



Do Individual Differences Modulate the Effect of Agency on Learning Outcomes with a Serious Game?

Julien Mercier^(✉), Kathleen Whissell-Turner, Ariane Paradis,
Ivan Luciano Avaca, Martin Riopel, and Mélanie Bédard

NeuroLab, University of Quebec at Montreal, 405 Sainte-Catherine East,
Montreal, QC H2L 2C4, Canada
mercier.julien@uqam.ca

Abstract. In order to support and promote new ways of learning, educational technology should be based on sophisticated theories and models of learning. Many issues are raised in the current understanding of learning by the constant evolution of educational technology and the burgeoning of educational contexts using these technologies. By examining the relation between agency and learning gains using a Serious Game for learning Physics, the present study focuses on a main issue of technology use: whether actively playing the game or watching someone play is beneficial for learning. Thirty-seven dyads participated in the study. Randomly assigned, one participant played a Serious Educational Game for learning Physics, *Mecanika* (Boucher-Genesse et al. 2011), for 120 min, while the other participant watched the player's gameplay in real-time on a separate screen. As pretest and posttest, the *Force Concept Inventory* (FCI; Hestenes et al. 1992) was administered to measure learning gains in Physics. Analyses of answers on the FCI demonstrate that a Serious Game, such as *Mecanika*, is beneficial to learning, regardless if learning is conceived as relatively coarse shifts from wrong to good answers (scientific conceptions) or as more nuanced shifts from fillers/misconceptions to scientific conceptions. Also, individual differences in learning gains across dyads were found, which can be explained by the gameplay of a dyad created by the active player. Furthermore, the effect of agency is systematic and not modulated by individual differences: watchers learn more than players. These results need to be further explained by modeling the learning process.

Keywords: Agency · Learning gains · Serious educational game

1 Introduction

According to Stephanidis, Salvendy and their group of experts (2019), how technology can be used to its full potential to foster learning is still vastly an open question after decades of research and technological evolution. Technology use for learning encompasses old as well as new issues including privacy and ethics, learning theories and models, and pedagogical aspects. These issues translate into current challenges for applied research.

With respect to learning theories and models, educational technology should strive to support and promote new ways of learning, creative learning and lifelong learning for all learners, avoiding to focus on tech-savvy generations. The field should aim to design technologies focusing on the needs of learners and educators that are gracefully embedded in the educational process and do not disrupt learners and teachers, and to design serious games featuring the appropriate balance between seriousness and fun, and be driven by tangible educational needs and not by new technological capabilities. The field should address unprecedented challenges related to privacy and ethics, such as the extensive monitoring of students (potentially under-age) by data-gathering software, sensors and algorithms, such as the parent issues of data ownership, and such as management as well as human rights concerns (e.g. potential for excessive control restricting the freedom of the individual). In terms of pedagogical aspects, key aspects still in need of improvement include the involvement of educators in the design of learning technologies, a serious and multi-faceted assessment of long-term impact of learning technologies, support for personalized creativity and for the amplification of human creative skills, for the entire spectrum of creative activities including in smart environments, blending digital and physical artifacts.

Despite more or less direct ramifications regarding the previous issues, the present work focuses on learning theories and models. In particular, the study presented examines the accepted notion that active learning fosters learning gains.

In learning about real-world phenomena from observations without having acquired scientific conceptions, one develops misconceptions which wrongly predict phenomena that do not correspond to current scientific knowledge. Such misconceptions are constructed through interactions with the environment, which do not provide all the information necessary to construct scientifically-valid explanations. In this context, Clark's (2013) prediction-action model posits that learning occurs when these misconceptions are surmounted by input that make prediction errors manifest. In this view, teaching and fostering learning involve providing input that will ultimately lead a learner to formulate predictions adequately representing the state of the world.

An interactive learning environment providing simulations of Physics phenomena should help learners test their worldview against scientific conceptions embedded in the simulations. In this context, agency should also be beneficial to learning because the possibility to control the simulations should optimize the testing of predictions. To the contrary, it was shown elsewhere that agency was detrimental to learning (Mercier, Avaca, Whissell-Turner, Paradis and Mikropoulos, this volume, submitted). The objective of this study is to refine the previous results by examining if a computer-based interactive learning environment is beneficial to learning, and to verify if the effect of agency on learning is modulated by individual differences.

