new players approach games with generalized, game-agnostic priors rather than sophisticated or even winning strategies derived from extensive search or experience.

Alternate models. Our Intuitive Gamer framework naturally extends to model players with greater or lesser competence and motivation. We can think of the Intuitive Gamer as operating under the same mechanisms as more expert models, but scaled down (offering, therefore, a path to scale back up). Specifically, we modulate parameters in our framework (value function and search depth) to instantiate a version of the kind of cognitive models that have been successful at capturing human experts in a single game 4-in-a-row (62). We implement an approximately depth-5 version of the Intuitive Gamer, which we refer to as the "Expert Gamer." Of note, as the value function and search depth are intimately tied (a temporally-local value function is reasonable with a single step of look-ahead but ill-suited to the standard *minimax* search over game trees (27); because it "forgets" prior successes and blunders as it searches for the optimal move), we introduce an inherited version of our heuristic that retains a proportion of the value of prior moves as it performs deeper search. This means the Expert Gamer is deeper and involves a more sophisticated value function, in line with prior work. More details in the Methods and ablations are in the Supplementary Information. We also quantitatively compare reasoning judgments to an even more computationally-intensive variant popular in AI: Monte Carlo Tree Search (MCTS) (16, 20, 52). On the other end of the spectrum, we consider unmotivated players who select actions uniformly at random (see Figure 2b). We also compare against models that do not involve explicit structured game simulation, operating either over pre-computed linguistic features from the game description alone (see Methods and Supplementary Information) or more flexible linguistic computable models, e.g., large language models (see Supplementary Information).

Resource-rational reasoning. The Intuitive Gamer occupies an interesting point on a frontier of compute efficiency and game reasoning sophistication. On the one hand, the Intuitive Gamer is orders of magnitude faster than more sophisticated game reasoners (MCTS and the Expert Gamer) in terms of wall clock time and number of board states evaluated (see Extended Data Table 2 and Supplementary Information). The Intuitive Gamer is approximately $700\times$ faster than the Expert Gamer in wall clock time, with approximately $500\times$ fewer board evaluation, and is nearly $40,000\times$ faster than MCTS with almost $10,000\times$ fewer node evaluations, as computed under self-play game simulations for the respective model. On the other hand, while not as close to game-theoretic optimal play as the Expert Gamer or MCTS, game evaluations under the Intuitive Gamer are still well correlated with game-theoretic optimal analyses and far more aligned with game-theoretic optimal play than random play (see Extended Data Table 3 and Methods). The Intuitive Gamer's mechanisms of fast, flat goal-directed probabilistic simulation

thus constitute a highly efficient and resource-rational approach to game reasoning, which we hypothesize people could plausibly and productively engage when encountering new games.

New games for studying novice game reasoning. To explore how people reason about new games within the bounds of a behavioral study, we need a broad collection of games that are *novel* without requiring lengthy demonstrations or explanations. To that end, we implement a wide range of two-player, grid-based strategy games derived from classic *M-N-K* games (examples include Tic-Tac-Toe and Gomoku)—where players take turns placing pieces on an $M \times N$ grid trying to make K pieces in a row—frequently studied in prior work (17, 3, 62). Our set of 121 distinct games covers a variety of environment specifications (e.g., the size and shape of the board), transition dynamics (e.g., the number of moves a player makes on their turn), and win conditions (e.g., whether completing a line results in a win or a loss). In addition, games vary in their duration and balance under optimal play (see Methods). In each case, however, the core mechanic of the game (i.e., placing a piece into an empty grid cell) remains both consistent and familiar. We provide examples in Figure 1a and Extended Data Table 1.

We next use these games to study how people evaluate games they have never played before Figure 1b and make decisions in new games—playing the games themselves (Figure 1c), predicting where others will play (Figure 1d), and evaluating whether games are worth continuing to play at all.

How do people evaluate new games before they have played them?

We first assess how people evaluate a game (Figure 1b), from "just thinking" about the game alone—before any play. We consider two kinds of game evaluation: how people assess the expected outcomes of a game and a more subjective assessment of whether the game is likely to be engaging (or "fun") to play.

Is the game likely to be fair? We recruited 238 participants to evaluate the expected outcomes of the game in a "zero-shot, zero-experience" experiment. Participants "just thought" about the games, answering questions about a subset of 10 of the novel strategy games (plus Tic-Tac-Toe) based solely on the natural language description of their rules. To minimize demands on spatial working memory, participants were given access to a self-play scratchpad (see Methods) which they could optionally use while thinking about a game. Participants estimated both the likelihood of a draw and the likelihood that the first player would win if the game did *not* end in the draw. These estimates define an expected payoff of the game from the perspective of the first player (see Methods), which in turn encodes an intuitive notion of fairness (i.e., whether a game is biased towards a particular player).

We assess how each characteristic of the Intuitive Gamer model (fastness, flatness, goal-directedness, and probabilistic simulation) contributes to people's payoff predictions by comparing to alternative models that increase or decrease the sophistication of each component (Figure 3a). The Intuitive Gamer model correlates well with human estimates ($R^2 = 0.81$ [95% CI: $0.77, 0.85$]), matching the total explainable variance from the human data as estimated by split-half correlations ($R^2 = 0.82$ [95% CI: $0.77, 0.86$]). The Intuitive Gamer fits significantly better fit than alternate models varying in sophistication. It outperforms the random model ($R^2 = 0.48$ [95% CI: $0.41, 0.54$]), Expert Gamer model ($R^2 = 0.66$ [95% CI: $0.61, 0.71$]), MCTS baseline ($R^2 = 0.60$ [95% CI: $0.55, 0.65$]), and a variant of the Intuitive Gamer that is not probabilistic ($R^2 = 0.53$ [95% CI: $0.48, 0.57$]).

To more directly probe each component of the Intuitive Gamer model, we conduct a series of single and combined lesions of key model components. The base Intuitive Gamer model can be parameterized by a series of binary choices: whether to be goal-directed in value assessments;

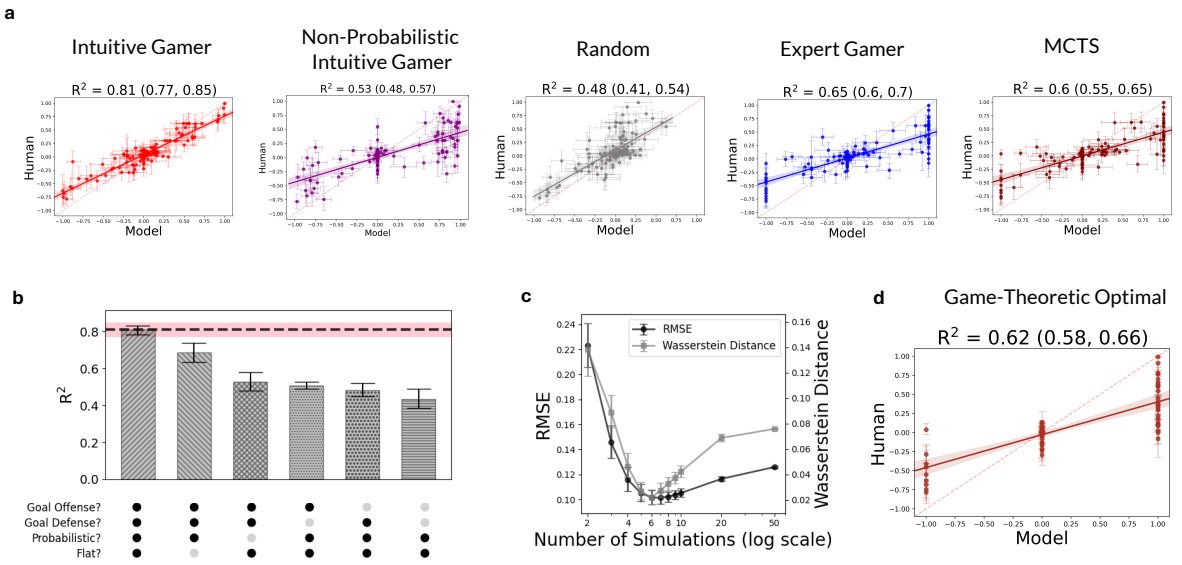


Figure 3: Evaluating games without ever playing. 238 participants judged the expected payoff of games drawn from our 121 game suite. **a,** The expected payoff computed under the Intuitive Gamer model captures human predictions well when compared to alternate models that scale up or down compute. Each point represents the payoff for one of the $n = 121$ game stimuli. Error bars depict 95% CIs around the mean estimated payoff from people and over $k = 6$ simulations for each model sampled from 20 simulated participants.; **b,** Comparing the full Intuitive Gamer against lesions of one or more critical components. Lesioning the flatness (increasing depth d to approximately 3), probabilistic nature of the game simulations (i.e., setting temperature of action selections to zero), and goal-directedness (ablating components of the value function) all degrade fit relative to the full Intuitive Gamer model. The dashed line indicates the mean and 95% CI split-half R^2 on the human-predicted payoffs, indicating the full Intuitive Gamer model captures essentially all of the explainable variance not due to noise. **c,** Only a small number of simulations (k approximately $5 - 7$, which we take as $k = 6$) are needed to well-capture the variance in participants' judgments, as measured by (1) the Root Mean Squared Error (RMSE) between participants' variance per game and the variance over model simulations per game and (2) the Wasserstein Distance between the distribution of variances across games which enables a finer-grained distinction between low variance from high k. Full scatterplots are shown in Extended Data Figure 1.; **d,** People's predictions are reasonable relative to the game-theoretic optimal values, but not as well-captured as the Intuitive Gamer predictions. Points show the average payoff per game, for the 78 of 121 games where an optimal payoff can be computed (either analytically or approximately with MCTS). Error bars depict 95% CIs around the bootstrapped mean human prediction per game.

whether to be probabilistic; whether to be flat. Ablating any of the factors impairs fit to human payoff judgments (see Figure 3b). "Fastness" can be further varied by modulating the number of simulations the Intuitive Gamer reasoning modules takes when assessing payoff. A small number of simulations (k more than one, but less than 10) best captures the variance in human prediction (Figure 3c and Extended Data Figure 1). These comparisons suggest that human payoff judgments of novel games are well-modeled by efficient and compute-limited search rather than purely random exploration of game states or deep and intensive simulation. For simplicity, this estimate assumes that all game simulations are independent and run to the end of the game; allowing game simulations to probabilistically terminate early (see Methods) yields very similar fits in the expected payoff evaluations and number of mental simulations per participant (see Extended Data Figure 2).

We were able to estimate the optimal game-theoretic payoff for many (78 of 121, see Methods) of the games, allowing us to compare people's subjective judgments to a fully objective game evaluation. Human judgments are reasonable relative to the game-theoretic optimal ($R^2 = 0.62$ [95% CI: $0.58, 0.66$]). While human judgments generally track the direction of the game-theoretic

optimal (e.g., estimating a payoff greater than zero when the first player should definitely win), the Intuitive Gamer provides a superior qualitative and quantitative fit to human judgments.

Is the game likely to be fun? We ran a parallel "zero-shot zero-experience" study described above on a new group of 246 participants who were asked to make a subjective estimate of how fun a game is likely to be, before ever playing, to test the hypothesis that the Intuitive Gamer's reasoning process can also account for how people make more subjective evaluations about new problems. Specifically, we hypothesized that several features of a game that can be read off easily from the model's fast and flat probabilistic simulations – how *balanced* a game is, how much a game *rewards thinking*, and how *long* a game is – could explain much (or even most) of the variance in people's funness judgments. We define a game's balance as the difference between the probability distribution of observed outcomes and an ideal outcome distribution relative to a game where exactly half of the games end in wins and losses and none end in draws (measured by the Earth Mover's Distance (47)). This feature captures the notion that players prefer decisive games (i.e., those that do not end in draws) that are also balanced across players. We define how much a game rewards thinking (13) as the relative advantage of playing strategically ("thinking") versus moving haphazardly ("without thinking"), which we measure as the proportion of simulated games won by the Intuitive Gamer model against a player that chooses actions uniformly at random from any legal move. Finally, our game length feature is modeled as an inverted U-shaped function of the expected number of turns until a game ends in win or draw, reflecting the intuition that longer and more involved games tend to be more fun, but the longer a game lasts without presenting proportionately more challenge, the less enjoyable it will be. (For further motivation and details of how these game features are computed, see Methods.)

Each of these three features computed under the Intuitive Gamer model captures significant variance in human funness judgments (see Figure 4a-c). A simple regression model which combines these features (including a quadratic game length term) attains $R^2 = 0.57$ [95% CI: $0.51, 0.63$]; Figure 4d) which approaches the total variance explainable in the human data as computed with bootstrapped split-half R^2 from participants over all games ($R^2 = 0.60$ [95% CI: $0.51, 0.68$]). Note there is generally greater variation across individuals' judgments of how fun a game is, relative to people's judgments of objected expected payoff, which is reasonable given that funness is both more vague and a more subjective evaluation.

To assess the sensitivity of these model fits to the choice of games, and compare alternative base game play modules from which to simulate play, we ran a generalization test wherein we fit regression models on 50% (20) of the 40 games studied and tested each model on the held-out 50% of games. Simulations under the Intuitive Gamer model predicted human funness evaluations significantly better than alternative models with greater or less compute resources (depth of thinking), as well as models that relied exclusively on linguistic features (see Figure 4e), and similarly captured most of the explainable variance in people's predictions as when these models were fit to the full set of 40 games. (For more details and analyses of these model fits, see Supplementary Information.)

Taken together, these two experiments in which participants "just think" in order to rate both objective measures of a game (e.g., its payoff or fairness) and more subjective evaluations (e.g., its enjoyability) lend support to our hypothesis that these kinds of novice judgments are best captured by a reasoning module that is goal-directed but limited in its computational cost (i.e. fast and flat) compared to alternatives that vary in more or less computational demands and game reasoning sophistication.

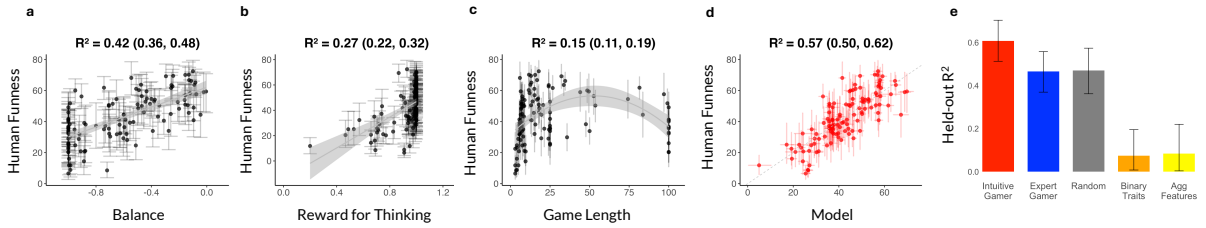


Figure 4: Evaluating whether games are likely to be fun, before ever playing them. 246 participants judged whether they find the games fun. The resulting human game fun ratings relate to several distinct game features **(a-c)** that can be read out under Intuitive Gamer model simulations: **a,** game balance, as predicted by the Intuitive Gamer model; **b,** predicted advantage over a random agent given the Intuitive Gamer model's game play (reward for thinking); and **c,** expected game length under the Intuitive Gamer model, as fit with a quadratic term to account for a non-linear relationship to funness. **d,** A regression model with all features predicted under the Intuitive Gamer model (balance, reward for thinking, and length) captures much of the explainable variance in the human funness ratings ($R^2 = 0.60$ [95% CI: $0.51, 0.68$]). Error bars depict 95% CIs around the human mean funness per game; error bars for the model depict 95% CI over the predicted mean for each of the fit models (models are fit to each bootstrap subsample). **e,** Simulations under the Intuitive Gamer model better capture human judgments than simulations under models varying the computational sophistication (random and expert game) in a generalization test, where the parameters of each model are fit on 50% of the games and tested on the other 50% of games. Funness models based on explicit simulation better capture human judgments than models based only on non-simulation based linguistic features (e.g., binary game traits—whether the game has asymmetric win conditions, non-square boards, etc) and aggregate features based binary traits (counts of how many are "on" to capture "novelty" relative to Tic-Tac-Toe; and board size, encoded as number of rows $\times$ number of columns). Error bars depict 95% CIs of the held-out R^2 over 500 bootstrapped train/test splits.

How do people make decisions playing a game for the first time?

Relative to the flat, single-step thinking that our Intuitive Gamer model posits people use to evaluate new games, conventional models of how people actually play games (62) perform much deeper tree search, simulating multiple future turns by both players in order to evaluate potential moves before choosing the best move at each step of a game. But these models have only been evaluated on players who are either experts in a game, or have at least played a game many times. What do human players do when they are playing a game for the very first time? Do they search deeply, as previous models might suggest, or do they use only a fast and flat decision mechanism as our Intuitive Gamer model does—and as our results suggest they do when merely imagining how a game might unfold before playing it?

What moves do people make? To test whether the Intuitive Gamer player module captures how people actually play a game for the first time, we recruited 302 participants to play each other in a subset of 40 the 121 game variations, plus Tic-Tac-Toe (Figure 1c). Each participant played only a single round of any given game and did so immediately after reading a brief description of the rules for the first time. During the game, participants took turns making moves and could elect to surrender or request a draw if they wanted to end the game.

Initial evidence that the Intuitive Gamer model captures not only how people think about new games but also how they play them for the first time comes from comparing the distribution of actual game outcomes (first or second player win, or draw) in this game play experiment with the outcome distributions to be expected if both players make move decisions according to the model. The model captures most of the variance in actual game payoffs with novice players ($R^2 = 0.72$ [95% CI: $0.67, 0.76$]) and predicts these payoffs as well as human estimates of expected game payoffs do, from our earlier game evaluation experiment (see Extended Data Figure 3).

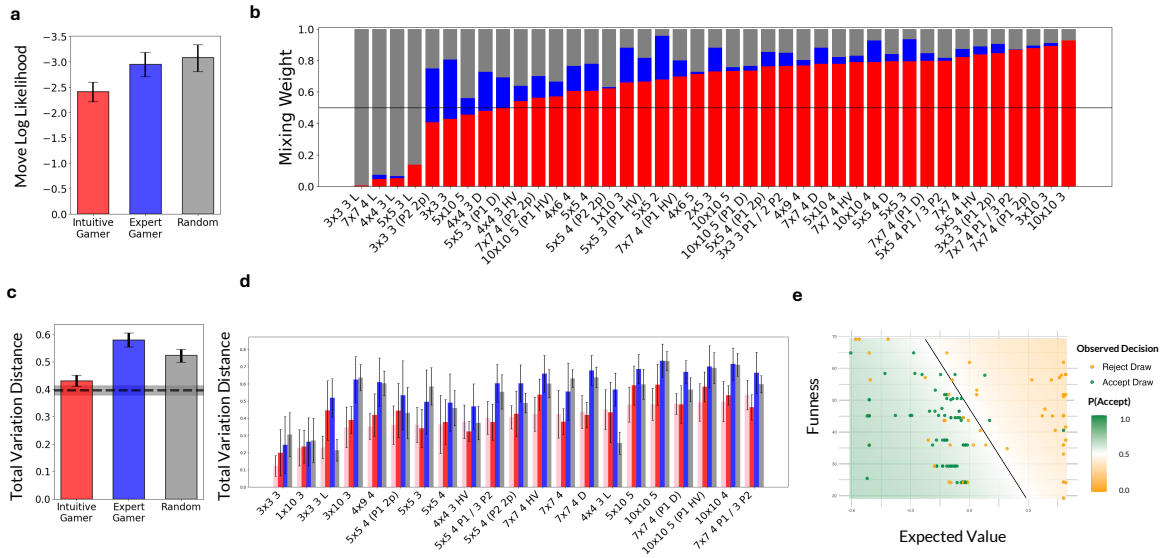


Figure 5: Modeling people's actions and distribution over predicted actions in the first encounter with a new game. a, The average log-likelihood of human moves under three variants of the game player module. Data come from 302 participants playing a single round each of 5 novel games (along with Tic-Tac-Toe); across participants 40 of our novel games were tested. The Intuitive Gamer model captures the probability of players' moves (under an α-softmax model marginalizing over α) significantly better than alternatives. Error bars depict 95% bootstrapped confidence intervals around the mean log-likelihood over all games and all players. **b,** Bar heights represent the estimated contribution of each model in a probabilistic mixture (admixture model, see Methods) fit to participants moves for each game. The Intuitive Gamer model is the component that best explains participants' moves in approximately 90% (37 of 41) games; games where people would not be novices (Tic-Tac-Toe) or some particularly atypical games (e.g., misere where the first person to get K in a row loses, listed with an "L" like "7 × 7 4 L") are comparatively less well captured by the Intuitive Gamer compared to alternate models. Game descriptions are lemmatized to show board size and key features (e.g., HV indicates only horizontal and vertical wins are allowed; P2 2p means Player 2 can move twice on their first turn). A full description of lemmatized game codes are in the Methods. **c,** The distribution over likely next moves under the Intuitive Gamer model also captures a new group of 314 participants' elicited belief distributions of where a player should move (over frozen videos of games played by participants in the prior experiment). Total Variation Distance (TVD; lower means closer to people) is computed between the aggregate participant suggested move distribution against the models' inferred move distribution, aggregated over all matches and marginalizing out α under an α-softmax model. The dashed line is the split-half human TVD (effectively the noise ceiling).; **d,** The Intuitive Gamer is generally aligned to the distribution of human judgments (and near at to the the split-half human TVD), with similar failures to the play data (e.g., for misere games). Error bars depict standard deviation over boards for each game. **e,** Decisions of whether to accept or reject a draw, when requested is captured by expected value of the player's current position, but participants tend to be willing to reject a draw and play out a game with slightly lower expected value if the game is judged to be fun.

A more direct test is to compare predicted move choices that the Intuitive Gamer model makes at each step of each game with the actual moves that new players made at the same game steps when they were playing these games for the first time. Because there is a strong element of randomness in people's move choices as well as the model's predictions, we evaluate the likelihood of people's moves under the Intuitive Gamer relative to alternative models that are either more or less sophisticated and computationally intensive, the Expert Gamer and Random models from our previous study, respectively (Figure 5a). The Intuitive Gamer model best predicts players' moves in aggregate, as measured either by average per-move log likelihoods or per-match log likelihoods (summed over all moves within a match), across 1808 distinct matches with a total of 9892 moves. All of these differences are highly significant as measured

by paired t-tests (mean differences in total per-match log likelihood for the Intuitive Gamer vs Expert model = 2.83, $t_{1807} = 29.4$, $p < 0.001$, vs Random model = 3.73, t_{1807} = 26.4, $p < 0.001$; mean differences in per-move log likelihood for the Intuitive Gamer vs Expert model = 0.51, $t_{9891} = 39.9$, $p < 0.001$, vs Random model = 0.68, t_{9891} = 42.3, $p < 0.001$).

We also fit these three models of action choice to the distribution of all moves people made in each individual game, and the distribution of all moves an individual player made across the different games they played, using probabilistic (admixture) models (see Supplementary Information). At a per-game level, we find that the Intuitive Gamer model generally but not always fits human play better than the Expert or Random models: it captures more than 50% of the probability distribution of moves in 32 out of 41 games, and a plurality of moves in 5 of the remaining games (Figure 5b). Individual player's distributions of moves are also best fit by the Intuitive Gamer: it captures more than 50% of the probability distribution of moves for 243 out of 302 players, and a plurality of moves in 16 of the remaining players (Extended Figure 4).

Post-hoc exploratory analyses looking at intermediate stages of gameplay suggest that the Intuitive Gamer model best fits early and mid-stage play, though in later stage play, greater depth variants are more comparable (see Supplementary Information). This could be due to some people being more motivated to think deeper at the end of the game or finding it easier to think deeper with a smaller space of moves to consider (as there are fewer legal moves at the end of the game).

Additionally, as in the game evaluation experiments, we find that both aspects of goal-directedness in the Intuitive Gamer value function (offensive progress towards your goal, and defensive assessment of how to block your opponents progress) matter for capturing human behavior in novel games (see Extended Data Figure 5). Overall, our results strongly support the hypothesis that when playing a novel game for the first time, people adopt a similar fast, flat and goal-directed approach to choosing their next move as they appear to do in mentally simulating games before they start to play.

What move should someone else make? In the previous experiment, as in almost all studies of game play, participants only made a single choice for each move, and they only played each game once. Yet our model posits that people think about new problems by running *probabilistic* mental simulations letting them form graded expectations about the value of a range of possible actions. To assess how well the Intuitive Gamer player module captures people's probabilistic expectations, we recruited a new group of participants to predict distributions of likely actions when watching videos of other novices playing these games (see Methods; Figure 1d).

The Intuitive Gamer model generally better captures the distribution of moves predicted by people compared to alternate models across the 249 game boards for each match and game stage, as measured by the Total Variance Distance (TVD) between model and people's predicted distribution, and is near the expected noise ceiling (computed via split-half) human judgments for most game boards (see Figure 5c). Differences in distance (where lower means closer to people's predicted action distribution) are highly statistically significant as measured by paired t-tests in the TVD per game board for the Intuitive Gamer vs Expert model = -0.15, $t_{248} = -15.0$, $p < 0.001$, vs Random model = -0.09, $t_{248} = -7.5$, $p < 0.001$. Exceptions for poor fits align with those found in the play experiments (e.g., on misere games; see Figure 5d and example boards in Figure 6). Fits are robust to choice of distributional measure (see Supplementary Information). People assign similar log probability on the move actually played by the human player to that of the Intuitive Gamer, further highlighting alignment of the Intuitive Gamer to people's probabilistic judgments (see Supplementary Information).

When is a game worth continuing? Game evaluation is important not only when deciding whether a game is worth playing but also for assessing—during play—whether a game is worth continuing to engage with. In our game play study, participants had the option to request a

Figure 6: Example human- and model-predicted distributions over the next action in real games. Each set of three boards (**a-j**) shows a representative state taken from a different game between two human players, along with the actual move one player took at the next step, and the predictions made by other humans and two models for how that move should have been made. Filled black and white circles show the moves both players took prior to this point in the game. The one open circle on each board indicates where the human participant whose turn it was on this move (white or black) actually played. The top board in each panel show the aggregated human-predicted distribution over next moves, as determined by participants in the watch-and-predict study. The middle and lower boards in each panel show the predictions of the Intuitive Gamer and Expert Gamer player modules, respectively. The purple-shaded colors of the empty board squares indicate the model- or human-predicted probability distributions over next moves; that is, the estimated probability that the player about to move would select that position on their current turn. Participants in the watch-and-predict study typically spread their probability judgments over a range of possible next actions (**a-e**), although sometimes predict just a single move with high confidence (**f, g**), and the Intuitive Gamer often captures these tendencies, assigning similar probabilities over an analogous range of "reasonable" moves. The Expert Gamer model fits human predictions less well. It tends to be overly confident, allocating more probability to a smaller set of moves that are high-utility and often counter-intuitive to naive players (**a-c, e**). The Expert Gamer can also fail by being too diffuse: in states where deep lookahead search indicates a sure loss under perfect play, the Expert Gamer model becomes defeatist and assigns equal probability to all actions, while human participants still make non-arbitrary decisions and highly non-uniform prediction distributions that the Intuitive Gamer captures (**d, f**). The Intuitive Gamer does not always perform best; occasionally the Expert Gamer's sharper confidence aligns better with human predictions (**g**) or the Intuitive Gamer model's distribution is overly sharp (**h**). Finally, in some cases both models make much sharper predictions than humans do. This can happen in crowded boards when people (but not models) are liable to miss a clear winning move or fail to block an oppponent's win (**i**), or in misere variants where a player who gets K in a row first loses (**j**), perhaps owing to these games' unfamiliarity or greater computational complexity. The full set of board predictions (for 249 moves across 41 distinct games) we modeled is included in the Supplementary Information.

draw at any point in play. If they received a draw request from another player, they had the choice of whether to accept the draw and immediately terminate play or reject the draw and keep playing. Over the 1808 matches participants played, players made 142 draw requests, of which 83 were accepted and 59 were rejected. Upon receiving a draw request, how do people decide whether to keep playing?

