"Not another Z piece!" Adaptive Difficulty in TETRIS

Katharina Spiel[2,1], Sven Bertel[1], Fares Kayali[2]
[1]Bauhaus-Universitaet Weimar, Germany
[2]TU Wien (Vienna University of Technology), Austria
katharina.spiel@tuwien.ac.at sven.bertel@uni-weimar.de fares.kayali@tuwien.ac.at

ABSTRACT
Difficulty in TETRIS is adjusted by adapting the speed with which blocks fall. In this contribution, we describe results of an exploratory study in which we investigated relationships between players' performance and their subjective assessment of difficulty and fun. We tested five different algorithms that, instead of adjusting game speed, adjust difficulty by choosing blocks based on the current game state. With our results, we establish pile height and bumpiness as parameters that indicate the performance of a player during a live game, discuss the inherent difficulty of different block choosing algorithms and show how the relationship between fun and perceived difficulty varies for distinct player groups. With regard to adapting difficulty, we argue that one can still teach an old dog such a TETRIS a lot of new tricks.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Tetris; Perceived Difficulty; Fun; User Study

INTRODUCTION
In order to be able to test the viability of alternative ways of adaptation in well known games, it is important to know how they each relate to fun and difficulty (performance-related and perceived). The micro-analysis of a well-known and beloved game sheds more light on the ongoing debate on the role of challenge for engagement in games (compare findings of [1], where high challenge yielded high engagement and [17], who report on the opposite effect). However, recent research indicates that there is no one-way relationship as it also, for example, matters how difficulty is adjusted [21].

For those readers who are unfamiliar with TETRIS, we will start by providing a short explanation of the game. We will then give an overview of parameters that can be suitably employed to describe individual TETRIS game states. Next, we will present a set of algorithms for choosing how blocks are spawned. We will then describe our study, and ensuing results, in which we asked about the perceived difficulty of the different block choosing algorithms. Results of the study show how suitable various game state parameters respectively are as a basis for adapting the game to a user's play. We will conclude with a discussion of results and the more general ramifications that they possess for games researchers and for adapting gameplay.

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05...$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025721

RELATED WORK
Adaptivity is a big concern of game research in order to create engaging games that are challenging for a range of different skill levels [18]. The effects and effectiveness of adaptivity are often investigated with games that target a limited group of players (e.g., educational/serious games in [19]) or are new to most participants in user studies (e.g., HEX in [25]). TETRIS is familiar to a large range of age groups, skill levels, gender populations and occupations. This means that there exist many players who have specific expectations on the behaviour of the game. When adapting the difficulty in TETRIS, this expected behaviour should still occur, so that players are not disrupted in their gameplay.

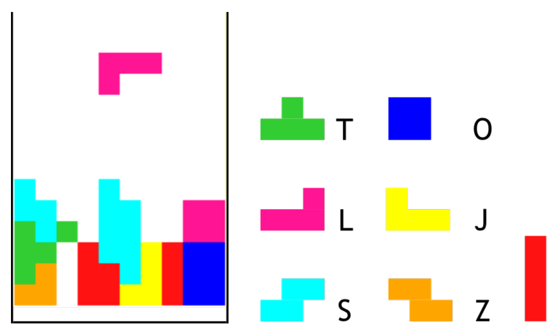

Figure 1. Illustration of a game of TETRIS in play (left) and the block types that can fall down during the game (right). In the text, block types are referenced by their corresponding letter.

TETRIS
Pajitnov and Gerasimov developed TETRIS in 1985 [24]. The game is played on a field that is 20 rows high and 10 columns wide. Small blocks – traditionally called *tetrominoes* – appear on top of the game field and fall down stepwise (see Figure 1). The time frame from the appearance of the block until it settles in a place is called an *episode*. A player needs to arrange the blocks in such a way that they create lines by filling a row; this

increases the score and full lines are then removed from the field. The more lines are removed in one episode, the higher is the awarded score. Due to the setup of the blocks, at most four rows can be cleared at once. Such an event is called a TETRIS. When the pile reaches the top of the field, the player loses. During gameplay, the speed at which blocks appear and move down the field traditionally increases with every ten rows cleared. Experienced TETRIS players tend to react with boredom, if the game's difficulty remains constant [6].