2 Theoretical Framework

2.1 A Prediction-Action View of Learning and Agency

In the context of a prediction-action framework, Lupyan and Clark (2015) provide a pivotal question for the design of serious games for learning by suggesting what one

knows ought to change what one sees. Globally, a prediction-action framework explains learning as the production of representations at multiple levels of abstraction, so that a given level predicts the activity in the level below it. Also, reducing prediction error in the present enable better predictions in the future.

That is, higher-level predictions currently used are informed by priors (prior beliefs, usually taking the form of nonconscious predictions or expectations) concerning the environment. Prior beliefs or lower-level neural expectations are statistically optimal in the sense that they represent the overall best method for inferring the state of the environment from the ambient conceptual and sensory evidence. A prediction-action framework seems to articulate perception and attention as optimal (Bayesian) ways of combining sensory evidence with prior knowledge in the process of learning. Predictive-processing models are based on an asymmetry between the forward and backward flow of information: The forward flow computes residual errors between predictions and the information from the environment, while the backward flow delivers predictions to the appropriate level. The forward flow also escalates high-information contents upward by pushing unexplained elements of the lower-level sensory signal upward so that the appropriate level selects new top-down hypotheses that are better able to accommodate the present sensory signal. One can sometimes become aware of them when they are violated (Lupyan and Clark 2015). This long-term error-reduction mechanism can be thought of as responsible for academic learning.

In sum, Lupyan and Clark (2015) propose that the learning of a symbolic language (verbal or mathematical) may modulate the recruitment of prior knowledge and the artificial manipulation, at any level of processing, of the relative influence of different top-down expectations and bottom-up sensory signals. These manipulations, which can be communicated to others, could selectively enhance or mute the influence of any aspect, however subtle or complex, of our own or another agent's world model. Exposure to and the acquisition or learning of a symbolic language (whether shared or self-produced) leverages the exploration and exploitation of our own knowledge as well as the knowledge of others. In the context of designer learning environments, the possibility to manipulate their features by directly interacting with them should optimize the prediction-action by improving the fit between the predictions formulated and the predictions tested by the learner. In contrast, without this possibility of agency, the fit between the predictions tested and the current predictions of a learner is necessarily lessened.

2.2 Learning Analytics and the Design of Serious Games for Learning

Learning Analytics and their Potential Uses. Serious Games have already showed their advantages in different educational environments (Stephanidis, Salvendy et al. 2019). Game Learning Analytics can further improve serious games, by facilitating their development and improving their impact and adoption (Alonzo-Fernandez et al. 2019).

Game Learning Analytics (hereafter GLA) is an evidence-based methodology based on in-game user interaction data, and can provide insight about the game-based educational experience and outcomes, like validating the adequation of the game

design to the educational goals. Besides providing a visual in-game follow-up, GLA can verify if the game in question is accessible to the target population as expected. Another great advantage of GTA that could redefine educational assessment is its capability to predict in-game learning gains.

Alonzo-Fernandez and her colleagues (2019) used GTA for serious games in order to evaluate game design and deployment processes. Alonzo-Fernandez et al. (2019) asserts that the design of games should be based on clear goals, and specify how the attainment of these goals are to be measured with interaction data adequately collected. Additionally, the use of learning analytics is greatly facilitated if, from the very beginning, games are designed so that data can be extracted from them and provide the information required to validate the games and assess the process and outcomes of learning with students using them. Early uses of GTA in designing Serious Games are also possible. Effectively GTA allows to remotely collect and analyze data and feedbacks during a beta testing on target users. This way, potential problems could be quickly solved and the game can be largely deployed. In addition, improvements for subsequent versions of an already deployed game could be leveraged by players' interaction data.

Optimal Serious Games' criteria emerged from Alonzo-Fernandez and her colleagues (2019) observations regarding assessment. A main recommendation is to keep in mind that Serious Games are not only created to play and learn, but ultimately to collect data on evidence-based learning assessment. To do so, conception of Serious Games need to take into account the types of data wished to be collected and the standard format of GLA data.

The work of Alonzo-Fernandez et al. (2019) showcases the importance of game learning analytics in different contexts, even when being used as the sole means to obtain players feedback. While his experiments of GLA were conducted in various real contexts pursuing diverse goals, Alonzo-Fernandez et al. (2019) suggest the exploitation of GLA in the context of Serious Educational Games.

