Bugs on the Brain: A Mental Model Matching Approach to Cognitive Skill Acquisition in a Strategy Game

Journal of Expertise
2019. Vol. 2(2)

ISSN 2573-2773

Joe A. Wasserman[1] and Kevin Koban[2]
[1]Department of Communication Studies, West Virginia University, USA
[2]Institute for Media Research, Chemnitz University of Technology, Germany

Correspondence: Joe A. Wasserman, joe.wasserman@gmail.com

Abstract
This study investigated the development of player skill and cognitive understanding of a game over repeated plays to (a) bridge separate research traditions on skill acquisition and games learning and (b) provide deeper insight into the process of developing mental models of games. 325 participants responded to an online questionnaire with questions concerning their experience with the game, Hive, as well as both open- and closed-ended items designed to compare their understanding of the game to an expert's understanding. Open-ended items were content analyzed and modeled as a latent variable. As predicted, both player skill and mental model matching were positively associated with number of plays. Additionally, while player skill had a curvilinear relationship with number of plays that indicated diminishing returns on additional plays, that between cognitive understanding and plays appeared to be linear. The implications of these findings for the cognitive underpinnings of player skill—and for mental model matching theory in particular—are discussed. Supplemental online material is provided here: https://osf.io/3yeg2/

Keywords
games and learning, mental models, skill acquisition, board games

Introduction
For well over 50 years, there has been interest in using games and simulations as vehicles for learning (Faria, Hutchinson, Wellington, & Gold, 2009). The potential of games—both digital and analog—for instruction has been successfully leveraged for a diverse range of educational content, from teaching fractions to fourth graders (Jiménez, Arena, & Acholonu, 2011) to teaching English, math, and science to undergraduates (Crocco, Offenholley, & Hernandez, 2016). Recent meta-analyses and systematic reviews further suggest that games are effective learning tools for many topics, in many contexts, and for many people (Boyle et al., 2016; Clark, Tanner-Smith, & Killingsworth, 2016). Better understandings of the cognitive mechanisms underlying learning from games would contribute to improved learning outcomes by facilitating better design of learning games, more effective integration into educational contexts, and more targeted interventions that best take advantage of games' strengths.

Two lines of research concerned with games and learning have developed, one concerned with players' development of skill and performance (e.g., Gobet & Charness, 2006), the other related to cognitive learning of academic content through gameplay (e.g., National

Research Council, 2011). Some have argued that games *de facto* entail learning (Bryant & Fondren, 2009; Gee, 2003). Recent theorizing has proposed that learning through gameplay is a process of (a) learners developing mental models of games and (b) transferring those understandings to academic contexts (Boyan & Sherry, 2011; Martinez-Garza & Clark, 2017). Because mental models are also thought to support players' effective decision-making in games (McGloin, Wasserman, & Boyan, 2018), there is the potential to integrate perspectives on learning games as skill acquisition and mental model development. This study's goals are two-fold: first, to illustrate a methodology for measuring mental models of games, and second, to initially apply this methodology to investigate relationships among individuals' experience with, performance in, and mental models of a recreational analog game.

The diversity of definitions of games attests to the difficulty of defining them, although Stenros (2017) identified 10 dimensions of games along which definitions vary, facilitating a definition that includes a broad range of activities, analog and digital, that challenge players to accomplish goals while constrained by rules. While a full discussion of game definitions is outside the scope of this paper, the role of game mechanics in particular is of central importance to the conceptualization of games in this study. Game mechanics exist at the intersection of active players, rules, and goals. In combination, rules delimit the full range of actions that are possible for players to perform within a game (Salen & Zimmerman, 2004). Game mechanics emerge from the interactions of a game's full constellation of rules, determining the consequences of players' in-game choices. For example, rules in a platformer like *Super Mario Bros.* govern the speed of jumping and descent, as well as the effects of colliding with various objects in the game. From the interaction of these more elementary rules, as well as the game environment, game mechanics emerge, such as jumping on a winged Koopa Troopa to clear a large vertical or horizontal space that would otherwise be impossible. In sum, game mechanics (a) afford and constrain players' abilities to pursue their goals and (b) structure causal relationships among elements of a game, including the influence of player actions and influences of game elements on each other.

Game Learning as Skill Acquisition

By drawing on the aforementioned conceptualization of games as comprising sets of interacting game mechanics that players manipulate in their pursuit of goals, learning in games can be seen as a process of developing abilities to more successfully manipulate game mechanics in the service of overcoming progressively more difficult challenges (Gray & Lindstedt, 2017). From this perspective, learning in games is a process of skill acquisition. As individuals repeatedly encounter the same or similar challenges, their skills improve, although their rate of improvement typically declines with practice (Ritter & Schooler, 2001). These skills range from relatively discrete (e.g., correctly executing a particular combination of moves in a fighting game) to relatively holistic (e.g., incorporating multiple combinations of moves into an effective strategy for defeating opponents). As the ability to overcome challenges, skill can be observed as individuals overcome challenges during gameplay, e.g., by defeating other players or achieving higher scores (Isaksen & Nealen, 2016). Game performance, therefore, can serve as an observable index of skill and learning. This approach has been used in much research on skill acquisition and expertise in games (e.g., Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Stafford & Dewar, 2014).

As players repeated play a game, their performance gradually and incrementally improves over time (Campitelli & Gobet, 2008; Stafford & Dewar, 2014). Research on skill acquisition in games has taken advantage of existing performance indices of skill, such as Elo ratings (Elo, 1978) in chess databases (Vaci & Bilalić, 2017) or similar rankings (e.g., TrueSkill) in multiplayer digital games (Stafford, Devlin, Sifa, & Drachen, 2017). Game scores can also function as performance indices of player skill (Stafford & Dewar, 2014). Findings in this area are fairly consistent: the more an individual plays a game, the better they perform in that game. Nevertheless, there is variability in learning trajectories between individuals on account of various factors, such as exploring different game strategies (Stafford & Dewar, 2014) or an individual's age (Vaci & Bilalić, 2017), among others.

Game Learning as Mental Model Matching

By focusing on performance measures of game learning, research on skill acquisition has substantially advanced understandings of skill acquisition processes. Skill acquisition research in games has been supplemented by research into the cognitive underpinnings of skilled performance. In particular, research on chess has identified a number of cognitive explanations for skillful performance, such as chunks (i.e., configurations of several pieces in a position) stored in long-term memory (Chase & Simon, 1973), templates of positions in long-term memory (Gobet & Simon, 1996), depth of search while solving problems (Charness, 1981), pattern recognition of the current position (Chase & Simon, 1973), iterative pattern recognition of positions generated in the mind's eye (Gobet & Simon, 1998), and recall of key positions (Cooke, Atlas, Lane, & Berger, 1993). Because there is limited research into the cognitive underpinnings of skill in other games (Gobet, de Voogt, & Retschitzki, 2004; e.g., Othello, see Wolff, Mitchell, & Frey, 1984), it is less clear whether and how particular cognitive skills identified in this research might generalize to games more broadly.

Further investigating the cognitive underpinnings of skill in addition to game performance per se has at least three potential benefits. First, in academic game-based learning contexts, cognitive understandings of gameplay, as opposed to game skill itself, are more proximal to game-based learning objectives (Boyan & Sherry, 2011), and are therefore of greater consequence for understanding the process of learning academic material from gameplay. Second, while cognitions about a game are expected to be related to game performance, there are some conditions in which cognition and performance may diverge. For example, it is possible for players to rely on relatively shallow heuristics that are nevertheless successful without developing deeper understandings of the game (Martinez-Garza & Clark, 2017). Alternatively, players may explore sub-optimal and ultimately unsuccessful strategies in order to deepen their cognitive understanding of a game (Gray & Lindstedt, 2017). Third, in competitions between novice players, game outcomes may be due more to chance or to opponents' mistakes than to differences in skill. In such a case, although performance indices among novice players may not be a good indicator of skill, they may nevertheless have identifiable differences in cognitive understandings. This study contributes to understandings of cognitive game learning and its relationship to skill by investigating a general cognitive mechanism that has been proposed to explain learning from games: mental model matching (McGloin et al., 2018).

