# Impact of different levels of difficulty on immersion in video games

Jim Sigailov-Lanfranchi

University College Dublin jimsigailov@gmail.com

**Abstract.** Twelve participants played three levels of Tetris that varied by difficulty and immersion was measured after each level with a survey. The levels corresponded to the scenarios: [skill of the player > challenge; skill = challenge; skill < challenge]. Flow levels of participants were measured as well. The question asked was whether different difficulties would influence how immersed players would be, hypothesizing that players would be more immersed when cognitively overloaded than when facing a challenge adapted to their skill. Results showed no significant difference between the different conditions but pointed towards less immersion when the players faced a challenge inferior to their skill.

Keywords: Immersion, Difficulty, Challenge, Flow, Videogames, Tetris

## 1 Background

Video games playing is becoming one of the major hobbies across the planet, with three-quarters of all Americans having at least one gamer in their household (ESA, 2019). Consequently, over the past decades, a substantive body of research on video games has appeared and a growing number of research papers are dedicated to describing and analyzing the multi-faceted phenomenology of gaming. Two of these facets that may variate the experience of the player are the difficulty of the game and how immersed the player feels. As many video games propose several levels of difficulty and many multiplayer ones propose matchmaking systems where the player faces an opponent of similar skill, the question of how immersion evolves throughout different levels is of interest not only to researchers but also to developers.

However, to my knowledge, there is currently very little literature on how different levels of difficulty influence how immersed people are when gaming. This may be due to how the concept of flow state is conceived amongst game researchers. Indeed, flow seems often conceived as "the optimal experience" for gamers (e.g. Brockmyer et al, 2009; Chen, 2007). Because flow, amongst other things, is characterized by a feeling of loss of concern for the self in the real world during the activity, a deep involvement with the task undertaken, and an altered sense of time (Cowley et al., 2008), without further research, common sense would make you think that immersion

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

is maximized (or at least very strong) during flow. Additionally, it is known that in the context of video games, flow varies relatively to the balance of [skill of the player/challenge presented by the game] as follows:

When difficulty is too low compared to the skill of the player, the game becomes boring, and cannot be flow-inducing. When difficulty matches the skill of the player, and the other conditions relative to the appearance of a flow state are met, the situation is supposed to be flow-inducing, and very immersive. Finally, when the challenge is too high for the player, the situation becomes anxiogenic and the conditions are not flow-inducing anymore.

Hence, we end up with a sort of bell-curve in which flow is maximized when [challenge=skill].



Fig. 1 The flow channel (from Keller & Bless, 2008)

At a first glance, it may seem that immersion would follow a similar pattern. Yet, let us study this more precisely. When the game is boring, it makes sense to suppose the game as not immersive, as the game is not very demanding, and the mind has cognitive room to wander freely. However, we hypothesize that when facing an extreme and anxiogenic challenge, the player would try to gather all attentional and cognitive resources. Consequently, the player could be as immersed, or even more immersed, than in the flow-inducing condition.

As aforementioned, previous literature on experimental manipulation of challenge and its impact on immersion is limited. Qin et al. (2010) showed that players tended to feel more immersed when the difficulty was changing up and down than when the difficulty was changing down and up or simply increasing continuously. They also showed that participants were more immersed when subject to a "medium" rate change in difficulty rather than an excessively fast or slow one. However, their paper studied dynamics of immersion throughout non-random changes of difficulty and cannot be extrapolated to determine in absolute terms whether, say, an 'easy' difficulty is more immersive than a 'hard' one. Cox et al. (2012) also showed that increasing physical demand (by requiring the participants to press more buttons) was not enough alone to increase immersion. Time pressure, on the other hand, by adding both physical and cognitive challenge, successfully increased immersion. In other words, they showed that some form of challenge can impact immersion.

In this context, it makes sense to ask: how do different levels of difficulty impact immersion?

This paper starts with 3 hypotheses:

- 1. Immersion is at its lowest point when the challenge of the game is below the skill of the player.
- 2. Immersion is high when the game presents a challenge that matches the skill of the player. When this is the case, the player enters a state of flow.
- 3. Immersion is even higher when the game presents a challenge that exceeds the skill of the player, as the player must gather all attentional and cognitive resources to face the challenge. This, in turn, creates deep immersion.