We modeled these choices as probabilistic value-based decisions trading off a player's expected reward of winning with the expected cost of continuing to play: that is, as the difference between the estimated probability of ultimately winning the game (and receiving a $\$0.50$ bonus; see Methods) and an opportunity cost based on the expected remaining length of the game, in terms of number of moves.

Both of these expectations were computed from the Intuitive Gamer reasoning module conditioning on the current board state (rather than a blank opening board, as in our earlier game evaluation studies). In exploratory analyses, we also hypothesized that a player might be willing to keep playing a game even if the expected cost of continuing outweighs the potential objective monetary reward, if playing provides an alternative subjective reward—namely, if it is likely to be fun, which we measured using the subjective funness evaluation from our earlier "just think" experiment. A logistic regression model fit to participants' draw request decisions using these three features (see Methods) captures the intuition that players' game-time evaluations are probabilistically rational under limited resources (time) and sensitive to subjective factors like the engagement of games in addition to their objective value (Figure 5e). Lesions of the Intuitive Gamer evaluation function lead to significantly worse qualitative and quantitative fits to human draw request choices (see Extended Data Figure 6 and Extended Data Table 4).

Discussion

Systems of rules and rewards govern many aspects of human behavior—from jobs and institutions to games and geopolitics. How can people think reasonably about any such system or problem for the first time, so flexibly and quickly? We formalized and empirically evaluated a computational model of an Intuitive Gamer, which draws on fast and flat goal-directed probabilistic mental simulations to make rich quantitative predictions about how novice players judge the objective value of a previously unseen game (how fair or balanced it is) as well as make more subjective evaluations (e.g., how much fun the game may be). Our model also captures how people act within these games (in both playing, and judging others' play), and does so better than alternative models with more and less computational power. Crucially, the same model generalizes across these different tasks and the many different games we study.

Our work differs from most prior studies of game play in cognitive science and AI which focus on how expert play is achieved, and typically consider only expertise in a single game. In contrast, we study a *class* of 121 strategic games with varying dynamics and highlight the phase of "pre-expertise": the critical first moments before people can rely on the benefits of repeated experience, but when they must nonetheless take reasonable actions and even decide whether to play at all, or continuing playing. Studying pre-expertise across varied domains is important not just for our understanding of cognition generally, but also to construct more accurate models of human behavior under novel economic mechanisms (25, 39, 37, 34, 35, 29) or alongside unfamiliar agents like AI systems (10, 68, 24, 56, 65, 15). From an AI standpoint, the human judgments that we collect could also form the basis of evaluations and benchmarks for researchers interested in assessing the human-likeness and capabilities of AI systems that must engage with novel problems (67), just as our Intuitive Gamer models could inform the design of more efficient AI systems that make reasonable judgments and decisions in new multi-agent settings despite using substantially less compute than conventional expert-level models.

Looking forward, we hope that our work will motivate studies of how people think about new games and novel problems well beyond those we have studied here, towards a general model of people's "intuitive theory of games" analogous to the computational models of intuitive theories of physics or intuitive theories of mind that cognitive scientists have constructed (5, 4, 61). In particular, future work should study how the mechanisms of fast and flat goal-directed probabilistic simulation might generalize to the vast range of competitive

board games human cultures have produced (with thousands now available in digital form (46)), as well as games of other kinds: collaborative games, cooperative games, and games that require balancing competition with cooperation in multi-player teams, of the sort that underlie much of human social life.

Additionally, there is a need for more fine-grained process-level and individual-level accounts of how people reason about novel games. While we presented initial analyses of the impact of early termination in the Intuitive Gamer model's simulations, more work is needed to understand people's mental simulations prior to any play—e.g., how people judge the expected outcome of a game if they stop imagining play before reaching a win or draw, or how their simulations may be temporally dependent rather than independent as our simple models assumed. Participants in our study were required to consider each game for 60 seconds before submitting their judgments, to encourage at least minimal thinking, but future work should explore how varying this and other task constraints might affect the amount and kind of mental sampling people do. Understanding the role of prior game experience will also likely be important for modeling variability in peoples' thinking, particularly for judgments about whether a game is likely to be fun which were more variable than people's evaluations of game fairness. People may also differ in their preferences for competitive games that demand strategic thinking (43) and in their assessments of the momentary funness of a game relative to the expected funness (48), which may shape their decisions of whether to play a game multiple times or continue playing in a particular match.

Finally, people not only think about games and problems that have been presented to them but also *make* new games and new problems, either by modifying the rules or rewards of existing ones to make them more fair or more fun, or creating entirely new games, goals, or pursuits. Children do this all the time as part of play, and adults do too, in ways that may be ultimately responsible for some of humanity's greatest discoveries and innovations (13). In ongoing work we are exploring whether Intuitive Gamer-like models based on fast and flat goal-directed simulation can also explain how people play these "meta-games": deciding to change a game, or constructing new games that they and others find rewarding to think about.

Science itself has been characterized as a family of games against nature (23), where scientists set the rules that lead to the most productive research (31). Mathematicians explore conjectures and proof strategies by imagining adding or relaxing the logical rules of how a problem is formulated (57). Suppose that these and other open-ended, creative activities at the frontiers of human knowledge do in fact share something deeply in common with intuitive reasoning about games, or even children's play. Then perhaps mechanisms like the fast probabilistic mental simulations studied here could even help to explain how researchers choose to approach a new problem, or which problems to think about in the first place. Whether scientific thinking can be meaningfully captured by such models is very much an open question, but one we are now better positioned to investigate. At the very least, it feels like a game worth playing.

References

[1] A. Almaatouq, J. Becker, J. P. Houghton, N. Paton, D. J. Watts, and M. E. Whiting. Empirica: a virtual lab for high-throughput macro-level experiments. *Behavior Research Methods*, 53 (5):2158–2171, 2021.

[2] I. Althofer. Computer-aided game inventing. *Friedrich Schiller University, Jena, Germany, Tech. Rep*, 2003.

[3] O. Amir, L. Tyomkin, and Y. Hart. Adaptive search space pruning in complex strategic problems. *PLoS Computational Biology*, 18(8):e1010358, 2022.

[4] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4):0064, 2017.

[5] P. W. Battaglia, J. B. Hamrick, and J. B. Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, 2013.

[6] C. B. Browne. *Automatic generation and evaluation of recombination games*. PhD thesis, Queensland University of Technology, 2008.

[7] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012.

[8] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu. Deep Blue. *Artificial Intelligence*, 134(1-2):57–83, 2002.

[9] X. Cao and Y. Lin. Uct-adp progressive bias algorithm for solving gomoku. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 50–56. IEEE, 2019.

[10] M. Carroll, R. Shah, M. K. Ho, T. Griffiths, S. Seshia, P. Abbeel, and A. Dragan. On the utility of learning about humans for human-AI coordination. *Advances in neural information processing systems*, 32, 2019.

[11] N. Charness. Expertise in chess: The balance between knowledge and search. In K. A. Ericsson and J. Smith, editors, *Toward a general theory of expertise: Prospects and limits*, pages 39–63. Cambridge University Press, Cambridge, 1991.

[12] W. G. Chase and H. A. Simon. The mind's eye in chess. In *Visual information processing*, pages 215–281. Elsevier, 1973.

[13] J. Chu, J. B. Tenenbaum, and L. E. Schulz. In praise of folly: flexible goals and human cognition. *Trends in Cognitive Sciences*, 2023.

[14] A. A. Cleveland. The psychology of chess and of learning to play it. *The American Journal of Psychology*, 18(3):269–308, 1907.

[15] K. M. Collins, I. Sucholutsky, U. Bhatt, K. Chandra, L. Wong, M. Lee, C. E. Zhang, T. Zhi-Xuan, M. Ho, V. Mansinghka, et al. Building machines that learn and think with people. *Nature Human Behavior*, 2024.

[16] R. Coulom. Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games*, pages 72–83. Springer, 2006.

[17] K. Crowley and R. S. Siegler. Flexible strategy use in young children's tic-tac-toe. *Cognitive Science*, 17(4):531–561, 1993.

[18] FAIR, A. Bakhtin, N. Brown, E. Dinan, G. Farina, C. Flaherty, D. Fried, A. Goff, J. Gray, H. Hu, et al. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

[19] M. Franke and J. Degen. The softmax function: Properties, motivation, and interpretation. OSF Preprints, 2023.

[20] M. Genesereth and M. Thielscher. *General Game Playing*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2014.

[21] F. Gobet, J. Retschitzki, and A. de Voogt. *Moves in mind: The psychology of board games*. Psychology Press, 2004.

[22] T. L. Griffiths, A. N. Sanborn, K. R. Canini, and D. J. Navarro. Categorization as nonparametric Bayesian density estimation. *The probabilistic mind: Prospects for Bayesian cognitive science*, pages 303–328, 2008.

[23] J. Hintikka. On the logic of an interrogative model of scientific inquiry. *Synthese*, pages 69–83, 1981.

[24] M. K. Ho, D. Abel, C. G. Correa, M. L. Littman, J. D. Cohen, and T. L. Griffiths. People construct simplified mental representations to plan. *Nature*, 606(7912):129–136, 2022.

[25] L. Hurwicz. The design of mechanisms for resource allocation. *The American Economic Review*, 63(2):1–30, 1973.

[26] T. Icard. Subjective probability as sampling propensity. *Review of Philosophy and Psychology*, 7(4):863–903, 2016.

[27] J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American mathematical society*, 4(3):502–506, 1953.

[28] L. Kocsis and C. Szepesvari. Bandit based Monte-Carlo planning. In *European Conference on Machine Learning*, pages 282–293. Springer, 2006.

[29] R. Koster, J. Balaguer, A. Tacchetti, A. Weinstein, T. Zhu, O. Hauser, D. Williams, L. Campbell-Gillingham, P. Thacker, M. Botvinick, et al. Human-centred mechanism design with democratic ai. *Nature Human Behaviour*, 6(10):1398–1407, 2022.

[30] I. Kuperwajs, M. K. Ho, and W. J. Ma. Heuristics for meta-planning from a normative model of information search. *Planning*, 1:a2, 2024.

[31] L. Laudan. *Progress and its problems: Towards a theory of scientific growth*, volume 282. Univ of California Press, 1978.

[32] H. Li. gobang: JavaScript Gobang AI based on Alpha-Beta Pruning. https://github.com/lihongxun945/gobang, 2025.

[33] R. D. Luce. *Individual choice behavior*, volume 4. Wiley New York, 1959.

[34] E. S. Maskin. Mechanism design: How to implement social goals. *American Economic Review*, 98(3):567–576, 2008.

[35] D. McFadden. The human side of mechanism design: a tribute to Leo Hurwicz and Jean-Jacque Laffont. *Review of Economic Design*, 13(1):77–100, 2009.

[36] M. L. Menéndez, J. A. Pardo, L. Pardo, and M. d. C. Pardo. The Jensen-Shannon divergence. *Journal of the Franklin Institute*, 334(2):307–318, 1997.

[37] P. R. Milgrom. *Putting auction theory to work*. Cambridge University Press, 2004.

[38] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[39] R. B. Myerson. Mechanism design by an informed principal. *Econometrica: Journal of the Econometric Society*, pages 1767–1797, 1983.

[40] M. R. Nassar and M. J. Frank. Taming the beast: extracting generalizable knowledge from computational models of cognition. *Current opinion in behavioral sciences*, 11:49–54, 2016.

[41] A. Newell and H. A. Simon. *Human problem solving*. Prentice-Hall, Englewood Cliffs, NJ, 1972.

[42] A. Newell, J. C. Shaw, and H. A. Simon. Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development*, 2(4):320–335, 1958.

[43] C. T. Nguyen. Games and the art of agency. *Philosophical Review*, 128(4):423–462, 2019.

[44] OpenAI Team, A. Jaech, A. Kalai, A. Lerer, A. Richardson, A. El-Kishky, A. Low, A. Helyar, A. Madry, A. Beutel, A. Carney, A. Iftimie, A. Karpenko, A. T. Passos, A. Neitz, A. Prokofiev, A. Wei, A. Tam, A. Bennett, A. Kumar, A. Saraiva, A. Vallone, A. Duberstein, A. Kondrich, A. Mishchenko, A. Applebaum, A. Jiang, A. Nair, B. Zoph, B. Ghorbani, B. Rossen, B. Sokolowsky, B. Barak, B. McGrew, B. Minaiev, B. Hao, B. Baker, B. Houghton, B. McKinzie, B. Eastman, C. Lugaresi, C. Bassin, C. Hudson, C. M. Li, C. de Bourcy, C. Voss, C. Shen, C. Zhang, C. Koch, C. Orsinger, C. Hesse, C. Fischer, C. Chan, D. Roberts, D. Kappler, D. Levy, D. Selsam, D. Dohan, D. Farhi, D. Mely, D. Robinson, D. Tsipras, D. Li, D. Oprica, E. Freeman, E. Zhang, E. Wong, E. Proehl, E. Cheung, E. Mitchell, E. Wallace, E. Ritter, E. Mays, F. Wang, F. P. Such, F. Raso, F. Leoni, F. Tsimpourlas, F. Song, F. von Lohmann, F. Sulit, G. Salmon, G. Parascandolo, G. Chabot, G. Zhao, G. Brockman, G. Leclerc, H. Salman, H. Bao, H. Sheng, H. Andrin, H. Bagherinezhad, H. Ren, H. Lightman, H. W. Chung, I. Kivlichan, I. O'Connell, I. Osband, I. C. Gilaberte, I. Akkaya, I. Kostrikov, I. Sutskever, I. Kofman, J. Pachocki, J. Lennon, J. Wei, J. Harb, J. Twore, J. Feng, J. Yu, J. Weng, J. Tang, J. Yu, J. Q. Candela, J. Palermo, J. Parish, J. Heidecke, J. Hallman, J. Rizzo, J. Gordon, J. Uesato, J. Ward, J. Huizinga, J. Wang, K. Chen, K. Xiao, K. Singhal, K. Nguyen, K. Cobbe, K. Shi, K. Wood, K. Rimbach, K. Gu-Lemberg, K. Liu, K. Lu, K. Stone, K. Yu, L. Ahmad, L. Yang, L. Liu, L. Maksin, L. Ho, L. Fedus, L. Weng, L. Li, L. McCallum, L. Held, L. Kuhn, L. Kondraciuk, L. Kaiser, L. Metz, M. Boyd, M. Trebacz, M. Joglekar, M. Chen, M. Tintor, M. Meyer, M. Jones, M. Kaufer, M. Schwarzer, M. Shah, M. Yatbaz, M. Y. Guan, M. Xu, M. Yan, M. Glaese, M. Chen, M. Lampe, M. Malek, M. Wang, M. Fradin, M. McClay, M. Pavlov, M. Wang, M. Wang, M. Murati, M. Bavarian, M. Rohaninejad, N. McAleese, N. Chowdhury, N. Chowdhury, N. Ryder, N. Tezak, N. Brown, O. Nachum, O. Boiko, O. Murk, O. Watkins, P. Chao, P. Ashbourne, P. Izmailov, P. Zhokhov, R. Dias, R. Arora, R. Lin, R. G. Lopes, R. Gaon, R. Miyara, R. Leike, R. Hwang, R. Garg, R. Brown, R. James, R. Shu, R. Cheu, R. Greene, S. Jain, S. Altman, S. Toizer, S. Toyer, S. Miserendino, S. Agarwal, S. Hernandez, S. Baker, S. McKinney, S. Yan, S. Zhao, S. Hu, S. Santurkar, S. R. Chaudhuri, S. Zhang, S. Fu, S. Papay, S. Lin, S. Balaji, S. Sanjeev, S. Sidor, T. Broda, A. Clark, T. Wang, T. Gordon, T. Sanders, T. Patwardhan, T. Sottiaux, T. Degry, T. Dimson, T. Zheng, T. Garipov, T. Stasi, T. Bansal, T. Creech, T. Peterson, T. Eloundou, V. Qi, V. Kosaraju, V. Monaco, V. Pong, V. Fomenko, W. Zheng, W. Zhou, W. McCabe, W. Zaremba, Y. Dubois, Y. Lu, Y. Chen, Y. Cha, Y. Bai, Y. He, Y. Zhang, Y. Wang, Z. Shao, and Z. Li. OpenAI o1 System Card, 2024. URL https://arxiv.org/abs/2412.16720.

[45] S. Palan and C. Schitter. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[46] E. Piette, D. J. Soemers, M. Stephenson, C. F. Sironi, M. H. Winands, and C. Browne. Ludii–the ludemic general game system. In *ECAI 2020*, pages 411–418. IOS Press, 2020.

[47] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pages 59–66. IEEE, 1998.

[48] R. B. Rutledge, N. Skandali, P. Dayan, and R. J. Dolan. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, 111(33): 12252–12257, 2014.

[49] A. N. Sanborn, T. L. Griffiths, and D. J. Navarro. Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, 117(4):1144, 2010.

[50] C. E. Shannon. Xxii. programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(314):256–275, 1950.

[51] K. Sheoran, G. Dhand, M. Dabas, N. Dahiya, and P. Pushparaj. Solving connect 4 using optimized minimax and monte carlo tree search. *Advances and Applications in Mathematical Sciences*, 21(6):3303–3313, 2022.

[52] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[53] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*, 2017.

[54] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

[55] H. Simon and W. Chase. Skill in chess. In *Computer chess compendium*, pages 175–188. Springer, 1988.

[56] M. Steyvers, H. Tejeda, G. Kerrigan, and P. Smyth. Bayesian modeling of human–ai complementarity. *Proceedings of the National Academy of Sciences*, 119(11):e2111547119, 2022.

[57] T. Tao. There's more to mathematics than rigour and proofs, 2007.

[58] G. Todd, A. G. Padula, M. Stephenson, É. Piette, D. Soemers, and J. Togelius. Gavel: Generating games via evolution and language models. *Advances in Neural Information Processing Systems*, 37:110723–110745, 2024.

[59] K. E. Train. *Discrete choice methods with simulation*. Cambridge University Press, 2009.

[60] J. W. Uiterwijk. Solving strong and weak 4-in-a-row. In *2019 IEEE conference on games (CIG)*, pages 1–8. IEEE, 2019.

[61] T. D. Ullman, E. Spelke, P. Battaglia, and J. B. Tenenbaum. Mind games: Game engines as an architecture for intuitive physics. *Trends in cognitive sciences*, 21(9):649–665, 2017.

[62] B. van Opheusden, I. Kuperwajs, G. Galbiati, Z. Bnaya, Y. Li, and W. J. Ma. Expertise increases planning depth in human gameplay. *Nature*, pages 1–6, 2023.

[63] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[64] E. Vul, N. Goodman, T. L. Griffiths, and J. B. Tenenbaum. One and done? Optimal decisions from very few samples. *Cognitive science*, 38(4):599–637, 2014.

[65] X. Wang, Z. Lu, and M. Yin. Will you accept the AI recommendation? predicting human behavior in ai-assisted decision making. In *Proceedings of the ACM Web Conference 2022*, pages 1697–1708, 2022.

[66] G. N. Yannakakis and J. Togelius. *Artificial Intelligence and Games*. Springer, 2018.

[67] L. Ying, K. M. Collins, P. Sharma, C. Colas, K. I. Zhao, A. Weller, Z. Tavares, P. Isola, S. J. Gershman, J. D. Andreas, et al. Assessing adaptive world models in machines with novel games. *arXiv preprint arXiv:2507.12821*, 2025.

[68] T. Zhi-Xuan, J. Mann, T. Silver, J. Tenenbaum, and V. Mansinghka. Online Bayesian goal inference for boundedly rational planning agents. *Advances in neural information processing systems*, 33:19238–19250, 2020.

[69] J.-Q. Zhu, A. N. Sanborn, and N. Chater. The bayesian sampler: Generic bayesian inference causes incoherence in human probability judgments. *Psychological review*, 127(5):719, 2020.

Methods

Game construction

We manually constructed the 121 two-player competitive strategy game variants played on $M \times N$ grids, where M are the number of rows and N are the number of columns. One goal in creating these games was to ensure there is enough diversity in board size and rule structure as well as systematic variance. To that end, we designed a series of variations on square boards: on 10×10 boards, we have have K-in-a-row for K from 2 to 10; for 3-in-a-row, we have have M by $N = M$ boards from 3 to 10; we also include a few 5×5 boards varying K. We additionally included a few other square boards varying in complexity, as well as rectangular boards ranging in size from from 1×5 to 5×10, and we have integer $2 \leq n \leq 6$ for K-in-a-row. To assess how people reason about games that are not physically realizable, we included three "infinitely" sized games with $K = 3, 5, 10$ for K-in-a-row. These categories give us 41 games with typical "M-N-K rules" (e.g., where the players take turns and have the same objective: "make K in a row, where horizontal, vertical, or diagonal all count). This set includes the standard Tic-Tac-Toe as well as 4×9, 4 in a row wins from (62). We then created a number of games with varied game rules. We kept the selection of board sizes and K-in-a-row fixed across categories (10 games within each category). The selection included 3×3, 4×4, 5×5, and 10×10 boards and $n \in \{3, 4, 5, 10\}$. We also designed games wiht more atypical rules, varying the winning conditions (e.g., K-in-a-row loses and diagonal connections do not count as wins) and first-mover dynamics (e.g., Player 1 can place two pieces on their first turn). These categories give us an additional 80 games, totaling the 121 in our dataset. We manually implemented an automated win condition checker that permits flexible game assessment over all game types.

Game name codes. Games are expressed in abbreviated form throughout the manuscript. Games are described by their board size (rows $\times$ columns) and the number K in a row to win. E.g., "$4 \times 4, 3$" means the game is played on a 4×4 grid and the first person to get 3 pieces in a row wins. Unless otherwise stated, horizontal, vertical, and diagonal all count. If only horizontal and vertical 3 in a row count, the game would be written as "$4 \times 4, 3$ HV". If only diagonal counts, then the game will be written as "$4 \times 4, 3$ D". If the constraint only applies to one player (e.g., P1 can only win horizontally and vertically, and P2 can win any way), that is represented as "$4 \times 4, 3$ (P1 HV)". If one player can go twice on their turn, e.g., the second player (P2) can play twice, that is written as "$4 \times 4, 3$ (P2 2p)". A misere game (where first to K in a

row loses) is written with an "L", e.g., "$4 \times 4, 3$ L". In our dataset, multiple rule modifications cannot co-occur, so each game can be expressed in this abbreviated way.

The full game natural language game descriptions were provided to participants (e.g., as shown in Table 1). The game codes are only used for ease of presentation in this manuscript.

Human experiments

All human experiments were conducted under prior approval from the Institutional Review Board (IRB) at the Massachusetts Institute of Technology through the Computational Cognitive Science Lab. All participants provided informed consent.

Zero-shot game outcome evaluation experiment. We recruited 238 participants from Prolific (45) to judge novel games. Each participant was randomly presented with 10 games sampled from our 121 diverse game stimuli, as well as regular Tic-Tac-Toe (won by making 3 *in a row on a* 3×3 *board*) to set baselines for game judgments. We collected approximately 20 judgments per game stimuli for each game reasoning query. Participants were paid at a base rate of \$12.5/hr with an optional bonus up to \$15/hr; the full experiment approximately took 25 minutes.

Participants were instructed to evaluate likely game outcomes. Specifically, participants produced judgments on a continuous 0 to 100 probability scale to predict the likelihood of a first player win (*if the game does not end in a draw, assuming both players play reasonably, how likely is it that the first player is going to win (not draw)?*) and a draw (*assuming both players play reasonably, how likely is the game to end in a draw?*). Judgments were made using sliders. Both game outcome question sliders appeared on the same page.

Participants produced judgments about each game based on a linguistic game specification. We additionally provided participants with an interactive scratchpad board that they were told they could, but were not required to, use to inform their judgments. The scratchpad was automatically sized to the board of the game specified; infinite boards were presented on a 13×13 grid with dashed lines indicating the board could continue. The scratchpad permitted automatically placing pieces of different colors ("red" and "blue" to simulate different players); participants could force the the next play to be made by the same player (e.g., color red twice in row) by pressing the spacebar. Two buttons appeared below the scratchpad, permitting the user to either undo their last move or clear the screen to begin a new "game". See Supplementary Information for examples of the experiment interface. Participants were required to consider each game for at least 60 seconds before being allowed to make their game judgments.

After answering all game reasoning queries, we additionally asked participants to create a new grid-based game variant that they would find fun. Participants wrote a linguistic game specification, describing the board size and win conditions. As in the game judgment queries, participants were again provided an optional scratchpad and required to spend 60 seconds before submitting a response. The scratchpad enabled participants to try out the game they intend to create. After specifying a game, participants were asked to answer the same game reasoning query (either game outcomes or game fun) about their own game. These game-generation responses were not used in this study, and are actively being explored in a follow-up study. Example screenshots of experiment interfaces are included in the Supplementary Information.

Zero-shot game funness evaluation experiment. We repeated the same methodology above with a new group of $N = 246$ participants, replacing the game outcome questions with a question about game funness. Participants instead assessed the expected funness of the game (*how fun is this game*?) on a confidence scale spanning 0 (*the least fun of this class of game*) to 100 (*the most fun of this class of game*).

Zero-shot human-human gameplay experiment. We recruited 302 participants to play these novel games in a pre-registered experiment. We selected a subset of 40 games from the full set of 121 to span a representative range of the game play variations (board shapes, board sizes, and win rules) in the original dataset while generally favoring games that would not take very long to play in a live experiment. We randomly constructed 8 batches of 5 of these games. Each participant additionally played one round of Tic-Tac-Toe. The order of games was shuffled for each new set of participants.

Participants were automatically paired with another player. We developed our interface using Empirica (1), which supports synchronous human-human pairing. Participants played one round from five different games. Players were informed they would get a bonus of $0.50 for every win. Participants had to spend at least 5 seconds reading the game description before they began. We appended "Horizontal, vertical, and diagonal all count" to all game descriptions where any direction was allowed after we noticed some participants in pilots were confused as to which line directions would result in a win. Players were randomly assigned to move either first or second and a corresponding piece color (red or blue). Players took turns making moves on the synchronous game interface. Players had no time limit on their turn.

Players were also allowed to request a draw or decide to surrender using buttons at the bottom of the interface. If a player surrendered, the game ended immediately (and that player lost). If a player requested a draw, the other player was allowed to either accept the draw (after which the game ended immediately and no player won) or reject the draw (leading the game to continue being played). Draw requests appeared as a popup banner for the other player. We include screenshots of the interface in the Supplementary Information.

The match ended when either a player won, a player surrendered, the board filled up completely (draw), or the players agreed to a draw. Both participants were informed about the game outcome. After each match, players made a judgment about either the expected outcomes of that game overall (with a new set of reasonable players) or the game's funness (in a new match against a new player). Each pair of players was randomly assigned to either the outcome or funness rating condition. Judgments were made on a slider. Players were also presented with a "frozen" version of the match on an example board with which they could replay all of the moves they and their opponent had made. Players also indicated how skilled they thought their opponent was at this game (*out of* 100 *other random new players, where do you think the opponent you just played would rank in skill for this game?*). After the judgments were made, the players continued to the next, new game. At the end of the study, they filled out a text-based survey providing general information on their strategy and how fun they found the experiment. We filtered out 18 participants who did not pass our quality control (that is, they provided judgments that were "standard" values (near 0, 50, or 100) on 80% or more of judgments) for a total of 284 subjects.

Watching and predicting play experiment. We recruited a new set of 314 participants, in a pre-registered experiment, to reason about the games zero-shot from only *indirect* experience: watching two other agents play. We selected a subset of 20 from the games from the previous human-human play study to ensure representation across game rules and dynamics. We also included Tic-Tac-Toe (totaling 21 games). Participants watched a series of videos of other agents' gameplay. Each video involved two humans playing each other, sourced from our live human-human gameplay experiment. We sampled 4 human-human played matches randomly from each of the 21 games[†], after filtering out any matches that ended preemptively from a draw request or surrender. For each match, we sampled three specific boards to be evaluated corresponding to the beginning, middle, and end of the match. For the beginning and end boards we randomly selected either the third or fourth move and the second to last or third to

[†]Due to a randomized batching error, only 3 unique matches were sampled for Tic-Tac-Toe; hence, 249 game boards over the matches from 21 games and three stages per game.

last move, respectively. For the middle board, we selected the median move. We filtered out any match that ended before eight moves. Participants watched one match from five different games, plus Tic-Tac-Toe.