Mental model matching. Mental models (Craik, 1943; Johnson-Laird, 1980) are a common explanation for learning in and from games (Clark, Nelson, Sengupta, & D'Angelo, 2009). A mental model can be defined as "a cognitive representation of situations in real or imaginary worlds (including space and time), the entities found in the situation (and the states those entities are in), the interrelationships between the various entities and the situation (including causality and intentionality), and events that occur in that situation" (Roskos-Ewoldsen, Roskos-Ewoldsen, & Dillman Carpentier, 2002, pp. 110–111). Although it has been suggested that mental models of games can include entities beyond game mechanics (Wasserman & Banks, 2017), mental models of game mechanics are most central to mental model matching theory (McGloin et al., 2018).

As an explanatory mechanism for learning from games, mental models have been incorporated into a more elaborated theory of mental model matching (McGloin et al., 2018). Mental model matching specifically focuses on developing mental models of a game's mechanics, collectively referred to as a game model, which structure the causal relationships among entities in a game and afford and constrain player actions as players pursue goals (Salen & Zimmerman, 2004). When first playing a game, individuals can draw on their existing mental models of other games (Koban, Liebold, & Ohler, 2015) or activities (McGloin & Embacher, 2018) to inform their initial engagements with the game model. During play, individuals run their mental models as cognitive simulations (Craik, 1943) to guide in-game decision-making (McGloin et al., 2018) by predicting future states (Battaglia, Hamrick, & Tenenbaum, 2013) and engaging in counter-factual reasoning to infer causality (Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017). If the predictions players generate from running their mental models do not match observed outcomes, they revise their mental models based on these discrepancies, potentially by seeking out additional information in the face of failure (Lee, Liu, Jullamon, & Black, 2017). As individuals repeatedly play a game, they iteratively develop and refine their mental models of the game model such that over time, mental models align with—or *match* (McGloin et al., 2018)—the game model to an increasingly greater extent (Landriscina, 2013). As applied to training and learning, individuals are hoped to apply their mental models of a game toward developing deeper understandings of external referents (Martinez-Garza & Clark, 2017)—e.g., learning history from *Civilization* (Black, Khan, & Huang, 2014), learning to pilot aircraft from flight simulators (Korteling, Helsdingen, & Sluimer, 2017), or understanding economics from a theme park management game (Foster, 2011).

Hypotheses

Both skill acquisition and mental model matching are gradual, iterative processes that develop over time as individuals repeatedly engage with a game. Because of the fundamentally longitudinal nature of these processes, H1 predicts the following: The number of plays of a game will be positively associated with (a) performance at that game and (b) mental model matching.

Both skill and mental model matching have upper limits: (a) absolute upper limits of the best possible performance or mental models that completely match a game model or (b) potentially temporary plateaus in individuals' skill acquisition and mental model development related to suboptimal strategies (Gray, 2017). As individuals approach these limits of skill or mental model matching, the rate of their gains from practice diminish (Destefano & Gray, 2016; Heathcote, Brown, & Mewhort, 2000; Stafford & Dewar, 2014). Because additional experience should therefore yield diminishing returns, H2 predicts the following: The magnitude of the association between plays and (a) performance and (b) mental model matching will diminish with a greater number of plays.

Method

Following approval from the university's Institutional Review Board, participants were recruited via board game discussion forums (BoardGameGeek, Reddit /r/boardgames), forums on websites for playing games online (BoardGameArena and BoardSpace), and social media (Twitter and Facebook) to participate in a confidential research study about experiences with the board game Hive (Yianni, 2001). Participants completed an online questionnaire which included questions about participants' experiences with Hive, closed- and open-ended items intended to measure mental model matching, and demographics. Participants were able to opt into a drawing for one $50 gift card—88.9% of participants did so.

Study Context: Hive

Hive is a tabletop game in which two players alternate actions, placing and moving hexagonal pieces with colorful bugs printed on them. There are five different types of bugs, and like chess, each has unique movement abilities. The goal is to be the first player to surround the opponent's Queen Bee. Hive was chosen as the research context for this study for four reasons. First, a Hive strategy book (Ingersoll, 2013) was available for use as an expert reference model to develop measures of mental model matching (see Measures). Second, online platforms for playing Hive have built-in ranking systems, which provided performance measures of Hive skill comparable to those used in classic skill acquisition studies. Third, as a turn-based game, performance in Hive should be independent of motor skill. Therefore, Hive should be conducive for investigating the cognitive components of mental model matching and performance in isolation from behavioral skills. Fourth, Hive has active online player communities. Therefore, it was deemed likely that a sufficient sample of recreational Hive players could be recruited.

Participants

Participants were 325 individuals who completed the online questionnaire. Of those who voluntarily reported demographic information through open-ended self-report, the average age was 32.9 (*SD* = 8.75, median = 31, range: 18–64, *n* = 318), 90.3% (*n* = 287/318) identified as male, and 84.7% (*n* = 255/301) identified as white or Caucasian. Of the 318 who reported their country of residence, the majority were from the United States (64.5%, *n* = 205), followed by those in various European countries (20.1%, *n* = 64). Of the 315 who indicated their education, 81.6% (*n* = 257) had completed a four-year degree or higher. Of the platforms available for playing Hive, 89.5% (*n* = 291) indicated that they played the physical tabletop game, 32.9% (*n* = 107) played on BoardGameArena, 10.2% (*n* = 33) on BoardSpace, 16.6% (*n* = 54) on a tablet computer, and 8.0% (*n* = 26) on Steam. These demographics were similar to those previously reported for board-gaming communities, in which 96% identified as male, 68% were North American and 25% European, the average age was 36, and 68% had an undergraduate degree or higher (Woods, 2012).

Measures

Game plays. Participants were asked to report the number of complete games of Hive that they had played, and if they recorded their plays, to look up that number. On average, participants had played Hive 79.5 times (*SD* = 272.0, median = 20, range: 1–3330, *n* = 325), with 75% having played 50 times or fewer, and 50% 20 or fewer. The sample distribution of plays was positively skewed (skew = 8.5) and highly leptokurtic (kurtosis = 84.9). Although only 35.1% (*n* = 114) of participants indicated they kept a written record of their plays, even this percentage should contribute to the reliability of estimation. As shown in Figure 1, participants who did not keep a written record of their plays appear to have been more likely to report plays that were multiples of 5, 10, or 25 more than participants who did record their plays.

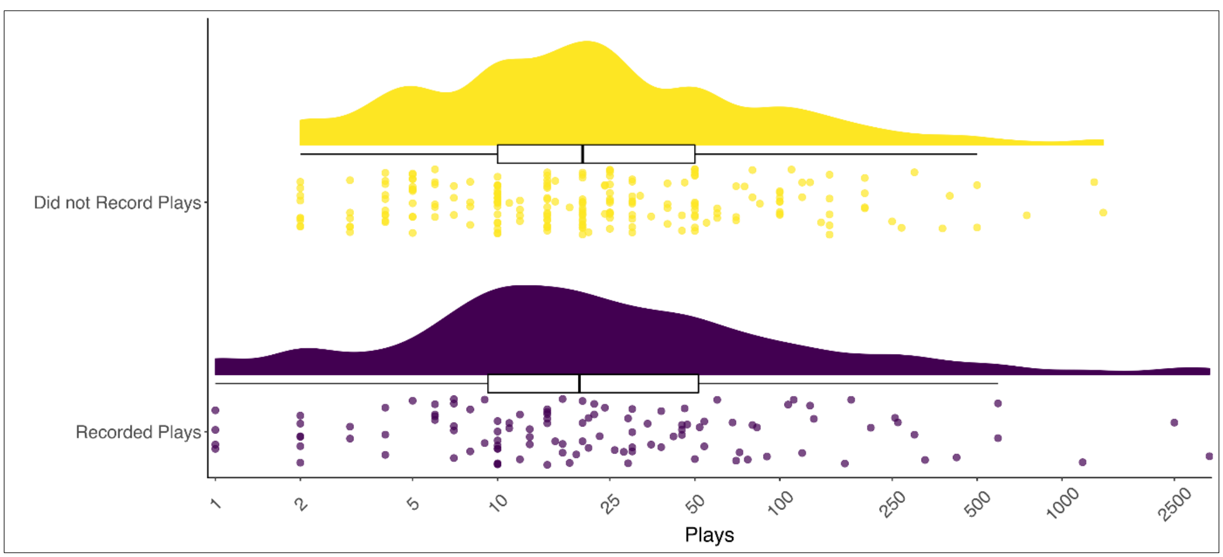

Figure 1. Raincloud plots of reported plays by whether participants recorded their plays
Note. For participants who did and did not record their plays of the game Hive, reported plays are visualized (adapted from Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019) as a density plot, boxplot (thick line: median; upper and lower hinges: 25th and 75th percentiles), and scatterplot (raw data). Data were $\log_{10}$-transformed before plotting.