Game State Analysis

While TETRIS is played in real time, the game is divided into distinct episodes, one for each block. Logging and analysing a game is easy, because snapshots of the game state suffice to extract meaningful parameters. This is important as 'the case for dynamic difficulty adjustment' [11] requires parameters assessing player competency during a live game. In TETRIS, this is particularly difficult as neither time played nor lines made (or score for that matter) are good indicators of the playing style of a player [14].

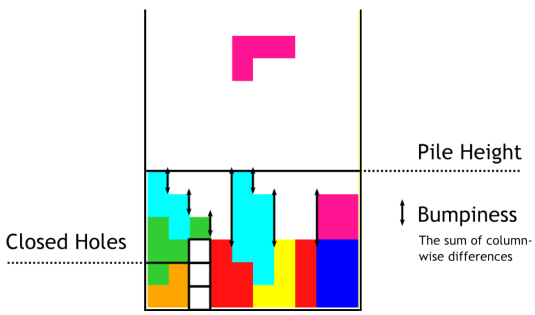

Figure 2. Parameters for TETRIS game state analysis

In order to enable artificially intelligent agents to play TETRIS, Fahey suggested two parameters, among others: *pile height*, which describes the height of the highest point of the contour and *number of closed holes*, which counts the number of unreachable areas with size $1x1$ under the contour [9]. Incidentally, minimising the pile height has been described as a successful strategy in playing TETRIS [8]. Flom and Robinson [10] added *bumpiness*, a measure accumulating the sum of height differences along the contour (see Figure 2 for illustrations). Other parameters that were put forth for AIs playing TETRIS (e.g., [4], [23] or [16]) are essentially derivations of the parameters discussed above. They help an AI to judge its performance in a game and change strategies. We investigate what they can tell about a human player's performance during gameplay in order to be able to adapt the difficulty in TETRIS during a live game. All of the discussed parameters depend on the current state of the game or, for lines cleared, on its past, so that they only have to be calculated once per episode. This not only allows for an episode-wise analysis of a game, but also for a comparatively fine-grained adjustment without disrupting single episodes. Since players expect a change of difficulty only after they have cleared a row, it might be confusing for players who already know the original game if the game reacts to anything else than to placing a block.

CHOOSING BLOCKS

The heart of any TETRIS game is the algorithm that chooses which blocks are spawned. The following paragraphs give an overview over traditionally used algorithms and a set of algorithms that we developed.

Traditional Algorithms

A commonly used algorithm for choosing blocks is TRUE RANDOM. The piece selection is random and independent. Very easy and very difficult series of blocks are equally likely.

GRAB BAG is the original TETRIS algorithm [13]. One instance each of all possible blocks are put into a bag and then drawn randomly from it without replacement, until the bag is empty. This is supposed to create a fair random game. With this algorithm, there are at most twelve pieces between two I blocks, and a maximum of four S or Z pieces can come in a row. That way, the chances of encountering a run of the same pieces are lowered (e.g., when compared to TRUE RANDOM).

Additional Algorithms

In order to be able to increase and decrease difficulty in more fine grained ways, we developed three additional algorithms. The conceptual aim of NICETRIS is to play the game with a minimal challenge. Through contour analysis this approach ensures that players always have a convenient option to place the current block and, hence, are able to clear rows quickly. The algorithm creates an array containing all shapes that fit into the current pile. If the contour is not fitting for any piece in particular, a generally well fitting block (O, I, or L) is chosen.

In order to increase the likelihood for an unbeatable game consisting of only S and Z pieces [5], SKEWED RANDOM assigns a 50% chance to either of the set of S or Z pieces, instead of $2/7 = 28.57\%$ as is the case for TRUE RANDOM.