Westera (2018) presents a computational model for simulating how people learn from serious games based on simulation studies across a wide range of game instances and player profiles for demonstrating model stability and empirical admissibility. While avoiding the combinatorial explosion of a game micro-states, the model offers a meso-level pathfinding approach, which is guided by extant research from the learning sciences. The model can be used to assess learning from (or with) the serious game and for investigating quantitative dependences between relevant game variables, gain deeper understanding of how people learn from games, and develop approaches to improving serious game design. With this in mind, simulation models can also be used as an alternative to using test players as the only source of information, to improve game design. To reduce overall complexity and to avoid model overfitting and the combinatorial explosion of game states and player states, the model focuses on meso-level aggregates that constitute meaningful activities. It accounts for discrete-time evolution, failure, drop-out, revisit of activities, efforts made and time spent on tasks.

According to Westera (2018), advances in learner data analytics, stealth assessment, machine learning and physiological sensors for capturing such data represent new additional ways to enrich the constituents of the model. Ultimately, the computational modelling approach would help to design serious games that are more effective

for learning. So far, however, conditions for empirical validation with real players are only partially met, since some of the learner model's variables are still hard to record without the appropriate interdisciplinary work in cognitive science merging psychology, neuroscience and education, such as the real-time progression of motivation and flow.

Real-world, authentic experiments should be based on discrete events at longer time scales and aggregates of finer-grained events constituting an approximation of the model rather than the full discrete-time evolution version (Westera 2018). At a given, typically longer, temporal grainsize, self-report instruments and performance measures can be used to capture additional data.

Modeling for Assessing Learning Gains. The main interlinked constituents of the model are the knowledge model, the player model and the game model. The productive outcomes of a serious game need to be expressed as knowledge gains Westera (2018). As such, they are the easiest aspect to model, requiring only the knowledge model. The knowledge model is generally expressed as a knowledge tree of operationalized learning goals or learning outcomes (e.g. skills, facts, competences) while child nodes in the tree have a precedence relationship with their parent nodes. While the game is represented as a network of meso-level activity nodes, each activity in the game is allowed to address one or more nodes from the knowledge tree. Each activity in the game is characterized by prior knowledge requirements and by an inherent complexity.

To explain learning gains, the player model accounts for the player's mental states, preferences and behaviors. Only few primary player factors will be taken into account: overall intelligence, knowledge state, and motivation. While both intelligence and prior knowledge refer to the player's learning capability, motivation is linked to personal attitudes, emotions and ambitions. Westera (2018) purports that these are exactly the key dimensions that reflect the potential of serious games: learning new knowledge from serious games, while benefitting from their motivational power. These elements can be further detailed to reflect the state of the art in the learning sciences.

Then, after updating the player's knowledge states, the whole cycle is repeated for the next time step, while the player progresses in the game. In each time step (Δt) the player's knowledge state must be updated to account for the knowledge gained during that period of time.

Following the classic approach in intelligent tutoring systems research, a node in the knowledge tree, as a parent, combines, integrates and extends all child nodes, the process of mastering a parent node inherently contributes to the further mastery of all subordinate nodes in the parent tree. Hence, the updating process should also be applied for updating the respective child node states and deeper subordinate levels. As a consequence, mastering a parent node in the game, be it partially, will directly contribute to knowledge gains in all conditional nodes in the knowledge tree.

Each activity is supposed to somehow contribute to the mastery of learning goals, which means that a mapping of knowledge nodes to game activities is needed. Such mapping is not always straightforward, depending on the type of serious game (educational simulations, Serious Games, Serious Educational Games). Educational simulations as interactive representations of real-world phenomena used to practice tasks to be eventually performed in the real world and may be the least easy to map. Serious

Games are designed to develop skills in performing tasks using realistic situations and may contain by design a mapping of the knowledge. Serious Educational Games may be easiest to map in that they are similar to Serious Games but incorporate specific a priori pedagogical approaches to not only develop skills but teach specific learning content as well (Lamb et al. 2018). In all cases, various methodologies are available for defining the mapping, for instance Evidence-Based Design and Bayesian nets for stealth assessment (Shute 2011).

Modeling for Improving Game Design. To bolster efficiency, serious game design should be paired with instructional design, to optimize the use of game mechanics from entertainment and instructional principles. Appropriate modeling is essential to grasp the complexity of learning from and with serious games, in order to construct the required research base so that game design and instructional design cease to be viewed as ill-structured, artistic domains.