Game skill. Indices of Hive skill were collected by asking participants who played Hive online to report their rank on BoardSpace (M = 1404.2, SD = 377.2, range: 0–1924, n = 29) and Elo rating on BoardGameArena (M = 259.8, SD = 188.7, range: 0–1077, n = 106). Because of the low number reporting BoardSpace rank, and because 18 reported both, only BoardGameArena Elo ratings were used for analyses (see Figure 2, middle-center panel for distribution of reported Elo ratings).

Mental model matching. Mental model matching was measured with two novel instruments that had been previously piloted with undergraduates playing Hive for the first time at a large mid-Atlantic university. An expert's understanding (Ingersoll, 2013) was used as a reference model, or benchmark, against which to compare participants. By reviewing Ingersoll (2013), closed-ended and open-ended items were developed to measure (a) understanding of the strengths and weaknesses of the five types of bugs in Hive and (b) strategic problem-solving ability in realistic Hive scenarios.

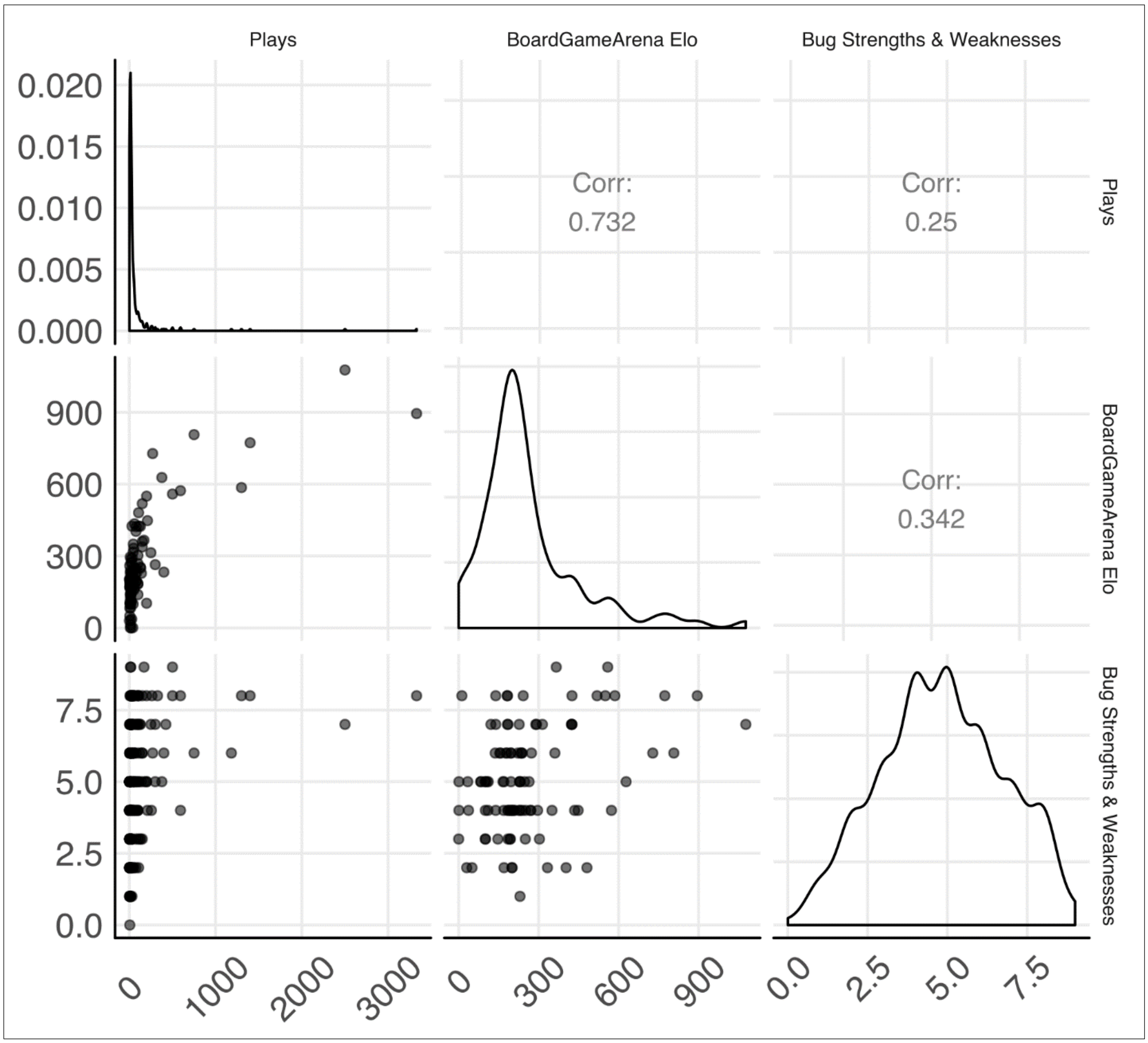

Figure 2. Scatterplots, density plots, and correlations of study variables
Note. Scatterplots are depicted in the lower half, density plots on the diagonal, and zero-order Pearson correlations in the upper half. Bug strengths and weaknesses are sum scores of observed responses, rather than estimated factor scores of the latent variable used in analyses.

Bug strengths and weaknesses. Given the conceptualization of mental models as comprising entities and their interrelations, greater mental model matching should be reflected by fuller understandings of the characteristics of important entities in a game model, as well as the relationships of those entities to others. To measure this aspect of mental model matching, participants responded to two prompts for each of the five types of bugs in Hive: “Please describe the strengths of the [bug] in as much detail as possible” and “Please describe the weaknesses of the [bug] in as much detail as possible.” To analyze this open-ended data, a codebook (see Supplement 1) of 15 dichotomous strengths and weaknesses was derived from Ingersoll’s (2013) descriptions of the strengths and weaknesses of each bug. This codebook was used with minor differences in the pilot study. All coding was conducted blind to all other participant characteristics. After 6.5 hours of iterative training, the first author and a research assistant coded a random subsample of 50 participants and achieved interrater reliabilities of $\kappa \geq .80$ on nine codes, $\kappa = .70–.80$ on five, and $\kappa = .65$ with 96% agreement on one code that occurred very rarely (<10%) in the data. Although $\kappa = .65$ was low, it was deemed acceptable because intercoder reliabilities that account for chance agreement strictly punish disagreements when rate of occurrence is low (see Gwet, 2002). Subsequently, half of the data were coded by each.

To allow each item to differentially contribute to scores on this measure, they were modeled as reflective indicators of a latent variable with dichotomous indicators, also known as item response theory (IRT) using weighted least squares estimation with diagonal weight matrix (WLSMV) in Mplus 8.1 (Muthén & Muthén, 2017). Item response theory models estimates a latent ability—in this case, mental model matching—in relation to observed responses to categorical items and the probability of a correct response as a function of ability (Raykov, 2017). A two-parameter IRT model estimates each item’s discrimination (i.e., how sharply it distinguishes between individuals of different ability) and difficulty (i.e., at what value of ability does it discriminate most sharply), whereas a one-parameter (i.e., Rasch) IRT model estimates each item’s difficulty but fixes all discriminations to be equal. A unidimensional, two-parameter IRT model was estimated using all 15 strengths and weaknesses and data from all 294 participants who responded to at least one of these items. Six items were dropped: two that over 90% of participants answered identically and four with less than 10% of variance explained, indicating poor local fit. This final nine-item measure, which had over 98% covariance coverage for all pairs of items, had good global fit, $\chi^2(27, n = 294) = 36.587$, $p = .103$. Local fit was also satisfactory (for all items, R^2 range: .14 – .45).[1] A second, more parsimonious one-parameter Rasch model was also estimated using WLSMV, which again had good global fit, $\chi^2(35, n = 294) = 44.569$, $p = .129$. The results of a Chi-square difference test[2] indicated that the two-parameter model was not a significantly better fit than the one-parameter model, scaled $\Delta\chi^2(\Delta df = 8) = 9.062$, $p = .337$. Therefore, the one-parameter model was retained for analyses (see Figure 2, bottom-right panel for distribution of observed sum scores). See Table S2 in Supplement 3 for complete parameter estimates with confidence intervals.