The BUST HEAD algorithm is inspired by BASTET [20], which relies on the analysis of relative heights of neighbouring columns. Whenever a deep hole is established, effectively no I blocks will be given anymore. Our version uses the inverted principles of the NICETRIS algorithm above. First, the algorithm checks which pieces do not fit the current contour. Then, a bag of these pieces plus the O piece is used to randomly choose the next block. If every possible block could fit the contour, the undesirable combination of O, S and Z pieces is used as the pool for drawing the next element.

STUDY

We conducted an exploratory study to establish the actual and perceived difficulty as well as associated reported fun when these algorithms are used. Additionally, we recorded indicators for performance during a live game.

Our hypotheses were:

1. Algorithms

    (a) The difficulty of algorithms can be established through the performance measure of 'lines made' to be in the following order:
    1. NICETRIS → 2. GRAB BAG → 3. TRUE RANDOM → 4. SKEWED RANDOM → 5. BUST HEAD. This

order is based on assumptions about difficulty that went into designing the algorithms.

(b) Players overall report the most fun for GRAB BAG and TRUE RANDOM algorithms.

2. Difficulty and Fun
   (a) The better a player performs, the lower their perceived difficulty.
   (b) The easier a player rates a game, the more fun they report.
   (c) The better a player performs, the more fun they report.
3. Performance Indicators
   (a) The better a player is, the lower their pile height, bumpiness and number of closed holes.
   (b) Pile height is the strongest indicator of a player's performance during the game.

Procedure

Participants were recruited on a voluntary basis through a public notice. They were free to choose between a lab setting with a university owned computer or their own computer and a home setting on their own computers. The diversity of settings was intentional, since test participants were encouraged to create a pleasant gaming context for them in order to counteract effects that might occur when gaming in laboratory settings (for a discussion of this issue, see [15]).

After giving their consent to an anonymous use of their data in subsequent analyses, participants first completed a questionnaire asking for demographic data. They then started playing ten TETRIS games (two games each per algorithm, in a random order). After every game, there was a short questionnaire asking how players rated fun and difficulty of the game they had just played. We refrained from using an established questionnaire such as PENS (Player Experience of Needs Satisfaction, [22]) or the GEQ (Game Experience Questionnaire, [12]) after each game, because even in shortened versions, those take quite some time to be filled in and would have interrupted the game flow. Players also had to play too many games to be able to differentiate them properly afterwards; this ruled out administering questionnaires only after all games had been played. Pauses between games were self-paced; however, it was suggested that after five games, test participants take a longer pause.

Results

In total, 16 participants took part in the study, resulting in 160 games played, 32 per algorithm. Eight participants identified as male, eight as female. Participants' ages ranged from 22 to 34 years (mean: 26 years). We initially planned with a lower number of participants as the number of games played would have been sufficient to answer most of our questions. However, due to the popularity of TETRIS it was remarkably easy to acquire additional voluntary test participants.

Eleven of the participants studied or worked in Computer Science, five had other occupations. All of them knew TETRIS and had played it before, albeit with different intensities. On a Likert scale from 1-10, test participants rated their expertise in playing TETRIS on average as 5.9 (selected range: [3..9], median: 6).

Algorithms

Table 1 shows how players performed in the study according to the number of lines cleared during the game. We chose this measure of performance, because there is no consistent mode in which to calculate points in a given TETRIS game. The difference between NICETRIS and GRAB BAG appears to be negligible. Between TRUE RANDOM and SKEWED RANDOM there is a larger difference than between the other algorithms. The performance distribution for each algorithm is significantly different from each other with a large effect of the algorithm on performance ($p < 0.001$, Spearman's $\rho = 0.524$). Hypothesis 1 (a) can be accepted in slightly modified form in that the actual order is:
1. NICETRIS = 1. GRAB BAG → 2. TRUE RANDOM → 3. SKEWED RANDOM → 4. BUST HEAD. Importantly, these findings are replicated by perceived difficulty rankings (see Figure 3).

Algorithm	Mean	Med.	sd
NICETRIS	34.19	32	13.585
GRAB BAG	34.22	32	14.524
TRUE R.	31.97	35	14.901
SKEWED R.	21.66	19.5	12.936
BUST HEAD	13.38	10.5	9.366
TOTAL	27.05	27	15.55

Table 1. Lines cleared for each block choosing algorithm.