Before each match, participants were informed of the game rules and required to think about the rules for 5 seconds before the video began. We again appended "Horizontal, vertical, and diagonal all count" to all game descriptions where any direction was allowed. Videos played forward at a fixed rate, as in (62). We chose two seconds per move to give viewers enough time to process each move without taking too long overall. Each video was stopped at the three time points described above. At each stopping point, participants indicated their belief over where they thought the acting player should move next. Participants were given five clicks which they could spread across the legal moves on the board to indicate their confidence that the player should move there. We chose five clicks to balance granularity of the elicited belief distribution against the burden on the participants. After each click, the opacity of the cell increased to indicate higher confidence. Participants were informed of the number of clicks they had left and could reset their clicks by clicking on a button below the interactive board.

After watching each video and indicating where they thought a player should move at each of the three timepoints, participants were then shown the remainder of the game as a board snapshot (cells indicated where players had moved and the order of play). Participants then answered either the same game outcome or funness judgments about the game overall, as described above. Judgments were made on a slider. Participants also indicated how skilled they thought each of the players was. We filtered out 10 participants who did not pass our quality control, leaving us with a total of 304 valid participants.

Intuitive Gamer model implementations

The Intuitive Gamer model consists of a *player* module that predicts individual actions in a given board state and a *reasoning* module that synthesizes simulated gameplay traces into predictions over game properties. We describe the details of each module below.

Intuitive Gamer player module. The player module functions as a one-step lookahead heuristic-based search. At a given board state s, we sweep over all legal actions a and compute the value $\tilde{\mathcal{V}}$ for each. The final action $\hat{a}$ is then sampled from the softmax distribution over each value with temperature τ. Concretely:

$$P(\hat{a} \mid s) = \frac{\exp\left(\tilde{\mathcal{V}}(s, \hat{a})/\tau\right)}{\sum_{a} \exp\left(\tilde{\mathcal{V}}(s, a)/\tau\right)}. \tag{1}$$

The value function $\tilde{\mathcal{V}}$ is computed as the linear combination of a set of three abstract utility functions: U_{self}, U_{opp}, and U_{aux} as described in the main text. These first two utilities are designed to capture the general intuition that game players aim to make progress towards their goal while preventing their opponents from doing the same. In addition, the latter (U_{aux}) encodes the "center bias"—a preference for making moves near the center of the board observed in other studies of similar games (62). We next describe our specific implementation of each utility function.

The first utility (U_{self}) computes a measure of intermediate "goal progress" of the active player, based on a feature n_1 defined as the largest line of contiguous pieces created by the action that could be extended to result in a win. This means that actions which extend a line in an illegal direction (e.g., a diagonal line in a game with only horizontal or vertical wins) or that are already blocked by an opponent's piece or the edge of the board do not contribute to goal progress at all. If the action would result in a win for the active player (i.e., $n_1 = K$) then an additional 1 is added to the feature to magnify the value of winning actions. We note that our

choice of feature only rewards actions that form *contiguous* lines of pieces—any piece which is not immediately adjacent to a previously-placed piece from the active player has $n_1 = 0$. While other formulations (such as rewarding actions that form the ends of a line that's empty in the middle or detecting particular patterns of pieces on the board) are sensible, we choose a simple and easy-to-calculate that function that reflects a straightforward intuition (i.e., making K in a row by first making $K-1$ in a row, $K-2$ in a row, and so on).

The second utility (U_{opp}) computes a measure of "progress blocked" for the opponent, based on a feature n_2 that is largely symmetric with the goal progress feature above: it is computed exactly as the goal progress the opponent *would* obtain if they made the same move (i.e. with respect to the opponent's allowed winning directions). We subtract 0.5 from n_2 to reflect people's tendency to weigh offense over defense (17), so blocking the opponent's $\hat{K}$ in a row is not as good as making a $\hat{k}$ in a row for oneself (but is better than making a $\hat{K}-1$ in a row). However, if n_2 equals the winning K, we do not subtract 0.5 in order to account for the importance of blocking winning threats. As above, there are other reasonable formulations of this feature that we leave to future work.

Finally, the third utility (U_{aux}) is computed as the normalized Euclidean distance between the position of the action and the center of the board, $\xi \in [0, 1]$. This reflects the intuition, applicable across our family of games, that people often place pieces around the center of the board. This tendency may in part be explained by the fact that placing pieces closer to the center of the board allows that piece to participate in the *most possible winning terminal states* for any *K-in-a-row* win condition.

As described in the main text, the heuristic value of a position a in board state s is computed as:

$$\tilde{\mathcal{V}}(s_t, a_t) = U_{\text{self}}(s_t, a_t) + U_{\text{opp}}(s_t, a_t) + U_{\text{aux}}(s_t, a_t). \tag{2}$$

We assume each utility function is represented as an exponentiation of the respective feature:

$$\tilde{\mathcal{V}}(s, a) = w_{\text{connect}} \times + w_{\text{block}} \times 2^{n_2} + w_{\text{center}} \times 2^{(1-\xi)}. \tag{3}$$

The choice to exponentiate some measure of progress for heuristic functions is common in gameplay modeling (e.g., 9, 51, 32). We chose base two based on light initial exploration (prior to collecting any human gameplay data), under the goal of keeping our intuitive model simple. Future work can explore other design choices for encoding and combining multiple utility functions in both a game-general and game-specific manner.

We fix temperature (τ) at 1 for our game reasoning and action experiments (with the exception of marginalizing over τ only for the admixture analyses in the "play" experiment). We set weights of each component (w) to 1 for all experiments. We use the same settings for the Expert Gamer model (as it uses the same value function and softmax-based action selection). We find that equal weights is a reasonable fit for the Intuitive Gamer model to human payoff predictions (see Supplementary Information). When lesioning a component of the value function, we set its weight to zero.

Intuitive Gamer reasoning module. The Intuitive Gamer player module is queried as part of the game reasoning module. For each game G, k game simulations are sampled, producing a set of k outcomes $o \in 1, 0, -1$ encoding whether the first player won (1), lost (-1), or the game ended in a draw (0). From these outcomes, we can compute a payoff of a game G:

$$\psi(G) = (1)P(\text{win} \mid \neg\text{draw}) \cdot P(\neg\text{draw}) + (-1) \cdot P(\text{lose} \mid \neg\text{draw}) \cdot P(\neg\text{draw}) + (0) \cdot P(\text{draw}). \tag{4}$$

$$\psi(G) = P(\neg\text{draw}) \cdot (P(\text{win} \mid \neg\text{draw}) - P(\text{lose} \mid \neg\text{draw}))\,. \tag{5}$$

Our game reasoning module is constructed to represent the reasoning of any one participant. Therefore, we draw k simulated matches for $N = 20$ simulated participants (as each game has, generally, 20 participant responses; see "Analysis methods" below for details on selecting k).

Partial game simulations. Our primary game reasoning module assumes mental simulations are conducted until the end of the game is reached: either a player attains the win condition, or the all open board positions are filled and the game is called a draw. While many of our games do not take many moves to reach an end, people may not mentally simulate all the way to the end of the game in each of their k simulations. We conduct a preliminary exploration into the impact of partial game simulations in fitting peoples' game fairness evaluations by modeling the possibility that people halt any one of their k simulations early. We sample from a uniform distribution over the board size and ending early if the game happens to halt before reaching that time. Early stopping requires determining how to assign the outcome. We consider a simple rule of assigning any early-stopped game to a draw (allowing us to explore encoding a weak prior towards games ending in draws). We repeat the same exploration of variance (see "Analysis methods" below) and find that the optimal number of samples is similarly more than one and less than ten, but may be closer to $k = 4$ (see Extended Data Figure 2). Understanding what distributional form over stopping times and how people decide what outcome to call for a game that did not reach termination are important questions for future work.

Funness model. To estimate a game's funness, we considered three features derived from playouts under the Intuitive Gamer player module: balance, reward for thinking, and game length. We measure game balance using the Earth Mover's Distance (47) between the observed outcome distribution (i.e. under simulations of the Intuitive Gamer model against itself) and an "ideal" distribution in which the first and second player each win half of the games and never draw—the larger this distance, the lower the value of the balance feature.

We make the assumption that in a fun game the two players should have an equal chance of winning and draw rarely happens, consistent with prior work (2, 6, 58). Reward for thinking is obtained from the expected win rate of the Intuitive Gamer model against a random gameplay agent. The intuition is that in a fun game *thinking* should matter, and thus an agent that performs computations should soundly defeat one that does not (13, 58). Finally, we measure game length using the same Intuitive Gamer-vs-Intuitive Gamer simulations used for the balance feature. Expected game length reflects the intuition that games that are too short or too long are not fun, which we encode with a quadratic term, as outlined below. We use 120 game simulations for each feature (to align with the $k = 6$ simulations for each of 20 participants), randomizing whether the Intuitive Gamer plays first against the random agent in the simulations used to assess reward for thinking. Game simulations are run to the end of the game (where either a player wins or the game ends in a draw); future work can explore variants of the funness model where features are computed under partial game simulations. We use the same features and number of simulations when comparing to alternate models (see below).

Alternate models

We next detail several alternate models.

Expert Gamer. We compare our Intuitive Gamer model against an "Expert Gamer" model that differs along two key dimensions: (1) sophistication of the value function, and (2) depth of search. This model is closely based on the model of human expertise for 4-in-a-row on a 4×9

board in (62), which empirically estimates tree search depth from human players after hours of continuous gameplay experience. As in van Opheusden et al., the depth is controlled by a probabilistic "stopping parameter" governing how many iterations over which they expand nodes in the search tree; we ran all simulations by expanding the search tree a fixed number ($k_{iterations} = 636$) of iterations, where $k_{iterations} = 636$ is the empirical mean value of this parameter estimated from the (62) data. This setting corresponds to approximately depth-5 search. On each iteration, the Expert Gamer selects a node (corresponding to a board state). We then expand this node, considering all playable pieces, wherein we compute the value of each piece as outlined below. The Expert Gamer model then probabilistically expands nodes in the search tree by repeatedly sampling actions from a softmax policy governed by temperature τ over its current action estimates based on these computations, expanding unexplored states in the search tree, and backpropagating utilities estimated at future states. As in van Opheusden et al., any node that ends in a definite win or loss is assigned $\pm\sigma$ based on whether the move would result in a win for the current play ($+$) or loss ($-$). We set $\sigma = 1000$.

We next detail the Expert Gamer value function. As in van Opheusden et al., we compute the value of any move by looking at features over the *entire* board state, rather than locally circumspect around the possible move position in question (like the Intuitive Gamer model). This is more computationally-intensive, particularly for larger boards. Specifically, for any open legal position p, we imagine a state s' that has that position played by the current player. We then sweep over all played positions with that board state. For each played position p', we compute the same novice, partial-progress value function. We then define the value of that state s' as the difference in cumulative value from the played positions by the current player minus the opponent. We set the value function feature weights (w_{connect}, w_{block}, w_{center}) using the fit values from the Intuitive Gamer model. It is worth noting that the Expert Gamer is itself an important contribution—it is a more general version of a relatively deep goal-directed model. We leave validation of how well the Expert Gamer captures human expert reasoning and play for future work.

Random. The Random gameplay model selects actions uniformly over the space of legal moves. Games are simulated to the end (e.g., til either player wins or the game ends in a draw).

Monte Carlo Tree Search. We additionally implement a standard upper-confidence-bound Monte Carlo Tree Search (MCTS) (16, 20) agent to act as an approximate gameplay "oracle" (pseudocode can be found in the Supplementary Information). Unlike the Intuitive Game and Expert Gamer models, MCTS does not use game-specific heuristic features and instead estimates intermediate utilities by expanding a search tree guided by repeated random rollouts to terminal states. MCTS algorithms are commonly used to approximate optimal gameplay in arbitrary games (20), as they are empirically both efficient and accurate. Here, we use a large number of tree expansions (1000) to model the behavior of a near-perfect player in our novel games. Due to computational costs, we estimate the expected payoffs using 50 simulations per game; these samples are then bootstrapped in their fit to people, as with the other models.

Lesions of the full Intuitive Gamer model. We compare the full Intuitive Gamer model against several variants, that modulate whether it is flat, goal-directed, or probabilistic. The non-flat version of the Intuitive Gamer model performs a deeper search over possible moves when selecting actions. The non-goal-directed version of the Intuitive Gamer model ablates one or more of the value function components (i.e. player progress or blocking progress—see Supplement for ablations of the center-bias component). The non-probabilistic version of the Intuitive Gamer model ablates the softmax element of action selection (i.e., selects with temperature zero). We vary fastness by modulating the number of simulations drawn (k) from

the gamer model, over which the game reasoning engine computes the expected payoff (see Extended Data Figure 1).

Game-theoretic optimal payoffs. Many of the games in our set have known game-theoretic ground truth outcome values assuming optimal play from both players. We describe how we filter which games of the 121 are estimatable via a game-theoretic optimal payoff. We first include those with known game-theoretic optimal payoffs (either definite Player 1 win, definite draw, or definite Player 2 win) according to (60). Then, to approximate a larger set of games' optimal payoffs, we additionally include games for which MCTS either converged to $\{-1, 0, 1\}$ in its payoff predictions (noting that MCTS estimates are not guaranteed to be perfect). This process results in a set of 78 out of the full 121 games for which we have estimated game-theoretic payoffs.

Analysis methods

We next provide additional details on experimental analyses.

Computing and comparing payoff predictions. Participants answered two questions in the game outcome reasoning experiment: how likely the first player wins given the game does not end in a draw ($P(\text{P1 wins} \mid \text{not draw})$), and how likely the game is to end in a draw ($P(\text{draw})$). Together, these responses yield all information we need to compute an expected payoff. We compare the expected payoff from participants against those of models and report the R^2. Specifically, we bootstrap subsample with replacement from the human participant samples, per game, and compare models against the mean payoff per sample. We also bootstrap subsample k for 20 simulated participants per model from the pool of model simulations. Unless otherwise stated, we run 1000 bootstrap samples.

Sensitivity analyses into the number of samples k in the reasoning module. We choose k (the number of simulations sampled for each simulated participant) by inspecting the variance of the payoff predictions over the simulated set of 20 participants against the empirical variance observed in the human data. We measure the Root Mean Squared Error (RMSE) between the model- and human-predicted variance for each game, as well as the Wasserstein Distance between the distribution over model and human variances (Figure 3c). We compute the latter to better capture the poor match of high k to people's variances (i.e., with increased k, variance collapses to zero). We find that that generally approximately $k = 5 - 7$ full game simulations reasonably well-captures human variance (see Figure 3c and Extended Data Figure 1) under both measures. We report all main results with $k = 6$.

Fitting regression models to funness judgments. We fit a linear regression model to these features using the `lmer` package in R. Features are normalized to zero-mean and unit standard deviation. We fit the models to 1000 bootstrapped subsamples of the human data for all games. Models are tasked with predicting the mean of participants' funness judgments for each game. Additionally, to assess generalization, we conduct a separate analysis where we fit to 50% of the games and test on 50% of the games. We report both results in Figure 4 and include additional details in the Supplementary Information. We assess potential non-linearities in the features by adding a quadratic component. We compare models via an ANOVA test and AIC in the Supplementary Information.

Assessing contribution of non-simulation-based features to funness judgments. In exploratory analyses, we assess the potential role of non-simulation-based features that can be

read off of the game description alone. We compute a series of binary game traits that capture ways in which a game may differ from the base Tic-Tac-Toe. For instance, a game may not be a 3×3 board; the game may end with $K \neq 3$ pieces in a row to win; the second player may have a different win condition than the first player. We consider the following binary features: if the game has constrained win conditions (e.g., no diagonals allowed); if the game has asymmetric win conditions (between Player 1 and 2); if the game has asymmetric play dynamics (e.g., Player 2 can place two pieces on their first turn); if the board is not square; $K \neq 3$ pieces in a row to win; if the board is larger than 3×3; if the game ends when the first player to achieve K in a row loses (i.e. misere variants). We combine the binary features into a single aggregated measure of "approximate novelty," which is the sum of the number of binary features present in any one game. We also explore the incorporation of board size (encoded as the number of rows $\times$ number of columns) as part of non-simulation-based game features that reasoners may consider when assessing the funness of games. We assess how well participants' funness judgments can be explained by these features alone by fitting bootstrapped sets of linear regression models to these features in the same $50/50$ train/test splits over the games. Additionally, we explore the relative benefit of adding all binary features or any of the aggregate features to the simulation-based model over the full set of games (see Supplementary Information). Additional analyses are included in the Supplementary Information.

Assessing game-time action selections. To assess how well models capture how people play, we condition the model in question on an intermediate board state of the actual human-versus-human game. We computed the model's predicted distribution over next moves. We assess the fit of this predicted distribution in two ways. First, we compute the log likelihood of the actual players' moves against those predicted by the model. To handle cases where any model may place at or near zero probability on any given position, thereby skewing the log likelihood, we incorporate a "slop" parameter α which captures the probability that a player makes a move from the primary model ($1 - \alpha$ that they make a random move on that turn (40)). We sweep over $\alpha \in \{0.5, 0.6, 0.7, 0.8, 0.9, 0.9999\}$ and compute the average log likelihood for each model over the settings of α. We show per-game stage and per-game results in the Supplementary Information. We conduct a one-sided paired t-test over each individual move as well as the aggregated move likelihood per game, comparing the Intuitive Gamer to the Expert Gamer and Intuitive Gamer to random, respectively.

Fitting admixture models to action selections. We then consider two different admixture models: one over all moves made for a game, and one over all moves made for each player. Admixture models captures the notion that a player may play according to either the Intuitive Gamer model, Expert Gamer, or Random on any turn. We jointly fit the weights of the Intuitive Gamer and Expert Gamer model and take the weight of random to be the remaining mass that brings the weights to 1, over bootstrapped samples over the entire population of players. As the weights are fit jointly across the models, we need to be particularly careful about the magnitude the results for the respective models. To that end, we jointly sweep over the temperature used in the action selection (with temperature ranging from 0.5 to 3.5 to produce a distribution over weights). We also fit the admixture over each individual, estimating the relative contribution of each model to the way they play. As each player only plays six games (and one round per game), we fit these weights to all of their moves across all games. We use `scikit-learn` for fitting, with the SLSQP optimizer.

Computing aggregate human distribution over predicted next action. For our "watch-and-predict" experiments, we have access to finer-grained predictions from any *one* participant. We again extract distributions over likely next moves, conditioned on the observed intermediate board state from the novice and alternate models; however, we now compare the distributions

directly to the suggested distribution of moves made by participants. We construct an aggregate move distribution over the participants, treating each of the 5 clicks per participant as contributing some mass to the "aggregate human" distribution.

Comparing model- to human-predicted distributions over next actions. We compare distributions using the Total Variation Distance (TVD); we fix temperature at 1 and sweep over α for both models and people. We conduct a similar one-sided paired t-test over the TVD across the three game stage boards seen per match and compare the Intuitive Gamer and Expert Gamer, and Intuitive Gamer and random. We compute split-half TVDs per stage, per match, per game by randomly splitting the participants' distributions who saw that board, averaging those boards into two aggregate distributions and computing TVD between them. This forms our split-half human estimate which we use to normalize the other models' TVD. We demonstrate the robustness of our results to our choice of distributional measure by repeating the analyses with the Jensen-Shannon Divergence in the Supplementary Information. We additionally run an admixture model over both participant- and game-level predicted watch distributions. These are optimized against the Jensen-Shannon Divergence.

Comparing model- to human-predicted distributions over next actions to people's played actions. We secondarily assess how well the distribution over suggested moves made by the watcher captures how people actually played by treating the suggested distribution of clicks by the human watcher akin to the move distributions from the models. We again assess the likelihood of the move played by the active player to the moves suggested by the predictors, the Intuitive Gamer model, and alternate models by sweeping over and averaging out a slippage parameter (α). We conduct these analyses in aggregate (over all board stages, matches, and games), as well as at a per game level (over board stages and matches).

Modeling draw requests. We take initial steps to analyze evaluations made during the game by looking at players' decisions of whether to accept or reject a draw, when offered. Draw requests could take place at any point during the game. For each board where a draw request was made, we run 40 simulation to the end of the game under the Intuitive Gamer agent model. We compute the expected payoff over bootstrapped subsampled sets of $k = 6$ outcomes (simulating a single player; previously computed under one of our $\psi(G)$ game reasoning queries), as well as the expected remaining length of the game ℓ from state s_t. We compute expected payoff here with respect to the player who is deciding whether or not to accept the draw request. Players receive a bonus payout of $\$0.5$ only if they win (nothing if the lose or draw); we can then compute an expected value that is the payoff $\times 0.5$. The expected payoff and length together define an "expected value of continuing" (C) from state s_t:

$$\mathrm{C}(s_t) = \beta_{\text{utility}} \cdot (\psi(G) * 0.5) - \beta_{\text{length}} \cdot \ell. \tag{6}$$

For each set of bonus-adjusted expected payoffs and expected game lengths per game, we fit a logistic regression model to predict whether a player would accept or reject the draw request, based on expected payoff and length, as well as the average predicted game funness predicted by people in the "just think" game reasoning experiment. We fit the logistic regression model using the `glm` R package. We also explore the same computations under different agent models varying in depth (e.g., the Expert Gamer model) and goal-directedness of the value function (e.g., ablating the goal progress component). We show the decision boundaries in Extended Data Figure 6 and report the bootstrapped parameter fits in Extended Data Table 4.

Acknowledgments

We thank Max Kleiman-Weiner, Laura Schulz, Rebecca Saxe, Mira Bernstein, Scott Stern, Ilia Sucholutsky, Lance Ying, Junyi Chu, Tracey Mills, Jacob Andreas, Tyler Brooke-Wilson, Judy Fan, Yuka Machino, Sydney Levine, Ionatan Kuperwajs, Umang Bhatt, Gabriel Poesia, and Ben Prystawski for comments on the manuscript or helpful conversations that informed this work. Thank you to the family and friends of the authors for many fun game nights. This work was supported by AFOSR (FA9550-22-1-0387), the ONR Science of AI program (N00014-23-1-2355), a Schmidt AI2050 Fellowship to JBT, and the Siegel Family Quest for Intelligence at MIT. KMC acknowledges support from the Cambridge Trust and King's College Cambridge. LCW acknowledges support from a Stanford HAI Fellowship. GT acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant DGE-2234660. TLG acknowledges support from the NOMIS Foundation. AW acknowledges support from a Turing AI Fellowship under grant EP/V025279/1, The Alan Turing Institute, and the Leverhulme Trust via CFI. This work is supported (in part) by ELSA - European Lighthouse on Secure and Safe AI funded by the European Union under grant agreement No. 101070617. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission.

Category	Example(s)	Count
M in a row on square boards	*3 pieces in a row wins on a 6×6 board; 7 pieces in a row wins on a 10×10 board*	20
M in a row on rectangular boards	*3 pieces in a row wins on a 1×5 board; 6 pieces in a row wins on a 5×10 board*	18
Infinite boards	*3 pieces in a row wins on an infinite board; 10 pieces in a row wins on an infinite board*	3
M in a row loses	*A player loses if they make 5 pieces in a row, on a 5×5 board.*	10
No diagonal wins allowed	*4 pieces in a row wins on a 10×10 board, but a player cannot win by making a diagonal row.*	10
Only diagonal wins allowed	*4 pieces in a row wins on a 5×5 board, but a player can only win by making a diagonal row.*	10
First player moves 2 pieces	*5 pieces in a row wins on a 10×10 board; the first player can place 2 pieces as their first move.*	10
Second player moves 2 pieces	*10 pieces in a row wins on a 10×10 board; the second player can place 2 pieces as their first move.*	10
First player handicap (P1 no diag)	*3 pieces in a row wins on a 3×3 board, but the first player cannot win by making a diagonal row.*	10
First player handicap (P1 only diag)	*4 pieces in a row wins on a 7×7 board, but the first player can only win by making a diagonal row.*	10
Second player needs M-1 to win	*The first player needs 4 pieces in a row, but the second player only needs 3 pieces to win on a 5×5 board.*	10

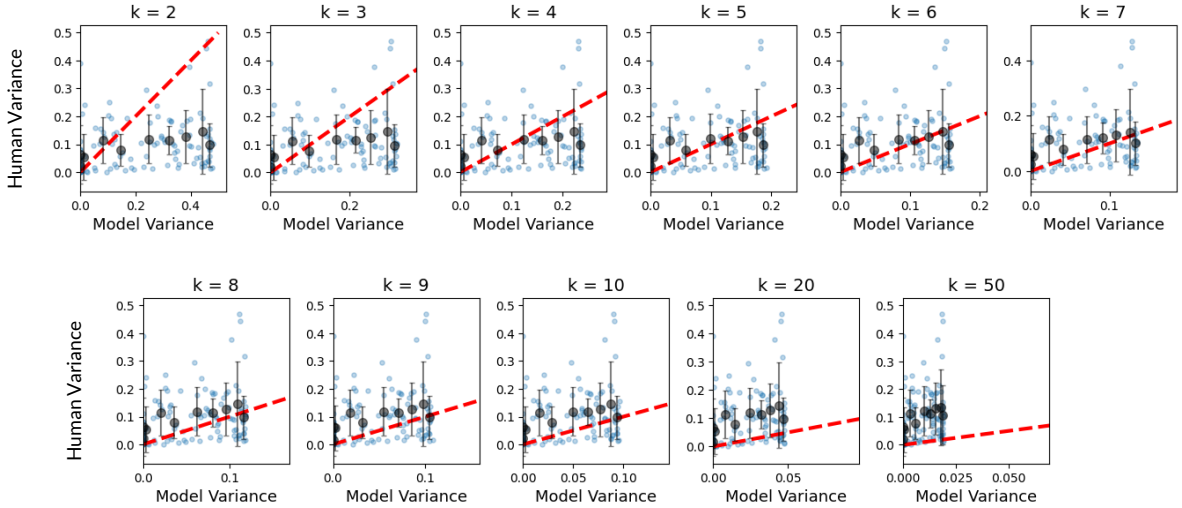
Extended Data Table 1: Diverse set of novel games. Wide variety of 121 two-player grid-based strategy games designed to study how people play and think about games when they are novices. Games vary in the board shape (e.g., square versus rectangular), win conditions (e.g., first to K in a row wins versus loses), and game dynamics (e.g., one player can play two pieces on their turn). Counts of games in each category, as well as one representative example game description for each.

Model	Time per Action (s)	States Evaluated per Action
Intuitive Gamer	0.0089 ± 0.0113	57.28 ± 33.18
Expert Gamer	6.48 ± 9.33	$31,417.80 \pm 22,145.58$
MCTS	342.40 ± 380.55	$523,297.22 \pm 348,852.03$

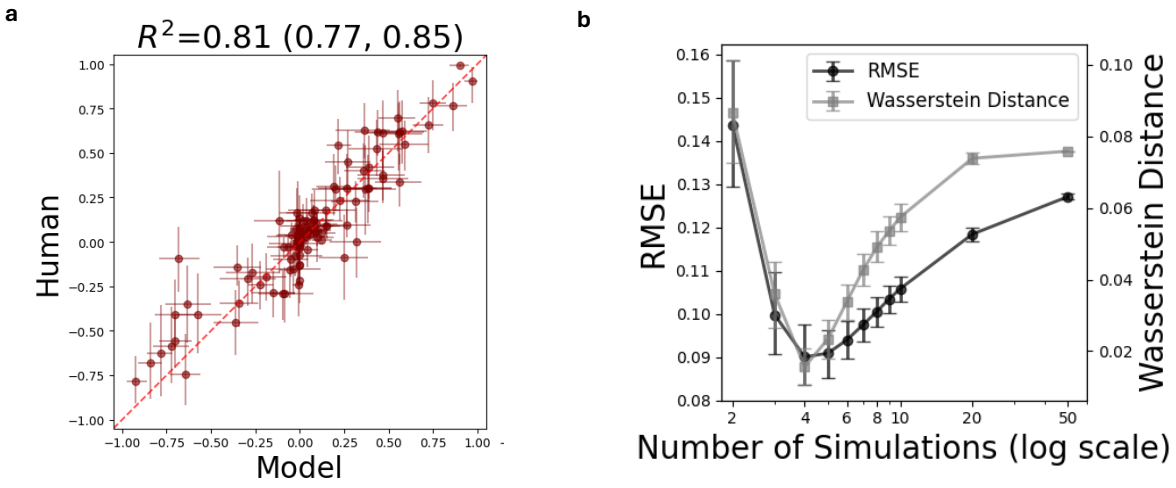
Extended Data Table 2: Efficiency of different game reasoning models. The Intuitive Gamer is more efficient than more sophisticated game reasoners, both in terms of average time to evaluate before deciding what action to select and average number of game states to evaluate before deciding what action to select. represent mean $\pm$ standard deviation over all moves when simulating evaluations for the game reasoning payoff predictions, for a subset of 41 of the 121 games. Additional details on cost comparisons are included in the Supplementary Information.