Improving game design through modeling is much more complex than modeling for assessing learning gains, requiring all three main interlinked constituents of the model (the knowledge model, the player model and the game model). The model Westera (2018) provides a proof of principle of a computational modelling methodology for serious games involving the three constituents that provides stable, reproducible and plausible results. Leveraging this approach however will require more detailed game states and player states, and careful tradeoffs to contain the combinatorial explosion of possible states.

Within the player model, players are characterized by intelligence, prior knowledge and susceptibility to flow, while their motivation and learning progress is evaluated and continually updated during their progression in the game. One limitation identified by Westera (2018) is that the model does not include cognitive models of human learning, but instead just relies on the phenomenology of the process of play. Connecting with existing models of human cognition would allow for including a multitude of psychological constructs, be it at the expense of simplicity.

Finally, the game model requires a knowledge tree indicating the learning objectives, and a set of game activities, possibly annotated with complexity and attractiveness indices. One limitation identified by Westera (2018) is that the model bypasses the complexities of instructional content and didactics by postulating that engagement in a game activity entails a productive learning experience. Second, although “game activities” are a key concept in the model, the meso-level indication does not say much about their grainsize. In fact, the model is ignorant and indifferent about the grainsize.

2.3 Hypotheses and Research Questions

In light of the previous considerations, this study investigates three hypotheses and two research questions:

Hypothesis 1: A serious game is beneficial to learning when learning is conceived of as shifts to scientific conceptions.

Hypothesis 2: A serious game is beneficial to learning when learning is conceived of as shifts among misconceptions, fillers and scientific conceptions.

Hypothesis 3: There are individual differences in learning across dyads.

Is the effect of agency on learning modulated by individual differences when learning is conceived of as shifts to scientific conceptions?

Is the effect of agency on learning modulated by individual differences when learning is conceived of as shifts among misconceptions, fillers and scientific conceptions?

3 Method

3.1 Sample

For this study, 82 paid volunteers (60\$) were recruited in undergraduate programs (primary education, special education, philosophy, and sociology) at University of Quebec at Montreal by the research coordinator who presented the research project in numerous classes. Only one participant is studying in another French-language university, University of Montreal, and was recruited by word of mouth. Since participants could volunteered individually or with a teammate, some participants were also recruited by snowball effect with respect of all including criteria. Participants volunteering in pairs formed a dyad for the experimentations, while individual volunteers were matched based on lab schedule and their respective availabilities. Students who attended a Physics class after high school, had severe skin allergies, had a pacemaker or had epilepsy were not included in the study. Moreover, the exclusion criteria included being under 18 years old or graduate students. Hence, 41 dyads of undergraduate students with novice background in Physics participated in the study.

The mean age of the remaining sample was 25.7 years old (ages ranged from 18 to 45 years old). There was 43 (58.10%) females and 31 males (41.89%). Most of the participants were right-handed (87.8%). Even though few players were left-handed, they reported using their right hand while using the computer mouse. A total of 22 players (29.73%) and 23 watchers (31.08%) declared a 10th grade knowledge level in Physics, and 13 players (17.57%) and 13 watchers (17.57%) in 11th grade. It was decided to kept the participant (player) who took a basic physic class at college in the sample, as well as the two participants (1 player and 1 watcher) who were previously educated in another school system because those two reported that they did not attend any Physics class after high school.

3.2 Task and Settings

Mecanika is a serious computer game developed by Boucher-Genesse et al. (2011). The game addresses 58 widespread misconceptions in Newtonian Physics (see Hestenes et al. 1992). Each of the 50 levels involves making an object move according to a given trajectory by applying different types of force to it. This involves choosing the right type(s) of force, the quantity of sources, and their appropriate positioning. The level is completed when the requested trajectory is entirely respected. In the view presented earlier, *Mecanika* is a tool to command a generative model of Newtonian mechanics and to test predictions about how physical objects behave.

In our paradigm, participants either played *Mecanika* or watched the player on a separate screen in real time. Their respective roles were randomly assigned. Participants progressed through the levels by achieving them or by being instructed to skip them after 20 min of play. This stop rule was apply in rare cases, in average, dyads skipped a level 4.3% of the total number of levels played. The task length was 2 h precisely (120 min); participants were stopped from playing without notice and research assistants entered the room, immediately talking to the participants and beginning their uninstallation. In average, each dyad played through 28.7 levels.