Strategy understanding. Because individuals draw on their mental models to guide their decision-making during gameplay, greater mental model matching should be reflected by better strategic decision-making. To measure this aspect of mental model matching, three closed-ended strategy puzzles were developed based on Ingersoll (2013). These puzzles asked participants about (a) the first bug they would play (First Bug), (b) the turn on which they would play their Queen Bee (Bee Turn), and (c) the space to which they would move an Ant to attack (Ant Attack). These items, including the ranking of responses from best to worst, were piloted with Hive players, including Ingersoll (personal communication, February 9, 2017), by asking them to respond to the items and indicate how well they thought each measured understandings of Hive. Typical responses ranged from moderately well to very well. Ingersoll verified the rank orders of response options (personal communication, February 9, 2017). Each item yielded an ordinal measure of three points (Bee Turn, Ant Attack) or four (First Bug). See Supplement 2 for questions as presented to participants, including images and square-bracketed ordinal values assigned to each response option before analyses.

A unidimensional IRT graded response model, which—akin to the aforementioned two-parameter model for dichotomous variables—estimates the likelihood of selecting a given category as a function of ability, was estimated using WLSMV for these three ordinal items. Although the bivariate covariance coverage of all items was over 99%, the model was just-identified and two of three items had very small R^2 values, indicating poor fit. See Table S3 in Supplement 3 for all parameter estimates with confidence intervals. As a follow-up analysis, Spearman's rank correlations among the three ordinal variables were performed, none of which indicated significant relationships (queen turn and first bug: $\rho = .10$, $p = .07$, $n = 324$; Queen Bee turn and Ant attack: $\rho = .05$, $p = .42$, $n = 323$; first bug and Ant attack: $\rho < .001$, $p = .99$, $n = 322$). Therefore, the unidimensional graded response model was rejected. Items were treated as ordinal variables, which Mplus models with probit regression as continuous, normally-distributed latent response variables under WLSMV.

Mental model matching measurement model. To integrate all measures of mental model matching, a model was estimated in which the final Rasch IRT model of bug strengths and weaknesses covaried with all three strategy items. This measurement model had good global fit, $\chi^2(62, n = 325) = 89.222$, $p = .013$, but one large modification index (MI = 13.116) suggested allowing the First Bug to covary with one indicator of bug strengths and weaknesses. Because this bug strength was one of the reasons to play it first, this post hoc correlation was deemed reasonable and included in the model. This final model had significantly better fit, scaled $\Delta\chi^2(\Delta df = 1) = 17.557$, $p < .0001$, and was used for analyses. See Table S4 in Supplement 3 for parameter estimates with confidence intervals.

Analysis Plan

To test H1, outcome measures were regressed on plays. To test H2, outcome measures were regressed on plays and plays-squared, expecting the linear regression coefficient to be positive (H1: increasing skill and mental model matching) and the quadratic regression coefficient to be negative (H2: diminishing rate of improvement). All models were estimated in Mplus 8.1, using MLR estimation for the continuous outcome, Elo rating, and WLSMV estimation for dichotomous and polytomous outcomes, mental model matching. To compare nested models, the quadratic term used to test H2b was calculated and included in the model for H1b; however, model parameters involving the quadratic term were fixed to zero for H1b.

Given planned comparisons between linear and quadratic models, outliers were identified in MLR as those with Mahalanobis distance Bonferroni-corrected $p < .0005$ in both MLR linear and quadratic models for skill and with Cook's $D > 1$ in both WLSMV linear and quadratic models for mental model matching. Models including and excluding these outliers were tested as a robustness check on results. For game skill, two outliers were identified, including both participants with the greatest number of plays. For mental model matching, two outliers were again identified, including the one participant with the greatest number of plays.

Results

See Figure 2 for scatterplots and zero-order correlations of observed variables. H1 predicted that the number of plays of a game would be positively associated with (a) performance at that game and (b) mental model matching. H1a was tested by regressing BoardGameArena Elo rating (for the subsample who reported one[3]) on plays (Table 1). In both linear models, regardless of whether univariate outliers were included, plays were a statistically significant positive predictor of Elo rating and explained a substantial amount of variance. Depending on the inclusion of outliers, for every 100 plays, participants' Elo ratings were predicted to be about 31 points higher (51 points excluding outliers). See Figure 3. To put these results in perspective, a player with a rating 31 points greater than their opponent's would be expected to win about 54% of their matches, with expected win percentage increasing by about 4% for each additional 31-point rating difference (Elo, 1978). H1a was supported[4].

H1b was tested by regressing (a) the latent bugs strengths and weaknesses variable (see Figure 4) and (b) the three ordinal strategy items on plays (Table 2). For most of these variables, regardless of the inclusion of univariate outliers, plays were a statistically significant positive predictor of greater mental model matching, with three exceptions: Ant Attack both with and excluding outliers, as well as First Bug excluding outliers. H1b was partially supported.

Table 1. Regression coefficients for BoardGameArena Elo rating on plays: linear and quadratic models

Parameter	*b*	*SE*	*p*	β	*SE*	*p*	R^2	*p*
				Linear (including outliers; *n* = 106)				
Intercept	211.32	12.84	<.001	1.13	0.22	<.001	.536	.003
Plays	31.24	5.80	<.001	0.73	0.12	<.001		
				Quadratic (including outliers; *n* = 106)				
Intercept	180.11	11.54	<.001	0.96	0.15	<.001	.656	<.001
Plays	73.70	11.11	<.001	1.73	0.39	<.001		
$Plays^2$	-1.59	0.35	<.001	-1.05	0.38	.005		
				Linear (excluding outliers; *n* = 104)				
Intercept	193.36	11.92	<.001	1.21	0.20	<.001	.468	.002
Plays	51.37	9.09	<.001	0.68	0.11	<.001		
				Quadratic (excluding outliers; *n* = 104)				
Intercept	159.05	11.57	<.001	1.00	0.14	<.001	.574	<.001
Plays	115.83	18.50	<.001	1.54	0.34	<.001		
$Plays^2$	-5.68	1.48	<.001	-0.92	0.33	.006		

Note. Plays were centered on 1; intercepts are predicted values for participants who have played Hive once. After centering, plays were divided by 100 for estimation and reporting. Because models are just-identified, no global fit statistics are available. See Tables S9–S12 in Supplement 3 for parameter estimates with confidence intervals. *b* = unstandardized, β = fully standardized.

H2 predicted that the magnitude of the association between plays and (a) performance and (b) mental model matching would diminish with a greater number of plays. H2a was tested by regressing BoardGameArena Elo on plays and a quadratic plays term, i.e., plays-squared (Table 1). Regardless of whether univariate outliers were included, the quadratic term was a statistically significant predictor of Elo. Every additional 100 plays made was associated with about 74 points greater Elo rating (116 points excluding outliers), while the magnitude of this association simultaneously decreased by about 1.6 points for each 100 plays (5.7 excluding outliers). See Figure 3. H2a was supported.