I will try to answer the question asked using an experiment where players face three different conditions that vary by difficulty, and measure immersion levels reached during the experience. A measure of flow will be used as a proxy to determine whether the medium level correspond to an adaptive condition where the challenge matches the skill of the player.

Keller & Bless (2008) successfully used a paradigm where they controlled the difficulty of the video game Tetris to show that some individuals were more sensitive than others to manipulation of the skill/challenge balance. Here, this paradigm is adapted for our purposes.

# 2 Methods

## 2.1 Participants

Data was recorded from 12 adults (5 females, 7 males) primarily postgraduate students, aged between 19 and 29 (mean=24.08, SD=2.39) years. All of the participants owned a personal computer and all but one generally played video games at least once a week. No reward was given for participation.

### 2.2 Design of the game

Three versions of Tetris were adapted from an open source code found online that replicated the design of the version of Tetris originally published by Nintendo for Game Boy in 1989. The goal of Tetris is to manipulate a random sequence of falling pieces (called Tetrominos) and arrange them as to complete lines at the bottom of the screen. The previously fallen pieces stack up at the bottom, and when a line is com-

plete, it disappears. The falling Tetrominos can be moved right and left and rotated by 90° in both directions using assigned keys (left, right and down directional arrows to move, "Z" and "X" to rotate. The up directional arrow also allowed for a clockwise rotation.). The upcoming Tetromino is shown on the right of the screen, as well as a score, the current level, and the number of lines completed.

Speed increases correspond to level 1 to 20, that is, falling pieces moving every {887,820,753,686,619,552,468,368,284,184,167,150,133,117,100,100,83,83,66,66, 50} milliseconds.

The Tetris music theme, originally played with the game, plays in the background.



Fig. 2 User interface and example of situation the player may encounter in the game

The three versions programmed tried to create a "boredom" condition (skill>challenge), a "flow-inducing" or "adaptive" condition (skill=challenge), and finally an "overload" condition (skill<challenge).

The first version programmed was characterized by having a constant level at what would be a reasonable/slow speed for my group of participants (Tetrominos fall every 619 ms, which corresponds to level 5 out of 20).

The second version, which is supposed to be flow-inducing has been programmed to try to adapt to the player and present them with a challenging yet comfortable pace. The blocks initially fall slowly (every 887 ms (level 1/20)) but the falling speed then accelerates every time the player completes a line. Additionally, two systems were added to counter the speed increase in case the challenge becomes too difficult for the player. First, if the player has stacked too many lines that have not disappeared, the speed will stop increasing until the player has gotten rid of the lines. Then, the system makes the speed gradually decrease if the player fails to complete a number of lines within a certain number of units of time (one unit of time corresponding to a new

Tetromino appearing). The number of units of time it would take until the game decelerate would decrease as the total height of the stack the player has formed increased. This means that the higher the stack (and so the lower the freedom of movement for the new pieces), the faster speed would decrease.

The third version corresponds to a more classic Tetris game, where speed increases every time two lines are completed (capped at a max level where falling blocks move every 50 ms) and never decreases. That ensures that when the player loses, the speed exceeds the skills of the player and the player has been cognitively overloaded.

Overall, our three versions fill our needs: one where the skill of the player exceeds the challenge (which will be called further the *boredom condition*), one where the skill should meet the challenge (which will be call further the *adaptive condition*), and one in which, at the end, challenge exceeds the skill of the player (*overload condition*). Please note that the overload condition, unless stated, refers to the experience at the end of the third version and not the experience throughout the third version.

When the game is lost, a message saying "Game over! Press OK for a new game" and an "OK" button pops up. The game resets when the "OK" button is pressed.

### 2.3 Measurement instrumentation

Two questionnaires were used, one to measure whether participants had entered a state of flow, and the other to measure how immersed they were. The immersion questionnaire is the second and refined version of the *Immersion Questionnaire* used by Jennet et al. (2006). This questionnaire was validated using a large sample (n=260) and factor analysis and has been used by other authors (e.g. Cairns et al., 2014; Herrewijn et al., 2013).