3.3 Measures

The *Force Concept Inventory* (FCI; Hestenes et al. 1992) is a widely used questionnaire designed to measure Newtonian's physic knowledge through six main concepts which are the three laws (first, second, and third), superposition principle, kinematics, and kinds of forces (see Fig. 1 for an example). Hestenes et al. (1992) has shown its equivalence to its predecessor, the *Mechanics Diagnostic*, and argued for its use as a diagnostic tool to identify misconceptions and for evaluating instruction, both in practical settings as well as research. The French adaptation of the FCI was administered immediately pre and post gameplay to establish learning gains attributable to the intervention.

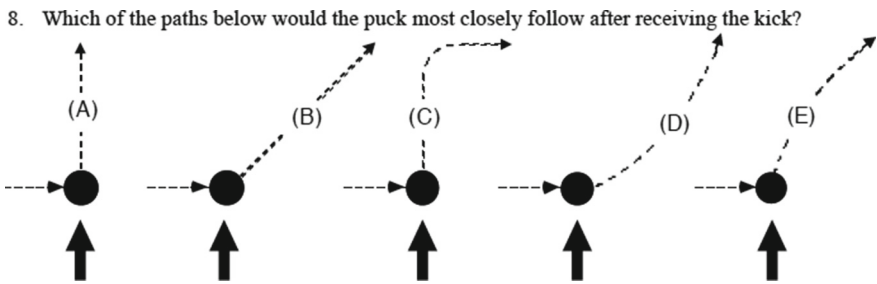


Fig. 1. Example of question from the *Force Concept Inventory* (FCI; Hestenes et al. 1992)

The FCI comprises 30 multiple-choice questions in which the choices reflect either the scientific conception underlying the question, documented misconceptions regarding the target knowledge, or fills that are wrong but not related to documented misconceptions. As can be seen in Figs. 2 and 3, the knowledge model of the game corresponds exactly to the structure of the FCI. It should be noted that each of the levels in *Mecanika* are specifically designed to address at least one conception/misconception within the FCI.

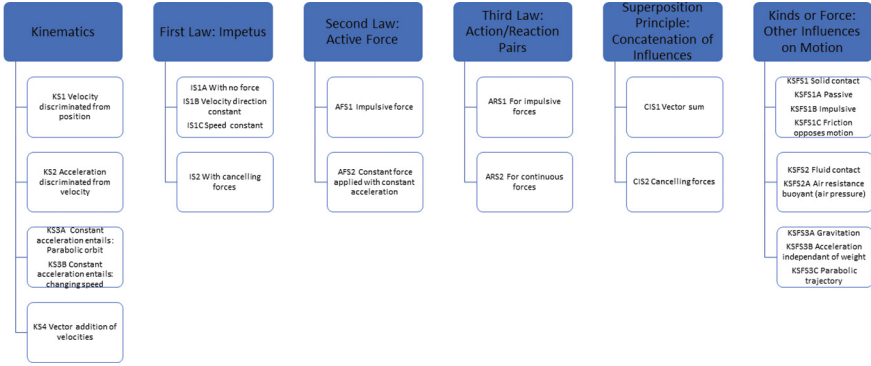


Fig. 2. The knowledge model underlying the levels in *Mecanika* – scientific conceptions.



Fig. 3. The knowledge model underlying the levels in *Mecanika* – misconceptions.

3.4 Data Preparation and Plan of Analysis

For each question in the FCI, transitions between the pretest and posttest were coded. Learning is operationalized in two manners in the present study, because both ways provide complementary evidence by either insisting on learning gains or on the nature of conceptual change.

In the first coding, the nine possible transitions from pretest to posttest were: *Fill to Fill*, *Misconception to Fill*, *Scientific to Fill*, *Fill to Misconception*, *Misconception to Misconception*, *Scientific to Misconception*, *Scientific to Scientific*, *Fill to Scientific*, and *Misconception to Scientific*.

Then, in the second coding, *Fills* and *Misconceptions* were collapsed as wrong answers (*Error*), leaving only the following 4 transitions: *Error to Error*, *Error to Scientific*, *Scientific to Error*, *Scientific to Scientific*.