Table 2. Regression coefficients for mental model matching on plays: linear and quadratic models

DV	Parameter	*b*	*SE*	*p*	β	*SE*	*p*	R^2	*p*
				Linear (including outliers; *n* = 325)					
Bugs	Plays	0.53	0.13	<.001	0.82	0.07	<.001	.674	<.001
Queen Turn	Plays	0.14	0.06	.027	0.36	0.16	.025	.123	.209
First Bug	Plays	0.25	0.10	.013	0.65	0.24	.006	.311	.072
Ant Attack	Plays	0.13	0.09	.151	0.35	0.22	.118	.112	.420
				Quadratic (including outliers; *n* = 325)					
Bugs	Plays	0.53	0.13	<.001	1.24	0.24	<.001	.251	.002
	$Plays^2$	-0.02	0.006	.004	-1.03	0.33	.002		
Queen Turn	Plays	0.14	0.06	.027	0.36	0.16	.025	.026	.192
	$Plays^2$	-0.004	0.004	.110	-0.03	0.02	.110		
First Bug	Plays	0.25	0.10	.013	0.65	0.24	.006	.117	.398
	$Plays^2$	-0.01	0.009	.559	-0.04	0.06	.559		
Ant Attack	Plays	0.13	0.09	.151	0.35	0.22	.118	.045	.587
	$Plays^2$	-0.002	0.007	.750	-0.02	0.05	.705		
				Linear (excluding outliers; *n* = 323)					
Bugs	Plays	0.59	0.17	<.001	0.76	0.09	<.001	.570	<.001
Queen Turn	Plays	0.21	0.08	.013	0.38	0.15	.011	.139	.152
First Bug	Plays	0.29	0.18	.112	0.56	0.26	.033	.244	.293
Ant Attack	Plays	0.14	0.09	.112	0.27	0.16	.096	.070	.393
				Quadratic (excluding outliers; *n* = 323)					
Bugs	Plays	0.59	0.17	<.001	0.99	0.38	.009	.266	.225
	$Plays^2$	-0.02	0.03	.478	-0.61	0.94	.518		
Queen Turn	Plays	0.21	0.08	.013	0.38	0.15	.011	.030	.197
	$Plays^2$	-0.009	0.006	.111	-0.04	0.02	.111		
First Bug	Plays	0.29	0.18	.111	0.56	0.26	.033	.190	.780
	$Plays^2$	-0.003	0.043	.953	-0.01	0.16	.952		
Ant Attack	Plays	0.14	0.09	.112	0.27	0.16	.096	.024	.292
	$Plays^2$	-0.004	0.01	.534	-0.02	0.02	.534		

Note. Plays were divided by 100 for estimation and reporting. Bugs: bug strengths and weaknesses. All models had good global fit, χ^2 $p \geq .19$. In Supplement 3, see Table 1 for global model fit indices and Tables S5–S8 for complete parameter estimates with confidence intervals. *b* = unstandardized. For Bugs, β = fully standardized; for all other, ordinal outcome variables, β = standardized *x*.

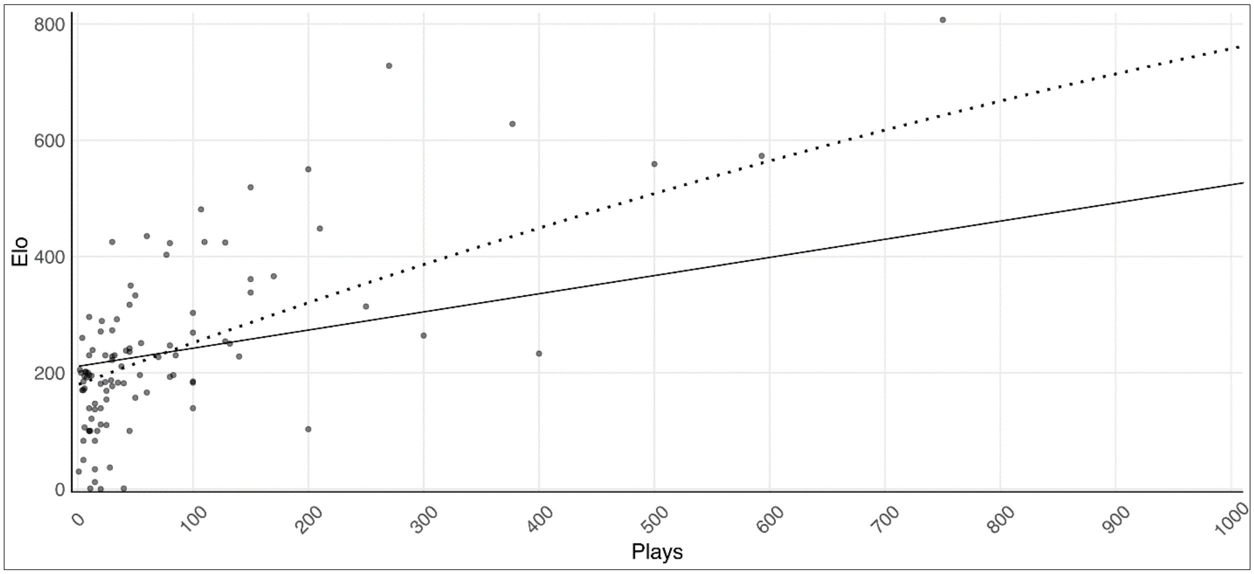

Figure 3. Elo regressed on plays, linear and quadratic models including outliers
Note. Linear (solid line) and quadratic (dotted line) models of Elo regressed on plays, including outliers (see Table 1), are overlaid on a scatterplot of the raw data up to 1000 plays.

H2b was tested by regressing mental model matching variables on linear and quadratic terms of plays (Table 2). Model comparisons indicated that the quadratic model had significantly better global fit only when outliers were included, scaled $\Delta\chi^2(\Delta df = 4) = 11.839$, $p = .019$, but not when outliers were excluded, scaled $\Delta\chi^2(\Delta df = 4) = 3.326$, $p = .505$. Closer inspection of regression coefficients indicated that the only significant quadratic term was that predicting understanding of bug strengths and weaknesses in the model including outliers, such that every additional 100 plays was associated with .53 *SD*s greater mental model matching, while the magnitude of this association simultaneously decreased by .02 *SD*s per 100 plays. See Figure 4. H2b was not supported.

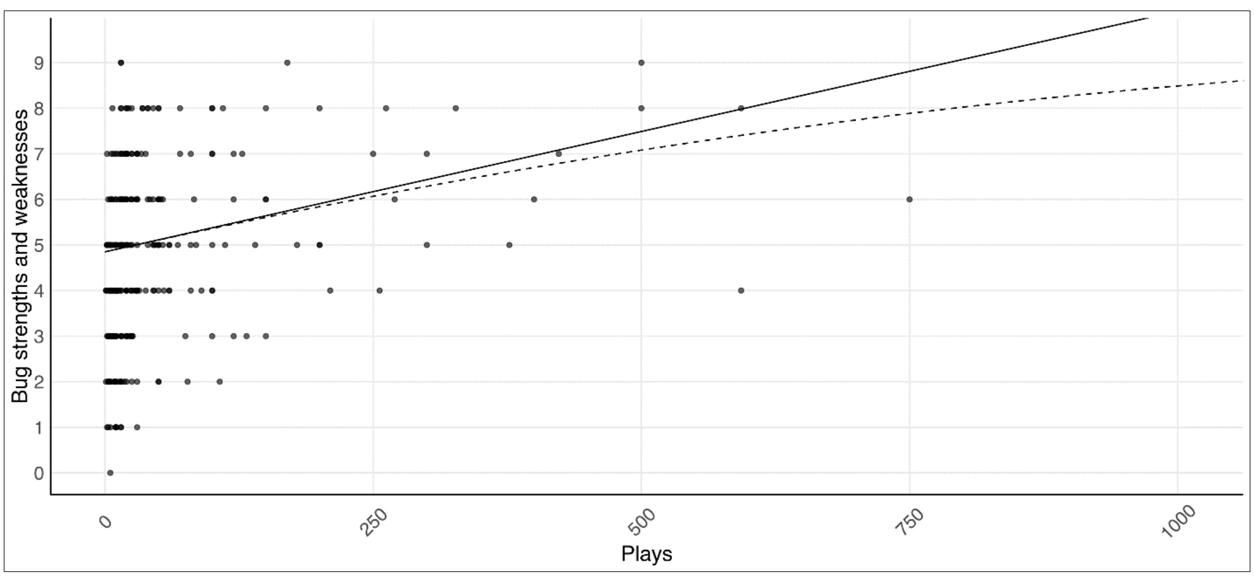

Figure 4. Bug strengths and weaknesses regressed on plays, linear and quadratic models including outliers
Note. Linear (solid line) and quadratic (dotted line) models of bug strengths and weaknesses regressed on plays, including outliers (see Table 2), are overlaid on a scatterplot of the raw data up to 1000 plays. In the scatterplot, bug strengths and weaknesses are sum scores of observed responses.

To further probe the results of the probit regressions of Queen Turn, First Bug, and Ant Attack on plays, estimated probabilities of each response as a function of plays were plotted using the parameters of the linear model including outliers (Figure 5). For all items, the probability of choosing the best response (category 4 for First Bug, 3 for both others) closely approached the asymptote of 1 as plays approached 1000. The worst response (category 1) was always the least likely to be selected, even for participants who had only played Hive once. The overall unlikelihood of the worst response (and second worst in the case of First Bug only) may have weakened these measures' sensitivity to game experience.

