UNIVERSITÉ DE GENÈVE

FACULTÉ DE PSYCHOLOGIE ET DES SCIENCES DE L'ÉDUCATION

tecfa

Technologies de formation et apprentissage

Embracing Complexity through Ill-Structured Problems – Play Traces Analyses of the Food Systems Game AL2049

Mémoire réalisé en vue de l'obtention de la maîtrise universitaire en sciences et technologies de l'apprentissage et de la formation

(MASTER OF SCIENCE IN LEARNING AND TEACHING TECHNOLOGIES, MALTT)

Kenneth Rioja

DIRECTION

Éric Sanchez, TECFA, FPSE, Université de Genève

Simon Morard, TECFA, FPSE, Université de Genève

JURY

Éric Sanchez, TECFA, FPSE, Université de Genève

Simon Morard, TECFA, FPSE, Université de Genève

Baltasar Fernández Manjón, e-UCM, Faculté en Sciences Informatiques, Université complutense de Madrid (UCM)

Genève, le 6 septembre, 2024

Word count : 15'754

UNIVERSITÉ DE GENÈVE
FACULTÉ DE PSYCHOLOGIE ET DES SCIENCES DE L'ÉDUCATION

tecfa
Technologies de formation et apprentissage

Déclaration sur l'honneur et auto-déclaration d'usage de l'intelligence artificielle (IA)

Je déclare que les conditions de réalisation de ce travail de mémoire respectent la charte d'éthique et de déontologie de l'Université de Genève. Je suis bien l'auteur de ce texte et atteste que toute affirmation qu'il contient et qui n'est pas le fruit de ma réflexion personnelle est attribuée à sa source ; tout passage recopié d'une autre source est en outre placé entre guillemets.

Au cours de la préparation de ce travail, j'ai utilisé 1) ChatGPT dans le but de rassembler des commandes et fonctions spécifiques sur R, pour une aide à écrire le résumé et une partie de la méthode (complexity) ; 2) Deepl pour de la traduction EN<->FR ; 3) Grammarly pour des corrections de grammaire et vocabulaire. Après avoir utilisé cet outil ou service, j'ai révisé et édité le contenu selon les besoins et assume l'entière responsabilité du contenu de la publication.

Genève, le 19 août 2024

Prénom, Nom : Kenneth, Rioja

Signature :

ACKNOWLEDGMENTS

Je tiens tout d'abord à remercier Éric Sanchez pour sa confiance et pour m'avoir donné cette rare opportunité de travailler sur ce set de données d'AL2049. Ce mémoire est seulement un morceau de ce tout ce que cache encore les données et j'espère avoir contribué à déchiffrer quelques-uns de ses *complexes* secrets. Je souhaite également remercier Simon Morard pour son temps, pour m'avoir guidé et conseillé à la suite des relectures des versions beta de mon mémoire. Thank you Baltasar for accepting to be part of the jury and for the opportunity to exchange about the present work. Je remercie aussi toute personne ayant contribué de peu ou de loin à la conception d'AL2049, en particulier Nicolas Godinot, Digital Kingdom et l'équipe du LIP sans qui je n'aurais jamais rédigé ce mémoire. Merci aux professeures et professeurs ainsi que les assistantes et assistants du MALTT, pour votre encadrement dévoué dans ce cursus unique, riche et rempli d'un potentiel inégalé. Je remercie aussi chaque étudiante et étudiant du master MALTT avec qui j'ai pu discuter ou travailler sur un des multiples projets du cursus. Une note spéciale pour les longues discussions à propos du jeu – et de la vie ! Merci Benoit Bediou pour m'avoir initié et généreusement aidé dans mes débuts sur R sans qui je n'aurais jamais pu faire les analyses suivantes. Je remercie également Daphné Bavelier qui a été, pour cette partie de ma carrière professionnelle, d'une aide incommensurable à la prise de décision et qui a abouti à un épanouissement intellectuel et professionnel dépassant mes attentes.

Merci à tous mes proches et ami·e·s qui ont écouté et qui se sont intéressé·e·s à ce que je faisais durant ces trois dernières années. Je remercie et je remercierai à jamais ma sœur, Kristine, mon père, Manuel et ma mère, Joy pour leur soutien et leur amour. Une pensée à Dylan.

Enfin, Estelle. Merci d'avoir été la voix de la raison et ma source de persévération quand j'ai failli, tremblé, pleuré, quand j'ai hésité à arrêter ceci ou cela. Pendant ces trois années à partager notre temps de qualité entre le master, l'école 42 et le labo en semaine, les travaux du Bellevue le week-end, le roller et les copains, tu m'as montré un soutien inconditionnel. Je te dédie cet énième accomplissement et les autres qui pourront en découler, c'est-à-dire toute la suite. *Profitons de nous et de la vie, comme on l'a toujours fait finalement*.

ABSTRACT

This master's thesis examines how players face ill-structured problems by interacting with a food systems game. The research investigates how a specific game design can nudge players to investigate multiple solutions to an ill-structured problem, rather than relying on a single approach. This work deciphers players' epistemic development (i.e., the ability to reshape prior knowledge and construct newer ones, Sanchez, 2022) through their exploration of the game's topic complexity. The study employs quantitative methods such as simple and multiple linear regression, correlation, clustering, and dynamic time series analysis to uncover patterns in play behavior. The analyses culminate in creating a complexity index that reflects players' epistemic development and their effectiveness in navigating complex scenarios within the game. This methodological framework provides a detailed set of indicators for each session, enabling the identification of epistemic development through the exploration of the game's representation of food system complexity.

Keywords: complexity; ill-structured problems; learning game; epistemic development; game learning analytics; dynamic time warping; ludicisation; food system; museum

TABLE OF CONTENTS

1 INTRODUCTION

Solving 21st-century problems in our ever-changing world is a hard matter. Why? Because our world is a complex system where every element, whether physical, digital or even ideological, has, to a certain degree, an impact on the rest of the system (e.g., living bodies, countries, society, etc.). It is thus a substantial challenge that ought not to be left out by citizens. Our current world is indeed facing more and more complex problems from tackling climate change to global inequality and poverty (see UNESCO 17 Sustainable Development Goals). As of today, we must have the cognitive and intellectual resources to think, prepare, and act for impact. Yet, it is not a new challenge. A quarter of a century ago, in 1999, Edgar Morin, Director-General of UNESCO said that "one of the greatest problems we face is how to adjust our way of thinking to meet the challenge of an increasingly complex, rapidly changing, unpredictable world. We must rethink our way of organizing knowledge" (p.7). To be able to live together in this rapidly changing world, we, citizens sharing this same world, must adapt to unpredictable outcomes. Training our ability to solve complex problems is a way to move forward, indeed, it is a critical soft skill to acquire and a duty to perform as citizens. By complex problems, we are pinpointing ill-structured problems: dilemmas with no verifiably correct solution (Jonassen, 2000).

"It is no longer possible for citizens to learn all they need to know about science in school or in higher education" (National Research Council, 2011, p.74). Squire and Patterson (2009, as cited by National Research Council, 2011) propose that informal science settings provide an important opportunity to improve the general scientific literacy necessary to respond to current social and scientific challenges (e.g., climate change, and pandemics). Museums seem to be the perfect informal science setting, and they are, by definition, places where we learn:

> "[It] was most traditionally the place consecrated to the Muses […], a mythological setting inhabited by the nine goddesses of poetry, music, and the liberal arts. 'They are called Muses', wrote the Chevalier de Jaucourt, 'from a Greek word which signifies "to explain the mysteries", […], because they have taught [people] very curious and important things which are from there brought to the attention of the vulgar.'" (Findlen, 1989, p.60)

Games can be a solution to even multiply this learning effect in museum settings and encourage visitors to solve ill-structured problems. They are now well known as a powerful medium to foster learning. Yet, learning games offer a non-negligible set of valuable features. As said earlier, they can train and engage players to solve ill-structured problems while collecting a tremendous amount of data for educational and research purposes. Indeed, data collected by learning games have the advantage to be objective, which is necessary to run sophisticated, trustful, quantitative analyses (Alonso-Fernández et al., 2019; Reardon et al., 2022). Indeed, the analysis of learning games (the field is called game learning analytics) seeks to enhance multiple facets of game learning like the creation of teacher-friendly roadmaps showing how to properly use the game to foster learning or to demonstrate the learning benefits and engaging experience of the games to cite a few.

The current work is part of the PLAY project, which sought to explore games for learning as a complete subjective experience. It seeks to include the context(s), the other player(s), and their interactions. It envisions learning games on a broader scale, as compared to gamification which solely adds game features on top of objects or contexts for engagement purposes. The first direction of the PLAY project is to design games that consider the subjective players' experience as a whole. Second, the project seeks to change the experience of museum visitors to encourage interaction with the museography, peers, and mediators. Third, the project wishes to approach the question of learning in the context of playful situations from the angle of the evolution of learners' relationship to knowledge. In short, the PLAY project aims to understand how, in museum scholar visits, learning games allow students to engage in solving complex problems.

AL2049 is one of those designed learning games made during the PLAY project. It is a learning game about food production that is exclusively played at the "Alimentarium", a food-themed museum based in Vevey, Switzerland. The game can be categorized as a resource management simulation game. Its development originates from design-based research (Sanchez & Monod-Ansaldi, 2015), that is, researchers, professional game developers (Digital Kingdom), teachers, and the curator of the museum worked together to create AL2049. One of the educational purposes of the game is to "understand the complex relationships between food production system components" (Oliveira et al., 2022). The game has been designed such that it allows players to solve an ill-structured problem about food system production in a museum about food. Ultimately, the game aims to develop players' skills to think and learn about food system complexity. The learning outcome is called epistemic development, it is the learning process in which individuals experience situations that lead them to question their prior knowledge and develop newer ones (Sanchez, 2022).

If AL2049 has been designed to make players think and refine their knowledge about the complexity of the food production system, what are the objective indicators of epistemic development *in-game,* and how to analyze them so that we can infer the resultant learning outcomes?

2 LITERATURE REVIEW

2.1 Museums, games, and learning

Engaging in museums' items on display is a step forward to a "critical reading of the world" as Borg and Mayo (2010, p. 42) said. Furthermore, the authors add: "The museums can be more representative in the forms of cultural production they display and the issues that they raise. They have the potential to capture the imagination of subaltern group members" (Borg & Mayo, 2010, p.42). Museums are, first and foremost, environments of cultural heritage coming with current issues that society is facing. This is why it is important to raise awareness to visitors by engaging them in the museography. Indeed, compared to a classical educational setting, museums have the advantage of allowing visitors to freely construct their knowledge. Moreover, this freedom allows agency which is one's ability to act in their current environment. The latter is known to be a core issue of any educational setting (Tchounikine, 2019). This sense of agency and control which is a known factor for increasing motivation and engagement (Ryan & Deci, 2000; Reid, 2012; Shu & Liu, 2019) can be further amplified by learning games. Gutwill and Allen (2012) have shown that practicing inquiry in a museum game-like setting allowed higher engagement levels. Moreover, the agency in games helps players feel safe to fail, persist, and feel ownership of their learning and is closely tied to self-efficacy (Bandura, 1977). This active learning monitors their understanding and helps them seek out opportunities to explore and apply what they discovered to shape their knowledge (Reardon et al., 2022), which, in turn, leverages meaningful learning (i.e. connecting new information with what you already know and with your relevant real-life personal experience, Brown, 2014, as cited by Reardon et al., 2022). Yet, meta-analyses on the effect of digital games on learning have heterogeneous outcomes. Clark and colleagues (2016) showed that digital games significantly enhanced students' learning compared to nongame conditions with a small effect size of 0.33. This view is supported by other meta-analyses (Barz et al., 2024; Riopel et al., 2019; and for critical thinking, see Mao et al., 2022). However, the effect sizes found in the literature can be heterogeneous and this may come from limitations in the meta-analyses as pinpointed by Riopel and colleagues (2019), involving factors such as the subject area and experimental simulation (Vogel et al., 2006) or the age of participants (Sitzmann & Ely, 2011). Moreover, when considering different types of moderators, the meta-analyses' positive effects on learning may drop to a null effect. This can be seen in Sitzmann and Ely (2011) with publication status (in favor of published articles) also no significant difference in learning was found when comparing active versus passive instructions (Wouters et al., 2013).

Museums are unique social environments to be able to learn and games can push this way. The more classical learning theories completely omit the social learning benefits (e.g., behaviorism, cognitivism, Piagetian constructivism) and affirm that learning is an experience printing a relatively permanent change in behavior (Vienneau, 2004). However, since our first minutes of life, we all learned from our relatives by observing, memorizing, and reproducing (Bandura & Walters, 1977). This later "social-cognitive" theory from Bandura spotlights "vicarious" learning (i.e., "experienced as a result of watching, listening to, or reading about the activities of other people, rather than by doing the activities yourself", Vicarious, 2024). Studies have shown that social interaction and collaboration foster learning (Doise et al., 2013; Roschelle & Teasley, 1995). When games are specifically designed to mediate and encourage social acts that constitute group learning, this can in turn lead to individual learning (Tchounikine, 2019). Museums are thus environments where learning can be fostered through games and social interactions with other visitors.

Games used in museums may also serve other aims. A meta-analysis of the use of serious educational games in museums listed three purposes for which serious educational games are designed (Wang & Nunes, 2019). The first is the communication of specific information about the museum environment. This provides visitors with another medium to show and exhibit the museum's collection of artifacts linked with a clear description of each object. This kind of learning is more formal and is equivalent to the transmission way of teaching. The second purpose of games is to train visitors in the acquisition of intellectual and physical skills. This is done through a well-defined learning objective that the game will try to solve through its learning outcome. Games can indeed have an impact on cultural, social, artistic, and commercial realities (Ermi & Mäyrä, 2005). The third purpose is experiential learning. It is less prescriptive than the previous way of doing and offers more opportunities for exploration. The study of Nelson and colleagues (2020) corroborates this view, their results show that visitors asked a significantly higher number of questions about the museum's topic when provided with a game compared to a more classic app. Indeed, experiential learning aims at providing interactive and complex environments. Above all, games offer a tailored solution in terms of game mechanics and aesthetics combining educational outcomes and museology features. Games thus hold a pivotal position for the maximization of the visitor's experience and museum impact (Paliokas & Sylaiou, 2016).

2.2 Ludicisation

Gamification, defined as "the use of game design elements in non-game contexts" (Deterding et al., 2011, p.10) is now widely used to make any setting or interaction more "fun" – at least to make it more engaging. In the past years, we can observe gamification as a motivational layer gently put over well-structured, deterministic problems leading to this famous image of a chocolate-dipped broccoli (Bruckman, 1999).