Reasoner	Acc to Game-Theoretic (GT)	R^2 to GT	Abs Diff from GT
Human	0.69 (0.65, 0.73)	0.62 (0.58, 0.67)	0.32 (0.31, 0.34)
Intuitive Gamer	0.75 (0.72, 0.78)	0.69 (0.66, 0.72)	0.25 (0.24, 0.26)
Expert Gamer	0.92 (0.91, 0.92)	0.87 (0.85, 0.88)	0.08 (0.08, 0.09)
MCTS	0.91 (0.90, 0.92)	0.89 (0.88, 0.91)	0.06 (0.06, 0.07)
Random	0.57 (0.55, 0.59)	0.39 (0.34, 0.44)	0.43 (0.41, 0.44)

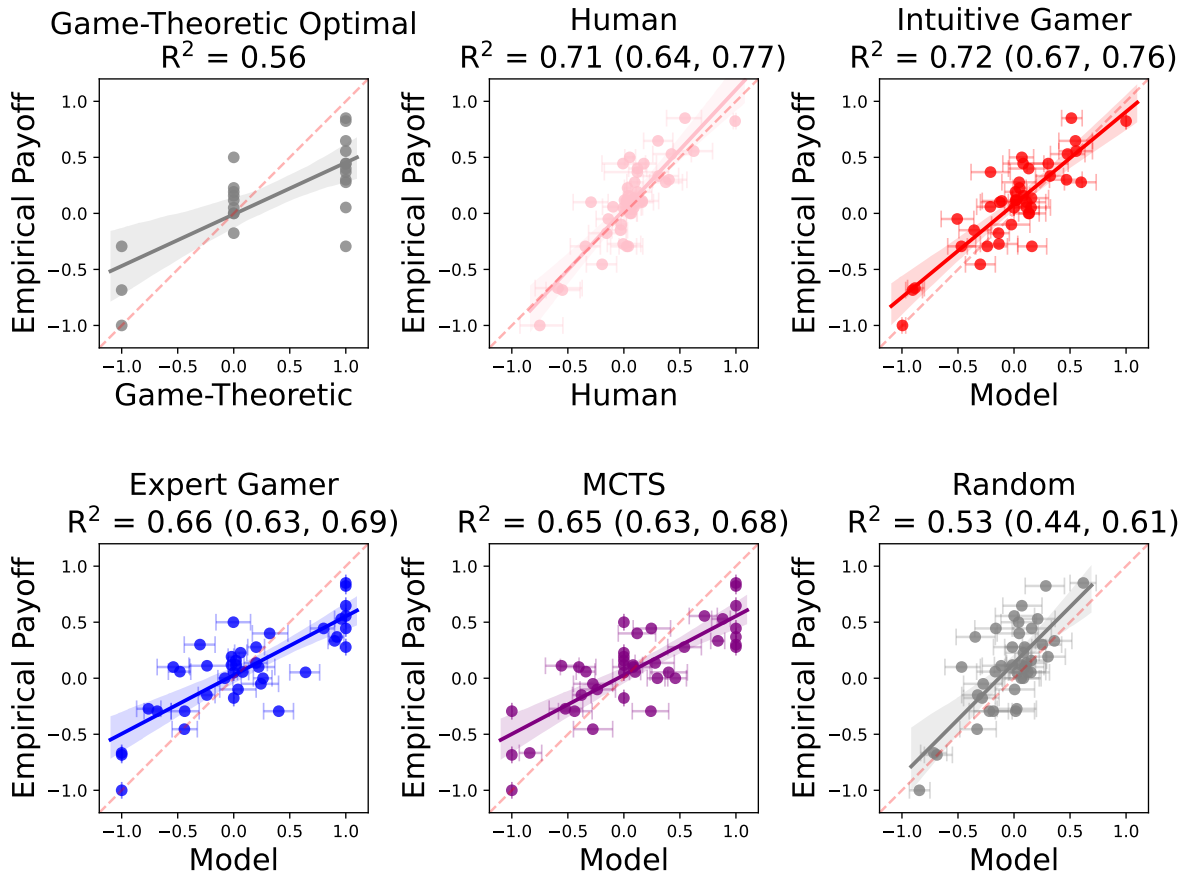
Extended Data Table 3: Model and human predictions relative to the estimated game-theoretic optimal payoff. Human and model judgments are compared to the 78 of the 121 games where the game-theoretic optimal payoff is estimatable. Accuracy between predicted payoff and the approximate game-theoretic optimal is computed by taking the predicted payoff as "correct" if the game reasoner predicted a payoff > 0.5 and the expected payoff is 1; correct if the predicted payoff is < -0.5 and the expected payoff is -1; correct if the predicted payoff is between -0.5 and 0.5 and the game is expected to end in a draw. R^2 correlation is computed between the raw predicted payoffs and the game-theoretic optimal values, as well as the average absolute difference between the expected predicted payoff and approximate game-theoretic payoff (lower is closer to "correct"). Bootstrap 95% confidence intervals (CIs) are shown in parentheses, where bootstraps are over bootstrapped samples of participants (with replacement) for people or model simulations.



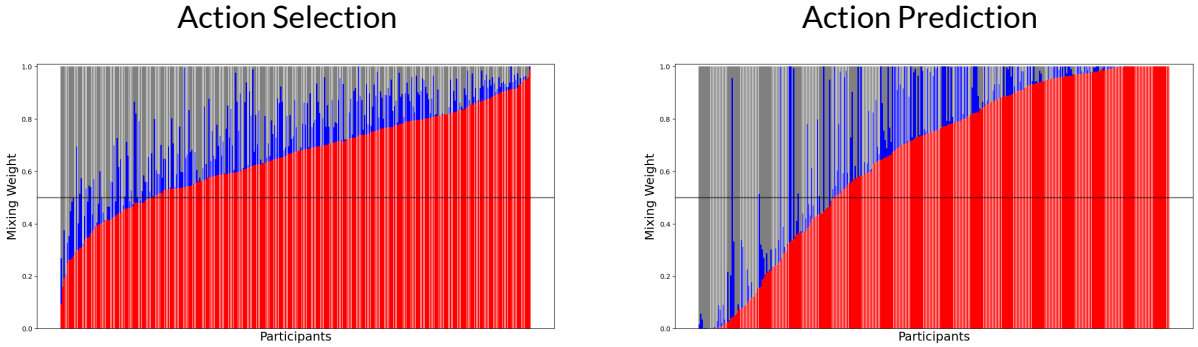
Extended Data Figure 1: Variance of human and model payoff predictions with varied compute budget (number of simulations, k) for the Intuitive Gamer model). The variance over each approximately 20 human participant payoff judgments per game and compare against the average variance under simulated sets of $N = 20$ simulated participants each drawing k simulations from the model. Model variance is binned into quintiles along the horizontal axis and the average human variance is computed for points in that bin (the black dot overlayed on the scatterplot). Error bars are standard deviation of the variance for points in that bin.



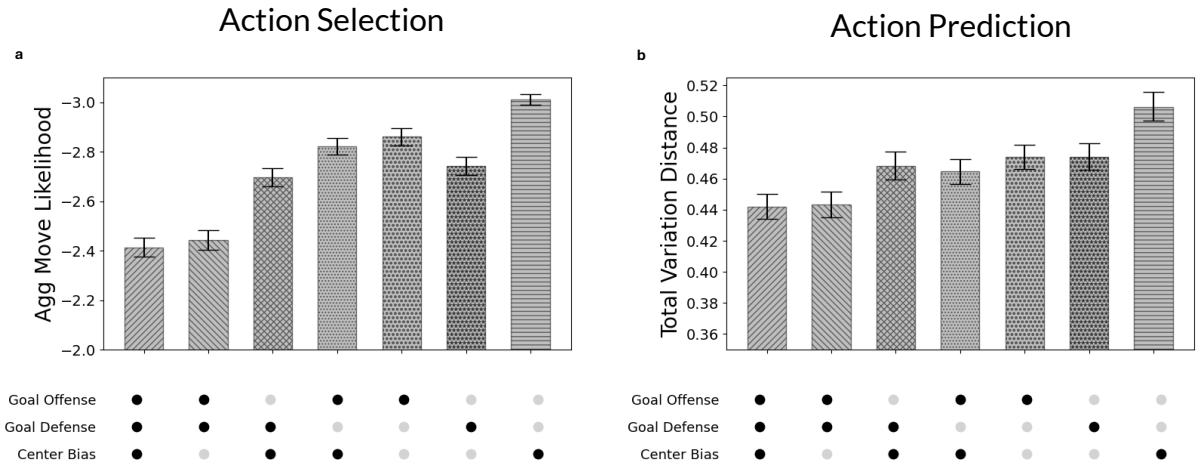
Extended Data Figure 2: Partial game simulations in the Intuitive Gamer reasoning module. All main results are reported under the assumption that the game reasoning module runs simulations to the end. While most games do not involve a large number of moves, under the Intuitive Gamer player module, for the game to end—it is possible that people are engaging even cheaper partial simulations when evaluating novel games for the first time. **a,** Exploratory analyses into running the Intuitive Gamer reasoning module under a probabilistic stopping rule equivalently captures the majority of the explainable variance in human payoff judgments as running the Intuitive Gamer reasoning module under full simulations (see Figure 3a). For direct comparison with the full simulation model, $k = 6$ partial simulations were run to estimate game payoff. The partial simulations were computed as follows. For each simulated match of a game, a stopping time was sampled from a uniform distribution over board size and terminated early if the game did not end before that time. If the game terminated before a winner or draw was called, the game counted as a draw. **b,** To assess the impact of compute budget in the number of simulations used to estimate payoff, the variance of payoff under different number of partial simulations k was again compared with the variance in people's payoff evaluations (as in Figure 3c and Extended Data Figure 1). The best fitting number of samples (k) is similarly greater than one and less than ten when using partial simulations, and perhaps even a bit less (e.g., $k = 4$) compared to using full game simulations (see Figure 3c).



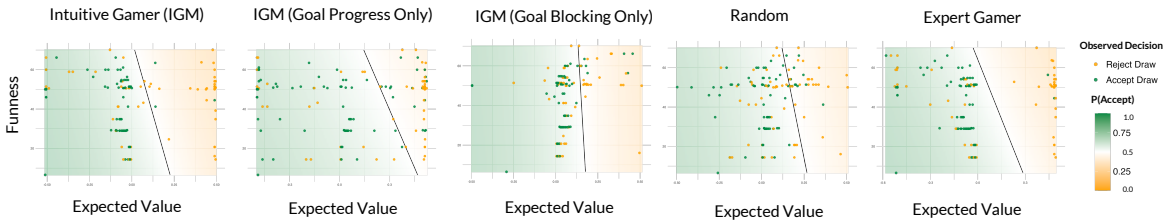
Extended Data Figure 3: Empirical observed payoffs in actual human-human play against human- and model-predicted payoffs. The empirical payoffs (based on the observed human-human play game outcomes) generally follow the payoffs predicted by people and the Intuitive Gamer model in the "just think" experiment (without any live play). The empirical payoffs do not align with what one would expect under optimal play, nor play under relatively more sophisticated expert models (MCTS or the Expert Gamer) or game agents that use less compute (random).



Extended Data Figure 4: Modeling individual game players. A probabilistic mixture model (admixture) is fit for each individual participants' action choices (left) and action predictions (right). Bar heights represent the estimated contribution of each model in the probabilistic mixture. The Intuitive Gamer is the dominant component for most participants; however, there is variability across individual participants' behavior—some participants make action choices more aligned with a Random or more sophisticated Expert Gamer model.



Extended Data Figure 5: Heuristic value function lesions. To assess the importance of components of the Intuitive Gamer model to fitting participant data, key features of the full Intuitive Gamer value function $\tilde{\mathcal{V}}$ are lesioned by setting the component weight to zero and leaving the others fixed at one. **a,** Aggregate log probability under each variant of the human played moves from the "play" (action selection) experiment; **b,** aggregate TVD comparing the model and human predictor distributions per game stage made in the "watch-and-predict" experiment. Error bars depict bootstrapped 95% CIs over the mean measure per setting. Ablating the connect weight or the block weight has major effects on how well the model explains human data in all three experimental conditions. However, ablating the center weight only has effects in the play experiment. The lack of impact of the center bias in the "watch-and-predict" data may arise from the center bias only mattering in the first moves (and participants are only asked to provide judgments over the third move onwards for the "watch-and-predict" task).



Extended Data Figure 6: Modeling people's decisions about whether to keep playing when they receive a draw request. Logistic regression models predict whether a player would accept or reject a draw request if received are fit to the expected value of continuing to play under the respective player module variants, and the expected funness of games (as predicted by people in the just-think experiments). Decision surfaces are plotted, averaged over 100 bootstrapped samples of expected value calculations over $k = 6$ simulations. The Intuitive Gamer and Expert Gamer qualitatively encode the intuition that games that have higher expected value from continuing or are expected to be fun are games where players are more open to continuing to play.

Variant	Avg Log Lik	Utility	Game Length	Funness
Intuitive Gamer	-0.56 (-0.59, -0.54)	-4.08 (-4.81, -3.37)	0.023 (0.010, 0.034)	-0.0135 (-0.0159, -0.0102)
Only Offense	-0.61 (-0.62, -0.60)	-2.14 (-2.33, -1.91)	-0.011 (-0.018, -0.002)	-0.0109 (-0.0127, -0.0090)
Only Defense	-0.65 (-0.66, -0.62)	-4.08 (-8.36, -1.69)	-0.016 (-0.018, -0.015)	-0.0028 (-0.0048, -0.0001)
Random	-0.63 (-0.66, -0.60)	-2.99 (-4.65, -1.55)	-0.009 (-0.018, 0.003)	-0.0069 (-0.0116, -0.0010)
Expert	-0.58 (-0.59, -0.56)	-3.31 (-3.73, -2.87)	0.027 (0.018, 0.034)	-0.0164 (-0.0181, -0.0146)

Extended Data Table 4: Draw request model parameter fits. Coefficients for draw request models, depicting 95% CIs over bootstrapped subsamples of $k = 6$ game simulations. The actions taken by a player (whether to accept or reject a draw) are assessed under the models by computing the log likelihood of doing the action, given the features computed over the particular board. The model was trained to predict whether a draw request would be accepted; therefore, negative weights for features mean that a higher value of that feature is more predictive of rejecting a draw (i.e., wanting to continue a game). Log likelihood is computed as the averaged likelihood assigned to each decision and averaged over bootstrapped model fits. Higher likelihood means better fit. Using the expected value of continuing as computed under Intuitive Gamer and deeper Expert Gamer simulations similarly explain people's decisions of whether to accept or reject a draw. Lesioning either or both of the goal-directed components of the Intuitive Gamer value function substantially impairs fit to people's decisions, particularly lesionining the offensive component. This may be because, upon receiving a draw request, a player decides whether to accept or reject from a more offensive stance.

Supplementary Information

2 Model pseudocode

We include pseudocode for our primary and alternate models.

2.1 Intuitive Gamer player module

Algorithm 1 Intuitive Gamer: MakeMove(s)

$n_1 \leftarrow$length of longest *live* connection extended
if n_1 is winning length **then**
  $n_1 = n_1 + 1$
end if
$n_2 \leftarrow$length of longest *live* opponent connection blocked $- 0.5$
if n_2 is winning length **then**
  $n_2 = n_2 + 0.5$
end if
$\epsilon \leftarrow$ Euclidean distance to center of board, normalized so $d = 1$ for the corner pieces
$\tilde{\mathcal{V}}(a, s) = 2^{(1-\epsilon(a))} + 2^{n_1(a)} + 2^{n_2(a)}$
return Sample(Softmax($\{a : \tilde{\mathcal{V}}(a, s)$ for $a \in s.actions\}$)

As explained in our main text, our Intuitive Gamer player module is flat, goal-directed, and probabilistic. It incorporates shallow search and considers only a constrained set of heuristics ($\tilde{\mathcal{V}}$) for measuring intermediate game states. Actions are evaluated based on their proximity to the center of the board d, the length of the live connections they extend n_1, and the length of the opponent's live connections that they block. By *live connections* we mean connected pieces that have enough open positions on the corresponding direction such that they can be extended to form a winning configuration. Actions are then sampled from a softmax distribution over open-board states, weighted by the heuristic value computations.

2.2 Expert Gamer model

We provide pseudocode for the Expert Gamer model (designed to generalize the 4-in-a-row model of human gameplay from (62)). Its algorithm repeats three sub-procedures, SelectNode, ExpandNode, and Backpropagate, within the main procedure MakeMove, which ultimately makes a move by sampling from a softmax distribution.

Algorithm 2 Expert Gamer: MakeMove

$root \leftarrow$ Node(s)
for $i = 1$ to num_steps **do**
  $n \leftarrow$ SelectNode($root$)
  ExpandNode(n)
  Backpropagate($n, root$)
end for
return Sample(softmax($\{c : c.val$ for $c \in root.children\}$)

The function Node is the constructor for the nodes of the tree search. As described in the "Methods" section, in our simulations we repeat the procedure in MakeMove for $num\_steps = 636$ iterations, drawing on results from an empirical parameter fit in (62) which found that expert human players (playing a specific 4-in-a-row variant) were well modeled by a heuristic search procedure whose number of iterations is geometrically distributed with a mean stopping parameter of $1/636$. Our model does not use a stochastic stopping parameter and just runs

deterministically to $num\_steps = 636$ iterations for all simulations. In the following, *player_type* $\in \{X, O\}$ and *move_number* is the index of the current move over the entire game (to account for variants which specify that a given player goes twice as their opening move).

Algorithm 3 Expert Gamer: SelectNode(n)

```
root ← n
while n.children ≠ ∅ do
    if rand()< ϵ then
        n = unirandom({c : c.val for c ∈ n.children})
    else
        if n.player_type = root.player_type then
            n = rand-argmax({c : c.val for c ∈ n.children})
        else
            n = rand-argmin({c : −c.val for c ∈ n.children})
        end if
    end if
end while
return n
```

Algorithm 4 Expert Gamer: ExpandNode(n)

```
s ← n.state
for all m ∈ LegalMoves(s) do
    b ← PlacePiece(s, m, n.player_type)
    i ← n.move_number
    n.children.append(Node(b, i + 1, Ṽ^EG(s))
end for
```

Within the procedure ExpandNode, the function $\tilde{\mathcal{V}}^{\text{EG}}$ is an extension of our Intuitive Gamer heuristic function $\tilde{\mathcal{V}}$ defined above, and it represents a more sophisticated heuristic. To set a frame of reference, suppose player X is evaluating the utility of board state b. The heuristic $\tilde{\mathcal{V}}^{\text{EG}}$**(b)** is evaluated as

$$\tilde{\mathcal{V}}^{\text{EG}}(s) = \sum_i (2 \times \mathbb{I}\{\text{the } i\text{'th move is done by } X\} - 1)\tilde{\mathcal{V}}(p_i, s_i). \tag{7}$$

where b_i is the i'th board observed in the game and p_i is the i'th move played in the game. The value function is defined symmetrically when O acts. We make this modification so that $\tilde{\mathcal{V}}^{\text{EG}}$ incorporates information from all potential future actions on the board, unlike $\tilde{\mathcal{V}}$ which only incorporates local information about an action-state pair (p, b).

Algorithm 5 Expert Gamer: Backpropagate($n, root$)

```
if n.player_type = root.player_type then
    n.val ← max_{c∈n.children} c.val
else
    n.val ← min_{c∈n.children} c.val
end if
```

2.3 Monte Carlo Tree Search

Similar to the "Expert Gamer" model, Monte Carlo Tree Search (MCTS) repeats four sub-procedures: SelectNode, ExpandNode, DepthCharge and Backpropagate, but the SelectNode,

ExpandNode, and Backpropagate sub-procedures are different from their corresponding counterparts in the human expert model. We run MCTS for $10,000$ steps, which we find balances a close approximation of the optimal for many games while keeping the runtime of any single game trial to less than two days.

Algorithm 6 MCTS: MakeMove

$root \leftarrow$ Node(s)
for $i = 1$ to num_steps **do**
  $n \leftarrow$ SelectNode$(root)$
  ExpandNode(n)
  $game\_outcome =$ DepthCharge$(n, root)$
  Backpropagate$(n, root, game_outcome)$
end for
return $\arg\max_{c \in root.children} c.val$

Algorithm 7 MCTS: SelectNode(n)

while $n.children \neq \emptyset$ **do**
  $n = \arg\max_{c \in n.children} c.UCB$
end while
return n

To balance exploring and exploiting in the MCTS SelectNode sub-procedure, we use a standard upper confidence bound (UCB) (7).

$$c.UCB = \arg\max_{c \in root.children} \left(\frac{c.val}{c.visits} + \sqrt{2 \frac{\log(c.parent.visits)}{c.visits}} \right).$$

Algorithm 8 MCTS: ExpandNode$(n, root)$

$s \leftarrow n.state$
for all $m \in$ LegalMoves(s) **do**
  $b \leftarrow$ PlacePiece$(s, m, n.player_type)$
  $i \leftarrow n.move\_number$
  $n.children.append$(Node$(b, i + 1)$)
end for

Algorithm 9 MCTS: Backpropagate$(n, root, game_outcome)$

$n.visits \leftarrow n.visits + 1$
if n.parent **then**
  if n.parent.player_type == root.player_type **then**
    $n.val \leftarrow n.val + game\_outcome$
  else
    $n.val \leftarrow n.val + (1 - game\_outcome)$
  end if
  Backpropagate$(n.parent, root, game_outcome)$
end if

Finally, as noted above, our implementations of the Expert Gamer and MCTS algorithms always parameterize simulation with a formal specification of the rules of any given game

Algorithm 10 MCTS: DepthCharge(n)

$s \leftarrow n.state$
$i \leftarrow n.move\_number$
while not HasTerminated(s) **do**
  $player\_type \leftarrow n.player\_sequence[i]$
  $m \leftarrow$ RandMove(s)
  $i \leftarrow i + 1$
  $s \leftarrow$ PlacePiece($s, m, player\_type$)
end while
if IsWin($s, root.player\_type; game\_rules$) **then**
  return 1
else if IsDraw($s; game\_rules$) **then**
  return 1/2
else
  return 0
end if

(e.g., restrictions on whether a given player can win only in certain directions on the board). We implement $\tilde{\mathcal{V}}$ and DepthCharge as described in the main text so that intermediate value functions and win conditions are calculated with respect to the specific game dynamics and win conditions for each game variant. *MakeMove* similarly takes in the current *move_number* move index to account for games in which player order is not strictly alternating, such as games in which the first or second player goes twice as their opening move.

3 Additional model details

We report additional details on the models, including compute usage, model-model simulations, and other design choices informing the Expert Gamer model.

3.1 Computational cost of models

We assess the relative compute demands of different gameplay models on two measures: (1) the time it takes each model to select an action and (2) how many board states have their value assessed in the process of the model deciding what action to take. We measure the first metric as the average time per move counts the time, in seconds, that each model takes selecting an action of where to move. This quantity is then averaged across all games, game trials, and actions within a game. Table 6 shows the average wall-clock time per move for the three models broken down by game. We note that wall-clock time is potentially limited by the exact implementation and hardware used. To reduce some of these confounds, all models utilize a shared codebase and are run on the same hardware. We measure the second quantity by counting the number of nodes in the constructed tree search for each move. This quantity is then averaged across game trials and moves within a game trial. Table 5 shows the average number of nodes whose heuristic values are evaluated in the search tree for each move for the Intuitive Gamer, Expert Gamer, and MCTS models. Each row in the table corresponds to one of the 41 games used in the "human-human play" experiment. All game simulations involve self-play with that same model.

Game	Intuitive Gamer	Expert Gamer	MCTS
3x3 3 P1/2 P2	7.71 ± 1.09	340.79 ± 421.19	$6,946.28 \pm 7,373.94$
3x3 3	6.72 ± 2.00	489.21 ± 477.21	$11,769.89 \pm 13,908.48$
3x3 3 (P2 2p)	6.41 ± 2.23	$923.50 \pm 1,062.07$	$12,168.06 \pm 13,930.55$
3x3 3 (P1 2p)	6.73 ± 1.99	152.83 ± 73.01	$13,445.12 \pm 14,038.39$
3x3 3 L	6.46 ± 2.17	380.69 ± 547.15	$13,618.89 \pm 15,778.72$
2x5 3	7.04 ± 2.43	218.78 ± 172.61	$18,019.79 \pm 18,641.68$
1x10 3	6.93 ± 2.50	259.58 ± 219.57	$18,860.72 \pm 19,800.13$
4x4 3 L	11.63 ± 3.43	$3,618.14 \pm 2,540.70$	$62,150.06 \pm 44,828.95$
4x4 3 D	11.45 ± 3.70	$3,367.23 \pm 2,507.85$	$64,164.73 \pm 41,121.09$
4x4 3 HV	12.87 ± 2.90	$3,171.31 \pm 3,045.33$	$74,368.59 \pm 41,785.08$
5x5 2	24.02 ± 0.81	606.18 ± 470.70	$75,410.55 \pm 56,977.89$
5x5 3 L	18.23 ± 5.03	$7,547.74 \pm 3,772.27$	$124,690.63 \pm 71,792.33$
4x6 5	16.41 ± 5.70	$6,569.44 \pm 3,984.28$	$127,823.20 \pm 62,653.84$
4x6 4	16.75 ± 5.56	$6,310.25 \pm 4,502.73$	$129,860.14 \pm 64,085.15$
5x5 4 D	17.20 ± 5.83	$7,599.20 \pm 4,259.91$	$134,135.65 \pm 64,384.30$
5x5 4 HV	17.93 ± 5.46	$6,464.76 \pm 4,240.71$	$136,459.28 \pm 64,148.43$
5x5 4 (P2 2p)	17.89 ± 5.60	$7,103.19 \pm 5,034.45$	$136,702.71 \pm 64,837.41$
5x5 4	18.05 ± 5.36	$6,618.10 \pm 4,702.53$	$144,846.18 \pm 62,310.97$
5x5 3 (P1 HV)	22.57 ± 1.86	$6,168.12 \pm 6,553.56$	$153,006.93 \pm 71,992.53$
5x5 3 (P1 D)	22.18 ± 2.22	$6,355.83 \pm 6,316.61$	$153,930.07 \pm 71,405.38$
5x5 4 (P1 2p)	19.14 ± 4.72	$6,984.52 \pm 5,777.07$	$155,149.51 \pm 57,176.56$
5x5 3	22.54 ± 1.93	$6,167.83 \pm 6,553.92$	$157,519.28 \pm 71,077.72$
5x5 4 P1/3 P2	21.94 ± 2.29	$7,475.19 \pm 6,392.12$	$176,543.76 \pm 45,694.49$
3x10 3	27.25 ± 2.24	$7,697.46 \pm 7,894.15$	$204,352.35 \pm 82,146.78$
4x9 4	26.77 ± 7.70	$13,698.40 \pm 7,066.52$	$240,662.55 \pm 85,097.34$
7x7 4 L	33.72 ± 11.05	$17,258.99 \pm 7,808.52$	$301,968.11 \pm 119,135.63$
5x10 5	34.18 ± 11.61	$17,794.77 \pm 8,438.64$	$313,634.43 \pm 120,188.42$
7x7 4 HV	38.99 ± 9.33	$17,756.62 \pm 8,252.03$	$377,773.04 \pm 101,370.85$
7x7 4 D	36.39 ± 10.42	$18,122.45 \pm 8,075.41$	$378,955.54 \pm 111,335.85$
7x7 4 (P1 D)	42.87 ± 4.78	$20,734.93 \pm 10,316.27$	$387,077.75 \pm 89,789.16$
7x7 4	42.49 ± 5.05	$21,326.28 \pm 10,600.34$	$387,483.47 \pm 110,703.17$
7x7 4 (P1 HV)	41.06 ± 6.16	$19,509.22 \pm 11,271.34$	$387,637.26 \pm 100,578.80$
7x7 4 (P2 2p)	42.86 ± 5.22	$20,767.11 \pm 10,223.01$	$388,333.98 \pm 109,289.19$
7x7 4 (P1 2p)	43.00 ± 4.69	$17,274.72 \pm 12,929.50$	$397,684.27 \pm 92,618.19$
5x10 4	42.84 ± 5.23	$21,163.57 \pm 10,751.05$	$397,818.92 \pm 109,009.24$
7x7 4 P1/3 P2	45.42 ± 2.84	$16,165.24 \pm 13,332.40$	$399,809.50 \pm 91,272.35$
10x10 5 (P1 D)	82.90 ± 13.98	$44,812.99 \pm 16,198.57$	$865,140.51 \pm 103,573.52$
10x10 3	97.43 ± 1.99	$33,495.23 \pm 25,525.04$	$872,437.53 \pm 177,821.03$
10x10 5 (P1 HV)	77.77 ± 17.70	$44,346.05 \pm 16,328.17$	$874,752.68 \pm 111,721.08$
10x10 5	83.64 ± 15.43	$48,697.30 \pm 16,778.13$	$885,028.95 \pm 117,608.47$
10x10 4	92.99 ± 6.52	$43,112.32 \pm 24,438.41$	$897,954.52 \pm 136,325.31$

Table 5: Efficiency, measured by number of states evaluated, across game reasoning modules. Average game board states explored per move across all game configurations. Values represent mean $\pm$ standard deviation.