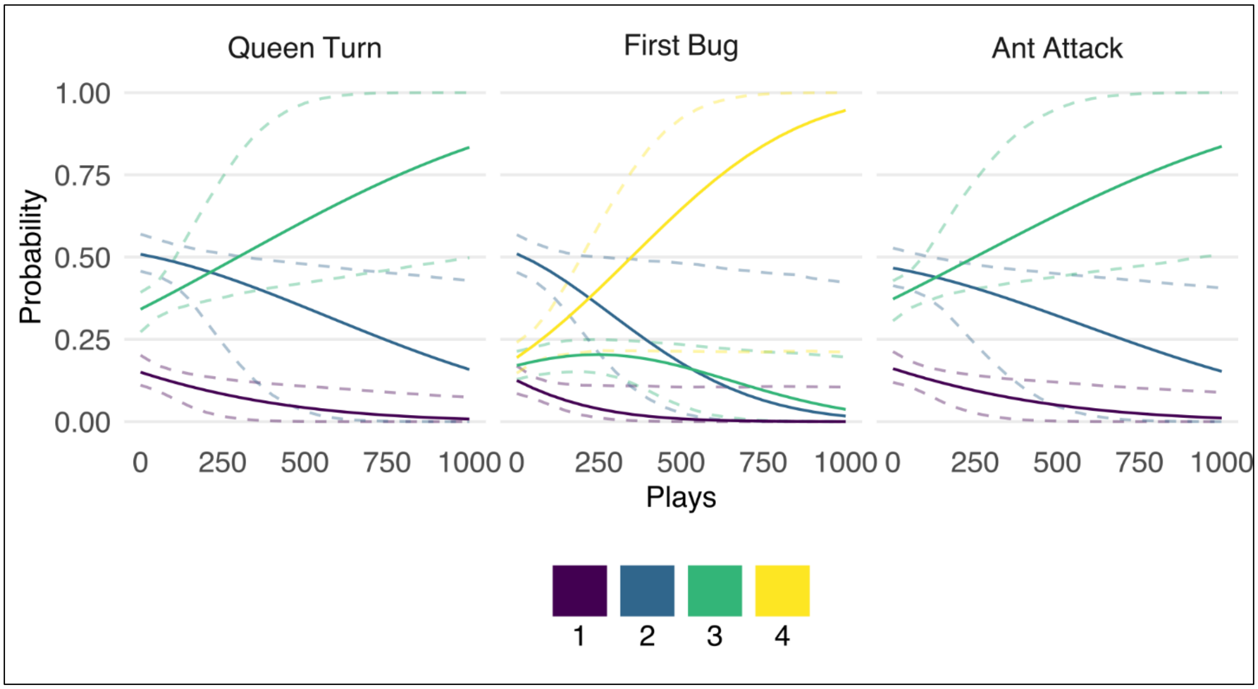

Figure 5. Estimated probabilities of responses to strategy understanding questions
Note. Expected probability of selecting each successively better response to strategy understanding items, from worst (1) to best (3 or 4, depending on the item), as a function of plays in the linear model including outliers (see Table 2). Dashed lines are non-symmetric bootstrapped 95% confidence intervals with 5000 bootstraps.

Discussion

This study examined the relationship between individuals' game experience, game skill, and understanding of the game mechanics in the context of the tabletop game Hive. As predicted, skill (as measured by Elo rating) as well as understanding of the game mechanics (conceptualized and measured as mental model matching) were positively associated with reported game experience. Also as predicted, the magnitude of the association between plays and Elo rating diminished with additional plays. In contrast, only one measure of mental model matching exhibited similar diminishing returns, and only when outliers were included. These findings are largely consistent with established findings in skill acquisition, are broadly consistent with mental model matching theory, and provide initial support for the approach developed for measuring mental model matching.

This study found support for a curvilinear relationship between game skill, operationalized as Elo rating, and game experience, such that more plays were associated with greater skill, but also found that this positive association simultaneously diminished with additional plays. This finding is consistent with the well-established finding that skill acquisition occurs over repeated engagements with a game (Stafford & Dewar, 2014) in a new context: the modern tabletop board game Hive. Simply put, the more an individual plays a game, the better they are likely to perform at it. Furthermore, the curvilinear relationship between experience and skill was re-confirmed, such that additional experience yields diminishing returns (Ritter & Schooler, 2001). Nevertheless, because of the sparseness of participants who reported playing Hive more than 50 times and the lack of longitudinal data, this study's observed curvilinear relationship between number of plays and game skill is suggestive but by no means definitive.

This study found that mental model matching—i.e., the accuracy with which individuals' cognitive representations of a game are aligned with the game model—was also positively associated with game experience. For most measures, a linear model appeared to better reflect the relationship between plays and mental model matching better than the expected curvilinear model including diminishing returns. This finding is still consistent with one of mental model matching's central propositions: that players iteratively refine their mental models of a game through repeated plays (McGloin et al., 2018). Through this process, the players' mental models come to more closely align with a game's functional structure that emerges from its mechanics. The gradual nature of this process may explain why multiple gameplay sessions are needed to effect significant learning gains from playing educational games (Clark et al., 2016): Single sessions do not enable learners to build robust mental models. Again, however, these results are merely suggestive, not only because of aforementioned scarcity of participants with many plays and cross-sectional data, but also because (a) more research is needed to evaluate the mental model matching measurement method developed here and (b) alternative cognitive explanations cannot be ruled out, as discussed in the limitations below.

Measuring Mental Model Matching

Methodologically, this study suggests that operationalizing mental model matching by comparing individuals' understandings to a reference model—in this case, an expert's understanding of the game—could be a fruitful approach across different games. The open-ended questions on bug strengths and weaknesses reflect declarative knowledge of the game, paralleling the multiple-choice chess knowledge test developed by Pfau and Murphy (1988), which was found to positively correlate at least moderately with Elo rating and three established measures of chess skill. The closed-ended strategy puzzles developed in this study bear similarity to existing measurement strategies in research on chess skill that ask participants, for instance, to evaluate board positions (e.g., Holding, 1979) or to suggest a best subsequent move (de Groot, 1978). Similar items can be designed for any (turn-based) strategy game, both tabletop and digital, once a respective game model is defined as a reference.

This method can be applied to games with very different game mechanics. Irrespective of a game's unique mechanics, its game model could be specified, for example, by collecting and analyzing concurrent and retrospective verbal protocols of highly proficient players (see, e.g., Boot, Sumner, Towne, Rodriguez, & Anders Ericsson, 2017). Based on these game models, open- and closed-ended questions, as well as codebooks for analyzing performance records, can be developed to compare individuals' understandings to these reference models. Valid insights into individuals' current or progressing understandings of different games can inform learning research in a meaningful way. For example, using these methods, researchers could begin to pinpoint the elements of a game model that are most crucial for transferring declarative and procedural knowledge from games to academic and other real-world contexts.

Nevertheless, the incidental finding that the closed-ended strategy measures employed in this study did not fit a unidimensional IRT model indicates the complexity of this task. There are several potential explanations for this finding, each of which should be attended to in future research. It is possible that mental model matching is hierarchical, e.g., starting from more elementary understandings of a game's elements in isolation before incorporating the relationships among them in more holistic mental structures. Considering the measures used in this study, it is possible that understanding individual bugs' strengths and weaknesses—which are relatively discrete—are foundational to developing predictively or practically effective mental models needed to solve more holistic strategy problems. Alternatively, it may be that individuals learn sub-components of an overall game before integrating them into a cohesive understanding of the whole (on part-whole transfer, see Proctor & Vu, 2006). In this case, it would be likely that some individuals with a good understanding of one part of the game could have a weak understanding of others. Such possibilities would necessitate multidimensional measures of mental model matching that encompass several components of individuals' understandings of the game model (e.g., opening/midgame/ endgame, pattern recognition).