Ludicisation goes beyond by converting an ordinary situation (e.g., I am an individual visiting a museum) into a perceived play situation (e.g., I am part of a group of scientists which aims at feeding the local population by installing food facilities in specific museum spaces; Genvo, 2013). The model of ludicisation proposed by Bonnat and colleagues (2020) aims at converting a "target domain" which includes complex and multidisciplinary knowledge (i.e., the domain to be learned, e.g., food system), into a "source domain" that takes the form of a playful learning experience (i.e., the learning situation, the game). Ludicisation can thus allow players to engage in solving complex problems with no straightforward solutions to develop skills that are embedded in this specific context. The play experience is fundamental here, it must engage players in taking part in the construction of this experience by bringing their desires, anticipations, and previous experiences to reflect the experience in that light. (Ermi & Mäyrä, 2005). Ludicisation indeed spotlights the entire situation, involving the player(s), the context(s), and their interactions. It can thus create situations leading users to adjust their relationship to knowledge by evaluating the consequences of their behavior (Sanchez & Pierroux, 2015). In the work of Sanchez and colleagues., 2015, they showed that *Classcraft*, an epistemic role-playing game designed to manage classrooms in secondary school, employed ludicisation in the way that it metaphorically represented the class through a battle combining collaboration and competition. Indeed, it allowed engagement in the classroom through the respect of the game rules (i.e., class rules). The game also offers an individualized way to be continuously rewarded with feedback, a limitation when one teacher needs to manage an entire classroom. This continuous feedback thus developed a feeling of competence for the students. Moreover, the implementation of such a game in a classroom should consider the teachers' appropriation (Torrente et al., 2010) as well as the influence of the institutional context (Bonvin et al., 2019). Once play is made central, players' engagement is met and this situation, in return, alters the players' relation to knowledge.

2.3 Ill-structured problems, epistemic games, and complexity

Games are fascinating mediums that can offer a singular perspective on the kind of problem to be solved. One of the dimensions Jonassen (2000) raises in his typology of problem-solving concerns the problem type, opposing well-structured problems versus ill-structured problems.

Well-structured problems are finite problems in which they present all elements of the problem to the learner as well as a limited number of rules and principles. As can be seen in schools or universities, the problem is organized in predictive ways and has knowable solutions. Whereas ill-structured problems cannot be solved with a high level of certainty, moreover several solutions may be proposed, none of which is certain or verifiable, even by experts (Sanchez, 2022). For Jonassen (2000), a problem appears ill-structured for three reasons. First, because its problem elements are unknown or not known with any degree of confidence (Wood, 1983, as cited by Jonassen, 2000). Second, it can offer multiple solutions, solution paths, or no solutions at all (Kitchner, 1983, as cited by Jonassen, 2000). Third, there are multiple criteria for evaluating solutions, so there is uncertainty about which concepts, rules, and principles are necessary for the solution and how they are organized. More importantly, they often require learners to make judgments and express personal opinions or beliefs about the problem (Meacham & Emont, 1989, as cited in Jonassen, 2000), constructing worlds based on implicit or explicit values, visions, and ideas as can be seen in serious open-ended games (Squire, 2007).

> "In school, when students fail to have a feeling for the whole system which they are studying, when they fail to see it as a set of complex interactions and relationships, each fact and isolated element they memorize for their tests is meaningless. Further, there is no way they can use these facts and elements as leverage for action – and we would hardly want them to, given that acting in complex systems with no understanding can lead to disasters. Citizens with such limited understandings are going to be dangers to themselves and others in the future" Gee (2005, p14).

Ill-structured problems and how to solve them can find their roots in seminal works of the philosopher John Dewey (1933). He describes problems as "whatever perplexes and challenges the mind so that it makes belief at all uncertain" and engaging in solving the problem involves "(*a*) a state of perplexity, hesitation, doubt; and (*b*) an act of search or investigation directed toward bringing to light further facts which serve to corroborate or to nullify the suggested belief" (Dewey, 1933, p. 10). This last point in of major importance for us here: how one can, when feeling the urge of alleviating an itchy level of uncertainty, investigate and explore complexity to solve the problem? Moreover, solving ill-structured problems is an important skill for pupils to acquire because it allows them to defend reasoned and argued solutions, to revise their relationship with knowledge and ultimately to identify themselves as capable of devising solutions to that problem. Indeed, problems are found to be first initiated by reflexive thinking (Toussaint & Lavergne, 2005) and this change in relationship to knowledge leads to epistemic development. The latter is defined as our ability to construct, evaluate, reshape and use knowledge (Greene & Yu, 2016; Sanchez, 2022). Games fostering epistemic development are called epistemic games, they aim at developing ways of reasoning, acting, and communicating that are equivalent to those of professionals in a specialized domain (Shaffer, 2006). The objective is not to suggest a particular career trajectory but to facilitate the emergence of disciplinary thinking and its transfer it to other contexts (Rupp et al., 2010). For Sanchez (2022), epistemic games are designed to be more precise than serious games, in the sense that they emphasize the perceived play situation, as shown earlier with ludicisation. Thus, considering the play subjectivity and its interactions between player(s) and context(s). Greene and Yu (2016) show that epistemic development predicts academic outcomes such as critical thinking and argumentation, which is essential to address our societal complex problems.

But what is complexity? Ladyman and colleagues (2013) reviewed various attempts to define complexity and broke down the concept of complexity into seven features. The first feature, nonlinearity, refers to when small changes in the input can lead to unreasonably large changes in the output, or vice versa. The second feature is feedback which can be referred to as positive or negative loops in a system leading to an amplified or damped behavior. It emphasizes the interaction and communication between elements of a system, you can think of birds adjusting their course depending on the proximity and bearing of the other birds around it. Spontaneous order, the third feature in complexity, arises from the emergence of organized patterns and structures of a very large number of uncoordinated interactions between individual elements. The fourth characteristic is robustness and a lack of central control. As a distributed system, it maintains stability when subjected to perturbations. You can think about a flock of birds staying together and not being disrupted by wind or a random elimination of one or many birds. Emergence, the fifth complexity characteristic, refers to the phenomenon where local interactions of simpler elements lead to the creation of new, often unexpected, structures and functions at the macro level. The sixth feature is called hierarchical organization. You can think of a multi-level entity with structures and properties that interact with the levels above and below. The best example of such a system is an ecosystem or the whole system of life on Earth. And finally, complexity can be defined through numerosity, all the discussed features only happen if many parts are engaged in many interactions.

Taking a reductionist point of view, to understand complexity, we isolate individual elements to be able to progressively link them to understand the relationships between them. Indeed, any complex system in contemporary science will be explained through the lens of a non-exhaustive model to try to reproduce the larger complex system. (Lemire, 2008, as cited by Trestini, 2019). And by modeling reality, why not use learning games?

The current review showed us the importance of having epistemic games in museums to be able to train visitors' ability to tackle complex, ill-structured problems. Yet, we have not found scientific work on ill-structured problems or epistemic games about complexity. AL2049 seeks to "understand the complex relationships between food production system components" (Oliveira et al., 2022), but the real question is: how players tackle complexity in AL2049?

3 RESEARCH QUESTIONS

Solving complex problems in our current world is of capital importance in the 21st century. The project PLAY aims to understand how, in museum scholar visits, learning games allow students to engage in solving complex problems. AL2049, one of these games, sought to develop players' skills to engage, learn, and think about food system complexity.

Being able to visualize our world as a system of systems allows learners to tackle problems in their globality and to take action in a more impactful and meaningful way. Current classical educational settings are hardly teaching those skills mainly because of assessment issues, institution prioritization of the knowledge and skills to teach, or teachers' time to fit all learning material in a school year, etc. Museums are thus informal science settings where such knowledge can be offered to their visitors.

Indeed, previous findings show that museums and learning games form a brilliant mix supporting learning outside the classical educational settings. When combined, they merge rewarding needs such as competence, relatedness, and autonomy while engaging visitors in an attractive, and pleasing playful experience (Ryan & Deci, 2000). Moreover, learning games are a powerful, yet adaptable medium to guide players in tackling ill-structured problems. To solve complex problems, one must construct, reshape, and create new knowledge about the subject of interest (Sanchez, 2022). This epistemic development is known to be the premise of critical thinking and argumentation (Greene & Yu, 2016). Furthermore, we saw that the definition of complexity included multiple features (Ladyman et al., 2013), some of which are thought to be present in the game design of AL2049.

Is AL2049 an ill-structured problem and how is complexity defined in AL2049? From the seven complexity features, an *a priori* analysis was performed. This led us to narrow down our definition of complexity to its *nonlinearity* feature (see the *a priori* analysis in the appendix). Therefore, this master's thesis attempts to answer to this question: How AL2049 allows players to tackle food system complexity and enhance their epistemic development?

Find below our hypotheses:

H1: Players tackle food production as a system in AL2049

H2: Players tackle complexity in AL2049 through *nonlinearity*, a complexity indicator.

4 METHODS

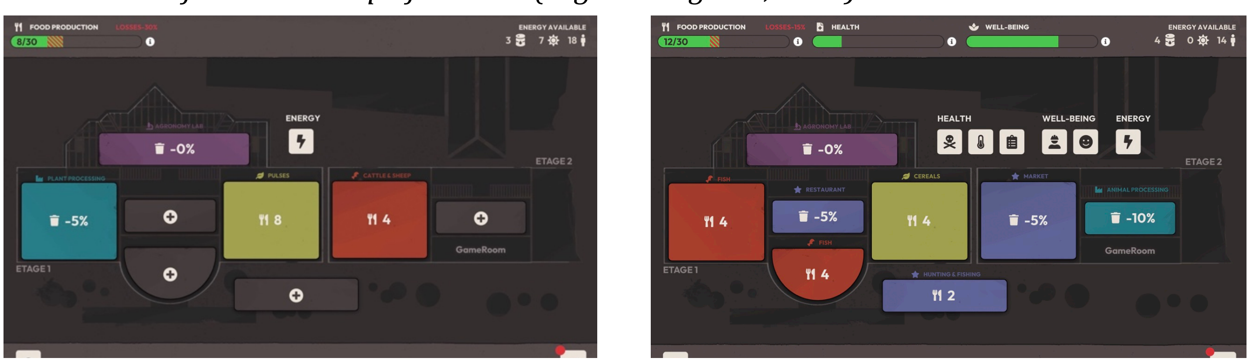
4.1 AL2049 game session

The game is played on tablets held by groups of three to four players and is supervised by a game master, usually a museum worker who knows the game and is responsible for giving the instructions. We call a 'session' each game played on a tablet. At the beginning of the game, the game master presents herself as a scientist working for the museum with one main goal: to feed thirty individuals living at the museum (which is the size of a classroom). The game first appears to players with a map of the museum (see Figure 1, left). However, spaces are locked. Players must physically walk to the different places of the museum to take a picture of the place and unlock the space. Once done, players must assign a function to the space such as livestock farming, crops, processing facilities, research labs, and food services. Assigning a function requires an amount of ten energy units, thus the player must choose to consume either fossil fuel (equal to ten units), renewable energy (three units), or human labor (one unit). During the entire session, the player has a finite number of four fossil fuels, ten renewable energies (plus three if a specific bonus is chosen), and thirty human labor units. The player can see on the top left corner of the screen a gauge of thirty individuals to feed which varies depending on the spaces' function assigned. A space's function can be deleted giving back the allocated energy units and allowing players to rebuild something else.

Once the eight spaces are assigned with functions (mostly happening after twenty minutes of play), players come back to the game master, and she allows the players to move ten years forward in time. A recap of the total number of fed individuals is displayed on-screen as well as two other gauges, a health level (how sick/healthy the population is) and a well-being level (how depressed/happy the population is). Therefore, the game implicitly swaps its main goal of "feeding thirty individuals" to "feeding thirty individuals and maximizing their health and well-being levels". Getting back to the map, the two new gauges are displayed next to the food production gauge as well as other buttons to display what spaces are positively and negatively involved in terms of health, well-being, and energy allocation

Figure 1

Screenshots of the main map of AL2049 (Digital Kingdom, 2019)



Note. Phase I is shown on the left and at phase II on the right. On the top left of the HUD (Heads-Up Display), we can see the number of individuals fed. Only in phase II, the Health and Well-Being levels are shown next to the individuals fed.

(see Figure 1, right). The play session stops either when players are satisfied with their museum's configuration or when the time to visit the museum for pupils is over.

4.2 Data description

Data, code, and supplementary materials can be found on osf (https://osf.io/7ytqf/). Data came from the PLAY project. Players are either pupils ranging from eleven to fifteen years old or adults. Data was collected at the Alimentarium (Vevey, Suisse) museum in an ecological context. Log data was collected at each confirmation of actions performed on-screen and not each screen tapping (i.e., event-based, Serrano-Laguna et al., 2017), this allowed players to continuously play without being interrupted by any kind of assessment (i.e., stealth assessment, Shute, 2011). Raw data (see TBALL_EventVariable_start_until2023_07_18_TRIEES.xlsx) were pre-processed before being analyzed; code and detailed procedure can be found on osf (https://osf.io/7ytqf/, see al2049_master_maltt_pre.R). Pre-processed "cleaned" data can be found in osf (https://osf.io/7ytqf/, see df_al2049_master_maltt_202408.csv). Here is the list of the cleaning steps:

- Split – The session ID "79e90533-9391-403a-b575-e4ca0657749d" had two games which is shown by having the verb "Start" twice. The decision has been to split it into two distinct session ID: "79e90533-9391-403a-b575-e4ca0657749**d**" and "79e90533-9391-403a-b575-e4ca0657749**e**".
- Game states transcription – A Game State is a string, e.g., *0#None;12!400;6!420!Soj;1!420;10!500;3!220;None;13!600;#[b14]*[b06,b03,b01,b13,b12,b10,b04]* (see "0.1_gamestate_decoding_EN_al2049_202408.txt", section I.) containing values about different variables. Added variables are listed in file "gamestate_decoding_EN_al2049_202408.txt", section V.
- Long to wide format for the list of variables – At each recording of the game state, the game wrote each variable in the table (see "0.1_gamestate_decoding_EN_al2049_202408.txt", section IV) and their modified value. To keep a uniform structure such as t_n = n^{th} action = n^{th} line, with each n^{th} line having all recorded variables at t_n, these variables were added as columns, and values were spread vertically in the table until the next change.

- Convert data type – Phase, human, green, and fossil variables were converted from char to num. Time was converted as POSIXct to be able to compute playtime.
- Remove sessions – The session "9a6a73fa-3808-4d7d-8f7b-ee7cc6a36403" has been removed due to having large play times.
- Rename sessions – Instead of having this long random session ID, we chose, after checking for non-duplicated names, to shorten the session ID to their first two and last two characters, making it easier to read.

This dataset gathers 174 sessions collected between September 2022 and June 2023. A session lasted 66.44 minutes on average (SD = 22.18), split between phase I (M = 45.95, SD = 23.63) and phase II (M = 22.79, SD = 7.68).

4.3 Variables

The following analyses focus on the primary mechanic of the game which is the resource allocation to build functions (that will, in turn, allow to feed individuals and/or to increase health and/or well-being levels).