NICETRIS and GRAB BAG received the highest fun ratings – contrary to hypothesis 1 (b), which predicted a focus towards the algorithms providing slightly more difficulty according to our formal analysis.

Difficulty and Fun

Over all games and results, performance-related and perceived difficulty show a negative weak correlation ($r = -.27$). This follows exactly what we predicted in hypothesis 2 (a). Perceived difficulty and fun also show a weak, negative correlation (Spearman's $\rho = -.255, p < .01$): players tended to have more fun in TETRIS the easier they perceived the game to be (in alignment with hypothesis 2 (b)).

Interestingly though, individually, only eleven out of the sixteen players found the game more fun when it was perceived as less difficult. The others attributed more fun to algorithms they perceived as more difficult, indicating that engagement and enjoyment are linked differently for different types of players (cf. [3]).

Measured performance and fun show a weak, positive correlation (Spearman's $\rho = .193, p < .001$), indicating players had more fun when they performed better (in line with hypothesis 2 (c)).

Performance Indicators

In order to adapt TETRIS dynamically, it is vital to not only know *how* to adapt, but also *which* parameters are indicative of

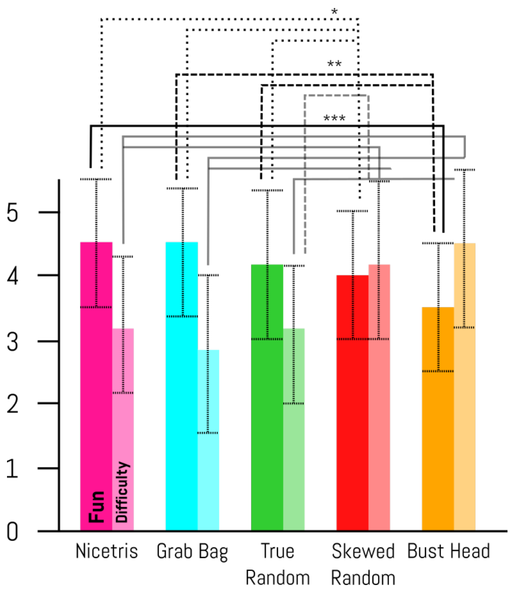

Figure 3. Perceived Fun and Perceived Difficulty for Different Types of Algorithms for Choosing Blocks in TETRIS as Reported in the Questionnaires. The level of significance is denoted as $*p < 0.05, **p < 0.01$ **and** $***p < 0.001$**. Significance of difference was tested using U-tests. Error bars indicate one standard deviation.**

player competency in order to make informed decisions about appropriate adaptation procedures. In order to understand this relationship we analysed different potential performance indicators, namely pile height, bumpiness and closed holes, and their predictive power towards final lines made.

Table 2 details the relationship between total lines made in a game and the overall values for pile height, bumpiness and closed holes. A good game with many lines made is expected to be on a low average pile height, show a low value of bumpiness and a lack of closed holes. The strong to moderate negative correlations let us confirm hypothesis 3 (a) – predicting that all of these parameters will be low when a player performs well in a game. Since pile height shows the strongest negative correlation it is also the strongest indicator of a player's performance (agreeing with hypothesis 3 (b)). It should be noted, though, that bumpiness is the most fine-grained value representing the overall look of the contour. Even if the pile height is high, a low bumpiness value indicates that this could be resolved within the next couple of blocks falling down.

Indicator	Mean	std	r	p
Pile Height	7.86	1.94	-.61	$< .001$
Bumpiness	18.14	2.97	-.54	$< .001$
Closed Holes	3.75	1.97	-.52	$< .001$

Table 2. Performance indicators as averages over a game in relation to the classical static measure of lines made (after the game). Significance of differences established through Student's t-test.

DISCUSSION

One of the two core results to be drawn from our exploratory study is that the block choosing algorithms differ from each other with respect to how difficult the ensuing games are. While the performance measures confirm this, we showed that the effect also holds for perceived difficulty. The effective differences between NICETRIS and GRAB BAG appear to be non-existent, but we identified a gap between TRUE RANDOM and SKEWED RANDOM. An in-between version like MILD SKEWED RANDOM that sets the likelihood of S or Z pieces at 39% can smooth the transition between TRUE RANDOM and SKEWED RANDOM during a live game.