Linear Versus Curvilinear Relationships between Practice and Mental Model Matching

The lack of statistically significant curvilinear relationships between plays and mental model matching bears further mention. Because participants were recruited primarily from board-gaming communities, even Hive novices among these participants may have had atypically good understanding of the game, thus creating a floor effect and limiting the range of responses. This possibility is supported by the very low expected probability of selecting the worst responses to strategy understanding items (see Figure 5). At the opposite end of the experience spectrum, potential diminishing returns may be most noticeable at a number of plays greater than that reported by the majority of participants, of whom 75% had played 50 times or fewer, and 50% 20 or fewer. These data might not have been dense enough at higher levels of expertise to reveal a curvilinear relationship—or, more fundamentally, to provide definitive evidence for any specific functional form. Additionally, the positive skew of the sample distribution of plays might indicate a dropout bias, in which data overestimate the positive linear relationship between skill practice and performance. Of particular relevance to this study, individuals who drop out of a game early often do so for performance-related reasons, such as after failing to improve sufficiently within a number of plays (Steyvers & Benjamin, 2018). Because learning functions aggregate many individual learning curves, systematic dropouts of individuals who fail to improve can exaggerate the effect of continued playing. This dropout bias typically occurs in datasets without strict experimental control, making the cross-sectional data obtained in this study vulnerable. Finally, measures may have been too insensitive to identify diminishing returns. The three ordinal items in particular were rather coarse-grained, which might have been insufficient to cover the breadth of strategic actions Hive allows and identify subtler differences in strategic understanding among players. For these reasons, this study, while being able to confirm a positive relationship, provides only preliminary evidence for a curvilinear relationship between mental model acquisition and game experience.

Limitations and Future Research

The measures of mental model matching used in this study were novel and were not straightforwardly unidimensional. Moreover, these measures were based on the understanding of a single Hive expert (i.e,, Ingersoll, 2013). It is possible that different results would have been obtained if a different expert's understanding had been used as the reference model. However, as a first attempt, these measures appear to have some construct validity and generally accorded with theoretically-derived and empirically-driven predictions. Given the difficulty of identifying unidimensional IRT models of mental model matching, future research should investigate the potentially hierarchical or part-whole nature of mental model matching. Greater insight into the structure of this process would facilitate better measurement and yield actionable empirical insights for applying games in learning contexts. For example, for a given game and learning objective, the ability to identify a particular part of a mental model or stage in hierarchical mental model matching would allow educators to use games more efficiently to teach. Furthermore, it should be possible to construct a common reference model that synthesizes several experts' understandings, which may enhance the validity of mental model matching measurement.

Although the methodological approach outlined in this study drew directly from mental model matching theory, the items developed to measure individuals' declarative and strategic knowledge of the game model do not directly address how this information is cognitively encoded. While mental models are traditionally considered to be structured analogously to the phenomena they represent (Craik, 1943; Johnson-Laird, 1980), other researchers have suggested that information is encoded in a propositional format (Pylyshyn, 2002), depictive format (Kosslyn, Ganis, & Thompson, 2003), or a combination of both (and potentially more) formats (Pearson & Kosslyn, 2015). Consequently, this study contributes to previous work establishing mental model matching theory as a fruitful approach to describe and explain skill acquisition in games (e.g., Boyan & Sherry, 2011; Martinez-Garza & Clark, 2017; McGloin et al., 2018); however, it does not provide evidence of the *structure* of knowledge or of mental models, nor can it establish that skill acquisition is identical to mental model matching.

Demographically, participants were relatively homogeneously male, white, and educated. However, as aforementioned, these demographics are not dissimilar from those previously found among board-gamers. Perhaps more critically for this study, players' experience with Hive was skewed toward novices. Such a distribution is typical for many games: Most individuals play very few times, while a small number of more dedicated players play many times (e.g., Pirker, Rattinger, Drachen, & Sifa, 2016; Stafford & Dewar, 2014). Participants who reported BoardGameArena Elo ratings were also more experienced than others and were limited to 106 participants. Future cross-sectional research on mental model matching and game skill development should attempt to recruit participants to represent more evenly a broader range of expertise to facilitate more precise model estimation across a larger range of experience.

Finally, this study was cross-sectional, limiting the ability to make causal claims about playing and mental model matching. Even so, this study's findings were largely consistent with a mental model matching explanation for skill acquisition in games. Future research should experimentally and longitudinally investigate mental model matching processes in order to (a) test model matching as an explanatory mechanism for skill acquisition; (b) microgenetically study individuals' learning curves, which have been found to have distinct dynamics from aggregated learning curves (Donner & Hardy, 2015); and (c) explore potential moderators that may enhance or inhibit mental model matching, such as gameplay motivations (Sherry, Lucas, Greenberg, & Lachlan, 2006), epistemic stances (Martinez-Garza & Clark, 2017), and specific learning strategies like making guided errors (Lorenzet, Salas, & Tannenbaum, 2005), deliberate practice and training (Campitelli & Gobet, 2008), or observing effective opponents (Weintrop & Wilensky, 2013).

Conclusion

This study's findings suggest that game skill and game understanding, conceptualized as mental models, proceed largely in tandem with game experience. Additionally, the method for measuring game understanding developed in this study holds promise for studying mental model matching across a variety of game types. Further research is merited into mental model matching as an explanatory mechanism underlying game performance, as well as into the potential for individuals to transfer mental models of games to other contexts.

Endnotes

1. Additional global fit information for all models is reported in Table S1 in Supplement 3.
2. Satorra-Bentler scaled Chi-square difference tests (Satorra & Bentler, 2010) were used for all model comparisons among those estimated via WLSMV.
3. This subsample (n = 106) played more games of *Hive* (M = 156.2, SD = 442.4) than the overall sample.

4. Post hoc subsample analyses were performed for all hypothesis tests to compare results for participants who recorded their number of plays (n = 114) and participants who only estimated how often they had played (n = 211). Overall, results appeared to be similar between these groups, but these conclusions are tentative because of small group sizes. For detailed results, see Tables S13 and S14 in Supplement 3.

Acknowledgements

The authors would like to thank Ashleigh C. Swain for coding open-ended data, Nicholas David Bowman for feedback on earlier versions of this manuscript, Emily B. Louk for running sessions and coding data during the pilot, and Tom Stafford and Guillermo Campitelli for their constructive feedback during the review process at Journal of Expertise.

Authors' Declarations

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that they conducted the research reported in this article in accordance with the Ethical Principles of the Journal of Expertise.

The authors declare that they are not able to make the dataset publicly available but are able to provide it upon request.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*(63). doi:10.12688/wellcomeopenres.15191.1

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*, 18327–18332. doi:10.1073/pnas.1306572110

Black, J. B., Khan, S. A., & Huang, S.-C. D. (2014). Video and computer games as grounding experiences for learning. In F. C. Blumberg (Ed.), *Learning by playing: Video gaming in education* (pp. 290–301). New York, NY: Oxford University Press.

Boot, W. R., Sumner, A., Towne, T. J., Rodriguez, P., & Anders Ericsson, K. (2017). Applying aspects of the expert performance approach to better understand the structure of skill and mechanisms of skill acquisition in video games. *Topics in Cognitive Science*, *9*, 413–436. doi:10.1111/tops.12230

Boyan, A., & Sherry, J. L. (2011). The challenge in creating games for education: Aligning mental models with game models. *Child Development Perspectives*, *5*, 82–87. doi:10.1111/j.1750-8606.2011.00160.x

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. doi:10.1016/j.compedu.2015.11.003

Bryant, J., & Fondren, W. (2009). Psychological and communicological theories of learning and emotion underlying serious games. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 103–116). New York, NY: Routledge.

Campitelli, G., & Gobet, F. (2008). The role of practice in chess: A longitudinal study. *Learning and Individual Differences*, *18*, 446–458. doi:10.1016/j.lindif.2007.11.006

Charness, N. (1981). Aging and skilled problem solving. *Journal of Experimental Psychology: General*, *110*, 21–38. doi:10.1037/0096-3445.110.1.21

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, *19*, 151–165. doi:10.1002/acp.1106

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81. doi:10.1016/0010-0285(73)90004-2

Clark, D. B., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). Rethinking science learning through digital games and simulations: Genres, examples, and evidence. In *Learning science: Computer games, simulations, and education*. Washington, DC.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*, 79–122. doi:10.3102/0034654315582065

Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). Role of high-level knowledge in memory for chess positions. *The American Journal of Psychology*, *106*, 321–351. doi:10.2307/1423181

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge, England: Cambridge University Press.

Crocco, F., Offenholley, K., & Hernandez, C. (2016). A proof-of-concept study of game-based learning in higher education. *Simulation & Gaming*, *47*, 403–422. doi:10.1177/1046878116632484

de Groot, A. D. de. (1978). *Thought and choice in chess* (2nd ed.). The Hague, Netherlands: Mouton Publishers.