The variables used throughout the analyses are 1) the proportion of fossil energy allocated ranging from 0 to 1, 1 corresponds to 4 fossil energies allocated; 2) the proportion of green energy allocated ranging from 0 to 1, 1 corresponds to 13 green energies allocated; 3) the proportion of human energy allocated ranging from 0 to 1, 1 corresponds to 30 human energies allocated; 4) the proportion of individuals fed ranging from 0 to 1.47, 1 corresponds to 30 individuals fed; 5) the health level, ranging from 0 to 1, 1 being a level of 9; 6) the well-being level, ranging from 0 to 1, 1 being a level of 13; 7) play time in minutes; 8) the total number of changes for a type of energy during a phase, labelled as n_{change}, e.g., if the players have changed four times fossil in phase I, then the indicator of n_{change} is equal to 4; 9) the amplitude of changes which will be the sum of the square differences between t_n and t_{n+1} in one type of energy during a phase, labelled as a_{change}; and 10) an indicator of global change multiplying the number and the amplitude of change, labelled as $n*a_{change}$.

We used proportions of energy allocated instead of raw values to ease reading. Therefore, a proportion of one equals to either the maximum value (4 for fossil, 13 for green, 30 for human energy), or the expected maximum value (30 for individuals fed), or the maximum value as found in the dataset (9 for health level and 9.4 for well-being level).

4.4 Quantitative analyses

4.4.1 To uncover player behavior

When faced with a large amount of play and game data, many authors use machine learning. Indeed, this technique was successful in describing many students' skills and knowledge (Martinez-Garza & Clark, 2017). Alonso-Fernández and colleagues (2019) reviewed data science techniques used in game learning analytics and showed that linear regressions, correlation, and clustering were the topmost used techniques in the field, respectively in 18, 17, and 16 out of 87 identified studies. We focus here on clustering which allows the exploration of data to identify groups of players with similar behaviors, as defined by patterns, or detect features that constitute such behaviors (Bauckhage et al., 2015). It has also been used to segment the target audience for user-oriented testing (Drachen et al., 2014). Bauckhage and colleagues (2015) listed seven key concerns encountered when using cluster analysis to evaluate player behaviors, including a) validation; b) interpretation and visualization, with an emphasis on feature selection to avoid low practical value results; c) time-dependency, players' behavior, thus clusters, may evolve; d) progress dependency, we should take into account the different progression between players; e) high dimensionality and big data, some clustering algorithms do not allow large scale data analyses; f) data type mixing, researchers should have a clear normalization strategy when using different type of data; g) feature selection, when having a higher order variable, find causes of an increase or a decrease of such variable.

Cluster interpretability is of major importance to accurately encapsulate the behaviors of players (Bauckhage et al., 2015). Indeed, the study of Drachen and colleagues (2012) supports this view. Titles and characteristics were given to behavioral clusters to allow better interpretability by the game designers of contemporary major commercial games ("AAA"-level). This can be done through the assignment of a simple expressive label to each found basis centroid (each player profile) to be easily interpreted by the readers (Drachen et al., 2012); Seif El-Nasr, Gagné et al., 2013). Drachen and colleagues (2014) ran a study comparing different unsupervised clustering methods like k-means, Non-negative Matrix Factorization (NMF), Principal Component Analysis (PCA), and Archetypal Analysis (via Simplex Volume Maximization). Using playtime and leveling speed data from World of Warcraft, they show that Archetypal Analysis is the best interpretable clustering method due to its real representation of players' "archetypical", yet extreme, behavior. When using k-means, players' representations (i.e., centroids) do not necessarily reside on existing data samples thus creating a mean of players' data making interpretation less easy. NMF and PCA show non-interpretable clusters, in their case the authors found decreasing player levels which is non-sense game-wise.

The previous studies used non-time-oriented data (Bauckhage et al., 2015; Drachen et al., 2012; Drachen et al., 2014) whereas Saas and colleagues (2016) used time series clustering. This study classified players by pattern shape and compared multiple techniques to cluster time series (e.g., Dynamic Time Warping, Correlation-Based Measure, Temporal Correlation and Raw Values Behavior Measure, Complexity-Invariant Distance Measure, Discrete Wavelet Transform, Symbolic Aggregate Approximation Related Function Measure and Trend Extraction). The choice of the clustering technique was argued and tailored to their application and business interest. Learning analytics researchers who emphasize the importance of temporal analysis consider learning as a process occurring over time (Molenaar & Wise, 2022). Moreover, it is important to have, before the analyses, a thoughtful conceptualization and operationalization of learning constructs concerning temporality (Knight et al., 2017). Indeed, the step of operationalization is a major step in the analysis of players' behavior, since constructs are unmeasurable, Landers and Bauer (2015) emphasized that the goal of operationalization is to ensure that the way we measure the construct is an accurate representation of the construct itself which leads in turn to precise, objective operational definitions. Following Reardon and colleagues (2022), game data comprise detailed information about players' frequence decision making which are tied to the specifics of the game world, making them more contextualized – and thus arduously generalizable to other contexts *and games*. The latter author listed four fundamental principles of game-based learning and how to analyze them: agency, engagement, growth, and social connection. For agency, they recommend categorizing players by recognizing their play styles and identifying goals or plans that players are pursuing. For engagement, they propose to measure the amount of gameplay, means, and variance in gameplay time or play time per session and analyze it using clustering to measure engagement. For growth, they suggest using cluster analysis of gameplay data that can be used to identify different learning phases such as exploration, tinkering, and refinement (EXTIRE framework, Berland et al., 2013, as cited by Reardon et al., 2022). Finally, for social connection, they pinpoint the importance of having mixed methods, gathering quantitative and qualitative methods since the social interactions mostly take place off-screen and are hard to fully capture through gameplay data.

The current work will use the aforementioned game-learning analytics techniques (Alonso-Fernández et al., 2019) such as linear regression to uncover the directionality of players behaviors and correlation to highlight the links between different game variables. We will also use non-time-oriented clustering (Bauckhage et al., 2015; Drachen et al., 2012; Drachen et al., 2014) as well as Dynamic Time Warping (Saas et al., 2016) to explore whether players behave in a similar way or not.

4.4.2 In epistemic games

Many books have been written about analytics on entertainment games (Gašević & Merceron, 2022; Lankoski & Bjork, 2015; Seif El-Nasr, Drachen et al., 2013; for a recent comprehensive review and classification of game analytics, see Su et al., 2021), but none have focused on the analytics of epistemic games. Yet, this field is not new and originates from the seminal work of Shaffer (2006) which uses Epistemic Network Analyses (ENA). Epistemic development can thus be analyzed through qualitative analyses, (Barzilai, 2017; Cabellos & Pozo, 2023; Hu et al., 2019), quantification of coded data through ENA (Rupp et al., 2010; Shaffer et al., 2016; Sweet & Rupp, 2012), or experimental design (Hu et al., 2019; Ke, 2019; Wang & Wang, 2017).

Another field of analytics, learning analytics focuses on learning as an outcome (Gašević et al., 2015). It is the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (Clow, 2013; Lang et al., 2022). The field has been found to miss a comprehensive understanding of how players interact with games and what tools are needed to understand the learning impact on players (Freire et al., 2016).

Few studies approached quantitative analyses of ill-structured problems to analyze players' behaviors in epistemic games. Kang and colleagues (2017) used sequential pattern mining, on an open-ended serious game. The study listed different patterns of sequential actions across different stages of the game, allowing the categorization of students (high-/low-performing groups) along with path diagrams to visualize the learning processes at play. Novoseltseva and colleagues (2022) focused on continuous sequences of actions of log data, leaving aside time. They also analyzed data using clustering and outlier detection. The study of Martinez-Garza and Clark (2017) sought to analyze players' epistemic stance. They used clustering which allowed them to revise their model called the Two-Stance Model framework (Martinez-Garza & Clark, 2016, as cited by Martinez-Garza & Clark, 2017). This model defines two different states for the student when playing an educational game, a Player state and a Learner state. The findings corroborate the two states, and they used their result to refine the model by adding the importance of prior knowledge, even if they did not support the back-and-forth between the two states.

4.4.3 In AL2059

We saw that the quantitative analysis of players' behaviors from an epistemic game offering an ill-structured problem is still an uncharted field. The present work focuses on analyzing play traces (i.e., pre-processed raw log data) of AL2049. It follows up Oliveira and colleagues' (2022) work where they called for quantitative analyses to see how the game is used by the players and how their behaviors evolve. Since log data are important to trace learning processes (Sanchez, 2022), the methodological purpose of this work sought to find indicators of epistemic development from behavioral data, and ultimately to show that tackling complexity as found in AL2049 fosters learning. How can we quantitatively interpret interactions with this game regarding the exploration of the complexity of the game's subject? The current work seeks to bridge the gap between game analytics and epistemic development analyses. We focus here on the quantitative analysis of players' behaviors during the game to identify epistemic development indicators and ultimately infer epistemic development processes (i.e. learning).

In phase I, the goal of the game is to “feed thirty individuals”, in phase II, the goal implicitly changes to “feed thirty individuals while keeping maximum health and well-being levels”. As said earlier, the current work will focus on the analysis of the primary mechanic of the game which is energy allocation.

A first set of analyses concerning the directionality of energy allocation within phases will be run. In phase I, we think that the players will allocate all kinds of resources with no distinction over the kind of energy used (fossil, green, human), the aim is solely to feed individuals using energy. In other words, we think that each type of energy allocation is increasing during phase I. However, in phase II, since the gauges of health and well-being are now displayed to the players, ineluctably changing the goal of the game, we are expecting a different manner of allocating resources. We are expecting, in phase II, a decrease in fossil energy usage as the game is coded to have fossil energy negatively impacting health (i.e., players ought to have high health levels) whereas green and human energy usage will fluctuate with no clear direction. Directions (i.e., increasing and decreasing) will be given through linear regression analyses. First, we will compute a multiple linear regression taking into account phase in the model to determine whether there is a difference of slopes between phases in each of the three energies, then we will compute simple linear regressions between play time and each of the three energies (fossil, green, human) for phase I and phase II to see whether participants are increasing or decreasing their usage within one phase. Finally, we will quantify at the session level whether players tend to increase or decrease their usage of energy by performing t-tests on slopes for each participant (see section 5.2).

After that the directions are given, how do the players allocate energies within each phase in a more precise and timely speaking fashion? We will look at how energy is used throughout the playtime within phases using Dynamic Time Warping (DTW). This will allow us to discriminate between different types of players through their energy allocation temporality, adding more information on top of their directionality (see section 5.3).

Once having in-depth details about one’s way of allocating energy, we will now be able to link this effective energy allocation with the coded game outcomes (i.e., game design). AL2049 has been designed so that players “understand the complex relationships between food production system components" (Oliveira et al., 2022), in other words, the game design should lead to a systemic understanding of food production. Correlating effective energy allocation (i.e., play behaviors) and game outcomes may confirm, first, that AL2049 has been played the way it has been designed (play behaviors confirming game design) and second, that players' behavior is indeed reflecting an understanding of systemic thinking (confirmed game design confirming pedagogical goal/food production as a system). This systemic thinking has been implemented in the game's functions, and especially the following functions involving energy allocation: more fossil energy allocated leads to a decrease in health; more human energy allocated leads to a decrease in well-being. Therefore, we are expecting, phase I, to have very high correlations between each of the energies and the number of people fed (primary goal in phase I, e.g., allocating resources to build functions to feed people). If the players have indeed played the way the game has been designed, we should see these significant high correlations. However, for phase II, we are expecting no clear link between the number of people fed and each of the three energies since the goal has shifted towards health and well-being levels. For correlations between energies, health, and well-being, we expect different outcomes. In phase II, health should be negatively linked with the use of fossil energy and well-being should be negatively linked with the use of human energy. Spearman correlations will be run to see these links (see section 5.4).

To quantify exploration in the game, we will look at the game's variables of change (i.e., how many times and to what extent energy allocation is changed). We will first check whether the number of changes is linked with the amplitude of change, indeed, one may change a lot while changing within a restricted range to create an indicator of change. We will then create a change indicator gathering the two previous variables. Compared to the directionality shown before, this indicator of change only attests to whether they explored many energy allocation configurations of the game or not. These analyses will be added to the previous ones to decipher whether the players are indeed exploring *in the right direction*. This change indicator will discriminate between players having equivalent exploration levels, while categorizing them in different bins, clustering methods will be used (see section 5.5).

Analyses of directionality with linear regressions (section 5.2), temporality with DTW (section 5.3), and game design confirmation with correlations (section 5.4) will answer our first hypothesis, whether players tackle food production as a system.

Analyses of change amplitude as shown by clusters (section 5.5) will answer to our second hypothesis, whether players tackle complexity in AL2049 through *nonlinearity*.

Our analyses will be finalized by the creation of a final indicator, an index reflecting epistemic development on the food system complexity, which we will call complexity index. It will be made post-analyses, and it will merge the previously found results (see section 5.6).

4.5 Analyses

Data were processed, analyzed, and visualized with R (v.2023.12.1.402, 2023.12.1.402`; Posit Team, 2023), using the following packages beepr (Bååth & Dobbyn, 2024), cluster (Maechler et al., 2023), dplyr (Wickham, François, et al., 2023), dtw (Giorgino, 2009), factoextra (Kassambara & Mundt, 2020), ggplot2 (Wickham, Chang, et al., 2024), NbClust (Charrad et al., 2014), pacman (Rinker, 2024), papaja (Aust & Barth, 2022), proxy (Meyer & Buchta, 2022), purr (Wickham, Henry, et al., 2023), rbbt (Dunnington, 2024), rstudioapi (Ushey et al., 2024) and tidyr (Wickham, Vaughan, et al., 2024). Spearman correlations were calculated with the function stats::cor(…, method = spearman). The optimal number of clusters was computed with factoextra::fviz_nbclust(…, cluster::pam, method = "gap_stat"). Partitionning Around Medoids (PAM) algorithm was performed using cluster::pam(…, n, metric = "manhattan"), n being the optimal number of clusters. The "metric" argument was set to "manhattan" because Manhattan distance is known to give more robust results if the data contains outliers, whereas Euclidean would be influenced by unusual values (Kassambara, 2024)

4.5.1 Linear regressions

Multiple linear regression considering phase in the model will allow us to determine whether there is a difference in players' energy allocation slopes across play time between phase I and phase II.

Since the multiple linear regression does not inform on how the slopes differ between phases, we will compute simple linear regressions between each of the three energies (fossil, green, human) and play time individually for phase I and phase II. We will thus have a more detailed view to see whether players are increasing or decreasing their energy allocation within one phase. This will give us information on the general slope direction within each energy and each phase.

We will finally look at the session level, whether one player increases, decreases, or has a flat curve of their usage of energy. This will allow us to be more detailed and quantify the previous analysis by pinpointing and counting players that increase, decrease, or have a flat curve.

Data was formatted so that the first time point reflected the time when the recording of the game was on (t=0, which does not mean that they made an action in the game). Plus, each time point reflected a change in the {energy}. And we finally kept the last value of {energy} when the phase ended. So for example, if we have a session ID in phase I using 3 renewable energy in t=5, 1 in t=8, and ending phase I at t=20, its data set would be [energy,time,0,0,3,5,1,8,1,20]. For phase II, we modified 'playTimeInMin' to reflect the time played since the beginning of phase II. Instead of having phase II beginning at t=25 and having green energy allocation of 2 at t=26, 1 at t=28, and ending phase II at t=30, its data set will be [energy,time,1,0,2,1,1,3,1,5]. Every model took each type of energy as a dependent variable, and time played in minutes as an independent variable. Phase was added as an independent variable for the multiple linear regression.