To determine whether the adaptive challenge would match the skill of the player, flow was measured and used a proxy. We hypothesized that if a participant can successfully be put in state of flow, this would indicate the challenge matched their skill. Hence, the second one is the *Flow State Scale-2* from Jackson & Eklund (2002), a scale used widely in the literature (e.g Procci and Bowers, 2011; Hamari & Koivisto, 2014) and translated into several languages, to measure after an experience how much participants entered the flow state in the said experience.

## 2.4 Procedure

Each participant played at home in the environment in which they would normally play video games. Although this may have resulted in some differences between participants, this choice was justified by ecological validity and considering this study focused on the differences between different playing experiences under the same conditions for each participant rather than between participants.

Each level was played for around 10 minutes. More precisely, at least 8 minutes and until they lost the game or 2 more minutes had passed, with the exception of the overload condition in which the participants played at least 8 minutes and always until they had lost.

An audio call connected the participant to the researcher for the purpose of giving oral explanations about the procedure and stopping them when time was up for each level (so that they would not have to pay attention to time themselves, which would have broken the immersion). Players were asked to reset the game through the "OK" button if they lost before they were interrupted.

The order in which they played the three versions was randomly assigned.

After being interrupted, for each level, players immediately answered the two questionnaires, starting with the FSS-2. Players were asked to focus on how they felt in the middle of the game for the boredom and adaptive condition, and how they felt at the end for the overload one.

All experimental procedures were approved by the UCD School of Computer Science Taught Masters Research Ethics Committee.

## 3 Results

The scores obtained on the questionnaires were submitted to two one-way repeated measures analysis of variance (ANOVAs), one for flow scores and one for immersion scores. Initial results showed no statistically significant differences for any of the different levels for a p-value <0.05, meaning the probability of the differences observed not being random is inferior to 95%, both for the flow scores [F(2,22) =1.00, p=0.386] and the immersion scores [F(2,22)=1.79, p=0.190]. Nonetheless, some differences were noticeable graphically, and the results were subjected to paired t-tests for further exploration. The table is presented at the end of this section. Fig 3. below (see next page) shows the means obtained for the flow scores.

## 3.1 Flow

The most noticeable feature of the data on flow collected is that participants seemed to have experienced a maximum level of flow during the adaptive condition (as was expected) (as well as during the boredom condition; more on that below). Additionally, flow levels as indicated for the overload condition are remarkably lower than for the two other levels, here again as expected, probably as players felt less in control in that condition. T-test on "adaptive vs overload" indicates there is an 89% chance of this difference not being due to randomness.

Interestingly though, the mean flow level reached in the boredom condition is roughly equal to that reached during the adaptive condition (t-test indicating 86.9% chances of this similarity not being random). A possible explanation for this is a flaw in the design on the first level. Indeed, instead of being forced into, and being stuck in a boring setup, players had the possibility to click on the down button to push the piece they were controlling down. By clicking this button repeatedly, they would effectively drop the piece, or in other words, accelerate its fall. This means that for the players that adopted this strategy, the falling speed was adapted to them, since they

were the one controlling it. As such, it makes sense that there would have achieved similar flow states in both conditions.



Fig. 3. Mean flow scores of the three levels of difficulty with statistically computed error bars corresponding to a 95% confidence interval.

Again, here, however, none of the noticeable differences or similarities were detected as significant by the ANOVA. [F(2,22) = 1.00, p = 0.39] This may be due to the fact that, for instance, although the mean of the boredom and adaptive conditions seem roughly equal, this hides some discrepancies amongst participants : indeed, although some had roughly equal scores in both conditions, some had a higher score in the boredom one, and some others a higher score in the adaptive condition. These scores on different participants may have cancelled each other and yielded roughly equal means, yet making stating with certainty that participants had similar scores in both conditions a fallacy. The analysis of variance serves to control for this by taking into account the evolution of the scores for each participant.

More importantly, these flow scores beg the question of whether the design of our game was successful in creating an adaptive situation. Although none of the differences observed were significant, and the boredom condition was as flow-inducing as the adaptive condition, we can conclude in view of the explanation given above than despite the lack of significant differences, the design of our game was somewhat successful in creating a condition in which the player meets a challenge roughly equal to his skill.