3.2 Heuristic quality of Expert Gamer model

The Expert Gamer model is more sophisticated than the Intuitive Gamer across two axes—search depth and heuristic sophistication. While it is easy to verify that the Expert Gamer model has a deeper search depth than the Intuitive Gamer, it is difficult to compare the heuristic sophistication of the two models without knowing the objective value of each board state of each game. To this end, we simulate game trials of the Expert Gamer model against a faithful implementation of the Intuitive Gamer model that incorporates search depth. We observe that the Expert Gamer model dominates regardless of whether or not it plays as Player 1 or 2. Our

Game	Intuitive Gamer	Expert Gamer	MCTS
3x3 3 P1/2 P2	$4.11 \times 10^{-4} \pm 1.02 \times 10^{-4}$	0.05 ± 0.02	16.25 ± 4.00
3x3 3 (P1 2p)	$4.10 \times 10^{-4} \pm 1.46 \times 10^{-4}$	0.03 ± 0.01	17.61 ± 5.01
3x3 3	$4.25 \times 10^{-4} \pm 1.73 \times 10^{-4}$	0.08 ± 0.03	18.85 ± 6.30
3x3 3 (P2 2p)	$4.02 \times 10^{-4} \pm 1.47 \times 10^{-4}$	0.09 ± 0.05	19.00 ± 6.41
3x3 3 L	$4.08 \times 10^{-4} \pm 1.47 \times 10^{-4}$	0.08 ± 0.03	19.08 ± 6.07
2x5 3	$3.17 \times 10^{-4} \pm 1.36 \times 10^{-4}$	0.04 ± 0.02	19.13 ± 6.33
1x10 3	$3.38 \times 10^{-4} \pm 1.40 \times 10^{-4}$	0.05 ± 0.02	19.39 ± 6.11
4x4 3 HV	$4.37 \times 10^{-4} \pm 1.19 \times 10^{-4}$	0.21 ± 0.10	38.03 ± 6.55
4x4 3 L	$7.27 \times 10^{-4} \pm 2.67 \times 10^{-4}$	0.45 ± 0.20	38.34 ± 12.44
4x4 3 D	$5.54 \times 10^{-4} \pm 1.79 \times 10^{-4}$	0.22 ± 0.07	38.95 ± 11.64
5x5 2	$1.85 \times 10^{-3} \pm 4.62 \times 10^{-4}$	0.13 ± 0.03	51.45 ± 5.11
4x6 5	$1.11 \times 10^{-3} \pm 3.40 \times 10^{-4}$	0.39 ± 0.12	54.35 ± 16.23
5x5 4 HV	$7.20 \times 10^{-4} \pm 2.58 \times 10^{-4}$	0.46 ± 0.15	68.63 ± 19.58
5x5 4 D	$9.78 \times 10^{-4} \pm 3.28 \times 10^{-4}$	0.83 ± 0.27	70.84 ± 21.44
4x6 4	$1.17 \times 10^{-3} \pm 3.44 \times 10^{-4}$	0.61 ± 0.20	71.58 ± 21.57
5x5 3 L	$1.58 \times 10^{-3} \pm 4.87 \times 10^{-4}$	1.30 ± 0.55	72.12 ± 24.70
5x5 4	$1.55 \times 10^{-3} \pm 4.84 \times 10^{-4}$	0.84 ± 0.36	76.41 ± 21.90
5x5 4 (P2 2p)	$1.43 \times 10^{-3} \pm 4.42 \times 10^{-4}$	0.82 ± 0.25	77.91 ± 22.59
5x5 3 (P1 HV)	$1.18 \times 10^{-3} \pm 2.00 \times 10^{-4}$	0.60 ± 0.18	78.96 ± 12.05
5x5 3 (P1 D)	$1.51 \times 10^{-3} \pm 3.69 \times 10^{-4}$	0.71 ± 0.22	79.25 ± 15.16
5x5 3	$1.81 \times 10^{-3} \pm 4.57 \times 10^{-4}$	0.83 ± 0.27	80.16 ± 10.07
5x5 4 (P1 2p)	$1.45 \times 10^{-3} \pm 3.56 \times 10^{-4}$	0.74 ± 0.28	85.30 ± 19.61
3x10 3	$1.91 \times 10^{-3} \pm 4.44 \times 10^{-4}$	0.87 ± 0.22	89.06 ± 13.60
5x5 4 P1/3 P2	$1.73 \times 10^{-3} \pm 4.55 \times 10^{-4}$	0.84 ± 0.22	89.83 ± 13.46
4x9 4	$2.30 \times 10^{-3} \pm 6.36 \times 10^{-4}$	1.59 ± 0.33	129.93 ± 35.15
7x7 4 HV	$2.37 \times 10^{-3} \pm 6.76 \times 10^{-4}$	2.02 ± 0.46	189.08 ± 43.82
5x10 5	$3.81 \times 10^{-3} \pm 1.20 \times 10^{-3}$	2.73 ± 0.70	211.20 ± 68.08
7x7 4 L	$4.58 \times 10^{-3} \pm 1.42 \times 10^{-3}$	4.41 ± 1.25	231.47 ± 75.59
7x7 4 (P1 HV)	$4.02 \times 10^{-3} \pm 9.90 \times 10^{-4}$	3.00 ± 0.58	235.27 ± 46.61
7x7 4 P1/3 P2	$6.14 \times 10^{-3} \pm 1.62 \times 10^{-3}$	3.67 ± 0.65	239.62 ± 40.60
7x7 4 D	$3.24 \times 10^{-3} \pm 9.26 \times 10^{-4}$	3.56 ± 0.99	243.65 ± 47.14
7x7 4 (P1 D)	$4.75 \times 10^{-3} \pm 1.11 \times 10^{-3}$	3.68 ± 0.58	261.60 ± 49.34
7x7 4	$6.53 \times 10^{-3} \pm 2.22 \times 10^{-3}$	4.30 ± 0.56	261.65 ± 43.19
5x10 4	$5.74 \times 10^{-3} \pm 1.44 \times 10^{-3}$	3.98 ± 0.81	266.12 ± 35.31
7x7 4 (P2 2p)	$6.45 \times 10^{-3} \pm 1.57 \times 10^{-3}$	4.68 ± 0.92	270.42 ± 41.26
7x7 4 (P1 2p)	$6.29 \times 10^{-3} \pm 1.87 \times 10^{-3}$	3.65 ± 0.90	291.32 ± 46.20
10x10 3	$2.78 \times 10^{-2} \pm 6.36 \times 10^{-3}$	17.34 ± 2.94	862.42 ± 66.75
10x10 5 (P1 HV)	$1.49 \times 10^{-2} \pm 3.54 \times 10^{-3}$	15.06 ± 2.14	953.86 ± 263.76
10x10 5 (P1 D)	$1.97 \times 10^{-2} \pm 4.86 \times 10^{-3}$	17.58 ± 2.61	987.14 ± 265.40
10x10 4	$2.50 \times 10^{-2} \pm 5.11 \times 10^{-3}$	20.28 ± 4.42	1015.55 ± 124.69
10x10 5	$2.13 \times 10^{-2} \pm 3.28 \times 10^{-3}$	21.28 ± 2.43	1151.12 ± 169.73

Table 6: Efficiency, measured by wall-clock time per action, across game reasoning modules. Average time per move (seconds) across all game configurations. Values represent mean ± standard deviation. The values are sorted by the average time per move of the MCTS model.

results are reported in Table 7. The Ablated Expert removes the mechanism that incorporates features of the global board state instead of the purely local features. In particular, the value function of the ablated expert differs from the value function of our expert model as given by Equation 7. The value function of the ablated expert is given by:

$$\mathcal{V}^{\text{EG, abl}}(s') = (2 \times \mathbb{I}\{\text{the } i\text{'th move is } X\} - 1)\mathcal{V}(a, s), \tag{8}$$

Matchup	Wins (%)	Losses (%)	Draws (%)
Expert vs Ablated Expert	43.0	16.4	40.5
Ablated Expert vs Expert	19.3	41.4	39.3

Table 7: Comparing the Expert Gamer with inheritance in the value function against a lesioned variant. Game outcome percentages for Expert Gamer versus an Ablated Expert Gamer, which does not incorporate the entire global state matchups in is computation of the heuristic value of each state. Results are based on 40 match simulations over the subset of 41 games used in the play experiment, showing win, loss, and draw rates of Player 1 averaged over all simulations and games. Rows are formatted as Player 1 model vs. Player 2 model.

ϵ	Expected payoff for the Expert Gamer
0.0	0.0831 ± 0.863
1e-4	0.0845 ± 0.861
1e-3	0.0683 ± 0.866
1e-2	0.0651 ± 0.865
0.1	0.0328 ± 0.879
0.2	-0.00396 ± 0.887

Table 8: Selection of ϵ in the Expert Gamer based on payoff in cross-model gameplay. Expected payoff and standard deviation for the Expert Gamer model, averaged across opponent and whether the expert played first or second.

where s is the board state that immediately preceded s' on the game tree and a is the action that takes board state a to state s'. The value function of the ablated expert is clearly very myopic as it depends only on the most recent action of the board state.

We incorporate an exploration component to our Expert Gamer model by setting a probability ϵ with which we explore a uniformly random child node. We have the Expert Gamer model play against the Intuitive Gamer (flat and depth-3 variants, as described at the end of the Supplement) on the subset of 41 games used in the play experiment with different values of ϵ' and set ϵ to be the ϵ' that yields the greatest expected advantage over the two models. Results are reported in Table 8, which shows the expected payoff averaged across the expert's opponent. Choice of ϵ has a minimal impact (Table 8); still, we take the maximum shows, this value turns out to be $\epsilon = 0.0001$.

3.3 Intuitive Gamer parameters

We additionally assess the sensitivity of the Intuitive Gamer model to choice of parameters (w) for the heuristic value function ($\tilde{\mathcal{V}}$). We fix the temperature $\tau = 1$ and vary w_{connect}, w_{block}, and w_{center}. We sweep over w_{connect}, w_{block}, and w_{center} between 0 and 2 in steps of 0.2 respectively, and run 50 simulations of our model for each setting (which we bootstrap subsample from 100 times, simulating 20 simulated participants running $k = 6$ mental simulations each). We then compute the R^2 under each variants' predictions relative to the human predicted payoffs. We observe in Figure 13 higher sensitivity to the parameterization of the goal-progress and goal-blocking components, highlight the importance of "goal-directedness" in the Intuitive Gamer model and appropriately balancing offense and defense. The sweeps also underscore that our selection of simply weighting all values with 1 is a reasonable modeling choice.

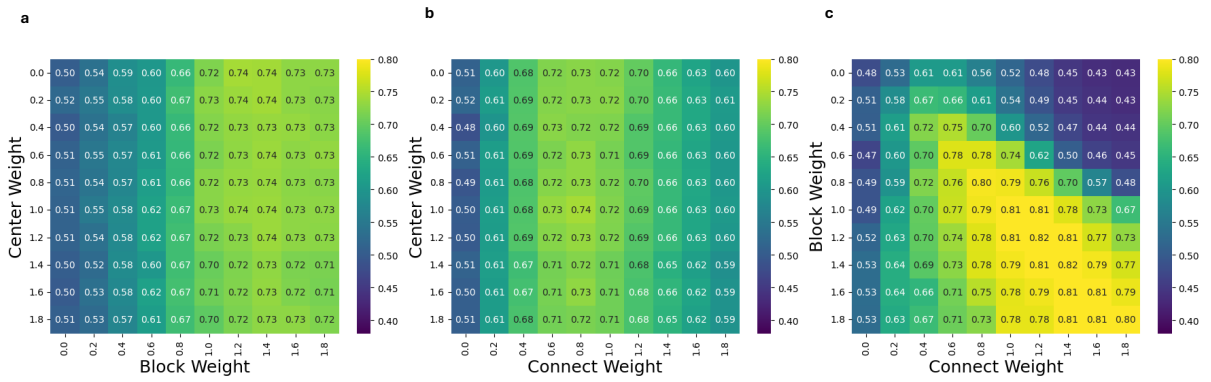


Figure 13: Parameter sensitivity analyses. Analyzing the impact of the goal progress (offensive "connect" component), goal blocking (defensive "block" component), and center weight on fit to people's judgments in the "just think" experiment, with subset of simulations for different settings of the heuristic value function weights. We run 100 bootstrapped subsamples of 20 participants from the "just think" experiment per game with $k = 6$, and compute R^2 between the average human payoff from the subset versus the average model payoff. Each cell averages over the held-out parameter (e.g., averaging over all settings of the connect weight in the leftmost plot). Fit to participants is substantially more impacted by the components controlling offense and defense, rather than center; interestingly, the defensive blocking component generally requires a value as high or more than the corresponding offensive component.

4 Example experimental interfaces

We include example interfaces for each of our main behavioral experiments. Figure 14 depicts example interfaces from the "just think" judging game outcome and game funness experiments. Figures 15 and 16 depict example interfaces from the human-human gameplay experiment. Figure 17 depicts examples from the indirect human watching and predicting play experiment.

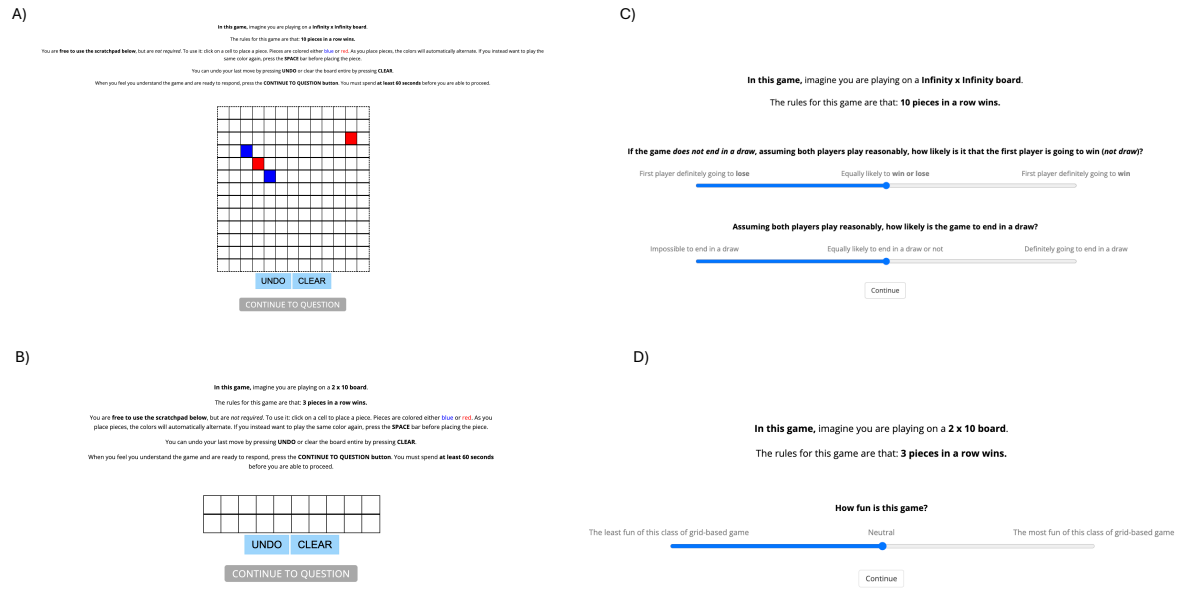


Figure 14: "Just think" evaluating games before any experience example interfaces. For each game, participants are given a scratchpad which they can optionally use while thinking about their response (**a-b**). Pieces will automatically change between blue and red upon each click, unless overridden (see Methods). Infinite boards **(a)** are shown with dashes along the edges. For each game, participants either judge game outcomes **(c)** or game funness **(d)** using sliders.

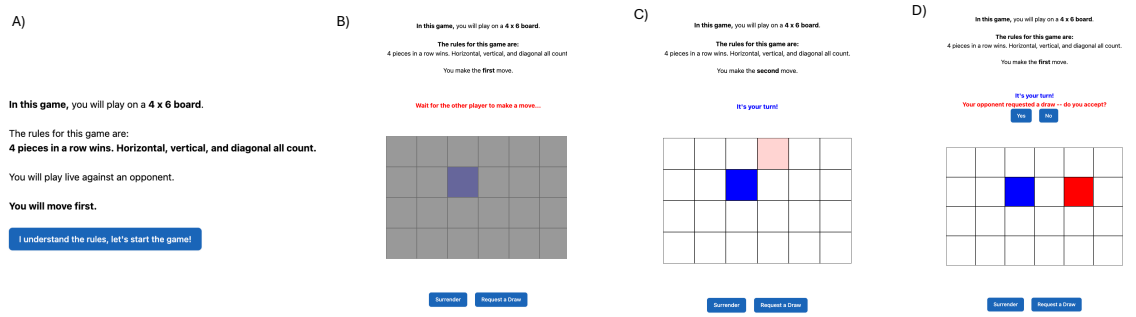


Figure 15: Example interfaces from the human-human gameplay study. **a,** Participants are first informed of the game rules and whether they will play first or second. **b,** The board is disabled and greyed-out when it is not their turn. **c,** When it is their turn, they see a hover (in the color of their piece) before they decide where to play. They decide to play by clicking on the board. **d,** When a player requests a draw, the other player is informed and allowed to accept or reject the draw.

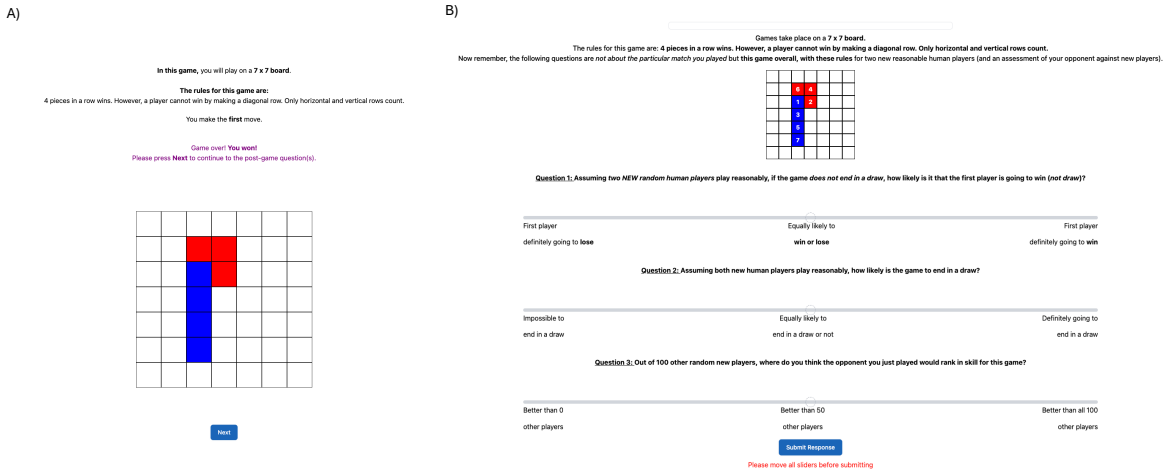


Figure 16: Additional example interfaces from the human-human gameplay study. **a,** After the game is over, both players are informed of the outcome. **b,** Each player then makes a series of judgments about the game via sliders and are shown a "snapshot" of how their match had unfolded.

5 Additional analyses into human and model game evaluation

5.1 Participant scratchpad usage

Before making their judgments, participants had the option to interact with an interactive version of the board to simulate self-play (a "scratchpad") to minimize demands on spatial working memory. Participants made on average 1.60 ± 1.46 SD rollouts for the payoff task and 1.36 ± 1.15 SD rollouts for the funness task. We define a rollout as the number of times a participant started interacting on a fresh board (i.e., zero rollouts indicates no interaction with the board; one rollout indicates on usage of the board; two indicates one restart of board clicks). We depict a histogram over rollouts per task in Figure 18. We do not draw a direct parallel between scratchpad rollouts and mental simulation (it is highly plausible that participants are doing more mental simulation than they are demonstrating on the scratchpad); we include the scratchpad was to the demands on spatial working memory of our task so that we can focus on studying reasoning. While from manual inspection, many participants using the scratchpads did appear to simulate play, some participants seemed to just click around without intention.

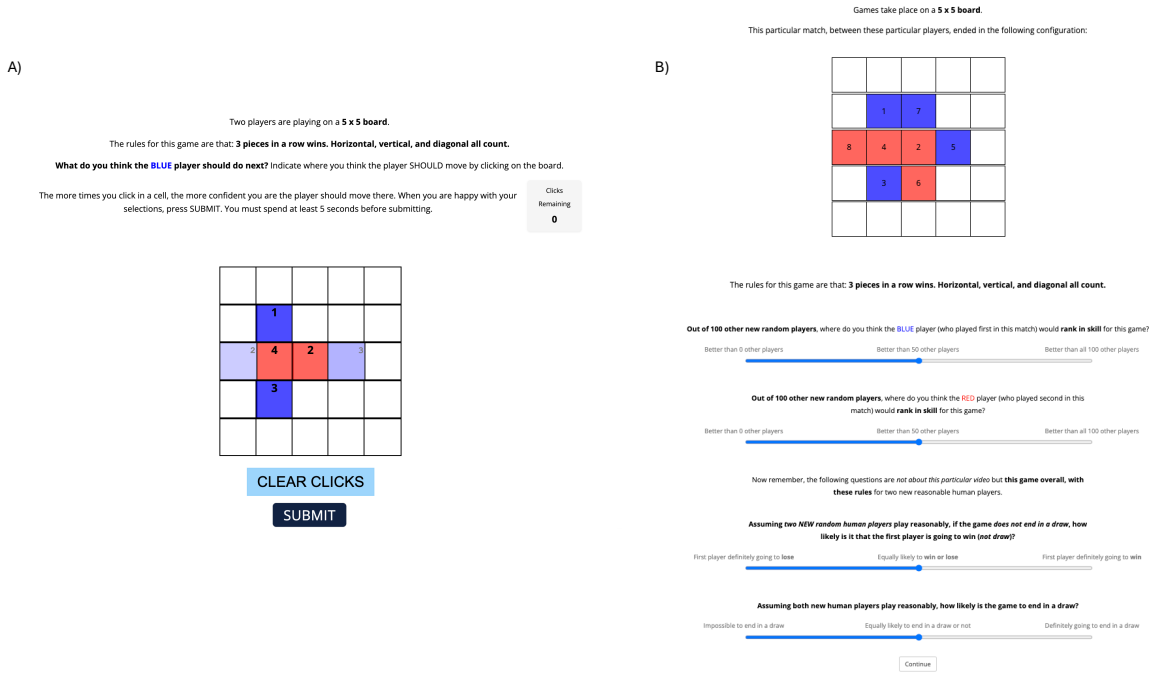


Figure 17: Example interfaces from the people watching others play and predicting moves study. Participants watch gameplay, which is frozen at three timepoints. At each timepoint, participants are tasked with predicting where a player should play by spreading 5 clicks over the board. **a,** The cells are colored in proportion to the number of clicks made by the participant, and the number of clicks remaining (out of 5) are shown to the participant. **b,** At the end of watching a full match, participants are shown a snapshot of the played match, and asked to make a series of judgments about the game via sliders.

5.2 Decomposed game outcome prediction tasks

Participants indicated their inferences about likely game outcomes about (1) if the game did not end in a draw, whether the first player was likely to win ($P(\text{P1 win} \mid \text{not draw})$); and (2) how likely the game was to end in a draw ($P(\text{draw})$). We compare the Intuitive Gamer model predictions against human judgments for these decomposed questions in Figure 19 and 20, respectively. The questions are critical together; e.g., if the participant thinks that the game will definitely end in a draw, the first question is ill-posed. This is accounted for in the payoff, and $P(\text{P1 win})$ when we combine $P(\text{P1 win} \mid \text{not draw}) \times P(\text{draw})$. We report participants' computed $P(\text{P1 win})$ and their reported $P(\text{draw})$ in Figures 19 and 20. We extract readouts of the corresponding questions to the the Intuitive Gamer model by counting the number of outcomes (in each bootstrapped set of $k = 6$ samples) that correspond to each query (e.g., the proportion of draws; or proportion of first player wins) and compare against alternate models using the same empirical outcome frequency in Figures 19 and 20, respectively. For cases where a game did not have any non-draw simulations, we take $P(\text{P1 win} \mid \text{not draw}) = 0.5$ (which is generally what we notice participants do; see Figure 21). In general, the Intuitive Gamer approaches the human split-half R^2 for the measures involving $P(\text{win})$ (i.e., split-half R^2 for $P(\text{P1 win}) = 0.82$ $[95\%\text{CI} : 0.77, 0.87]$; $P(\text{P1 win} \mid \text{not draw}) = 0.78$ $[95\%\text{CI} : 0.72, 0.83]$). However, the Intuitive Gamer (and alternate models) are all substantially sharper in $P(\text{draw})$ predictions than people and not near the noise ceiling (split-half human R^2 for $P(\text{draw}) = 0.78$ $[95\%\text{CI} : 0.72, 0.83]$). It is possible that people compute over a small sample of simulated games in a different way than counting outcomes, which may better correspond to the readouts they provided, or generally have a prior to believe that games are more likely to end in a draw.