4.5.2 K-medoid clustering

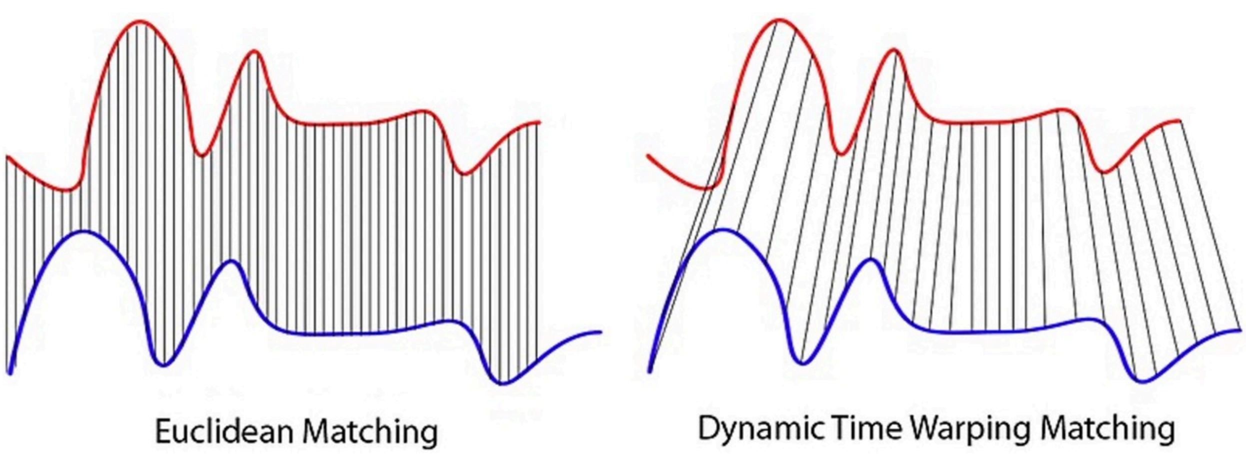
Clustering is an unsupervised machine learning technique, whose goal is to form homogeneous groups or clusters of objects having the least distance between data points of the same group and maximizing distance with out-group data points. The current work uses k-medoid algorithm instead of k-means as the clustering approach and has two advantages for our analyses. The first is to have clusters represented by the most central data point of the cluster, named cluster medoid, which is the representative example of the cluster. The second advantage is that this algorithm is less sensitive to noise and outliers, making it more robust than k-means (Kassambara, 2024). One of the most common k-medoids clustering methods is the PAM algorithm (Partitioning Around Medoids, (Kaufman & Rousseeuw (1990))

4.5.3 Dynamic Time Warping

Dynamic Time Warping (DTW) is a time-series analysis technique for measuring similarity between two sequences that may vary in time or speed (see Figure 2 for a graphical representation). Given two time series, the algorithm stretches or compresses the sequence locally to make one resemble the other as much as possible (Giorgino, 2009). In the context of play analytics, DTW is employed to compare sequences of gameplay events to identify patterns or similarities in player behavior and works particularly well to group similar player profiles with a shift on the time axis (Saas et al., 2016). This means that DTW can account for temporal variations, such as a player taking longer or shorter to complete certain tasks, while still capturing the overall similarity in gameplay patterns.

Figure 2

Distance measure comparison between Euclidean Matching and Dynamic Time Warping Matching (Commons Wikimedia – XantaCross, 2011)



Given two sequences $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_m)$, DTW finds an optimal alignment between them by minimizing the cumulative distance over all possible alignments (see equations in Giorgino, 2009).

To be able to use Dynamic Time Warping, we needed to arrange the data. When using raw time play in minutes for a given phase and energy and aggregating time series to have similar time points between sessions, it gave us too many time points (e.g., more than 800). Data arrangement thus consisted of 1) rounding each play time to minutes (e.g., if an action happened at 4 minutes and 39 sec, we chose to round the time at 5 minutes); 2) keeping the first time point and each time point where a change in energy allocation happened; 3) create the missing timepoints between each session to make them comparable and filling them with the previous value (e.g., session "001" has three timepoints at 2, 4 and 9 minutes with corresponding values 1, 2, 7 and session "002" has two time points at 2 and 8 minutes with values 4 and 8; they will both have time points at 2, 4, 8 and 9 minutes, "001" will have 1, 2, 2, 7 and "002" will have 0, 2, 8, 8). This means that each session in each phase for a given type of energy will have the same number of time points between the same phase and energy, but not necessarily between the other phase and energy (e.g., all sessions in phase II, green energy will have x time points, whereas all sessions in phase II, human energy will have y time points).

For each clustering, the order of the clusters and plots was manually manipulated to ease reading regarding our interpretations.

5 RESULTS

5.1 Time series

We first plotted the session-by-session time series, having in the x-axis the time played in minutes, and in the y-axis the energy allocation ranging from 0 to 1, with 1 being the maximum number of allocations for the specific type of energy, this is why for example, fossil energy allocation has steps of 0.25. Time series can be found in osf (https://osf.io/7ytqf/, see file "1.1_timeseries_complexity_al2049mastermaltt_202408.pdf" under the “supplementary_material” directory) and are ordered by session name in ascending number and alphabetically.

The plotted time series begins at the first players’ action, which is the first confirmed usage of any of the three types of energy. An effective session lasted in average 43.3 minutes (SD = 8.27), split between phase I ($M_{minutes}$ = 18.23, $SD_{minutes}$ = 4.71; $M_{actions}$ = 37.04, $SD_{actions}$ = 18.39) and phase II ($M_{minutes}$ = 22.78, $SD_{minutes}$ = 7.68; $M_{actions}$ = 78.19, $SD_{actions}$ = 39.38). When comparing the ratio of actions per minute between phase I and phase II, we can see that phase I is numerically handling fewer actions per minute (2.03) than phase II (3.43).

5.2 Directionality

5.2.1 Multiple linear regressions (phase level)

We found significant regressions for all of the three energies, for fossil energy allocation (F(3 ,2035) = 124.417, $p < 0.001$, $R^2_{adjusted}$ = 0.154, *equation*: fossil = 0.259 + 0.006 * (playTime) + 0.032 * (phase) - 0.006 * (playTime * phase)), for green (F(3 ,6355) = 665.628, $p < 0.001$, $R^2_{adjusted}$ = 0.239, *equation*: green = 0.334 + 0.003 * (playTime) + 0.36 * (phase) - 0.003 * (playTime * phase) as well as for human (F(3 ,10938) = 469.951, $p <$ 0.001, $R^2_{adjusted}$ = 0.114, *equation*: human = 0.506 + 0.003 * (playTime) + 0.282 * (phase) - 0.004 * (playTimeInMin * phase).

5.2.2 Simple linear regressions (phase level)

Computations show overall significant regressions. In phase I, fossil energy shows a positive slope, as seen with the beta (F(1 ,1072) = 150.611, $p < 0.001$, R^2 = 0.123, *equation*: $\text{fossil}_{(\text{phase I})}$ = 0.285 + 0.005 * (playTime), green energy shows a positive slope (F(1 ,1832) = 102.244, $p < 0.001$, R^2 = 0.053, *equation*: $\text{green}_{(\text{phase I})}$ = 0.345 + 0.003 * (playTime) and human energy shows a positive slope (F(1 ,3664) = 182.595, $p < 0.001$, R^2 = 0.047, *equation*: $\text{human}_{(\text{phase I})}$ = 0.512 + 0.003 * (playTime).

In phase II, fossil energy shows a statistically significant, negative slope (F(1 ,1289) = 47.368, $p < 0.001$, R^2 = 0.035, *equation*: $\text{fossil}_{(\text{phase II})}$ = 0.343 - 0.006 * (playTime), green energy shows a statistically significant positive slope (F(1 ,4824) = 203.665, $p < 0.001$, R^2 = 0.041, *equation*: $\text{green}_{(\text{phase II})}$ = 0.611 + 0.005 * (playTime). However, for human energy, the regression was not significant ($F(1, 7556) = 1.684$, $p = 0.194$), thus indicating no clear slope direction.

5.2.3 Simple linear regressions (session level)

The computed simple linear regressions show that over 174 sessions in phase I, for fossil energy, 73 (42%) have a significant positive slope ($p < .05$), and 0 sessions have a significant negative slope ($p < .05$). For green energy in phase I, 133 sessions (76%) have a significant positive slope, and 0 sessions have a significant negative slope. For human energy in phase I, 164 sessions (94%) have a significant positive slope, and 0 sessions have a significant negative slope.

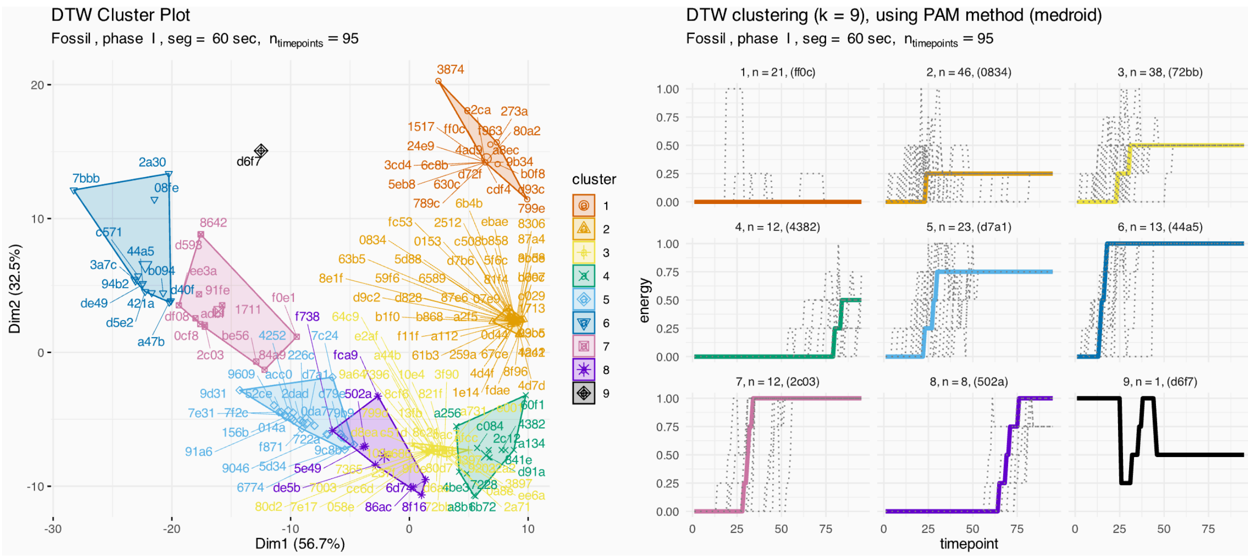
When looking at phase II, for fossil energy, 3 sessions (2%) have a significant positive slope, and 22 sessions (13%) have a significant negative slope. For green energy, 88 sessions (51%) have a significant positive slope, and 6 sessions (3%) have a significant negative slope. Finally, for human energy, 34 sessions (20%) have a significant positive slope, and 51 sessions (29%) have a significant negative slope. For a detailed view on the results for the 174 sessions for each energy and phase, tables can be found in osf (under supplementary_material, see CSV files from 2.1 to 2.6)

5.3 Temporality

5.3.1 Clusters of time series

Figure 3

Cluster plot for fossil energy in phase I using dynamic time warping (DTW) grouped by cluster highlighted with the associated medoid.



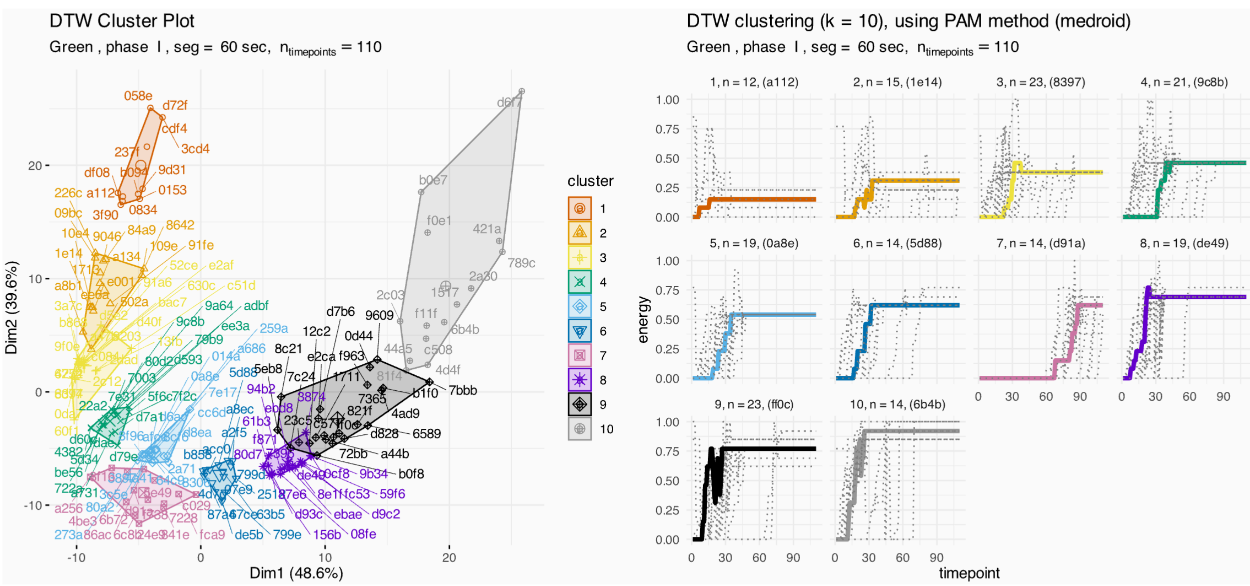
Note. On the left, clusters are shown by color, from 1 to 9 with colors varying from red to violet to black. On the right, ordered time series are grouped by cluster. *x*-axis shows time points (n = 95) and *y*-axis shows the proportion of used fossil energy. Above each grouping, is shown the number of the cluster, followed by the number of sessions grouped in this cluster and the session ID of the medoid between parentheses.

Concerning the consumption of fossil energy in phase I, we can see that, the optimal number of clusters is estimated to be nine. When looking at the cluster plot, it visually shows separated clusters, meaning that the algorithm successfully caught different patterns of fossil energy allocation over time. Figure 3 shows players' fossil energy allocation grouped by their cluster assignation. At first glance, the clusters seem to be mostly based on when and how much the player chose to consume fossil energy at the end of phase I. We remind the reader that the player has a stock of four fossil energies (thus, the steps at 0.25, 0.5, and 0.75).