In the case where this would not have happened, the core of our further results still stands, as a hierarchy where the first level is easier than the second and the second level is easier than the end of the third still exists.

# 3.2 Immersion



Fig. 4. Mean immersion scores of the three levels of difficulty with statistically computed error bars corresponding to a 95% confidence interval

Two features are striking on these results: the similarity of means of the adaptive and overload condition, and the mean of the boredom condition being noticeably lower.

As was to be expected, the boring condition is noticeably less immersive than the others (t-test indicating a 90% chance of the difference between boredom and overload to not be random). As players require less cognitive resources to play the game efficiently, their mind can somewhat freely daydream to other preoccupations.

Additionally, in accordance with our initial hypothesis, the overload condition seems to be as immersive as the adaptive one [t(11)=-0.17; p=0.869].

Table 1. t-test results on pairs of levels for both flow and immersion scores

	Flow	Immersion
Boredom vs Adaptive	t(11)=-0.17; p=0.869	t(11)=-1.32; p=0.214
Boredom vs Overload	t(11)=0.96 ; p=0.357	t(11)=-1.79; p=0.100
Adaptive vs Overload	t(11)= 1.73 ; p=0.112	t(11)=0.17; p=0.868

# 4 Discussion

From these results, it is unclear whether there is indeed an impact of difficulty on immersion or if the results obtained are random. One of the main reasons for this uncertainty is probably the rather low number of participants. As this experiment was done on a limited number of participants (n=12), statistical results are not very accurate. The confidence intervals are large, and results would need to be very contrasting to be detected.

Additionally, the participants were selected on the basis of their availability to the researcher, so it is rather likely that a sampling bias occurred.

If these results are replicated on a higher number of participants and are shown to be significative, the case of the boredom condition would be particularly interesting, as it would show that in some circumstances, players can be in a flow state without being immersed in the activity they are taking part in.

However, it is also possible that if these results were replicated on a higher number of participants, the same results would still be obtained: no significant differences between conditions, meaning that difficulty does not impact immersion. It would also be quite a finding for flow theory if it were shown that difficulty has no impact on flow theory, as the skill/challenge balance is considered a pillar of flow theory in video game research.

We propose below a series of effects that may have influenced our results, in one way or another, and perhaps, in making our different conditions less contrasting with each other than they would have been otherwise.

## 4.1 Limitations

The first limitation we have is due to the nature of the game itself. Tetris is not a game that allows for a strong empathetic connection with the game in most cases; there is no narrative structure or endearing characters. Yet, some researchers consider establishing an emotional connection with the media to be important to immerse oneself in the said media (Brown & Cairns, 2014). Indeed, a role-playing game where the player would control an avatar and follow a narrative arc would probably have a greater immersive character and would have been more suitable to the research question asked in this paper. Hence, if means permit, further research on this topic should try experimenting with this kind of game. The setting-up of the present experiment, however, was limited by constraints of time and resources and Tetris offered a first approach to the topic using an experiment that did not require extra learning time and thus reducing the total time of the experiment for each participant.

For many games, the average playing session by far exceeding 8 minutes. For example, Tarng et al. (2008) found that World of Warcraft players tended to play at least an hour over one session. First-person shooters (FPS) matches tend to last around 15 minutes and players tend to stay online for several matches. It is therefore likely that

this experiment did not replicate ecologically valid conditions of immersion in terms of time for many games. Indeed, although some immersion was recorded, it is possible that if players played for a longer time, higher levels of immersion would have been reached, and differences between the conditions would have been more contrasted.

Furthermore, the design of the experiment may be imperfect: it is possible that players lost in the third version for reasons other than being cognitively overloaded: wrong piece at the wrong time, strategy not working... Random number generation plays a role which could have been detrimental to our design.

Randomness is also a more global problem of our methodology: each participant had to deal with different pieces and as a result, each participant did the experiment in different conditions.

As previously mentioned, it is also possible that some players found the easiest version immersive and flow-inducing because they were able, by using the down button, to accelerate the falling speed of the Tetromino they were controlling, and as a consequence to increase the challenge by themselves, just as much as they needed until it became too difficult. In other words, instead of being stuck with a boring task, the players were able to adjust the task so it would match their skill.