Statistical tests according to the hypotheses and research questions were constructed using the loglinear approach since the data are categorical. The SAS CAT-MOD procedure was used (SAS Institute 2013), which provides a test of the influence of each factor, much like a usual factorial analysis of variance, but without an indication of effect size.

4 Results

The analysis shows that hypothesis 1 is accepted. A serious game is beneficial to learning when learning is conceived of as shifts to scientific conceptions. ($\chi^2_3 = 1558.02$, $p < .0001$). In addition to the 16.44% of answers already correct at pretest, combined for players and watchers, 12.04% of wrong answers at pretest transitioned to good answers at posttest.

Hypothesis 2 is also accepted. A serious game is beneficial to learning when learning is conceived of as shifts among misconceptions, fillers and scientific conceptions ($\chi^2_3 = 2808.86$, $p < .0001$). In addition to the 16.44% of answers already correct at pretest, wrong answers transitioned from fillers (1.36%) and misconceptions (10.69%) to good answers.

Finally, hypothesis 3 is accepted. There are individual differences in learning ($\chi^2_{81} = 149.22$, $p < .0001$). Dyads did not perform equally, and this is attributable to the unique experience created by the gameplay of the player of a given dyad.

The results for the question “Is the effect of agency on learning modulated by individual differences when learning is conceived of as shifts to scientific conceptions?” show that the effect of agency is not modulated by individual differences in learning ($\chi^2_{81} = 77.90$, $p = .58$). Systematically, watchers learn more than players (see Table 1).

Finally, the question “Is the effect of agency on learning modulated by individual differences when learning is conceived of as shifts among misconceptions, fillers and scientific conceptions?” could not be tested because of empty cells. Some of the nine possible transitions from pretest to posttest were too few or nonexistent.

Table 1. Transitions between pretest and posttest on the *Force Concept Inventory* (collapsing fills and misconceptions as wrong answers) by agency (player or watcher).

	Agency	Transitions between pretest and posttest				
		EtoE	EtoS	StoE	StoS	Total
Frequency	Players	786	134	95	197	1212
Percent		32.31	5.51	3.90	8.10	49.82
Row %		64.85	11.06	7.84	16.25	
Col %		50.51	45.73	51.63	49.25	
Frequency	Watchers	770	159	89	203	1221
Percent		31.65	6.54	3.66	8.34	50.18
Row %		63.06	13.02	7.29	16.63	
Col %		49.49	54.27	48.37	50.75	
Frequency	Total	1556	293	184	400	2433
Row %		63.95	12.04	7.56	16.44	100.00

E = Error
 S = Scientific conception

5 Discussion

The hypotheses and questions examined in this study jointly reveal that a serious game is beneficial to learning when learning is conceived of as shifts to scientific conceptions or as shifts among misconceptions, fillers and scientific conceptions. There are individual differences in learning in the sense that dyads did not perform equally, and this is attributable at least in part to the unique experience created by the gameplay of the player of a given dyad. Systematically, watchers learn more than players.

The findings regarding learning are in line with the theory and previous work. Because the impact of Serious Games may be highly dependent on the educational environments in which they are used (Stephanidis, Salvendy et al. 2019), it is interesting to note that the present learning gains attribution to a single, short gameplay episode complements a previous study by another team about the effectiveness of *Mecanika* (Boucher-Genesse et al. 2011), involving the same population used in this study but repeated play over weeks, which showed that playing the game was associated with a 30% gain in good answers on corresponding items of the same knowledge test, the *Force Concept Inventory*.

This clear and repeated empirical demonstration of the efficiency of the game is certainly facilitated by the clear relationship between game design and measures of knowledge gains in the form of the same knowledge model and supports Alonzo-Fernandez et al. (2019) assertion that the design of games should be based on clear learning goals.

By also stressing a need to specify how the attainment of the learning goals embedded in a Serious Game are to be measured, especially with interaction data, Alonzo-Fernandez and her colleagues (2019) corroborates the key elements required to further explain the patterns of results reported in the present study. This explanation is

especially required in the case of the present observation that active players systematically learn less than the passive watchers, which goes against prevalent intuitions in education as well as a major psychophysiological and cognitive theory of human functioning and learning that is highly relevant to the study of Serious Games, namely a prediction-action framework as recently discussed by Lupyan and Clark (2015).