Cluster #1 shows 21 players (12.1%) that chose to refrain from using fossil energy during the first phase, nonetheless, it can be found that players may use fuel energy *during* phase I. Cluster #2 shows 46 players (26.4%) using one unit at the end of the phase I. Two units of fossil fuels were used for 50 players (28.7%) categorized in clusters #3 and #4, the difference lies in the time frame in which they chose to do it, 38 players chose to use fossil fuel at the beginning of the phase while the rest chose to use it near the end of the phase. Cluster #5 groups 23 (13.2%) players who were using three units and 43 (24.7%) players were grouped reflecting the use of all the fossil units separated into three distinct clusters (#6, #7, #8), yet the difference between clusters, timely speaking, is in our opinion too small to be highlighted. Finally, there is one player that can be considered as an outlier, clustered alone in cluster #9.

Figure 4

Cluster plot for green energy in phase I using dynamic time warping (DTW) and time series grouped by cluster highlighted with the associated medoid.



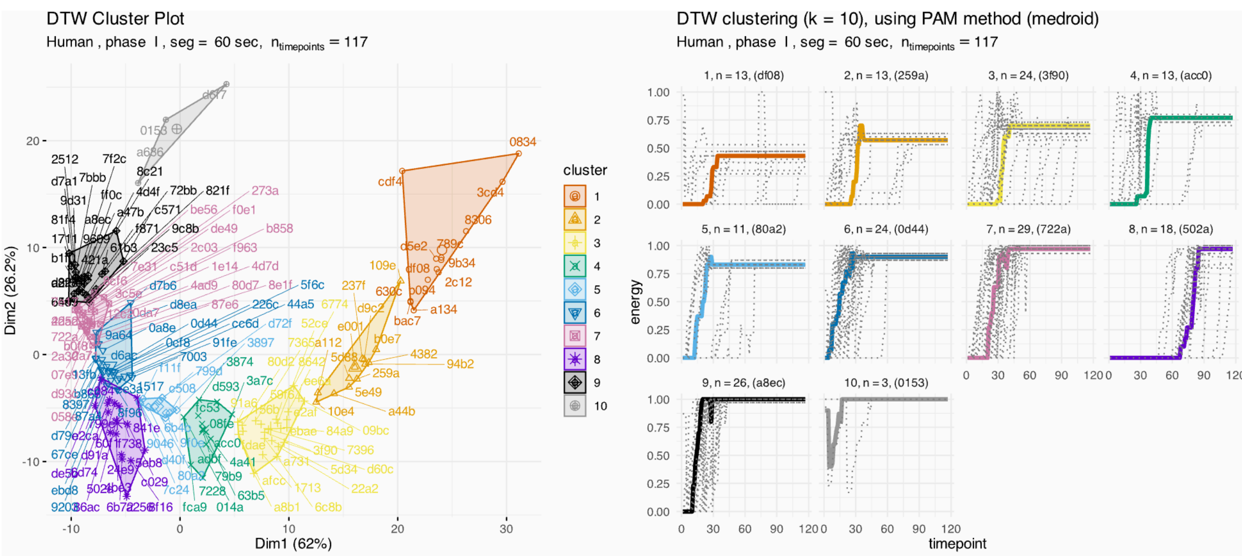
Note. On the left, clusters are shown by color, from 1 to 9 with colors varying from red to violet to black and grey. On the right, ordered time series are grouped by cluster. *x*-axis shows time points (*n* = 110) and *y*-axis shows the proportion of used green energy. Above each grouping, is shown the number of the cluster, followed by the number of sessions grouped in this cluster and the session ID of the medoid between parentheses.

For green energy allocation in phase I (see Figure 4), the optimal number of clusters is ten. On the one hand, the results show the same kind of results as for fossil consumption. We can see that from the first to the seventh cluster, the medoid picked shows players with increasing levels of green energy allocation over time. However, the last three clusters seem to show more variation than the previous clusters, with multiple ups and downs, before reaching a plateau.

For human labor in phase I, we computed an optimal number of clusters of ten (see Figure 5). From clusters #1 to #6, we have 98 sessions (56.3%) that overall increase their level of human consumption, clusters #7 and #9 seem to have more ups and downs than the other clusters. Cluster #10 show also three outliers, which are surprisingly beginning the phase with consumption of human energy reaching almost 100% at the first action.

Figure 5

Cluster plot for human energy in phase I using dynamic time warping (DTW) and time series grouped by cluster highlighted with the associated medoid.

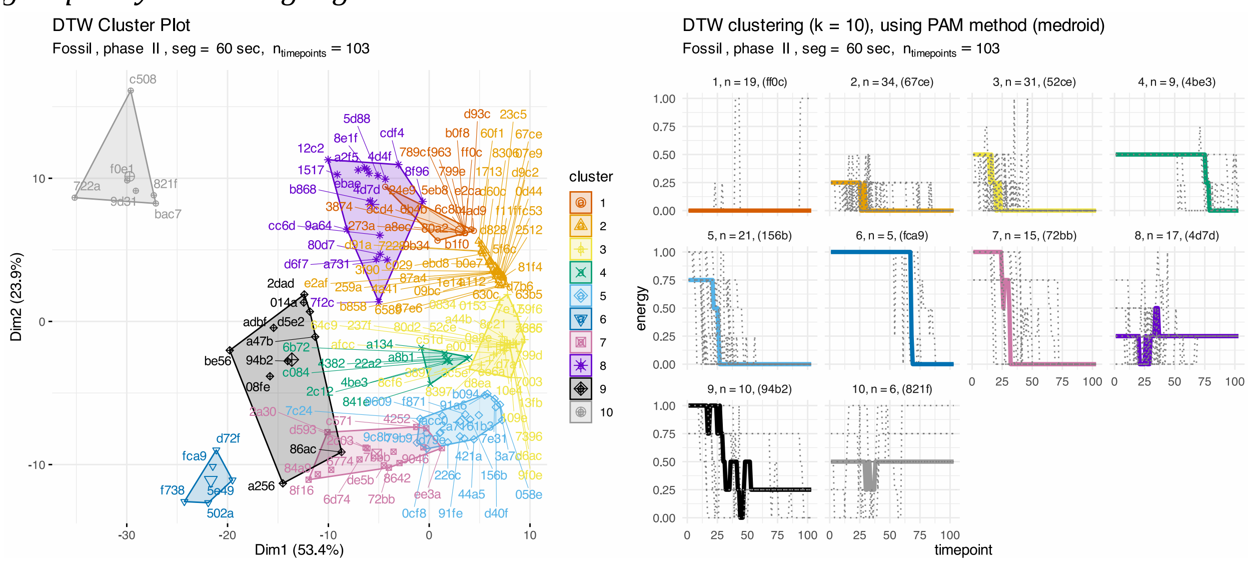


Note. On the left, clusters are shown by color, from 1 to 10 with colors varying from red to violet to black and grey. On the right, ordered time series are grouped by cluster. *x*-axis shows time points (n = 117) and *y*-axis shows the proportion of used human energy. Above each grouping, is shown the number of the cluster, followed by the number of sessions grouped in this cluster and the session ID of the medoid between parentheses.

For phase II, results were unanimous, the clustering method showed an optimal number of clusters of one (see the supplementary file, “4.0_dtw_al2049mastermaltt_202408.pdf”, pages 10 to 18, in osf https://osf.io/7ytqf/). These results may have been driven by outliers; this is why we decided to exclude sequentially the visually more distant sessions until having significant clusters. From the most to the less distant, the removed sessions for fossil were *0da7*, *1711*, *de49*, *df08*, *9203*, *fdae*, *5d34*; for green: *3cd4*, *9203*, *b0e7*, *789c* and for human: *3cd4*, *cdf4*, *d828*. The results shown below will reflect clusters without the previous sessions for each energy type.

Figure 6

Cluster plot for fossil energy in phase II using dynamic time warping (DTW) and time series grouped by cluster highlighted with the associated medoid.

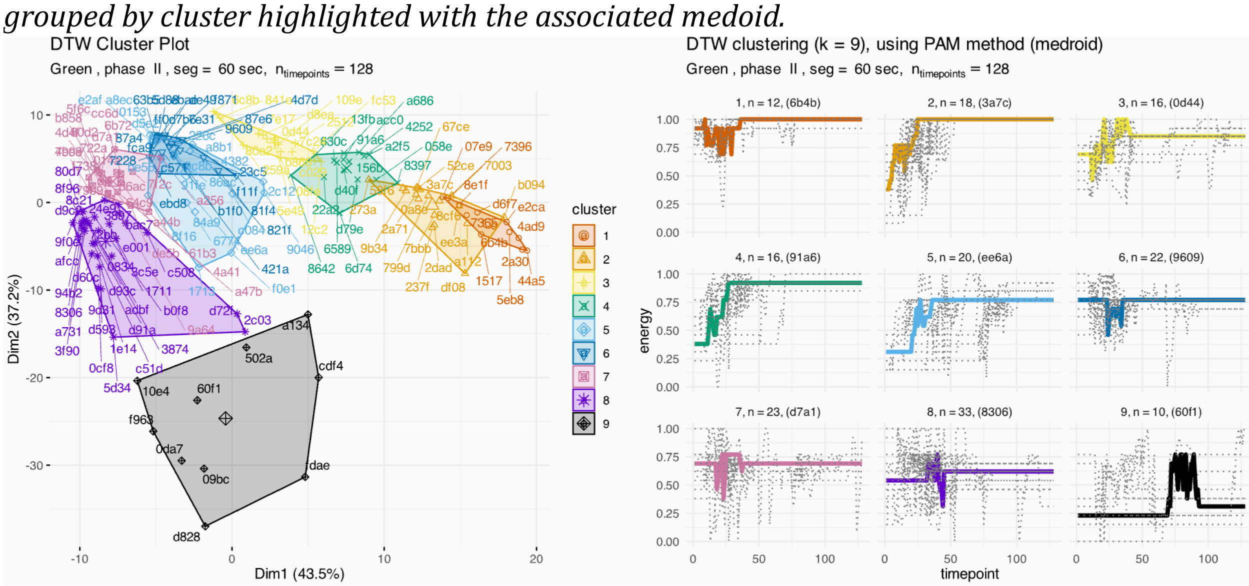


Note. On the left, clusters are shown by color, from 1 to 10 with colors varying from red to violet to black and grey. On the right, ordered time series are grouped by cluster. *x*-axis shows time points (n = 103) and *y*-axis shows the proportion of used fossil energy. Above each grouping, is shown the number of the cluster, followed by the number of sessions grouped in this cluster and the session ID of the medoid between parentheses.

For fossil energy in phase II, we have found an optimal number of clusters we can somewhat see the same patterns as fossil in phase I but in a reversed way (see Figure 6). Cluster #1 shows 19 sessions (11.3%) with a constant non-usage of fossil fuel. Most of the clusters (#1 to #7), representing 115 sessions (68.9%), show a decrease in consumption of fossil fuel ending at 0 at the end of phase II. Cluster #8 shows 17 sessions (10.2%) where they begin the phase with different levels of fossil consumption, and it seems that these sessions try to converge to a moderate use of 1 over 4 fossil energy. For cluster #9, we can see that most of the 10 sessions (6%) tend to decrease their consumption of fossil fuel energy, yet some sessions keep a small usage of it like the previous cluster. Finally, 6 sessions (3.6%) from cluster #10 seem to not follow any rule due to the high variations in terms of the level of fossil fuel at the beginning, during, and at the end of phase II. For clusters #7 to #10, we can see a lot of variation within each cluster during the entire phase. Cluster #8 and #9 tend to converge to low levels of fossil consumption unlike cluster #10 which has the inverse tendency to converge towards high values of fossil consumption. We can say that for fossil energy, the categories of players would be the green experts (#1), the fossil reducers (#2 to #7), and the convergers (#8 to #10).

Figure 7

Cluster plot for green energy in phase II using dynamic time warping (DTW) and time series grouped by cluster highlighted with the associated medoid.

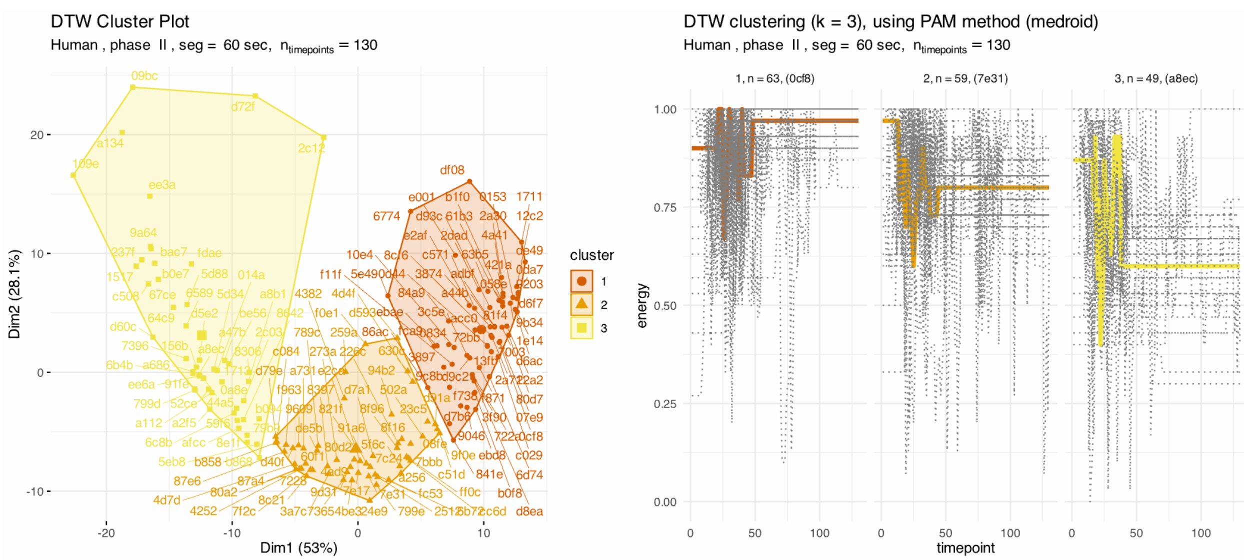


Note. On the left, clusters are shown by color, from 1 to 9 with colors varying from red to violet to black. On the right, ordered time series are grouped by cluster. *x*-axis shows time points (n = 128) and *y*-axis shows the proportion of used green energy. Above each grouping, is shown the number of the cluster, followed by the number of sessions grouped in this cluster and the session ID of the medoid between parentheses.

For green energy in phase II, the optimal number of clusters equals to nine (see Figure 7). Cluster #1 shows 12 sessions (7%) with already high levels of green energy and keeping themselves at the ceiling throughout the phase. Clusters #2 to #5 show a total of 70 sessions (41.2%) which are in general increasing their green energy allocation towards high levels of energy allocation. We will call them the "increasers". Clusters #6 to #8 show a total of 78 sessions (45.9%) which show either a convergence or a stabilization to medium-to-high consumption of green energy. By convergence, we mean no distinction between increasing or decreasing towards a certain level of energy allocation. We will call them the "mid-convergers". And finally, cluster #9 shows 10 sessions (5.9%) with no clear rules, the "no-rules". We can thus categorize the clusters themselves within four categories: the green experts (#1), the high-increasers (#2 to #5), the convergers (#6 to #8), and the no-rules (#9); clusters #6 to #9 can be also labeled as complexity players.