Another limitation is that the order in which participants played the different levels was not recorded, while this may have been a good predictor of their immersion or flow scores, in particular in view of Qin et al. (2010)'s findings (see 1. Background). By the time they would start the 3<sup>rd</sup> level, around 40 minutes since the beginning of the experiment would have started, and participants may have become tired. Likewise, players may have been more excited to play during the first level than the second.

Finally, personality is another factor that may have influenced the results and was not controlled for. In particular, in the overload condition, after discussing with participants, some reported feeling more combative when they realized they were going to lose soon while some, on the opposite, reported feeling helpless and giving up their efforts.

Further research shall aim to correct these limitations, control for these various factors and, as we suggested, more importantly, increase the number of participants. Ideally, a more "classical" RPG with an avatar controlled by the player should be used.

# 5 Conclusion

It would seem that difficulty impacts immersion. In particular, when the challenge faced by the player is lower than their skill, the game is less immersive than if the game presents a challenge that matches the skill of the player, or that exceeds the skill of the player. However, it is unclear if the results found are due to randomness or other various factors. Further research is needed to settle the question.<sup>1</sup>

## References

- Brown, E., & Cairns, P.:A grounded investigation of game immersion. In CHI'04 extended abstracts on Human factors in computing systems (pp. 1297-1300). ACM. (2004, April).
- Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. Journal of Experimental Social Psychology, 45(4), 624-634. (2009).
- 3. Chen, J. (2007). Flow in games (and everything else). Communications of the ACM, 50(4), 31-34.
- 4. Cowley, B., Charles, D., Black, M., & Hickey, R.: Toward an understanding of flow in video games. Computers in Entertainment (CIE), 6(2), 20. (2008)
- Cox, A., Cairns, P., Shah, P., & Carroll, M.: Not doing but thinking: the role of challenge in the gaming experience. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 79-88). ACM. (2012, May)
- Entertainment Software Association, Essential facts about the computer and video game industry (2019) [Fact sheet], https://www.theesa.com/wpcontent/uploads/2019/05/ESA\_ Essential facts 2019 final.pdf, last accessed 2019/08/10
- Jackson, S. A., & Eklund, R. C.: Assessing flow in physical activity: The Flow State Scale-2 and Dispositional Flow Scale-2. Journal of Sport and Exercise Psychology, 24, 133-150. (2002)
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A.: Measuring and defining the experience of immersion in games. International journal of humancomputer studies, 66(9), 641-661. (2008).
- Keller, J., & Bless, H.: Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. Personality and social psychology bulletin, 34(2), 196-209. (2008)
- Hamari, J., & Koivisto, J.: Measuring flow in gamification: Dispositional flow scale-2. Computers in Human Behavior, 40, 133-143. (2014).
- Herrewijn, L., Poels, K., & Calleja, G.: The relationship between player involvement and immersion: An experimental investigation. In *FDG* (pp. 364-367). (2013).

<sup>&</sup>lt;sup>1</sup> I would like to thank F. Costello and F. Cummins for their approval and guidance. I am also infinitely grateful to S. Barrau, K. Senfeld, S. Tsoutsou and my family for their time, help, advice, inspiration and support, without whom this paper would never have been. Finally, I would like to thank every one of my participants that accepted to lend me their time despite the absence of reward.

- 12. Procci, K., & Bowers, C. An examination of flow and immersion in games. In *Proceedings* of the Human Factors and Ergonomics Society Annual Meeting (Vol. 55, No. 1, pp. 2183-2187). Sage CA: Los Angeles, CA: SAGE Publications. (2011, September).
- 13. Qin, H., Rau, P. L. P., & Salvendy, G.: Effects of different scenarios of game difficulty on player immersion. Interacting with Computers, 22(3), 230-239. (2009)
- Tarng, P. Y., Chen, K. T., & Huang, P.: An analysis of WoW players' game hours. In Proceedings of the 7th ACM SIGCOMM Workshop on Network and System Support for Games (pp. 47-52). ACM. (2008, October)