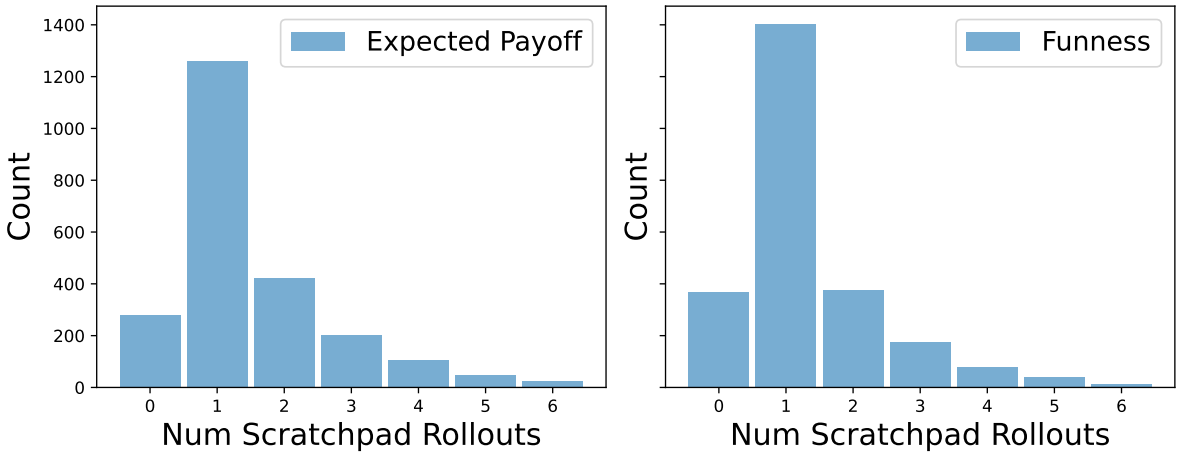


Figure 18: Scratchpad usage. Explicit scratchpad "rollouts" per game, per participant. A "rollout" is counted as any interaction with the scratchpad, involving at least one click. 2 rollouts means the participant pressed "RESET" once. Most participants engage at least once with any game, both when evaluating **a,** fairness and **b** funness; few conduct more than three explicit rollouts on the scratchpad.

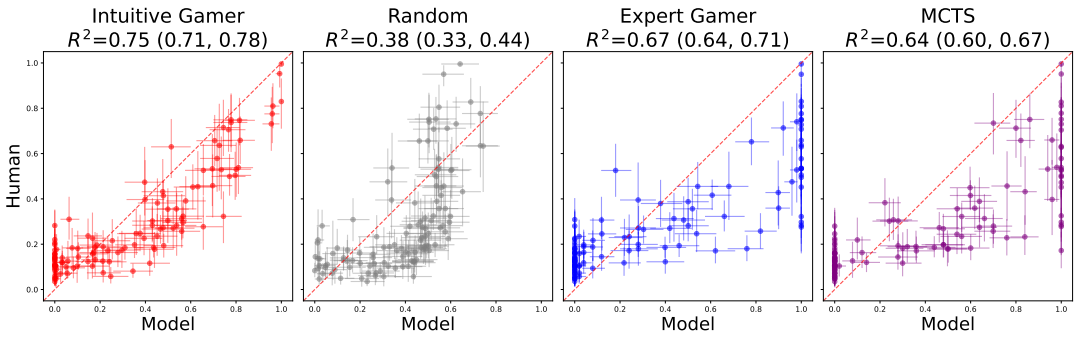


Figure 19: Empirical model comparisons for P(P1 win) against people. Bootstrapped 95% CIs over participants and samplings of $k = 6$ simulations for $N = 20$ simulated participants for each model.

Our early explorations into the impact over the game reasoning module running only partial simulations, where simulations that stop early are taken to be a draw, may encode such a prior and improves fit to human draw judgments (Figure 22). However, we leave a close analysis of partial versus full simulations for future work, as it requires a deeper cross-comparison with alternate models and other stopping rules (which added substantial complexity to this first exploration of an Intuitive Gamer model).

5.3 Predicting payoff from non-simulation based linguistic features

A key component of our hypothesis is that people assess new problems by drawing on fast probabilistic mental simulations. This hypothesis demands a comparison then against non-simulation-based alternate models. To assess the role of non-simulation-based features, we fit a linear regression model to the binary game traits (as introduced in the Methods). We fit to 70% of the games and test on the held-out 30%. We find that the model can capture some variance in human judgments $R^2 = 0.33$ [95% CI: $0.29, 0.37$] (see Figure 23) but comparatively less than models based on explicit simulation, as noted in the main text.

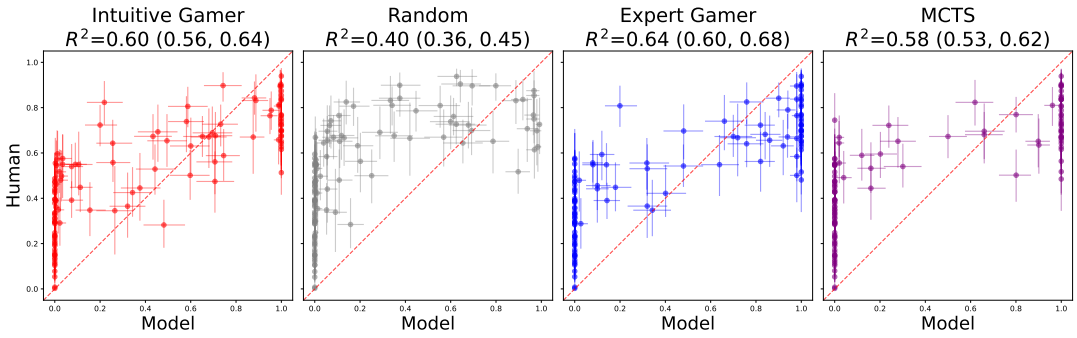


Figure 20: Empirical model comparisons for $P(\textbf{draw})$ against people. Bootstrapped 95% CIs over participants and samplings of $k = 6$ simulations for $N = 20$ simulated participants for each model.

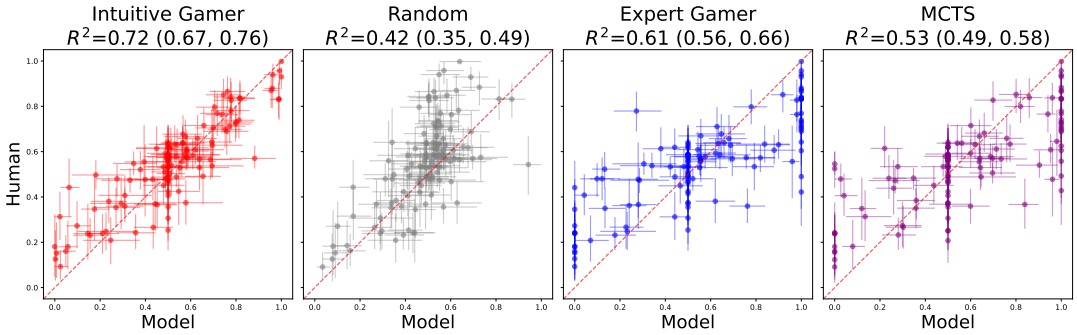


Figure 21: Empirical model comparisons for $P(\textbf{P1 win}|\textbf{no draw})$ against people. Games under the model that only have draws are imputed with 0.5. Bootstrapped 95% CIs over participants and samplings of $k = 6$ simulations for $N = 20$ simulated participants for each model.

5.4 Language model payoff and game-theoretic optimal comparisons

We also compare against a series of language models—varying in prompting type (directly asking about game outcomes, or permitting "chain of thought" reasoning)—as well as a state-of-the-art reasoning model, o1 (44). We compare predictions under the language models against human predicted payoffs Figure 24 and against the "optimal" game theoretic expected payoffs in Table 3. While o1 is closer to human judgments and a closer fit to the game-theoretic optimal values, it is still a ways off of "perfect" highlighting that the games we consider here are non-trivial to reason about.

5.4.1 Language model prompting

All language and reasoning models are prompted with a variant of the instructions provided in the human experiment. We prompt models to provide (in a single response) with our two game outcome reasoning questions—how likely the first player will win if the game does not end in a draw, and how likely the game will end in a draw—directly following the questions asked of participants. We provide the models with the same $0 - 100$ slider labels as in the human study and ask the model to provide a value in the same range per question. We sample 20 rollouts for each model. We sample all models at the default temperature of 0.7, except o1, which we sample using its default fixed 1.0 temperature.

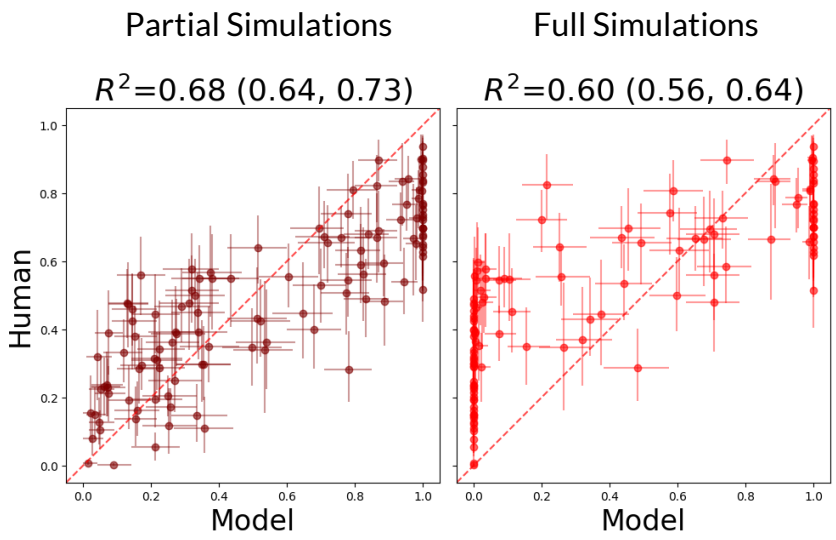


Figure 22: Full versus partial simulations to estimate P(draw). Comparing human and Intuitive Gamer model predicted P(draw) under partial versus full game simulations (both are computed with $k = 6$ simulations). Games that end early are deemed draws. Error bars depict 95% bootstrapped CIs over the average value for each game.

5.5 Reasoning about game funness

We next provide additional results into the funness model.

5.5.1 Parameter weights

As described in the Methods of the main text, we fit regression model weights for the funness components over repeated bootstrapped samples of $N = 20$ participants for each of the games. We can then inspect the regression weights to explain whether and how the factors relate to people's funness judgments. Regression model weights (Table 9) confirm that the most fun games are those which are balanced, demand thinking, and are neither too long nor too short (Table 9).

Parameter	Value (95% CI)
Balance	8.7 (7.5, 9.8)
Reward for Thinking	4.5 (3.6, 5.6)
Game Length (Linear)	9.9 (7.6, 12.3)
Game Length (Quadratic)	-4.4 (-5.8, -3.1)

Table 9: Funness model parameter fits. Bootstrapped mean and 95% confidence intervals for funness model parameter fits over 1000 fits to subsampled human funness predictions, over all games. Coefficients are over normalized features (see Methods).

5.5.2 Component ablations

To assess the impact of the relative contribution of each funness model component, we lesion each component. Analysis of Variance (ANOVA) and Akaike information criterion (AIC) assessments highlight that each component contributes to the fit to human data (see Table 10). We also assess the relative benefit for fitting each component with a non-linear spline in Table 11. We see a relative benefit only for the game length component. The balance feature collapses back to a single degree of freedom (linear fit).

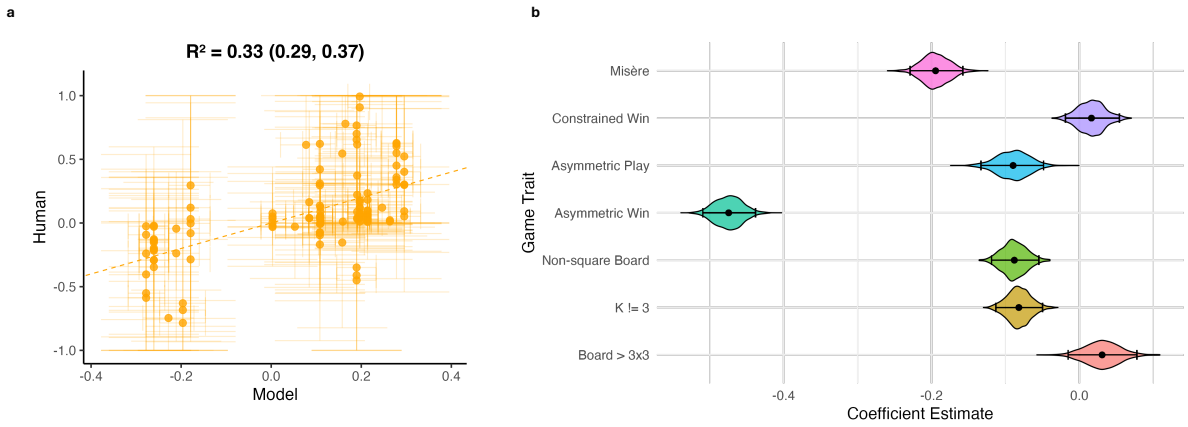


Figure 23: Predicting payoff from non-simulation game traits. Fitting a linear regression model to binary game traits captures some of the variance of payoff judgments. Each point in **(a)** is a game. The error bars are the 95% CIs over the human bootstrapped mean payoff and model-predicted mean payoff fit over bootstrapped subsets of 70/30% of games. **(b)** 95% CIs around the parameter fits per game trait.

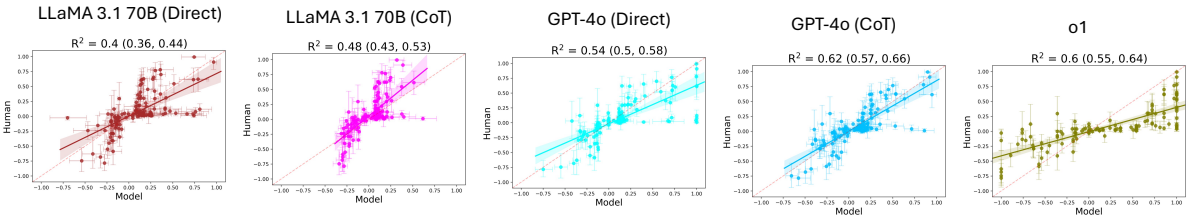


Figure 24: Comparing payoff predictions against language models. Correlations between bootstrapped samples of the human versus language (top) and "reasoning" (bottom) model predicted payoffs for the 121 games. 95% CIs are over 1000 bootstrapped subsamples.

Feature Removed	F Statistic	P Value	ΔAIC
Balance	80.517	6.79e-15	61.1
Reward for Thinking	18.521	3.57e-05	15.8
Game Length	10.099	1.91e-03	8.0

Table 10: Funness model feature lesions. Ablating each linear feature in the primary funness model reveals that all features matter. ΔAIC is $AIC_{ablated}$ - AIC_{full} (higher indicates better fit of the full model funness model; i.e., more negative impact on fit from the ablation).

Feature	F Statistic	P Value	ΔAIC
Balance	1.242	2.67e-01	-0.7
Reward for Thinking	1.082	3.00e-01	-0.9
Game Length	14.027	2.85e-04	11.8

Table 11: Impact of quadratic features in funness model. Comparing the inclusion of quadratic features on each of the simulation-based features in the funness model reveals that only a quadratic term for the game length feature significantly impacts fit to human judgments. ΔAIC is AIC_{lin} - AIC_{nonlin} (higher indicates better fit from non-linear).

5.5.3 Linguistic features

We additionally explore the role of linguistic non-simulation based features, as outlined in the Methods. Fitting a model to the binary traits alone captures a moderate amount of the variance in human funness judgments (Figure 26a) but comparatively less than the simulation-based models. Some of this may arise from relative correlations between the traits and the simulation-based features (Figure 26b), e.g., boards that have asymmetric win conditions are generally associated with unfair games (which could connect to the balance component). Adding in the binary traits to the simulation-based funness model does not parsimoniously improve fit to human data. We find minimal evidence for an effect from "novelty," computed as the number of binary traits (or deviations from Tic-Tac-Toe) that are "active." Adding in a linear component does induce a slightly better fit according to ANOVA and AIC tests (Table 12), but has a negligible on our generalization tests (Table 13)—that is, held-out fits when fitting to bootstrapped subsamples of 50% of the games. It is possible that some participants are primarily accounting for linguistic-only (non-simulation based) features when assessing funness; recall, we do see comparatively higher variability in participants' judgments for funness evaluations than payoff predictions. However, at the aggregate population level, accounting for simulation-based features dominates fits. We leave such participant-level modeling of funness for future work.

We also compare language- and reasoning-model predicted funness compared against human predicted funness in Figure 25. Models are prompted in the same way as in the payoff predictions described above; that is: using the default temperature of 0.7 and sampling 20 rollouts, under a slightly modified version of the full experiment instructions given to people, with the exception of o1 (sampled at its default 1.0). Most language models yield fits to human data that are comparably worse than the simulation-based features, with some exception to o1. We do not know what kind of simulation, if any, o1 is doing over these games; hence, the model at present does not support the kind of explanatory analyses we are interested in here. However, the comparisons point to potential value in using the human data collected here for benchmarking human-likeness of AI.

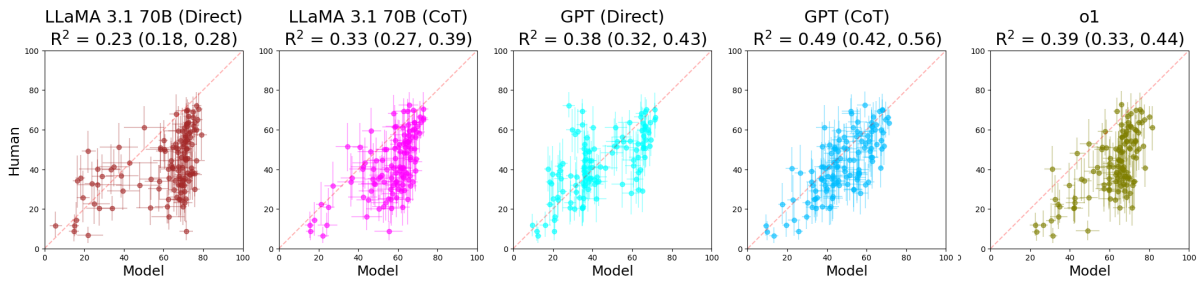


Figure 25: Language- and reasoning-model predicted funness per game, compared to people. Each point is one game. Error bars are bootstrapped 95% CI over the average human- and model-predicted funness.

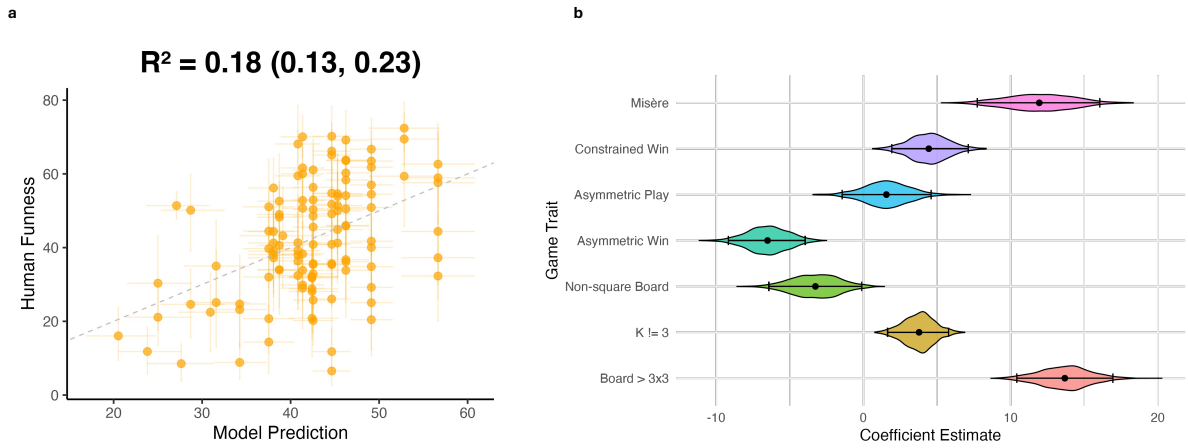


Figure 26: Predicting fun from non-simulation game traits. a, Modeling human funness judgments from non-simulation based binary game traits alone captures some of the variance in human judgments, however substantially less than the simulation-based model. **b,** Bootstrapped 95% CIs around the parameter fits from the funness model for each game trait.

Added Feature	F Statistic	P Value	ΔAIC
Board Size	0.340	5.61e-01	-1.6
Approx Novelty	4.765	3.11e-02	2.9
Binary Traits	1.224	2.96e-01	-4.8

Table 12: Additions of linguistic non-simulation features to the funness model. Comparing the inclusion of linguistic non-simulation based features to the funness model reveals only a weak potential effect of incorporating non-simulation based features into the model. ΔAIC is $AIC_{\text{sim-only}}$ - AIC_{expanded} (higher indicates better from including the additional feature).

6 Additional analyses into human and model action selection and prediction

We next present additional results from the "zero-shot human-human play" and "watch-and-predict" experiments. We report the full set of averaged log probabilities for each game in the human gameplay experiment in 27.

6.1 Action selection accuracy and rank

The Intuitive Gamer model generally is more accurate at capturing peoples' moves according to Top-1, Top-3, Top-5 accuracy of the moves people made (Figure 28). Accuracy is averaged

Variant	Held-Out R^2
Simulation Only	0.62 (0.50, 0.71)
+ Linear Novelty	0.62 (0.53, 0.70)
+ Quadratic Novelty	0.61 (0.50, 0.69)

Table 13: Generalization test when incorporating non-simulation based features into funness model. Comparing held-out fits (fit on 50% of games, test on 50% of games) when including a "novelty" (linear or non-linear feature) ontop of the base regression model featuring only simulation-based features (balance; reward for thinking; quadratic length). This reveals that the additional linguistic feature does not meaningfully impact generalization.

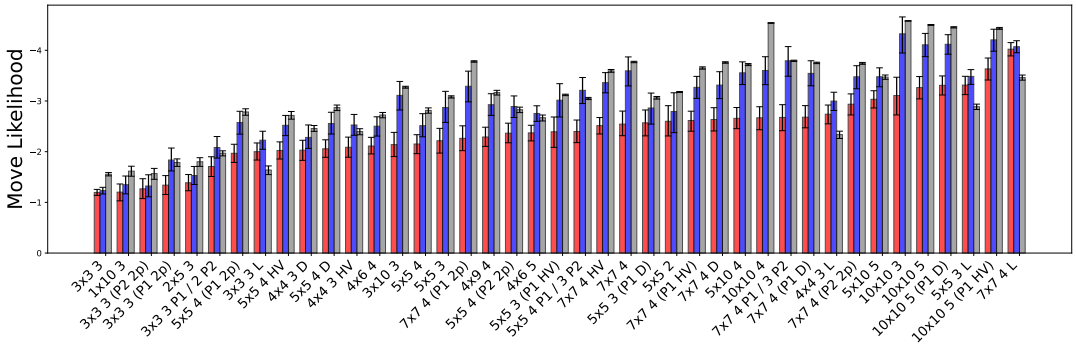


Figure 27: Aggregate move likelihood, per game, in the human-human play experiment. Aggregate move log probability (higher is better) for the moves people played in the human-human gameplay experiment under the Intuitive Gamer (red), Expert Gamer (blue), and random (grey) models, broken up by games. Error bars depict 95% bootstrapped confidence intervals around the mean log-likelihood over all moves for each game.

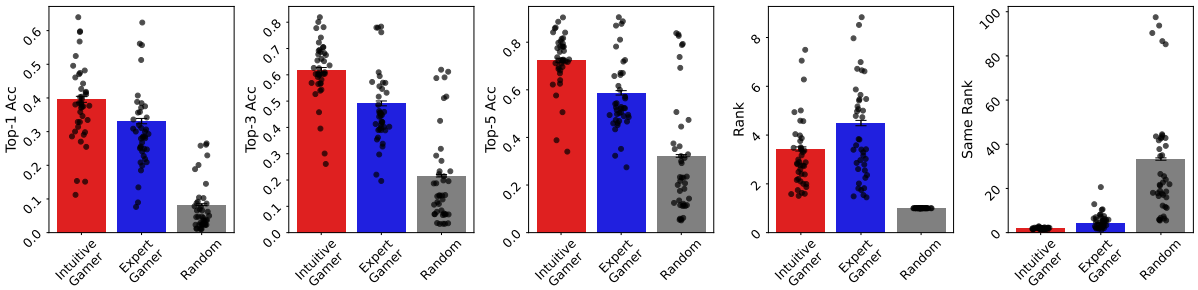


Figure 28: Accuracy and rank of human played move in human gameplay experiments. We additionally include additional statistics comparing the move distributions for the play experiments against observed play, particularly accuracy and the rank of the humans' played move relative to set of possible legal moves (ranks include ties). Each dot corresponds to the average for any one of the 41 played games.

over 100 bootstrapped samples to account for ties; for instance, to compute Top-1 accuracy, if the Intuitive Gamer assigns three moves to the same top probability, then accuracy is computed by sampling from that set. To account for ties, we also compute the average rank of the human played moves (including ties). However, random will appear best under this measure as all moves are assigned the same rank with uniform probability (therefore, appearing as if all played moves are assigned rank 1). To account for this, we also report the average number of moves assigned the same rank as the move played by people. Together, we see that generally the Intuitive Gamer assigns a relatively low rank to the played move (and not too many other moves to the similarly low rank).

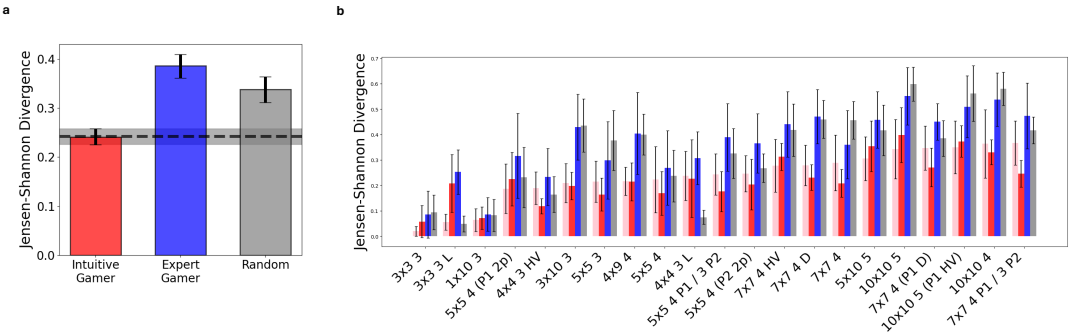


Figure 29: "Watch-and-predict" results are robust to the choice of distributional metric. Jensen-Shannon Divergence (lower is better) reveals that the Intuitive Gamer distribution are generally better aligned to the human watchers' distributions, over **a,** all game boards and **b,** at a per game level. Error bars in **a** depict bootstrapped 95%CI around the mean JSD for for all board; error bars in **b** depict standard deviation over boards per game.

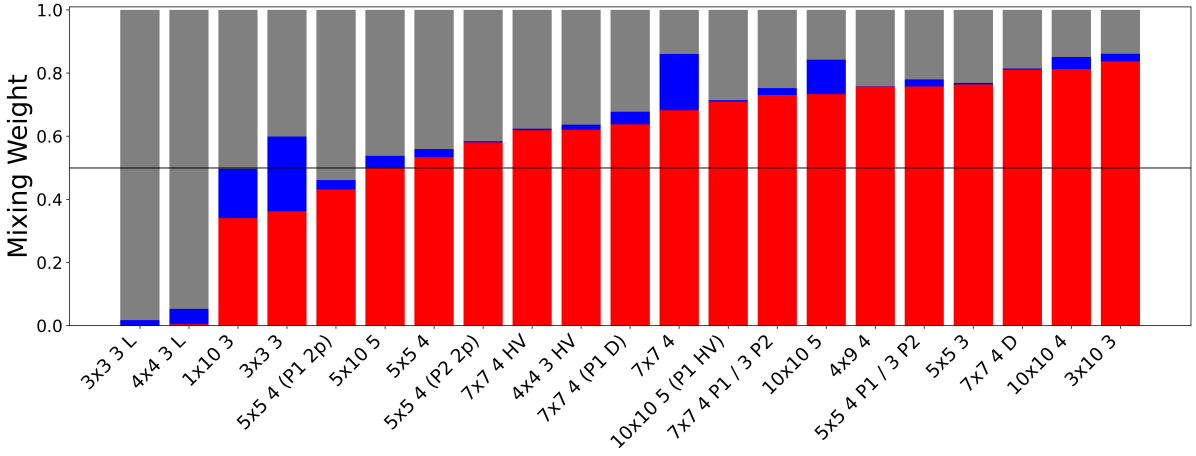


Figure 30: Average admxiture weights for each game, fit to the average human watch-and-predict distributions. The Intuitive Gamer is the dominant component for most games in an admixture fit over the models' distributions to the participants' distributions, optimized against the smoother Jensen-Shannon Divergence. Notable failure modes (e.g., misere games and comparable fits with the Expert Gamer model for Tic-Tac-Toe) align with places of weaker Intuitive Gamer model fits in the play data (see main text).