Since some dyads did significantly better than others in terms of individual learning, and thus that this learning is influenced by the common experience provided by the gameplay of the dyad's player, is it possible to pinpoint the influence of the gameplay on the individuals' learning by using a fine-grained record of affective and cognitive processes of the players and watchers, constructed from continuous psychophysiological measures of cognition and affect?

It was shown elsewhere (Mercier et al., in press) that comparing means of around 7200 data points per participant representing second by second variations in cognitive load and cognitive engagement between groups (players and watchers) using traditional analysis of variance methods was hampered by a lack of statistical power. This realization, along with the present results showing individual differences in learning and the commonalities of this learning within dyads sharing a common gameplay experience, points to the need for within-subject analytic approaches linking aspects of performance (number of trials, time on task, etc.) with psychophysiological indexes such as cognitive load and cognitive engagement measured during a *Mecanika* level that ultimately led to a shift to a scientific conception. These analytic approaches could lead to a better understanding of the gaming experience that, in turn, could lead to a deeper understanding of relevant issues about the design of serious games. Extending the model developed by Westera (2018) is key in this endeavor. Now that a computationally tractable model has become available, the next research steps involve strengthening the construct validity and ecological validity of the model by including theoretically-sound concepts from extent research from the learning sciences. Then, additional studies are needed for the empirical validation of the model across a wide range of serious games, along with serious game theory development and ultimately the development of predictive systems and tools.

Extensions of the present work should contribute to the aims identified by Alonzo-Fernandez et al. (2019) by thickening the information that can be collected online during gameplay. The use of self-administered questions measuring knowledge gains attributable to gameplay could be an additional source of precious data. In addition, the manipulation of the context of use (agency), there is also a need to improve upon data sources used in learning analytics. Gains in information about the learner and the learning process are generally accompanied by efforts in gathering, preparing, transforming and interpreting data. Following Westera (2018), the present study should be finally extended by explorations of the relationships between the available knowledge model, the player model currently under development and the game model, which remains underdeveloped at this time.

Acknowledgements. The authors would like to acknowledge the important technical development work of Anthony Hosein Poitras Lowen underlying the study reported in this paper. The study was conducted with the support of the Social Sciences and Humanities Research Council of Canada and the Canada Foundation for Innovation.

References

- Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., Fernández-Manjón, B.: Applications of data science to game learning analytics data: a systematic literature review. *Comput. Educ.* **141** (2019). <https://doi.org/10.1016/j.compedu.2019.103612>
- Boucher-Genesse, F., Riopel, M., Potvin, P.: Research results for Mekanika, a game to learn Newtonian concepts. In: Games, Learning and Society Conference proceedings, pp. 31–38, Madison, Wisconsin (2011)
- Stephanidis, C., Salvendy, G., et al.: Seven HCI grand challenges. *Int. J. Hum.-Comput. Interact.* **35**(14), 1229–1269 (2019). <https://doi.org/10.1080/10447318.2019.1619259>
- Clark, A.: Expecting the world: perception, prediction, and the origins of human knowledge. *J. Philos.* **110**(9), 469–496 (2013)
- Hestenes, D., Wells, M., Swackhamer, G.: Force concept inventory. *Phys. Teach.* **30**, 141–158 (1992)
- Lamb, R.L., Annetta, L., Firestone, J., Etopio, E.: A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. *Comput. Hum. Behav.* **80**, 158–167 (2018)
- Lupyan, G., Clark, A.: Words and the world: predictive coding and the language-perception-cognition interface. *Curr. Dir. Psychol. Sci.* **24**(4), 279–284 (2015)
- Mercier, J., Avaca, I.L., Whissell-Turner, K., Paradis, A., Mikropoulos, T.A.: Agency affects learning outcomes with a serious game. In: 22nd International Conference on Human-Computer Interaction, Copenhagen, Denmark (2020, in press)
- SAS Institute Inc.: SAS/ACCESS® 9.4 Interface to ADABAS: Reference. SAS Institute Inc., Cary (2013)
- Shute, V.: Stealth assessment in computer-based games to support learning. In: Tobias, S., Fletcher, J.D. (eds.) *Computer Games and Instruction*, pp. 503–523. Information Age Publishing, Charlotte (2011)
- Westera, W.: Simulating serious games: a discrete-time computational model based on cognitive flow theory. *Interact. Learn. Environ.* **26**(4), 539–552 (2018)