Figure 8

Cluster plot for green energy in phase II using dynamic time warping (DTW) and time series grouped by cluster highlighted with the associated medoid.



Note. On the left, clusters are shown by color, from 1 to 3 with colors varying from red to yellow. On the right, ordered time series are grouped by cluster. x-axis shows time points (n = 130) and y-axis shows the proportion of used human energy. Above each grouping, is shown the number of the cluster, followed by the number of sessions grouped in this cluster and the session ID of the medoid between parentheses.

Finally for human energy in phase II, the optimal number of clusters was set at three (see Figure 8). It is hard to discriminate between these three clusters since we see a lot of variation within a phase and between sessions. We will use the same labels as before. Cluster #1 shows 63 sessions (36.8%) with "high-convergers" cluster #2 shows 59 (34.5%) with high-to-mid-convergers" and cluster #3 shows 49 sessions (28.7%) with "mid-convergers".

5.4 Confirmation of the game design

For each phase and each session, Spearman correlations were run between each energy type and the people feed (main goal at phase I and shared goal at phase II) and the levels of health and well-being (shared goals in phase II).

5.4.1 Phase I, energies and people fed

In phase I, Spearman's rank correlation rho between fossil energy and people fed is positive, statistically significant, and very large (rho = 0.56, $p < .001$). Over 174 sessions, 67 sessions (38.5%) have significant positive correlations ($p < .05$) and 0 sessions have significant negative correlations ($p < .05$). The Spearman's rank correlation rho between green energy and people fed is positive, statistically significant, and very large (rho = 0.66, $p < .001$), 141 sessions (81%) have significant positive correlations, and 0 sessions have significant negative correlations. The Spearman's rank correlation rho between human energy and people fed is positive, statistically significant, and very large (rho = 0.77, $p < .001$), 170 sessions (97.8%) have significant positive correlations, and 0 sessions have significant negative correlations.

5.4.2 Phase II, energies and people fed

In phase II, the Spearman's rank correlation rho between fossil energy and people fed is positive, statistically significant, and small (rho = 0.17, $p < .001$), over 174 sessions, 33 sessions (19%) have significant positive correlations (p < .05) and 5 sessions (2.9%) have significant negative correlations (p < .05). The Spearman's rank correlation rho between green energy and people fed is positive, statistically significant, and large (rho = 0.39, $p < .001$), 80 sessions (46%) have significant positive correlations, and 4 sessions (2.3%) have significant negative correlations. The Spearman's rank correlation rho between human and people fed is positive, statistically significant, and very large (rho = 0.43, $p < .001$), 113 sessions (65%) have significant positive correlations, and 2 sessions (1.1%) have significant negative correlations.

5.4.3 Phase II, energies and health

In phase II, the Spearman's rank correlation rho between fossil energy and health is negative, statistically significant, and very large (rho = -0.61, $p < .001$), over 174 sessions, 1 session (0.6%) have significant positive correlations ($p < .05$) and 63 sessions (36.2%) have significant negative correlations ($p < .05$). The Spearman's rank correlation rho between green energy and health is positive, statistically significant, and large (rho = 0.33, $p < .001$), 64 sessions (36.8%) have significant positive correlations, and 8 sessions (4.6%) have significant negative correlations. The Spearman's rank correlation rho between human energy and health is positive, statistically significant, and near zero (rho = 0.06, $p < .001$), 43 sessions (24.7%) have significant positive correlations, and 30 sessions (17.2%) have significant negative correlations.

5.4.4 Phase II, energies and well-being

The Spearman's rank correlation rho between fossil energy and well-being is negative, statistically not significant, and tiny (rho = -0.05, $p = 0.076$), over 174 sessions, 7 sessions (4%) have significant positive correlations ($p < .05$), and 7 sessions (4%) have significant negative correlations ($p < .05$). The Spearman's rank correlation rho between green energy and well-being is negative, statistically significant, and tiny (rho = -0.04, $p = 0.010$), 20 sessions (14.5%) have significant positive correlations, and 24 sessions (13.8%) have significant negative correlations. The Spearman's rank correlation rho between human energy and well-being is negative, statistically significant, and very large (rho = -0.44, $p < .001$), 3 sessions (1.7%) have significant positive correlations, and 101 sessions (58%) have significant negative correlations.

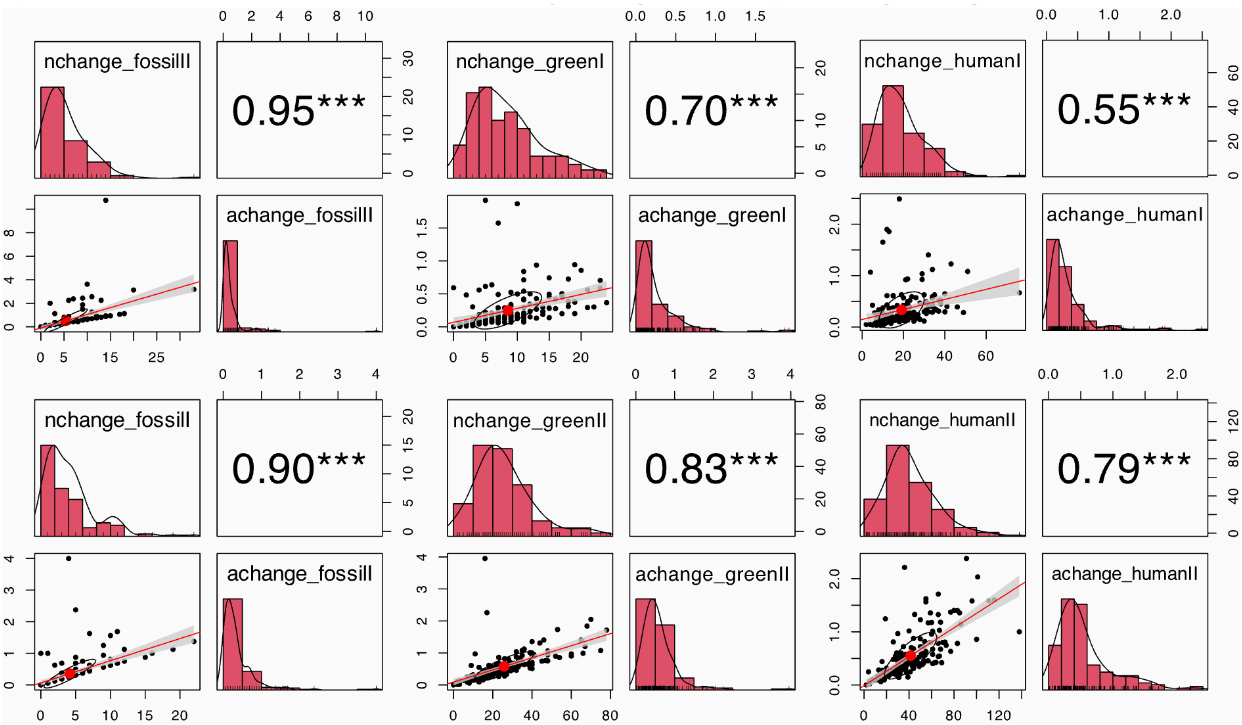
5.5 Change indicator to attest *nonlinearity*

5.5.1 Relations between number and amplitude of change

Spearman correlations between the number and the amplitude of changes were computed. The results show that every correlation between n_{change} and a_{change} in each phase and each energy is high, positive, and statistically significant (see Figure 9).

Figure 9

Spearman correlations between number of change and amplitude of change.

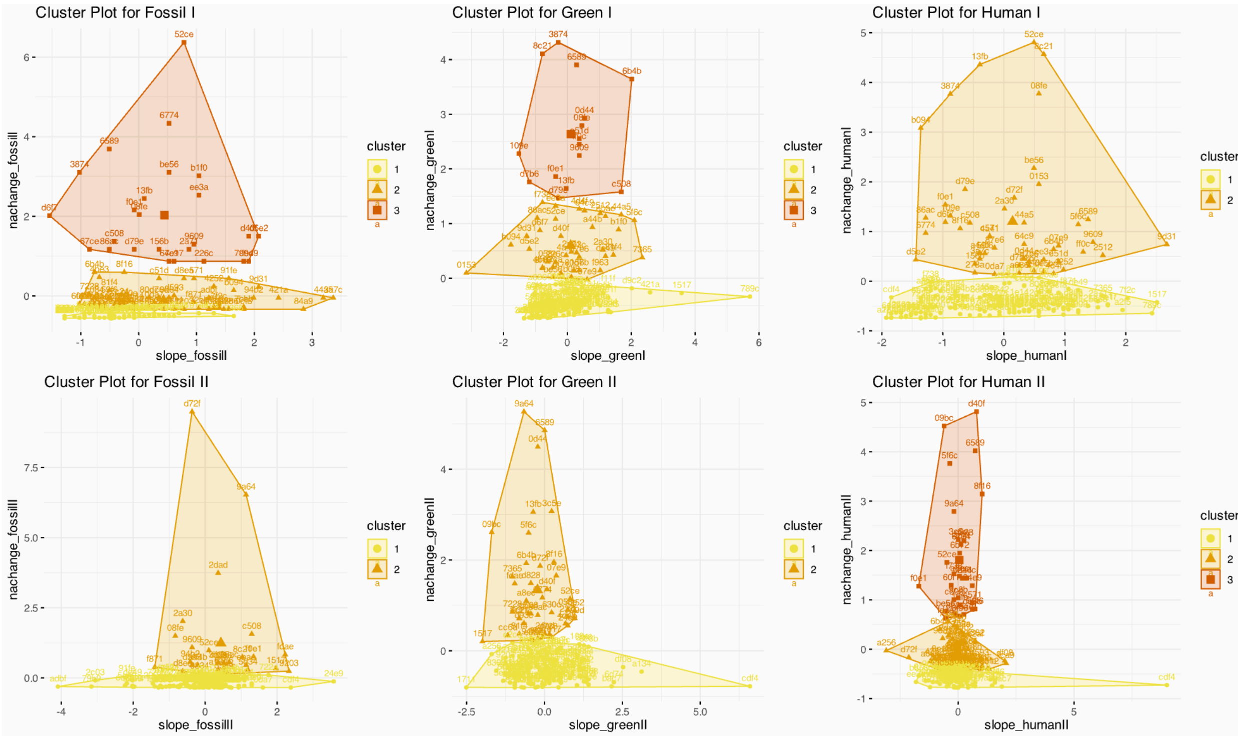


Note. *x*-axis shows the number of change (n_{change}) and *y*-axis shows the amplitude of change (a_{change}). First row shows phase I, second row shows phase II. First column shows fossil energy, second column shows green energy and third column shows human energy. Significance is shown by the asterisks (*** = $p < 0.001$).

5.5.2 Clusters of change

Figure 10

Cluster plots for energy allocation in each phase regarding slope and indicator of change.



Note. Clusters are shown by color, yellow shows cluster #1, orange shows cluster #2, red shows cluster #3. Medoids are shown by a bigger point in the scatter plot, a circle for cluster #1, a triangle for cluster #2 and a square for cluster #3. *x*-axis shows the slope and *y*-axis shows the indicator of change (n*a_{change}). First row shows phase I, second row shows phase II. First column shows fossil energy, second green energy and third human energy.

To better discriminate players' profile, we used Partitioning Around Metroid (PAM), an unsupervised learning k-medoid clustering algorithm. The clusters and graphs (Figure 10) show the cluster's medoid, which is a real session at the center of the cluster, this medoid/session. We computed the optimal number of clusters for each phase, for each energy, two for human energy at phase I as well as for fossil and green energies at phase II; and three for fossil and green energies at phase I, and human energy at phase II. At phase I, for fossil energy, we have 47.7% of the players categorized in cluster #1, 37.9% in cluster

#2, and 12.4% in cluster #3. For green energy, we have 69.5% of the players categorized in cluster #1, 21.8%, in cluster #2, and 8.6%in cluster #3. For human, we have 71.8% of the players categorized in cluster #1 and 28.1% in cluster #2. In phase II, for fossil energy, we have 83.3% of the players categorized in cluster #1 and 16.7% in cluster #2. For green energy, we have 85.9% of the players categorized in cluster #1 and 24.1% in cluster #2. Finally, for human energy, we have 40.8% of the players categorized in cluster #1, 42% in cluster #2, and 17.2% in cluster #3.

5.6 Complexity index

For each of the previously made clusters, i.e., for the DTW and change clusters, we labeled whether the attributed cluster group reflects complexity. First, we only took into account the clusters made in phase II where the problem involves multiple levels. We designed 6 criteria for each cluster in phase II, giving a total of, 2 (DTW and change clusters) * 3 (fossil, green, human), 6 criteria. Each criterion is a quantitative attempt to decipher epistemic development among players. From these criteria, we computed a complexity index which is the sum of each met criterion, find below the criteria:

- if, for fossil energy, the session was clustered in any group of **convergers,**
- if, for green energy, the session was clustered in any group of **convergers** or **no rule**,
- if, for human energy, the session was clustered in the group of **medium convergers**, since we only have three clusters and all show high levels of variability, we chose to make a distinction between the ones that are keeping high levels of human energy and the one who stayed relatively moderate in their use,
- if the session was clustered in the group of **high changers** for fossil, or green or human energy allocation.

Moreover, we set the complexity level to 0 if health and/or well-being level has not been modified throughout phase II.

The results show that on 174 sessions, 114 sessions (65.5%) meet at least 1 complexity criterion with the mechanic of energy allocation. We have 56 having a complexity index of 1 (32.2%), 23 at 2 (13.2%), 17 at 3 (9.8%), 12 at 4 (6.9%), 4 at 5 (2.3%) and 2 at 6 (1.1%).

6 DISCUSSION

6.1 Players tackling food production as a system

Our first analyses, as shown by multiple linear regressions, showed that players' energy allocation slopes significantly differ between phases. We can thus say that players do not use energy allocation the same way in phase I and phase II, and this can be explained by the change of goal between the two phases.

Our second set of simple linear regressions showed more detailed results on the difference in slopes between phases. We found positive slopes, yet small in effect sizes, for fossil, green, and human energy in phase I. This means that players in general are increasing their energy usage over time which is in line with the goal of the game at phase I: "feed 30 individuals".

Phase II, however, showed a different landscape. We found that fossil energy allocation is decreasing, green energy is increasing, and we found no statistically significant slopes for human energy. We can thus say that, in general, because players are decreasing their usage of fossil energy it must be compensated by an increase of their usage of green energy. Concerning human energy, we lack information at the player level on the absence of significant results. Are they consequences of *i)* different uses of human energy between players, some are increasing while some are decreasing (i.e., inter-session heterogeneity), or *ii)* is every session increasing and decreasing multiple times during the phase, leading to a general zero-slope (i.e., intra-session heterogeneity), or iii) are they not using human energy during phase II; or iv) is it a mix of all?

The third set of simple linear regressions allowed us to have a fine-grained view of each session. For phase I, we found high proportions of sessions having positive slopes (42%, 76%, and 94% respectively for fossil, green, and human energy).