6.2 Robustness to choice of distributional measure

To assess the robustness of our "watch-and-predict" experiment analyses to the measure of distributional similarity, we repeat our analyses using an alternative distributional measure: the Jensen-Shannon Divergence (36). The Intuitive Gamer again is generally at the human noise ceiling in average and similarly near the split-half noise ceiling for many (though not all, e.g., misere) games (see Figure 29). An admixture model over the model distributions to learn a mixing weights to fit the human watcher distributions (optimized using JSD) generally reveals that the Intuitive Gamer model is the dominant component across games—with exceptions for misere games and some other smaller-board games (see Figure 30), which warrants further investigation.

6.3 Modeling choices with a softmax-based model

As noted in the main text "Methods," when assessing how models capture the choices people actually made in games, we model participants' moves as some combination of the likelihood

under the core model of interest (the Intuitive Gamer or Expert Gamer model) as well as some probability that they play randomly. We sweep over $\alpha \in \{0.5, 0.6, 0.7, 0.8, 0.9, 0.9999\}$. An alpha of 0.5 means that any move is equally weighted between the primary model (Intuitive Gamer or Expert Gamer) and random; higher alpha puts more weight on the primary model. The Intuitive Gamer model generally leads to better fits, independent of the choice of α. However, both models yield a better fit with human play and human watch distributions when mixed in some proportion with random. This aligns with the admixture results in the main paper, which show that fits for most games and most people are best modeled by some random component. A lower alpha (more weight to random) leads to a comparatively better fit when modeling human play and watch judgments under the Expert Gamer model, further indicating that the Expert Gamer model is likely more sophisticated than how people actually reason in new games for the first time.

In the admixture analyses, we also repeat admixture fits for different temperatures (τ), ranging in $\tau \in \{0.5, 1.0, 1.5, ..., 3.5\}$ where τ controls the action choice distribution softmax. We refit the admixture at each setting of τ and apply the same τ to both the Intuitive Gamer and Expert Gamer model. We do this to ensure that the relative fits in the admixture are as fair as possible to each model as the Expert Gamer model is often relatively sharp. As temperature increases, all models collapse toward random; however, as temperature decreases, we lose the probabilistic nature that is core to our Intuitive Gamer hypothesis. For most settings of τ, the Intuitive Gamer is the dominant mixing component; however, a higher τ appears to yield a slightly better fit for the watch distributional fits to human watch distributions relative to the played moves, which may be due to the seemingly higher variability in participants' move distributions.

6.4 Predicted probability of played move relative to human predictions

In the "watch-and-predict" experiment, participants were asked to predict where a player should move. To assess whether the distribution of moves people predicted captures how the human player actually played, the same aggregate analysis (as in the main "play" experiment), computing the average log likelihood of a player's action under the move distribution predicted by participants watching that match versus the distributions predicted under the Intuitive Gamer model and alternate models along a spectrum of expertise. Human watchers place similar probability on the move people actually made as the Intuitive Gamer model (Figure 32).

6.5 Additional human- and model-predicted distributions

We depict all aggregate human- and model-predicted distributions from the "watch-and-predict" study (see Figures 33- 39). Four matches, paused at three timepoints (during early-, middle-, and late-stage play) were shown for each game.

6.6 Game lengths

We compare the empirical observed game lengths in the play experiment against the expected game length under model simulations. Intuitive Gamer simulations show the highest correlations with empirical human game length data, while Expert Gamer is more correlated than the random agent (Figure 40). We observe that the Intuitive Gamer simulations rarely "overpredict", and there are several outlier games where it takes all models significantly longer to simulate than humans to finish.

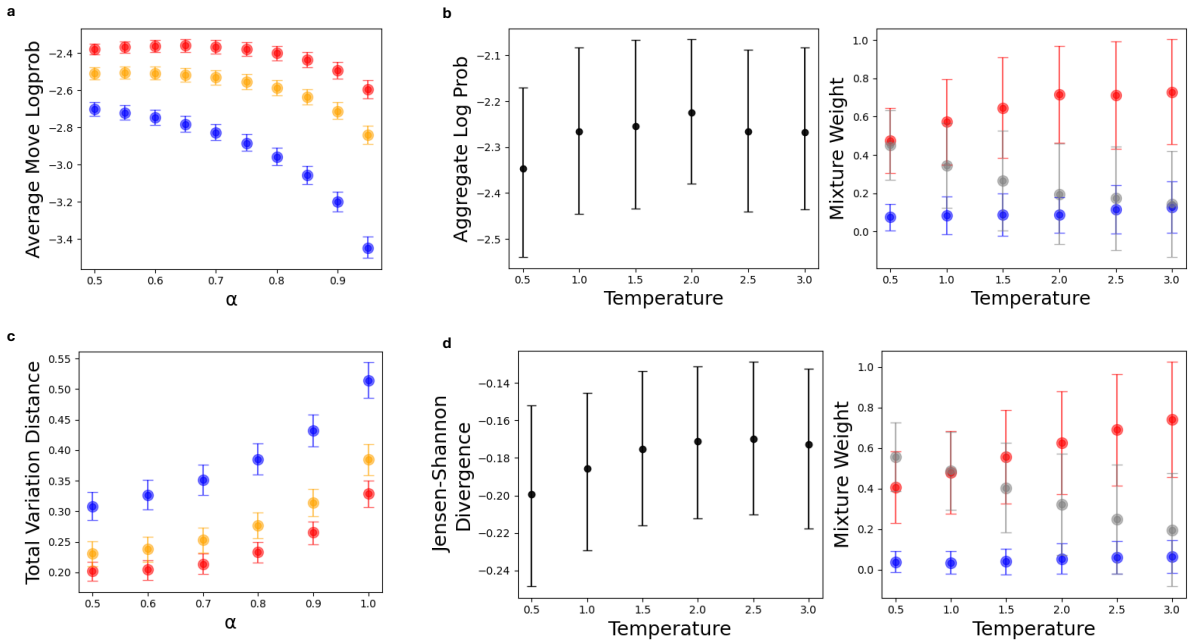


Figure 31: Mixture weight and temperature sensitivity in models' fit to people's action selection and prediction data. Varying the mixture weight α in the $\alpha-$softmax model for the core play **(a)** and watch **(c)** experiments. Error bars show 95% bootstrapped CI over the mean average move log probability (higher is better) and average Total Variation Distance (lower is better) relative to the human watcher move distribution. Red is the Intuitive Gamer model; yellow the depth-3 version of the Intuitive Gamer (non-flat); blue the Expert Gamer model. Admixtures are then fit over the flat Intuitive Gamer (red), Expert Gamer (blue), and random (grey) game agents for varied move distribution temperature for the play **(b)** and watch **(d)** experiments, respectively. Admixtures are fit for all data for each game. The left subplot depicts the average log probability of the moves in the resulting admixture for each temperature **(b)** and the resulting Jensen-Shannon Divergence (as the admixtures are optimized against JSD rather than TVD, given the smoothness of JSD) over the optimized mixture distribution. Error bars depict 95% CIs over the bootstrapped mean. The right subplot depicts the averaged mixing weights for each temperature value. Error bars depict standard deviation over the games.

6.7 Draw requests and surrenders

We next conduct exploratory analyses into draw requests and surrender decisions in individual matches against model and participant-predicted payoffs. Participants have high draw rates on matches that tend to be less fun and also less biased (see Figure 41a-b). This suggests that participants are sensitive to other features of games when making decisions about whether the request to end the game. Surrender rates are generally higher for less fun games, though show less of a trend based on game bias (see Figure 41c-d). Future work can work on richer process models of humans' draw and surrender decisions. The choice to request a draw instead of surrender is particularly worth investigating. If a player wanted to end the game, there should be no reason that they do not immediately surrender, as if they request a draw, the game only ends if the other player agrees to draw. Future modeling could explore other factors that people may consider when evaluating whether a game is worth playing (e.g., related to shame or dignity).

6.8 Evaluating games after a single exposure

After participants played a single match or watched and predicted actions in a single match, we asked them to either rate the expected payoff of the game overall or the funness. We also asked participants to rate the skill of their opponent (in the "play" condition) or the skill of both

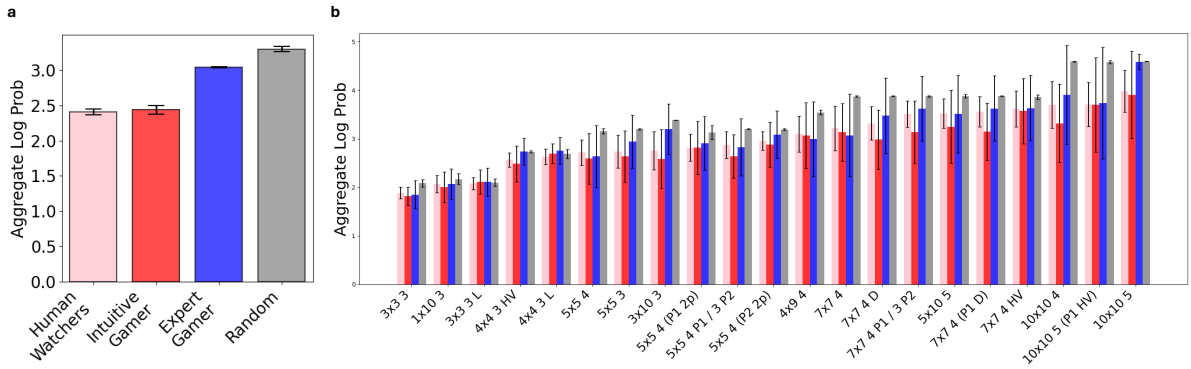


Figure 32: People's distribution of predicted moves capture real human player's actions comparably to the Intuitive Gamer model. a, From only a single indirect experience with a new game (watching one partial match), participants' distribution over predicted moves (pink) generally place similar probability on the move actually played to that predicted under the Intuitive Gamer model. The suggested moves made by the people predicting from watching videos (pink) to the actual move played by participants in each match and compare the log-likelihood under the human predictors' inferred move to the log-likelihood assigned by the respective models, marginalizing out α. Error bars depict bootstrapped 95% CIs over the averaged log-likelihood over boards, again sweeping over a "slop" parameter α (see "Methods"). **b,** Predicted move likelihood assigned to the move selected by the live player, broken down by game and ordered by lowest negative log likelihood (lower is better). Error bars depict standard deviation.

players (in the "watch" condition). We next conduct an initial exploration into these post-match game evaluations.

6.8.1 Evaluating games after one round of play

Participants' payoff predictions are highly correlated with their payoff predictions before any play and well-align with the Intuitive Gamer model (Figure 42). Participants' post-play funness judgments are less correlated with pre-play funness judgments (Figure 42) and are overall noisier (split-halfs $R^2 = 0.49$ [95% CI: $0.33, 0.64$] for post-play funness compared to a split-half of $R^2 = 0.69$ [95% CI: $0.56, 0.79$] for the same subset of 41 games rated on funness by participants who "just thought" about the games without any play). While we asked participants to assess the funness of the game overall—not just with respect to the match they played—it is possible people were biased in some form by the outcome of the game (Figure 44) or experience playing against the particular other player.

To assess the relative contribution of each simulated feature in our funness model, we rerun the same regression modeling analyses. The refit post-play funness model derived from the Intuitive Gamer features again reaches the split-half noise ceiling, capturing essentially all of the explainable variance in participants' judgments (Figure 43). Reward for thinking and game balance generally matter less for a fun game in people's post-play judgments compare to game length (Figure 43); however, we caveat that the post-play judgments are generally more variable (as noted above in the split-halfs).

6.8.2 Evaluating games after one round of watching

Participants' judgments after watching a single match are highly noisy for both payoff and funness queries (split-half $R^2 = 0.54$ [95% CI: $0.30, 0.76$] and $R^2 = 0.07$ [95% CI: $0.0001, 0.27$] for post-watch payoff and funness, respectively). In contrast to both the "just think" and post-play funness ratings which saw Tic-Tac-Toe rated around 50 (average fun rating of 51.4 ± 29.5 SD for "just think," and average fun rating of 49.4 ± 21.8 SD for post-play), the watchers

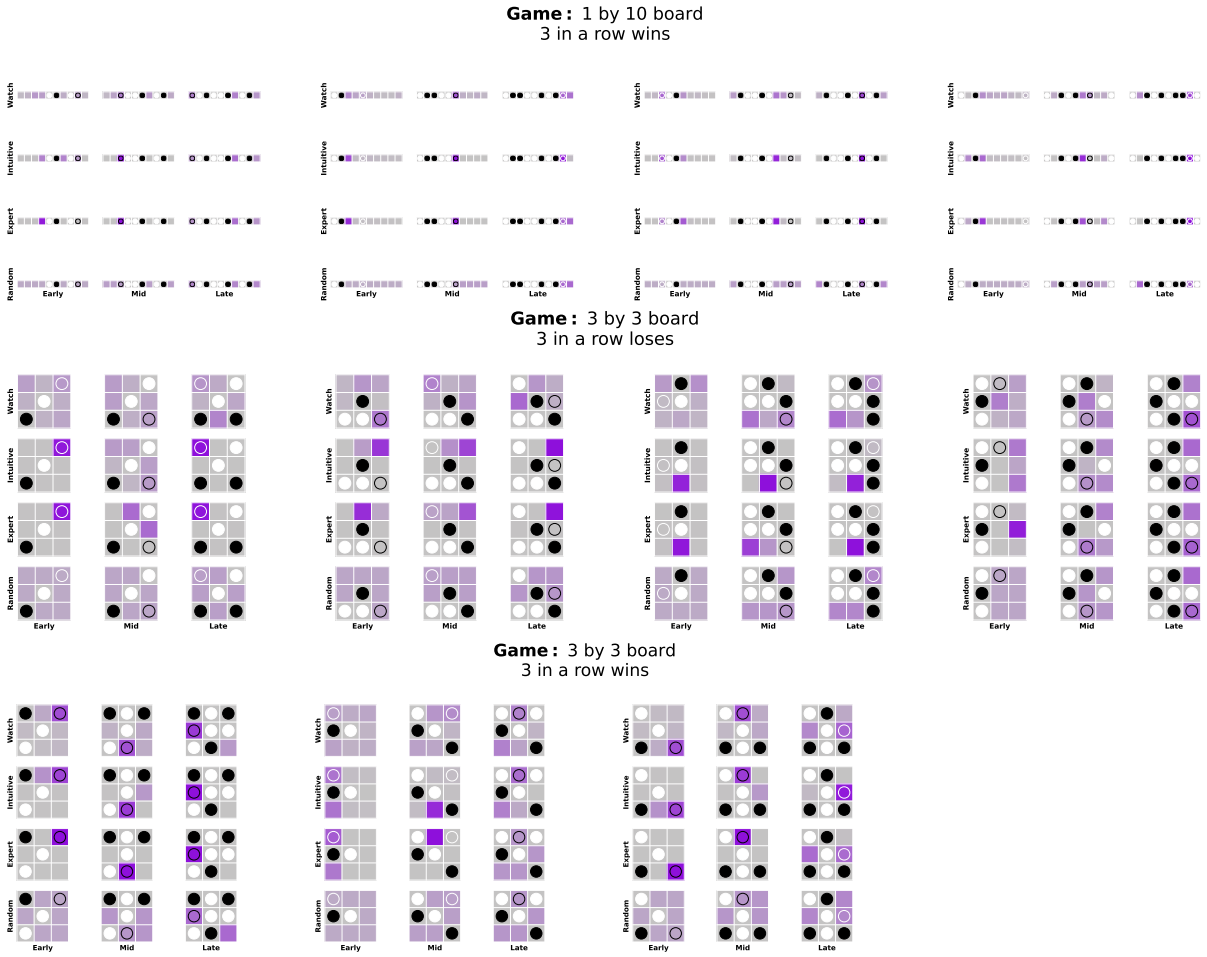


Figure 33: Full set of "watch-and-predict" distributions per match. As in the main text, the top entry in each panel depicts the aggregated human-predicted distribution over next moves, as determined by participants in the indirect play (watch-and-predict) study. The circle indicates where a person actually played. The color of each grid cell indicates where the player is predicted to move on that turn, under either the model- or human-predicted distribution; darker purple means a player is more likely to select that grid cell. We show all game boards and human- and model-predictions for each game type. The game board size and rules are written above each set of boards.

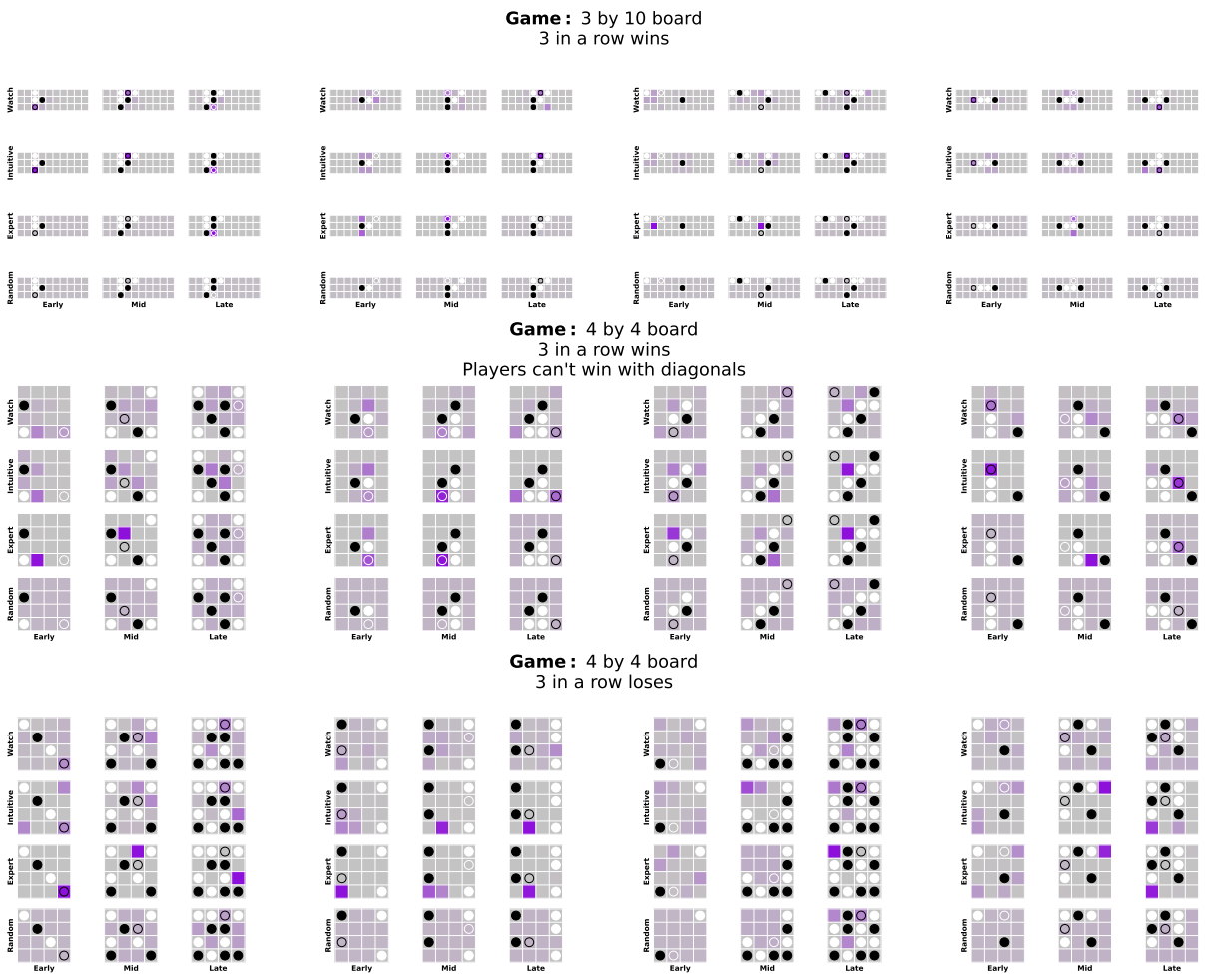


Figure 34: Full set of "watch-and-predict" distributions per match (continued).

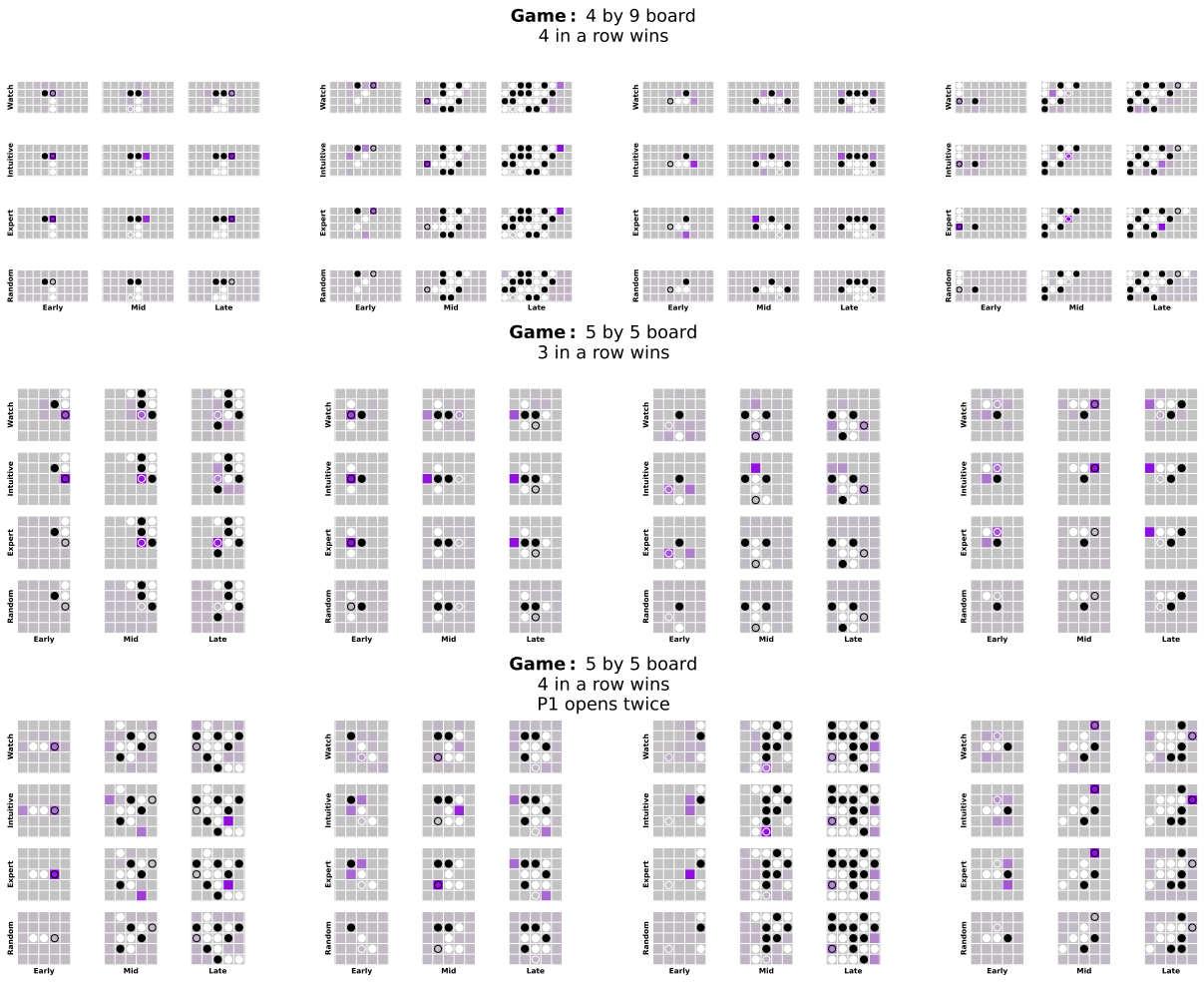


Figure 35: Full set of "watch-and-predict" distributions per match (continued).

Figure 36: Full set of "watch-and-predict" distributions per match (continued).

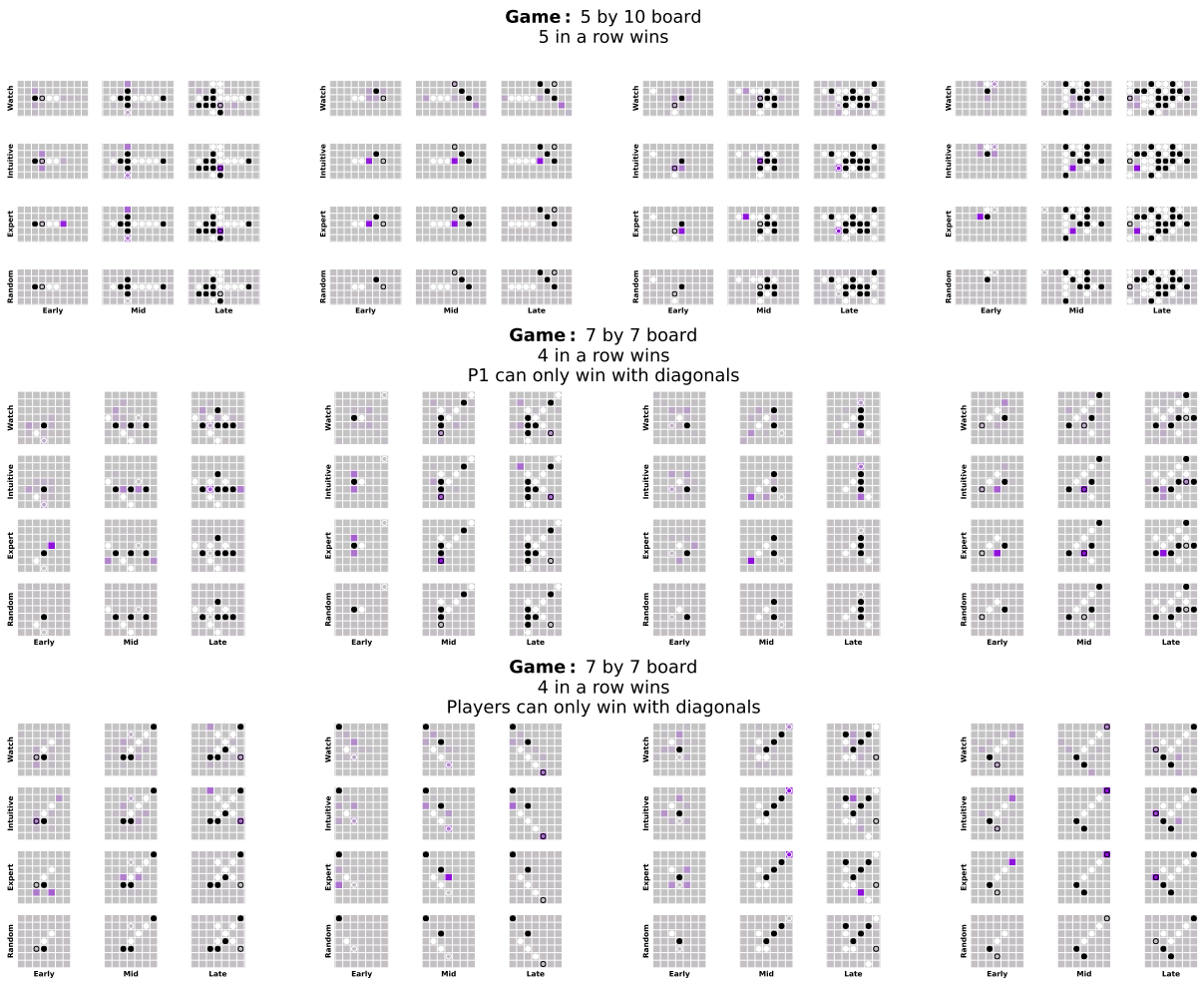


Figure 37: Full set of "watch-and-predict" distributions per match (continued).