NICETRIS does not perform as hypothesised. Since making a game easier than expected can create a situation in which players are under-challenged and their expert strategies are not suitable to what happens, they might perform less well than in more difficult games (see also the conceptual discussion of 'flow' [7]). Players who performed better tended to have more fun, but the more difficult players perceived the game, the less fun they reported – with large individual differences. This indicates that there might not be one single way players deal with challenge, but rather that there are types of players who attribute challenges more or less fun. Future work will tell whether this generalises to other games as well, but there is an indication in the work of [2] discussing player types and different sources of enjoyment for Multi User Dungeons.

The second core result of the study is that pile height is the most suitable parameter to capture player performance during a live TETRIS game, with bumpiness being a more fine grained, but slightly less indicative measure. Hence, these two methods for analysis can be used to adapt a live game to a player's current performance without relying on a past history of played games. These measurements enable a dynamically adapted version of TETRIS, since it answers the questions of how we can establish player competency in a live game.

CONCLUSION

We set out to give TETRIS – a well known game for which there are players of different skill levels easily available – a new twist and find out how to set different difficulty levels without relying on speed adaptation. By implementing five different block choosing algorithms and testing them in a user study with different parameters, we found that GRAB BAG, TRUE RANDOM, SKEWED RANDOM and BUST HEAD are sufficiently different from each other. With the addition of MILD SKEWED RANDOM we expect smooth transition between difficulty levels.

Also, we were able to establish pile height and bumpiness as suitable parameters to check for the performance of a player during a live game. This enables researchers to conduct double-blind comparative adaptivity studies in which players cannot easily guess whether they are playing an adaptive game or not.

Looking at the individual differences regarding the effect of an increased challenge on fun, we propose further research into different player types: those who engage more with an increased challenge (as in [1]) and those who engage less with an increased challenge (as in [17]).

ACKNOWLEDGEMENTS

We thank all participants in the study for engaging with the subject and donating us their time. We feel particularly grateful towards the anonymous reviewers for their insightful feedback. Additional thanks go to the Faculty of Media at the Bauhaus-University Weimar for supporting this work, the Human-Computer Interaction Group at Vienna Technical University for critiquing earlier versions of this paper and to Kearsley Schieder-Wethy for reading every version of a manuscript with critical enthusiasm and an eagle eye for grammar errors and typographical mistakes.