It also gives a more detailed view, yet ambivalent, on what is happening in phase II. Indeed, for fossil energy in phase II, only ten percent of the sessions have statistically decreasing slopes. The lack of significant positive slopes for fossil consumption, in both phases, comes from the fact that many sessions were not playing much with fossil energy. This is due to the number of solely three available fossil energies compared to thirteen and thirty green and human energies respectively. Green energy usage is increasing in half of the sessions; however, we can see that the other half of the sessions have slopes close to zero *with a high number of time points*. Finally, we saw that human energy is more prone to every possible change in a balanced way. One-fifth of the sessions increased their use of human energy, almost a third decreased their usage while the remaining half of the sessions seem to show a null slope. This allows us to say that the non-significant slope that we previously found for human labor at phase II, is at least due to inter-session heterogeneity (point i). We cannot now say for intra-session heterogeneity (point ii) but we can exclude the third point (iii) due to the high numbers of data points (see degrees of freedom in file "2.6_slr_session_p2h_al2049mastermaltt_202408.csv", as found on osf https://osf.io/7ytqf/).

Clustering using Dynamic Time Warping in phase I, for all the energy types, shows clusters with sessions linearly increasing their energy allocation. However, there are two clusters for green energy (8 and 9, 10% of the players) and three clusters for human energy (7 and 9, 31.6% of the players) which show medoids with variations of increasing and decreasing periods of energy allocation. This result coupled with the previously found high number of positive slopes for each energy type at phase I shows that players' behaviors are directed towards one single goal: to increase their energy allocation. This makes phase I a well-defined problem: "feed thirty individuals", and this is achieved by the finite rule of increasing each energy allocation (Jonassen, 2000).

However, this is not as simple for phase II. Here, clusters do not show clear linearly increasing energy allocation. First, we see for fossil energy at phase II, in the seven first clusters a decrease in fossil energy allocation, while in clusters 8 to 10 (13.2% of the players) we see players who are converging towards medium levels of fossil energy. These last clusters suggest that players did include fossil energy in their representation of a food system, instead of completely removing it. For green and human energies, it is also clear that players' behavior varies within this second phase, we see here that players are trying multiple solutions, thus making phase II an ill-structured problem (Jonassen, 2000).

Players converging to medium values of energy allocation (cluster #6 to #9 for green energy and cluster #3 for human energy) are also players considering green and human energy in a system, compared to players strictly using (or never using) one or the other. Since we have a limited number of resources to allocate, players who perceive food production as a system may moderately use each of the energy at their disposal. This way they may include the benefits of each energy allocation in a system, compared to only wanting to use one.

When coming back to the causes of the null slope found in the simple linear regression for human energy at phase II (section 5.2.2.), here the plots can visually illustrate the intra-session heterogeneity – not yet quantifying it. We see a distinct increase/decrease of human energy allocation within phase II. This is a hint towards the fact that the null slopes found in 5.2.2. for human energy at phase II are due to a mix of inter- and intra-session heterogeneity.

DTW clustering is also a method that moderately captures players' environmental convictions/prior knowledge (Martinez-Garza & Clark, 2017). Indeed, we found clusters showing players not using energy fossil at all and using green energy a lot, they have been labeled as "Green Experts" ('ge' is the plots) in that sense. However, this conviction for less or no human labor is not found in our current dataset (e.g., philanthropists or humanists).

Moreover, this method is showing that players are tackling food production as a system. From homogeneous increasing usage of each energy type in phase I, to different play behaviors in phase II, we can say that phase I is a well-defined problem and phase II is an ill-structured problem (Jonassen, 2000).

Now for the correlations between play behaviors and game design, in phase I, the general large correlations between the number of people fed and each of the three types of energy coupled with the high number of sessions with significant positive correlations show that energy allocation is positively linked with number of people fed. This adds up to the results of positive slopes found with the simple linear regressions (see section 5.2.2). We can now safely say that the players used the energy allocation mechanic the way the game has been designed: increase your energy allocation to feed the 30 individuals. A note on the lack of significant correlations in fossil energy, this outlines again the very small number of observations for fossil energy, thus not providing enough statistical power.

For phase II, we can see numerically lower significant positive correlations than for phase I for individuals fed. This can suggest two things: a) the goal of feeding people is still a significant goal for players facing the ill-structured problem. This is still in line with the game design where players are told to keep feeding individuals while keeping maximum levels of health and well-being. b) this can also show a tendency of players to refuse to solve the ill-structured problem in phase II, thus keeping high positive correlations and making the general correlations at medium effect sizes. We should be able to track the latter looking at the player level and see if they have high correlations in feeding individuals but low correlations in health and well-being levels.

Again, for phase II, concerning correlations between each energy type and health level, the large negative statistically significant correlations between fossil energy and health level confirm the expected use of the game. However, we were not expecting that more usage of green energy would lead to higher health levels, this may be the repercussion of less using fossil energy, which needs to be replaced (e.g., replacing one fossil energy corresponds to 3 green energies in the game). The null correlation for well-being was expected since it does not affect health as found in the game design.

Finally, for correlations between each energy type and well-being levels in phase II, we found that fossil energy, as well as green energy, are not linked to well-being, but we have a very large significant negative relationship between human energy and well-being, which is in line with the game design.

These last results of phase II show that the game has successfully been played the way game designers intended in phase II. Combined with the fact that the game at phase II is shown to be an ill-structured problem, our results show that AL2049 indeed allows players to tackle food production as a system (Gee, 2005), thus corroborating our first hypothesis.

6.2 Players tackling nonlinearity

Spearman correlations between the number of changes and the amplitude of change in each energy and each phase showed that when players are changing a lot in a phase and in energy, they tend to have high amplitudes of changes as well (see Figure 9). We have now quantitatively shown intra-session heterogeneity. The visualized high number of directional changes as shown in DTW clusters is indeed illustrated by the current result showing high amplitudes in changes as well.

To be able to transfer this intra-session heterogeneity to an indicator, we chose to create the indicator of change n^*a_{change}. Clustering sessions through this indicator allowed us to better discriminate players with low, medium, and high changes. The labels of different clusters can be attributed as low changers (cluster #1), mid changers (cluster #2 when 3 clusters), and high changers (cluster #2 when 2 clusters and cluster #3 when 3 clusters). Sessions categorized in the group of high changers in phase II (cluster #2 for fossil, cluster #2 for green, and cluster #3 for human) are composed of players trying multiple configurations of energy allocation. If they did so, we can infer that they are facing the *nonlinearity* aspect of the food system complexity (Ladyman et al., 2013). As seen earlier, AL2049 has been designed with energy allocation having different effects on levels of health and well-being. However, these effects on the system can intersect with the other game design features such as the effect of the allocated function in a space. For example, one might deallocate fossil fuel, leading to increasing health, but still assign a function that decreases health, thus resulting in a null health outcome. Players then experience the *nonlinearity* of the system by confronting this unexpected result (null outcome on health) and they will try other solutions to finally lead to their expected result (positive outcome on health). This way, we can say they experienced food system complexity as not being solely impacted by one factor, i.e., *nonlinearity*. We thus corroborate our second hypothesis.

6.3 Players tackling food system complexity

To sum up, analyses showed that if we see clusters having visually high levels of change while having moderate levels of consumption for each energy, we would be more confident in saying that these players have tackled complexity *as a system* (see section 6.1). Furthermore, if someone has quantitatively high levels of change, it means that the players have experienced *nonlinearity* (see section 6.2). The complexity index merges this system's understanding of food production and *nonlinearity* to be able to attest to the players' epistemic development. Other than tackling complexity through the sole change of energy allocation, the complexity index incorporates clues of more reasonable usage of energies. The latter is important in the food system production as depicted by AL2049. Players should be able to understand that the food system's complexity lies to a certain extent in a fragile equilibrium between the consumption of fossil, green, and human energy to be able to have optimal levels of people fed, health, and well-being. Results show that almost two-thirds of the players have tackled food system complexity, and thus have developed a sense of complex reasoning, or epistemic development (Greene & Yu, 2016; Sanchez, 2022). Looking at their behaviors on energy allocation, we can finally say that players can explore food system complexity in AL2049 which lead to epistemic development.

6.4 Limitations

The current analyses only focused on one sole mechanic of the game, which is energy allocation. What if players did not play with energy allocation? Indeed our work promptly requires more analyses, with other mechanics of the game such as more incorporating the levels of People Fed, Health, and Well-being in relationship with the type of function allocated in the space.

Concerning clusters, this work shows that we can use clusters to discriminate unconventional data, see clusters 9 for fossil I and 10 for human I. Also, we want to make the reader cautious about the results given, when looking at inter-individual differences, we can see that within one cluster there may be players following another path than the medoid shown. Indeed, it does not consider the heterogeneity during a phase nor *when* the changes are during the phase, it only catches the overall 'silhouette' of the time series. If DTW cluster finds no cluster, it means that there is no similar pattern over time when players play in phase II, however, it does not say that players do not play with the same strategy. Indeed, clustering is only here to categorize the time series, but it cannot explain in depth what each player has done at each *t* point of the phase (e.g., what are the levels for each energy allocation at a given time point), nor how many "round trips" they have done in the usage of a specific energy.

The clusters found for DTW at phase II are the consequence of removing seven, four, and three players, respectively for fossil, green, and human energy. Our rationale was to find interpretable clusters and player behaviors. Further analyses should be able to find DTW clustering methods that include these outliers (e.g., by giving them less weight) instead of roughly removing these precious, objective data points.

Our complexity index is not meant to identify learning as a performance; however, it measures the way learners may tackle food system complexity *in-game*. Players changing a lot of time their way of allocating resources does not yet show that they are exploring complexity the *right* way. One may change a lot of their energy allocation without being in line with the game's goal, thus not learning. As Sanchez (2022) says, "the game is only a simplified model of the world, and learning through play is only possible according to this model" (p.64, free translation). Moreover, epistemic development is found to be a process, thus hard to capture with solely behavioral data. A note of caution is made here, we want to emphasize the fact that the complexity index reflects a relative position one has with another within the 174 sessions. One might be seen as having an important number of changes, but since the clusters were made using the current 174 sessions, the session in question might not be categorized as having a high number of changes. Furthermore, this index only finds the players tackling food production complexity through the usage of energies, one can have high variability on their levels of health and/or well-being but without having a high complexity level because these two outcomes have intricate coupling with other game mechanics (e.g., the chosen function).

This work has only been quantitative and ought to be mixed with qualitative studies. As the data was collected in an ecological setting (i.e., directly at the museum), we did not have any control nor visibility on the interactions between player(s), with the teacher or the game master that could happen before, while, and after the play session. As found by Morard and colleagues (2023), oil barrels as shown by pictograms could be understood as water barrels needed for food production. Indeed, we miss more qualitative data to explain how players tackle food system complexity and challenge their epistemic development.

Finally, we do not have information on the follow-up of this experience nor whether a debriefing session was always happening or not. "We do not learn by playing, we learn by reflecting on a playful experience, a metaphorical experience of life, an experience of ourselves." (Sanchez, 2022, p.161, free translation).

6.5 Recommendations

Our results have been restricted by some quantitative limitations (e.g., lack of data points, and variable time series). Future epistemic game designers, if they expect their game to be analyzed by quantitative means should have the following recommendations in mind.

6.5.1 To redesign AL2049

1. The game must have a gameplay allowing multiple actions or use of the same variable over each phase of interest. This has been found in the lack of significant results in fossil energy allocation, where the sole number of three fossil energies could not allow an in-depth description, quantitatively speaking, of how users interact with this kind of energy.
2. The game, if it is sought to compare different phases of interest (e.g., phase I and phase II), should restrict the time played to similar ranges. We have observed sessions with less time played in phase I compared to phase II, thus making comparison between the two phases and thus interpretations of the results more difficult.
3. Concerning the re-design of AL2049, to assess players' ability to offer multiple solutions to an ill-structured problem, we would suggest adding a new secondary mechanic in the game to 'capture' their optimal solutions allowing players to compare between solutions. It would in turn give objectively chosen game snapshots to learning analysts to say that at these times, players made the decision that this was their 'best' current solution. Also, it would be considered as a time of contemplation, reflection, and knowledge reframing about the chosen solution.

6.5.2 For future work on AL2049

Future work can explore sequential pattern mining as found by Kang and colleagues (2017) or sequence mining as found in Martinez-Garza and Clark (2017) to find patterns in sequence data or using Hidden Markov Models as found in Tissenbaum and colleagues (2016). We would be curious to know what chunks of actions are done inside each phase, and if those are predicting a certain type of player or player's complexity index.

6.5.3 In the field of game learning analytics

1. Learning game designers should have a clear game design document (commonly known as GDD), to share along each researcher, even the one late to the party. This could be presented as a centralized document where we can find – along the game design document can offer: the list of collected variables in log data is present (e.g., data dictionary), the exact verbatim of the game master (e.g., what she says before and after the game and what she is allowed to say, or not), the expected timeline of the game, the list of the published papers, the *why* of the game development, etc.
2. The methodology significantly differs when data has already been collected compared to when we plan what analyses will be done once the data is collected. When designing a learning game, researchers should be aware of game design features such as adaptation and assessment which are essential from a pedagogical point of view (Moreno-Ger et al., 2008). Building on top of Serrano-Laguna and colleagues' work (2017), a standard to add in GLA/LA procedures is to plan how the collected data should be analyzed (i.e., preregistering analyses while designing the game). This is mandatory for the assessment of the student's skills, but it is crucial to corroborate or refute prior educational purposes in line with the game design.
3. Building trust in employing learning games in educational settings is key and it is the main hobbyhorse of the field. Our contribution to this is to have open data and transparency about the analyses. AL2049 can have effects on epistemic knowledge and this entire work has been dedicated to that and is freely available. We also encourage future work to follow the same path.

6.5.4 To teachers

The game is found to be an engaging learning material for students to explore food system complexity, as depicted by the numerically higher number of actions in phase II showing a growing interest in the topic. Students can experience awareness that food production is a complex question hinging on multiple factors. This effectiveness is due to the unique environment in which the game is included (i.e., ludicisation): the game is in a museum, allowing students to physically and freely explore the museum, it is supervised by a game master for the story and students play by groups to allow even more interaction.

7 CONCLUSION

Since Mystery at the Museum, one of the first famous games implemented for the Museum of Science in 2005 (Klopfer et al., 2005), learning games flourished in museums' settings. Indeed, this medium is now accepted as a learning tool and can express complex ideas (Squire, 2021). Yet very few games offer the opportunity to face ill-structured problems, that is, problems that are intricately inserted in a specialized domain where several solutions may be proposed, none of which is certain or verifiable (Sanchez, 2022). This type of game is called an epistemic game, it aims at leveraging students' ability to reason, act, and communicate in the same ways as professionals (Shaffer 2006, as cited by Sweet & Rupp, 2012).

This work aimed at understanding how an epistemic game allows players to tackle food system complexity and enhance their epistemic development. The game is called AL2049, and it offers a playful opportunity to tackle food production systems challenges. Indeed, the game's educational purpose was to "understand the complex relationships between food production system components" (Oliveira et al., 2022). Analyses were run on a dataset of 174 sessions of AL2049 and the methods used were simple and multiple linear regressions, correlations, k-medoid clustering, and dynamic time warping. This work, one of the first in the field, aims at finding quantitative indicators of epistemic development.