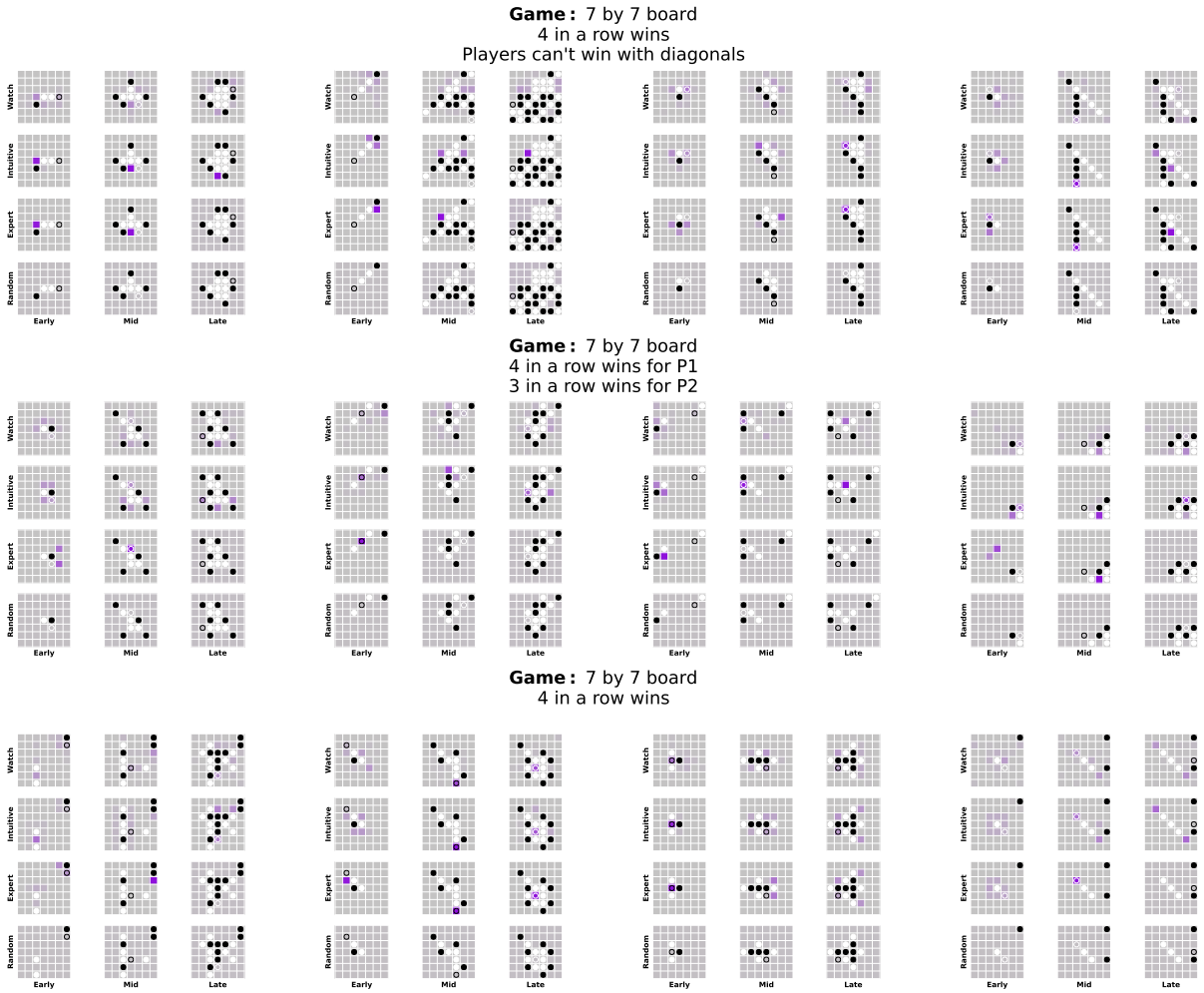


Figure 38: Full set of "watch-and-predict" distributions per match (continued).

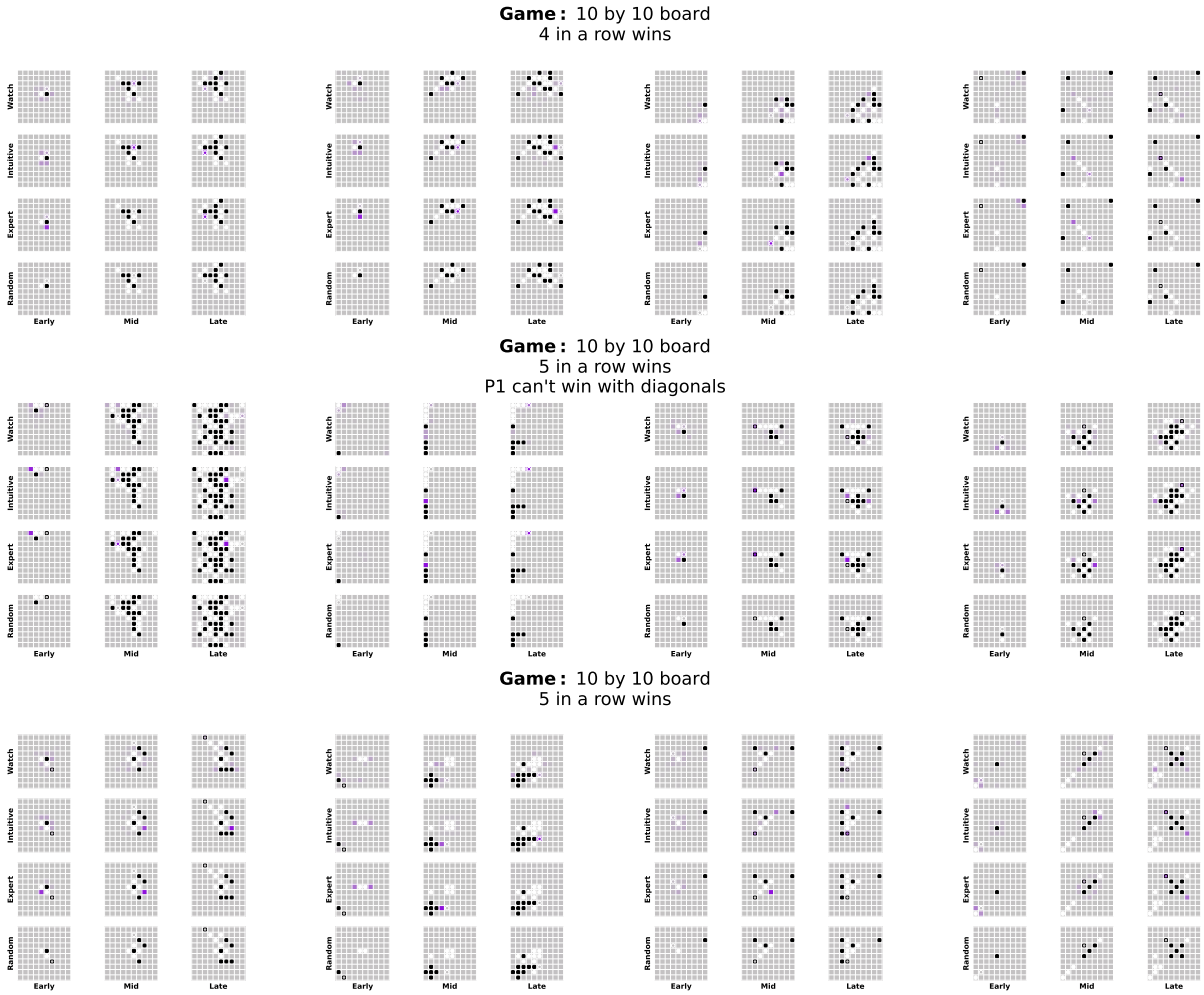


Figure 39: Full set of "watch-and-predict" distributions per match (continued).

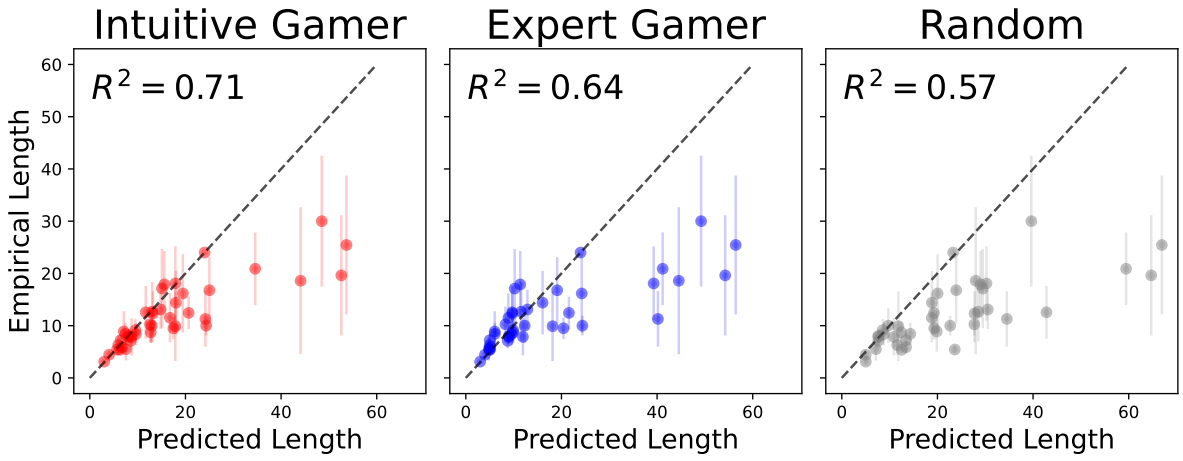


Figure 40: Empirical human- vs. model-predicted expected game length. Comparing the observed game length in human-human played games against the expected game length under simulations from the various game reasoning models. Game simulations are conducted between the same agent type (e.g., Intuitive Gamer against another instanced of itself). Empirical game length is coded as whenever the game ended: either from a player winning, the game ending in a draw. Games that ended early from an accepted draw request or a player surrendering are excluded.

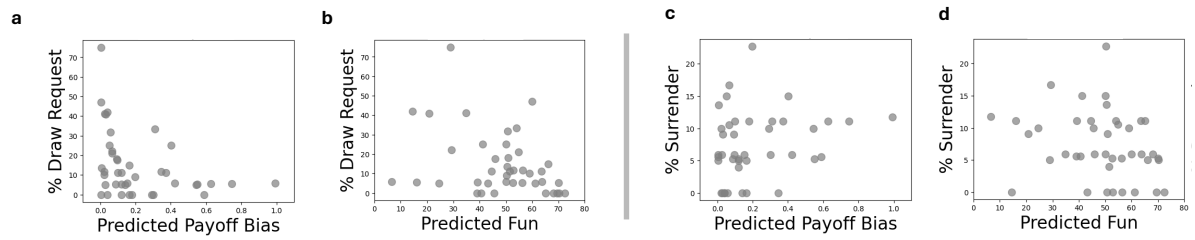


Figure 41: Draw and surrender rates and people's "just think" game evaluation. Frequency of draw requests (**a-b**) and surrenders (**c-d**) across all played matches per game (each point is a game). **a,** Games that are more biased (absolute difference in payoff from zero, under the averaged human-predicted payoffs in the "just think" study) are generally associated with higher draw rates. **b,** Games that are more fun (under the people's averaged funness judgments in the "just think experiment before any play) are generally associated with fewer draw requests. **c,** There is not a clear relationship between people's evaluation of game fairness (before any play) and surrender rates; **d,** in contrast, there games that are less fun tend to have somewhat higher surrender rates.

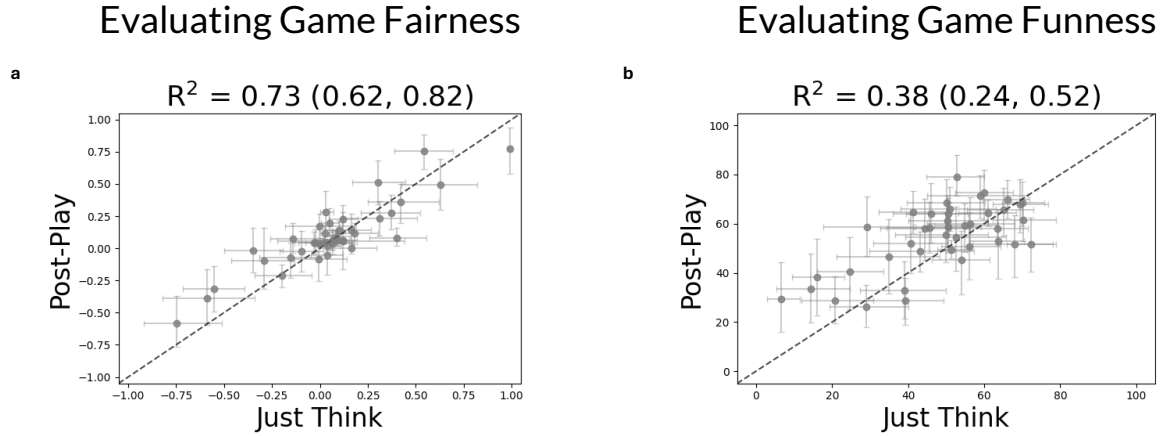


Figure 42: Impact of a single instance of play experience on judgments. **a,** bootstrapped mean predicted payoff for "just-think" versus "post-play" participant judgments; **b,** bootstrapped mean predicted funness from "just-thinking" versus "post-play".

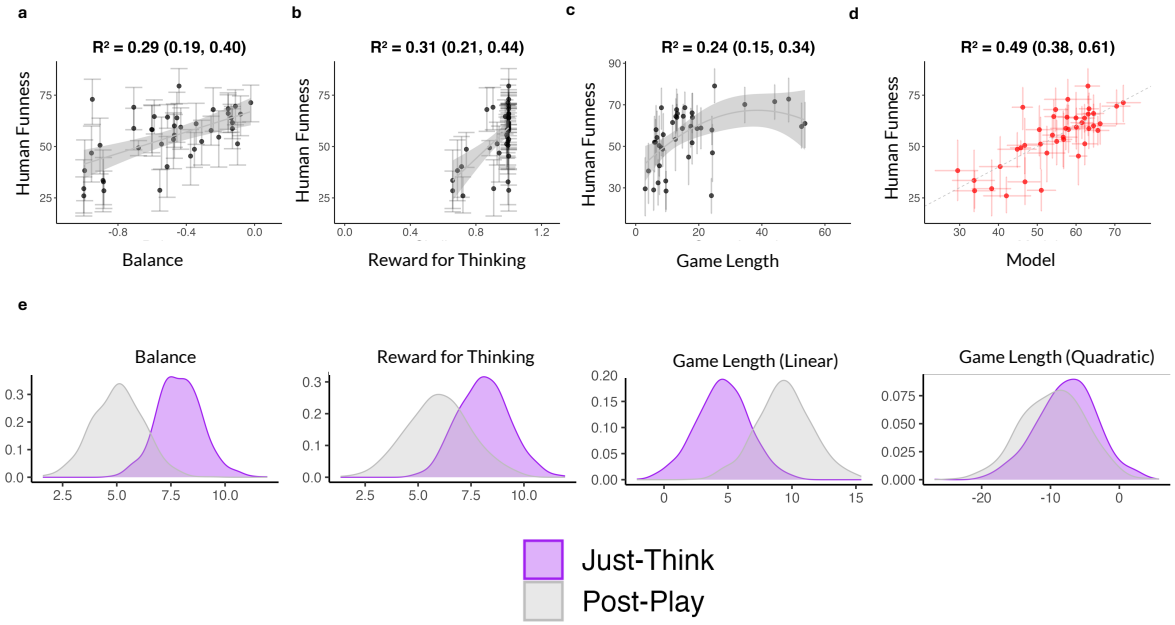


Figure 43: Post-play funness modeling. **a-c,** Features derived from the Intuitive Gamer model compared against post-play participant funness judgments. **d,** Regression model fit to the post-play data. 95% CIs are bootstrapped over participant ratings per games. **e,** Comparison of bootstrapped funness model parameter fits for the 41 games using the just-think human ratings (purple) versus the ratings provided by participants after one round of play (grey).

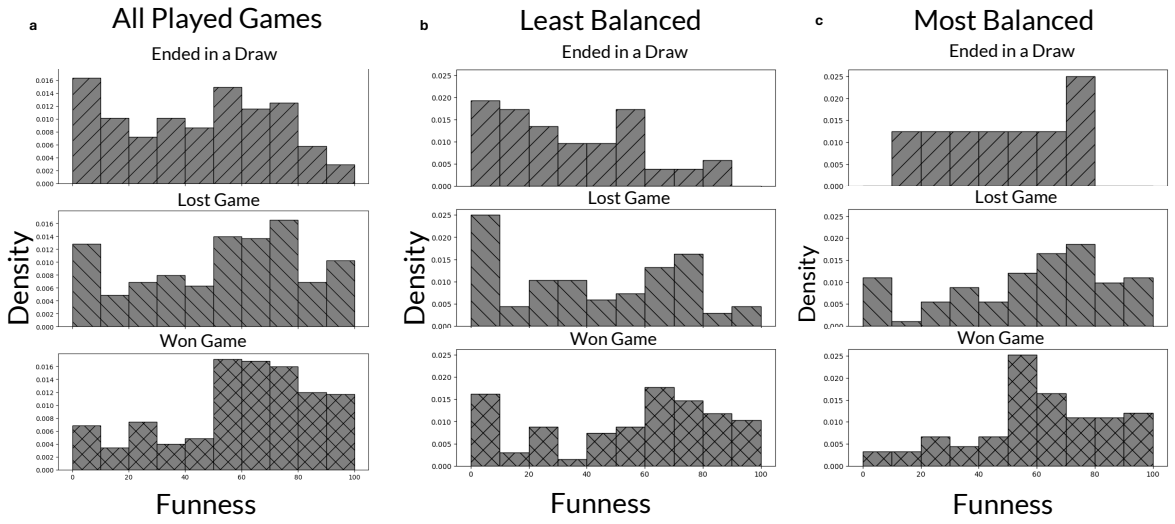


Figure 44: Relationship between attained outcome and funness judgments, after a single round of play in a new game. Funness judgments made by players after each match, broken down by whether they drew (top), lost (middle), or won (bottom). Even in cases where players lost, many still found the game fun. **a,** Funness for all matches for all played games. **b-c,** Post-play funness judgments for matches in the least balanced games (according to the Intuitive Gamer "balance" feature; the bottom 25 percentile; **b**) and the most balanced games (upper 75 percentile; **c**). Winning in an unbalanced game (**b**) still tends to lead to somewhat higher funness ratings; losing in a balanced game can also still lead to relatively high funness ratings.

Added Feature	F Statistic	P Value	ΔAIC
Board Size	7.651	9.00e-03	6.2
Approx Novelty	1.838	1.84e-01	0.1
Binary Traits	2.043	8.32e-02	2.5

Table 14: Assessing impact of non-simulation based features on funness model. Comparing the inclusion of non-simulation based linguistic features into the funness model fit to peoples' post-play funness judgments. ΔAIC is $AIC_{\text{sim-only}}$ - AIC_{expanded} (higher indicates better from including the additional feature). There is a slight effect from incorporating either board size or all binary game traits (in contrast to adding an aggregate "count" of the number of traits that are "on" for any game, i.e., "approximate novelty"). It is possible the potential benefit of incorporating board size is related to the seemingly higher role game length plays in people's funness judgments post-play.

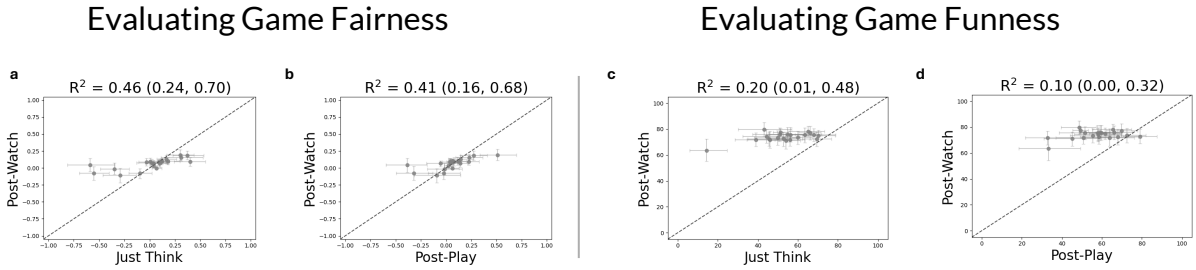


Figure 45: Judgments after watching a match. Comparing post-watch judgments of predicted payoff against those made in the **a,** "just think" and **b,** "play" experiments, on only the subset of 21 games considered in the watch experiment, as well as the predicted funness for **c,** "just-think" and **d,** after play. Participants in the post-watch experiment are generally more variable with each other; the noisier responses may be due to not having been held for 30 to 60 seconds before submitting their judgment like the other studies.

rated Tic-Tac-Toe at an average of 77.6 ± 16.6 SD on funness. We observe that most funness scores are inflated and payoffs collapse towards zero (Figure 45). We posit that the increased variability in the post-watch participants' judgments may have arisen from participants not having external pressure to "force" some degree of thinking. Notably, participants in the watch study were not held from submitting their answers until one minute passed as in the "just think" experiment (nor given an optional scratchpad) before making their judgments, which may have impacted effort. Participants in the post-play experiment also were required to spend time before responding (at least 30 seconds); however, post-play participants were not given an optional scratchpad after they played their game. Future work should explore the role of forced response time on effort in encouraging thinking about game evaluations.

7 Exploratory analyses with a intermediate depth model

While our primary model focuses on highly computation-bounded, single-step look-ahead, our model class can naturally be extended to capture behavior of some game reasoners who may think more. We next consider a variant of the Intuitive Gamer model that incorporates the potential heuristic value of likely subsequent states in addition to the current state. This intermediate depth (Intuitive Gamer depth-3) model—so-called because it considers the current state, the opponent's likely response, and the possible responses to those actions—represents a moderate increase in computational load, though still requires far less computation than the Expert Gamer model.

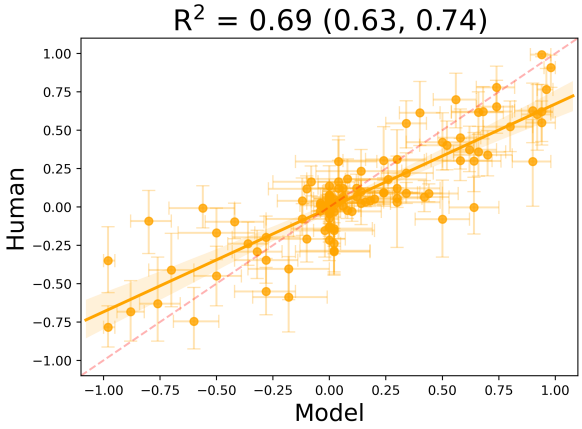


Figure 46: Intermediate depth payoff predictions. Non-flat (depth-3) variant of the Intuitive Gamer model's predicted payoffs for the 121 games in the "just think" experiment, compared against people's predicted payoffs. Error bars depict 95% CIs over bootstrapped human means and bootstrapped model-predicted payoffs under $k = 6$ simulated games for 20 simulated participants.

7.1 Model definition

The primary difference of the Intuitive Gamer depth-3 model and our primary model is the value function. Concretely, the value assigned to a position under the Intuitive Gamer depth-3 model is as follows:

$$\tilde{\mathcal{V}}^3(s_t, a_t) = \tilde{\mathcal{V}}(s_t, a_t) + \beta \times \mathcal{V}^{\text{next}}(s_t, a_t). \tag{9}$$

To compute $\mathcal{V}^{\text{next}}(s_t, a_t)$ we first apply the action a_t to obtain the subsequent state s_{t+1} and pass the turn to the opponent. We then assume that the opponent takes the most likely action under the Intuitive Gamer model, or the action that maximizes $\tilde{\mathcal{V}}(s_{t+1}, a_{t+1})$. This results in the state s_{t+2} and returns play to the original player. We then say that

$$\tilde{\mathcal{V}}^{\text{next}}(s_t, a_t) = \max_{a \in \mathcal{A}} \tilde{\mathcal{V}}(s_{t+2}, a).$$

In special cases where either the original player or opponent move twice in a row, we either decrease or increase the depth of the simulation by one move. In each case, s_{t+2} is the state in which the original player takes another turn. Actions are selected via softmax sampling over $\mathcal{V}^3$, as in the standard Intuitive Gamer model. For our experiments, we set the discount factor β to 0.5.

7.2 Results

We report predicted payoffs under the Intuitive Gamer depth-3 model in Figure 46, as well as aggregate match log likelihoods and measure of distributional alignment (under TVD) to the play and "watch-and-predict" data, respectively (Figure 47 and Figure 48). As expected, this depth-3 model is a slightly worse fit to human behavior than the standard flat (shallow; depth-1) Intuitive Gamer, though it still outperforms the expert model in many cases. Notably, the differences between the depth-3 and depth-1 models largely disappear in board states toward the end of a game (see Figure 47, right). This might be due to the fact that there are fewer possible moves in such states (making a deeper search easier to complete) or because the final outcome is more salient (encouraging the expenditure of more computational resources on search). We leave further analysis of these possibilities to future work.

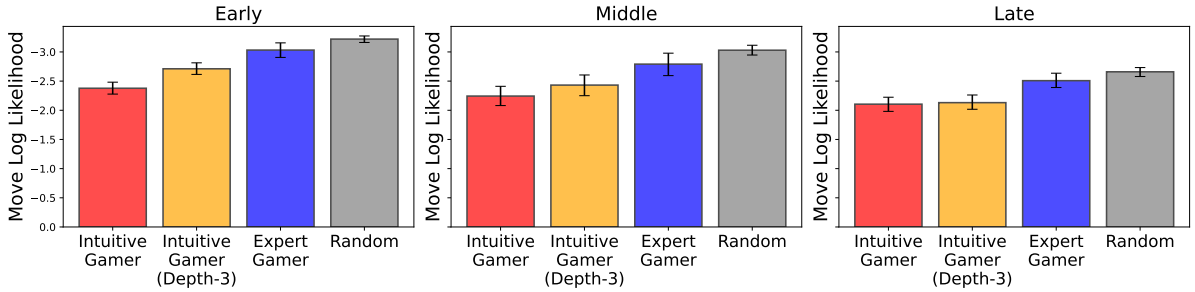


Figure 47: "Human-human play" analyses broken down by game stage. The bar plots show the log likelihood of an observed subjects' move under our different models. Error bars depict 95% bootstrapped confidence intervals around the mean log-likelihood over all games.

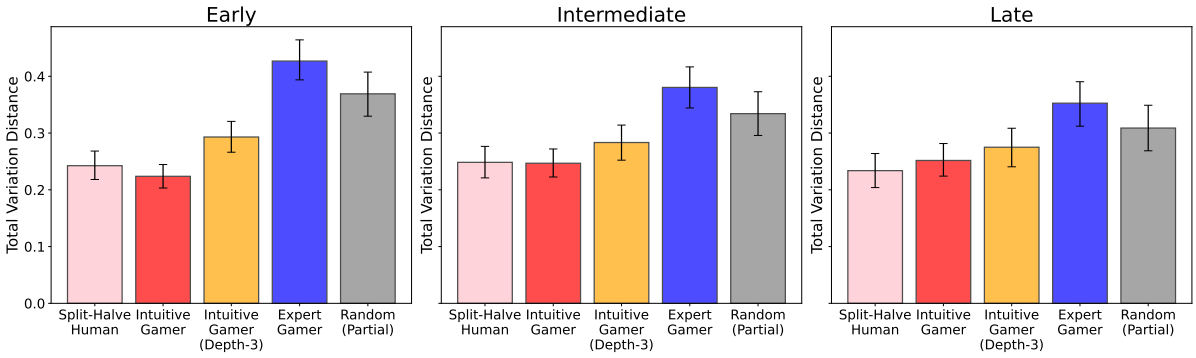


Figure 48: "Watch-and-predict" analyses broken down by game stage. The bar plots show the Total Variation Distance (lower is better) of model distributions to the human watchers' distribution, as well as compared to the split-half human watcher TVD.