Destefano, M., & Gray, W. D. (2016). Where should researchers look for strategy discoveries during the acquisition of complex task performance? The case of Space Fortress. In A. Papafragou, D. Grodner, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Donner, Y., & Hardy, J. L. (2015). Piecewise power laws in individual learning curves. *Psychonomic Bulletin & Review*, *22*(5), 1308–1319. doi:10.3758/s13423-015-0811-x

Elo, A. E. (1978). *The rating of chess players, past and present*. New York, NY: Arco Publishing.

Faria, A. J., Hutchinson, D., Wellington, W. J., & Gold, S. (2009). Developments in business gaming: A review of the past 40 years. *Simulation & Gaming*, *40*, 464–487. doi:10.1177/1046878108327585

Foster, A. N. (2011). The process of learning in a simulation strategy game: Disciplinary knowledge construction. *Journal of Educational Computing Research*, *45*, 1–27. doi:10.2190/EC.45.1.a

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science*, *28*, 1731–1744. doi:10.1177/0956797617713053

Gobet, F., & Charness, N. (2006). Expertise in chess. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 523–538). Cambridge, England: Cambridge University Press.

Gobet, F., de Voogt, A., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. New York, NY: Psychology Press.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1–40. doi:10.1006/cogp.1996.0011

Gobet, F., & Simon, H. A. (1998). Pattern recognition makes search possible: Comments on Holding (1992). *Psychological Research*, *61*, 204–208. doi:10.1007/s004260050025

Gray, W. D. (2017). Plateaus and asymptotes: Spurious and real limits in human performance. *Current Directions in Psychological Science*, *26*, 59–67. doi:10.1177/0963721416672904

Gray, W. D., & Lindstedt, J. K. (2017). Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cognitive Science*, *41*, 1838–1870. doi:10.1111/cogs.12412

Gwet, K. (2002). Inter-rater reliability: Dependency on trait prevalence and marginal homogeneity. *Statistical Methods For Inter-Rater Reliability Assessment Series*, *2*, 1–9.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207. doi:10.3758/BF03212979

Holding, D. H. (1979). The evaluation of chess positions. *Simulation & Games*, *10*, 207–221. doi:10.1177/104687817901000205

Ingersoll, R. (2013). *Play Hive like a champion: Strategy, tactics and commentary* (2nd ed.). Port Orange, FL: Randall M. Ingersoll.

Isaksen, A., & Nealen, A. (2016). A statistical analysis of player improvement and single-player high scores. In *Proceedings of the First International Joint Conference of DiGRA and FDG, 13*.

Jiménez, O., Arena, D., & Acholonu, U. (2011). Tug-of-War: A card game for pulling students to fractions fluency. In C. Steinkuehler, C. Martin, & A. Ochsner (Eds.), *Proceedings of the 7th International Conference on Games + Learning + Society Conference* (pp. 119–127).

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*, 71–115. doi:10.1207/s15516709cog0401_4

Koban, K., Liebold, B., & Ohler, P. (2015). Good newbie or poor newbie? Determinants of video game skill acquisition at an early stage. *Learning and Collaboration Technologies*, 608–619. doi:10.1007/978-3-319-20609-7_57

Korteling, H. J. E., Helsdingen, A. S., & Sluimer, R. R. (2017). An empirical evaluation of transfer-of-training of two flight simulation games. *Simulation & Gaming*, *48*, 8–35. doi:10.1177/1046878116671057

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2003). Mental imagery: against the nihilistic hypothesis. *Trends in Cognitive Sciences*, *7*, 109–111. doi:10.1016/S1364-6613(03)00025-1

Landriscina, F. (2013). *Simulation and learning: A model-centered approach*. New York, NY: Springer.

Lee, A., Liu, C., Jullamon, M., & Black, J. (2017). How'd that happen?! Failure in game spaces to prepare students for future learning. In K. E. H. Caldwell, S. Seyler, A. Ochsner, & C. Steinkuehler (Eds.), *GLS Conference Proceedings 2017* (pp. 119–128). Pittsburgh, PA: Carnegie Mellon ETC Press.

Lorenzet, S. J., Salas, E., & Tannenbaum, S. I. (2005). Benefiting from mistakes: The impact of guided errors on learning, performance, and self-efficacy. *Human Resource Development Quarterly*, *16*, 301–322. doi:10.1002/hrdq.1141

Martinez-Garza, M. M., & Clark, D. B. (2017). Two systems, two stances: A novel theoretical framework for model-based learning in digital games. In P. Wouters & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 37–58). doi:10.1007/978-3-319-39298-1_3

McGloin, R., & Embacher, K. (2018). "Just like riding a bike": A model matching approach to predicting the enjoyment of a cycling exergame experience. *Media Psychology*, *21*, 486–505. doi:10.1080/15213269.2017.1311269

McGloin, R., Wasserman, J. A., & Boyan, A. (2018). Model matching theory: A framework for examining the alignment between game mechanics and mental models. *Media and Communication*, *6*, 126–136. doi:10.17645/mac.v6i2.1326

Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.

National Research Council. (2011). *Learning science through computer games and simulations* (M. A. Honey & M. L. Hilton, eds.). doi:10.17226/13078

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences*, *112*, 10089–10092. doi:10.1073/pnas.1504933112

Pfau, H. D., & Murphy, M. D. (1988). Role of verbal knowledge in chess skill. *The American Journal of Psychology*, *101*, 73–86. doi:10.2307/1422794

Pirker, J., Rattinger, A., Drachen, A., & Sifa, R. (2016). *Analyzing player networks in Destiny*.

Proctor, R. W., & Vu, K. L. (2006). Laboratory studies of training, skill acquisition, and retention of performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 265–286). Cambridge, England: Cambridge University Press.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*, 157–182. doi:10.1017/S0140525X02000043

Raykov, T. (2017). *Item response theory*. Philadelphia, PA: Statistical Horizons.

Ritter, F. E., & Schooler, L. J. (2001). Learning curve, the. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 8602–8605). doi:10.1016/B0-08-043076-7/01480-7

Roskos-Ewoldsen, D. R., Roskos-Ewoldsen, B., & Dillman Carpentier, F. R. (2002). Media priming: A synthesis. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 97–120). Mahwah, NJ: Erlbaum.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference Chi-square test statistic. *Psychometrika*, *75*, 243–248. doi:10.1007/s11336-009-9135-y

Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. (2006). Video game uses and gratifications as predictors of use and game preference. In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 213–224). New York, NY: Routledge.

Stafford, T., Devlin, S., Sifa, R., & Drachen, A. (2017). Exploration and skill acquisition in a major online game. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 3249–3254).

Stafford, T., & Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological Science*, *25*, 511–518. doi:10.1177/0956797613511466

Stenros, J. (2017). The game definition game: A review. *Games and Culture*, *12*, 499–520. doi:10.1177/1555412016655679

Steyvers, M., & Benjamin, A. S. (2018). The joint contribution of participation and performance to learning functions: Exploring the effects of age in large-scale data sets. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-018-1128-2

Vaci, N., & Bilalić, M. (2017). Chess databases as a research vehicle in psychology: Modeling large data. *Behavior Research Methods*, *49*, 1227–1240. doi:10.3758/s13428-016-0782-5

Wasserman, J. A., & Banks, J. (2017). Details and dynamics: Mental models of complex systems in game-based learning. *Simulation & Gaming*, *48*, 603–624. doi:10.1177/1046878117715056

Weintrop, D., & Wilensky, U. (2013). Know your enemy: Learning from in-game opponents. In *Proceedings of the 12th International Conference on Interaction Design and Children*, 408–411. doi:10.1145/2485760.2485789

Wolff, A. S., Mitchell, D. H., & Frey, P. W. (1984). Perceptual skill in the game of Othello. *The Journal of Psychology: Interdisciplinary and Applied*, *118*, 7–16. doi:10.1080/00223980.1984.9712586

Woods, S. (2012). *Eurogames: The design, culture and play of modern European board games*. Jefferson, NC: McFarland.

Yianni, J. (2001). *Hive* [Board game]. Hertfordshire, United Kingdom: Gen42 Games.

Received: 15 January 2019
Revision received: 17 April 2019
Accepted: 20 April 2019