REFERENCES

1. Sami Abuhamdeh and Mihaly Csikszentmihalyi. 2012. The Importance of Challenge for the Enjoyment of Intrinsically Motivated, Goal-Directed Activities. *Personality and Social Psychology Bulletin* 38, 3 (2012), 317–330. DOI: http://dx.doi.org/10.1177/0146167211427147
2. Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research* 1, 1 (1996), 19.
3. Sven Bertel. 2014. Individual Cognitive Abilities and Styles in HCI: Three Main Challenges and a Tiered Adaption Model. In *HCI Engineering 2014: Charting the Way towards Methods and Tools for Advanced Interactive Systems. Workshop at the 6th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '14)*. ACM, New York, NY, USA.
4. Niko Böhm, Gabriella Kókai, and Stefan Mandl. 2005. An Evolutionary Approach to Tetris. In *6th Metaheuristics International Conference (6th Metaheuristics International Conference Wien August 22-26, 2005)*. MIC'05, Vienna, Austria, 1–6.
5. Heidi Burgiel. 1997. How to Lose at Tetris. *Mathematical Gazette* 81 (1997), 194–200.
6. Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2008. Boredom, Engagement and Anxiety As Indicators for Adaptation to Difficulty in Games. In *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era (MindTrek '08)*. ACM, New York, NY, USA, 13–17. DOI: http://dx.doi.org/10.1145/1457199.1457203
7. Mihaly Csikszentmihalyi. 1991. *Flow: The Psychology of Optimal Experience*. Vol. 41. HarperPerennial, New York, NY, USA.
8. Erik D Demaine, Susan Hohenberger, and David Liben-Nowell. 2003. Tetris is Hard, Even to Approximate. In *Computing and Combinatorics*. Springer, Heidelberg, Germany, 351–363.
9. Colin Fahey. 2012. Tetris. (July 2012). http://colinfahey.com/tetris/tetris.html
10. Landon Flom and Cliff Robinson. 2005. Using a Genetic Algorithm to Weight an Evaluation Function for Tetris. (2005).
11. Robin Hunicke. 2005. The Case for Dynamic Difficulty Adjustment in Games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '05)*. ACM, New York, NY, USA, 429–433. DOI: http://dx.doi.org/10.1145/1178477.1178573
12. Wijnand IJsselsteijn, Karolien Poels, and Yvonne De Kort. 2008. The Game Experience Questionnaire: Development of a self-report measure to assess player experiences of digital games. (2008).
13. Mitu Khandaker. 2011. Column: "Gambrian Explosion": Games, Randomness, and The Problem with Being Human. (April 2011). http://www.gamesetwatch.com/2011/04/column_gambrian_explosion_game.php
14. David Kirsh and Paul Maglio. 1992. Reaction and Reflection in Tetris. In *Artificial Intelligence Planning Systems: Proceedings of the First Annual International Conference (AIPS92*. Morgan Kaufman.
15. Robert Ladouceur, Anne Gaboury, Annie Bujold, Nadine Lachance, and Sarah Tremblay. 1991. Ecological Validity of Laboratory Studies of Videopoker Gaming. *Journal of Gambling Studies* 7, 2 (1991), 109–116.
16. John Lindstedt and Wayne Gray. 2013. Extreme Expertise: Exploring Expert Behavior in Tetris. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (Cooperative Minds: Social Interaction and Group Dynamics)*. Cognitive Science Society, Berlin, Germany, 912–917.
17. Derek Lomas, Kishan Patel, Jodi L. Forlizzi, and Kenneth R. Koedinger. 2013. Optimizing Challenge in an Educational Game Using Large-scale Design Experiments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 89–98. DOI: http://dx.doi.org/10.1145/2470654.2470668
18. Ricardo Lopes and Rafael Bidarra. 2011. Adaptivity Challenges in Games and Simulations: A Survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 2 (June 2011), 85–99. DOI: http://dx.doi.org/10.1109/TCIAIG.2011.2152841
19. Brian Magerko, Carrie Heeter, Joe Fitzgerald, and Ben Medler. 2008. Intelligent Adaptation of Digital Game-based Learning. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share (Future Play '08)*. ACM, New York, NY, USA, 200–203. DOI: http://dx.doi.org/10.1145/1496984.1497021
20. Frederico Poloni. 2012. Notes on the Bastet Algorithm. (January 2012). http://fph.altervista.org/prog/bastetalgo.html
21. Hua Qin, Pei-Luen Patrick Rau, and Gavriel Salvendy. 2010. Effects of different scenarios of game difficulty on player immersion. *Interacting with Computers* 22, 3 (2010), 230 – 239. DOI: http://dx.doi.org/10.1016/j.intcom.2009.12.004

22. Scott Rigby and Richard Ryan. 2007. The player experience of need satisfaction (PENS) model. *Immersyve Inc*. (2007).

23. Elad Shahar and Ross West. 2010. Evolutionary AI for Tetris. (2010). http://www.cs.uml.edu/ecg/pub/uploads/AIfall10/eshahar_rwest_GATetris.pdf

24. David Sheff. 1993. *Game Over: How Nintendo Zapped an American Industry, Captured Your Dollars, and Enslaved Your Children.* Random House Inc., New York, NY, USA.

25. Stefanie Wetzel, Katharina Spiel, and Sven Bertel. 2014. Dynamically Adapting an AI Game Engine Based on Players' Eye Movements and Strategies. In *Proceedings of the 2014 ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '14)*. ACM, New York, NY, USA, 3–12. DOI: http://dx.doi.org/10.1145/2607023.2607029