The analyses showed that players are changing their behaviors between phase I and phase II, going from a well-structured problem to an ill-structured problem. This second phase, where an ill-structured problem is depicted, showed more actions from the players, leading to an increase of engagement in the subject matter. Moreover, the analyses found that the game successfully transmitted this sense of system to players, it has been shown by the confirmation of the game design through the actual players' behaviors. AL2049 has been found to tackle food production complexity through one of its complexity features, *nonlinearity*. Players experience the fact that one change in the system can lead to unexpected outcomes by showing a high number of changes in their behavior. Finally, our analyses ended with an index of epistemic development, concluding that two-thirds of the players have signs of embracing AL2049 food production complexity, reshaping their knowledge on the specific topic.

This master's thesis contributed to the emerging field of Game Learning Analytics by leveraging quantitative techniques such as clustering and Dynamic Time Warping. The current work sought to find quantitative indicators of epistemic development, and we showed that it is possible, yet interpretations should gather more information on other game mechanics (e.g., not only relying on one only) and underline the importance of qualitative data (e.g., user experience, focus group, video data) to be able to arrive to our conclusions safely.

8 REFERENCES

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, *141*, 103612. https://doi.org/10.1016/j.compedu.2019.103612

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown* (0.2.0) [HTML]. https://github.com/crsh/papaja (Original work published 2014)

Bååth, R., & Dobbyn, A. (2024). *beepr: Easily Play Notification Sounds on any Platform* (2.0) [Computer software]. https://cran.r-project.org/web/packages/beepr/index.html

Bandura, A. (with Internet Archive). (1977). *Social learning theory*. Englewood Cliffs, N.J. : Prentice Hall. http://archive.org/details/sociallearningth0000band

Bandura, A., & Walters, R. H. (1977). *Social learning theory (Vol. 1).* Prentice Hall: Englewood cliffs.

Barz, N., Benick, M., Dörrenbächer-Ulrich, L., & Perels, F. (2024). The Effect of Digital Game-Based Learning Interventions on Cognitive, Metacognitive, and Affective-Motivational Learning Outcomes in School: A Meta-Analysis. *Review of Educational Research*, *94*(2), 193–227. https://doi.org/10.3102/00346543231167795

Barzilai, S. (2017). "Half-reliable": A qualitative analysis of epistemic thinking in and about a digital game. *Contemporary Educational Psychology*, *51*, 51–66. https://doi.org/10.1016/j.cedpsych.2017.06.004

Bauckhage, C., Drachen, A., & Sifa, R. (2015). Clustering Game Behavior Data. *IEEE Transactions on Computational Intelligence and AI in Games*, *7*(3), 266–278. https://doi.org/10.1109/TCIAIG.2014.2376982

Bonnat, C., Marzin, P., Luengo, V., Trgalová, J., Chaachoua, H., & Bessot, A. (2020). Proposition d'un modèle pour la compréhension des décisions didactiques d'un enseignant. *Éducation et didactique*, *14–3*, Article 14–3. https://doi.org/10.4000/educationdidactique.7793

Bonvin, G., Gonçalves, C., & Sanchez, E. (2019). *Ludicisation de la gestion de classe avec Classcraft : une étude empirique*. Communication présentée à 9e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Paris, France. http://hdl.handle.net/20.500.12162/319

Borg, C., & Mayo, P. (2010). Museums: Adult education as cultural politics. *New Directions for Adult and Continuing Education*, *2010*(127), 35–44. https://doi.org/10.1002/ace.379

Bruckman, A. (1999). Can educational be fun. *Game developpers conference.* https://faculty.cc.gatech.edu/~asb/papers/conference/bruckman-gdc99.pdf

Cabellos, B., & Pozo, J.-I. (2023). Can Video Games Promote Moral Cognition? Supporting Epistemic Play in Papers, Please through Dialogue. *Education Sciences*, *13*(9), Article 9. https://doi.org/10.3390/educsci13090929

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*, 1–36. https://doi.org/10.18637/jss.v061.i06

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*(1), 79–122. https://doi.org/10.3102/0034654315582065

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D.C. Heath and company.

Doise, W., Mugny, G., James, A. S., Emler, N., & Mackie, D. (2013). *The Social Development of the Intellect*. Elsevier.

Drachen, A., Sifa, R., Bauckhage, C., & Thurau, C. (2012). Guns, swords and data: Clustering of player behavior in computer games in the wild. *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, 163–170. 2012 IEEE Conference on Computational Intelligence and Games (CIG). https://doi.org/10.1109/CIG.2012.6374152

Drachen, A., Thurau, C., Sifa, R., & Bauckhage, C. (2014). *A Comparison of Methods for Player Clustering via Behavioral Telemetry* (arXiv:1407.3950). arXiv. https://doi.org/10.48550/arXiv.1407.3950

Drachen, A., Thurau, C., Togelius, J., Yannakakis, G. N., & Bauckhage, C. (2013). Game Data Mining. In M. Seif El-Nasr, A. Drachen, & A. Canossa (Eds.), *Game Analytics* (pp. 205–253). Springer London. https://doi.org/10.1007/978-1-4471-4769-5_12

Dunnington, D. (2024). *Paleolimbot/rbbt* [R]. https://github.com/paleolimbot/rbbt (Original work published 2018)

Ermi, L., & Mäyrä, F. (2005, May 30). *Fundamental Components of the Gameplay Experience: Analysing Immersion*. DiGRA Conference. https://www.semanticscholar.org/paper/Fundamental-Components-of-the-Gameplay-Experience%3A-Ermi-M%C3%A4yr%C3%A4/881b17d6c9c6627af30a2eea9aabd332f1bc05ab

Findlen, P. (1989). The museum: Its classical etymology and renaissance genealogy. *Journal of the History of Collections*, *1*(1), 59–78. https://doi.org/10.1093/jhc/1.1.59

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game Learning Analytics: Learning Analytics for Serious Games. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_21-1

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gašević, D., & Merceron, A. (2022). *The Handbook of Learning Analytics* (C. Lang, G. Siemens, & A. F. Wise, Eds.; 2nd ed.). SOLAR. https://doi.org/10.18608/hla22

Gee, J. P. (2005). Learning by Design: Good Video Games as Learning Machines. *E-Learning and Digital Media*, *2*(1), 5–16. https://doi.org/10.2304/elea.2005.2.1.5

Genvo, S. (2013). Penser les phénomènes de ludicisation à partir de Jacques Henriot. *Sciences du jeu*, *1*, Article 1. https://doi.org/10.4000/sdj.251

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, *31*, 1–24. https://doi.org/10.18637/jss.v031.i07

Gray, G., & Bergner, Y. (2022). A Practitioner's Guide to Measurement in Learning Analytics: Decisions, Opportunities, and Challenges. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 20–28). SOLAR. https://doi.org/10.18608/hla22.002

Greene, J. A., & Yu, S. B. (2016). Educating Critical Thinkers: The Role of Epistemic Cognition. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 45–53. https://doi.org/10.1177/2372732215622223

Gutwill, J. P., & Allen, S. (2012). Deepening Students' Scientific Inquiry Skills During a Science Museum Field Trip. *Journal of the Learning Sciences*, *21*(1), 130–181. https://doi.org/10.1080/10508406.2011.555938

Hu, D., Chen, K., Leak, A. E., Young, N. T., Santangelo, B., Zwickl, B. M., & Martin, K. N. (2019). Characterizing mathematical problem solving in physics-related workplaces using epistemic games. *Physical Review Physics Education Research*, *15*(2), 020131. https://doi.org/10.1103/PhysRevPhysEducRes.15.020131

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63–85. https://doi.org/10.1007/BF02300500

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, *72*, 757–770. https://doi.org/10.1016/j.chb.2016.09.062

Kassambara, A. (2024). K-Medoids in R: Algorithm and Practical Examples. *Datanovia*. https://www.datanovia.com/en/lessons/k-medoids-in-r-algorithm-and-practical-examples/

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (1.0.7) [Computer software]. https://cran.r-project.org/web/packages/factoextra/index.html

Ke, F. (2019). Mathematical problem solving and learning in an architecture-themed epistemic game. *Educational Technology Research and Development*, *67*(5), 1085–1104. https://doi.org/10.1007/s11423-018-09643-2

Klopfer, E., Perry, J., Squire, K., Jan, M.-F., & Steinkuehler, C. (2005). Mystery at the Museum – A Collaborative Game for Museum Education. In *Computer Supported Collaborative Learning 2005*. Routledge.

Knight, S., Wise, A. F., & Chen, B. (2017). Time for Change: Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics*, *4*(3), Article 3. https://doi.org/10.18608/jla.2017.43.2

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, *3*(1), 33–67. https://doi.org/10.1007/s13194-012-0056-8

Landers, R. N., & Bauer, K. N. (2015). Quantitative methods and analyses for the study of players and their behaviour. In *Game Research Methods* (pp. 151–173). ETC Press.

Lang, C., Wise, A., Merceron, A., Gašević, D., & Siemens, G. (2022). What is learning analytics. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 8–18). SOLAR. https://doi.org/10.18608/hla22.001

Lankoski, P., & Bjork, S. (Eds.). (2015). *Game research methods: An overview*. ETC Press.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2023). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.6. [Computer software]. https://cran.r-project.org/web/packages/cluster/index.html

Martinez-Garza, M. M., & Clark, D. B. (2017). Investigating Epistemic Stances in Game Play with Data Mining: *International Journal of Gaming and Computer-Mediated Simulations*, *9*(3), 1–40. https://doi.org/10.4018/IJGCMS.2017070101

Mao, W., Cui, Y., Chiu, M. M., & Lei, H. (2022). Effects of Game-Based Learning on Students' Critical Thinking: A Meta-Analysis. *Journal of Educational Computing Research*, *59*(8), 1682–1708. https://doi.org/10.1177/07356331211007098

Meyer, D., & Buchta, C. (2022). *proxy: Distance and Similarity Measures* (0.4-27) [Computer software]. https://cran.r-project.org/web/packages/proxy/index.html

Molenaar, I., & Wise, A. F. (2022). Temporal Aspects of Learning Analytics—Grounding Analyses in Concepts of Time. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The Handbook of Learning Analytics* (2nd ed.). SOLAR. https://doi.org/10.18608/hla22.007

Morard, S., Sanchez, E., Oliveira, G., & Godinot, N. (2023). AL2049, a playful museum's visit to grasp the issues of complexity. *GSGS'23, 8th International Conference on Gamification & Serious Games*. https://drive.google.com/file/d/1p6E-H902UaFC9lcFz9gCjuLPsw5W-N7R/view

Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. *Computers in Human Behavior*, *24*(6), 2530–2540. https://doi.org/10.1016/j.chb.2008.03.012

Morin, E. (1999). *Les Sept savoirs nécessaires à l'éducation du futur—UNESCO Digital Library*. https://unesdoc.unesco.org/ark:/48223/pf0000117740_fre

Moshman, D. (2014). *Epistemic Cognition and Development: The Psychology of Justification and Truth*. Psychology Press.

National Research Council (2011). Honey, M., Hilton, M. L., & National Academies Press (Eds.). *Learning science through computer games and simulations*. National Academies Press.

Nelson, B. C., Bowman, C. D. D., Bowman, J. D., Pérez Cortés, L. E., Adkins, A., Escalante, E., Owen, B. L., Ha, J., & Su, M. (2020). Ask Dr. Discovery: The impact of a casual mobile game on visitor engagement with science museum content. *Educational Technology Research and Development*, *68*(1), 345–362. https://doi.org/10.1007/s11423-019-09696-x

Novoseltseva, D., Lelardeux, C. P., & Jessel, N. (2022). Examining Students' Behavior in a Digital Simulation Game for Nurse Training. *International Journal of Serious Games*, *9*(4), Article 4. https://doi.org/10.17083/ijsg.v9i4.543

Oliveira, G., Godinot, N., Sanchez, E., Bonnat, C., Morard, S., & Dall'Aglio, S. (2022). Game Design for a Museum Visit: Insights into the Co-design of AL2049, a Game About Food Systems. In K. Kiili, K. Antti, F. de Rosa, M. Dindar, M. Kickmeier-Rust, & F. Bellotti (Eds.), *Games and Learning Alliance* (pp. 22–31). Springer International Publishing. https://doi.org/10.1007/978-3-031-22124-8_3

Paliokas, I., & Sylaiou, S. (2016). The Use of Serious Games in Museum Visits and Exhibitions: A Systematic Mapping Study. *2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, 1–8. https://doi.org/10.1109/VS-GAMES.2016.7590371

Posit team (2024). RStudio: Integrated Development Environment for R. *Posit Software, PBC*. Boston, MA. http://www.posit.co/.

Reardon, E., Kumar, V., & Revelle, G. (2022). Game Learning Analytics. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 152–162). SOLAR. https://doi.org/10.18608/hla22.015

Reid, G. (2012). Motivation in video games: A literature review. *The Computer Games Journal*, *1*(2), 70–81. https://doi.org/10.1007/BF03395967

Rinker, T. (2024). *Trinker/pacman* [HTML]. https://github.com/trinker/pacman (Original work published 2012)

Riopel, M., Nenciovici, L., Potvin, P., Chastenay, P., Charland, P., Sarrasin, J. B., & Masson, S. (2019). Impact of serious games on science learning achievement compared with more conventional instruction: An overview and a meta-analysis. *Studies in Science Education*, *55*(2), 169–214. https://doi.org/10.1080/03057267.2019.1722420

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 69–97). Springer. https://doi.org/10.1007/978-3-642-85098-1_5

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered Design of Epistemic Games: Measurement Principles for Complex Learning Environments. *The Journal of Technology, Learning and Assessment*, *8*(4), Article 4. https://ejournals.bc.edu/index.php/jtla/article/view/1623

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Saas, A., Guitart, A., & Periáñez, Á. (2016). Discovering playing patterns: Time series clustering of free-to-play game data. *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, 1–8. https://doi.org/10.1109/CIG.2016.7860442

Sanchez, E. (2022). *Le paradoxe du marionnettiste—Jeu et apprentissage*. Octares. https://www.octares.com/accueil/291-le-paradoxe-du-marionnettiste.html

Sanchez, É., & Monod-Ansaldi, R. (2015). Recherche collaborative orientée par la conception. Un paradigme méthodologique pour prendre en compte la complexité des situations d'enseignement-apprentissage. *Éducation & didactique*, *9*(2), 73–94. https://doi.org/10.4000/educationdidactique.2288

Sanchez, E., & Pierroux, P. (2015). *Gamifying the Museum: A Case for Teaching for Games Based Learning*.

Sanchez, E., Young, S., & Jouneau-Sion, C. (2015). Classcraft: De la gamification à la ludicisation. In S. George, G. Molinari, C. Cherkaoui, & D. M. et L. Oubahssi (Eds.), *7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015)* (pp. 360–371). https://hal.science/hal-01405965

Seif El-Nasr, M., Drachen, A., & Canossa, A. (Eds.). (2013). *Game Analytics: Maximizing the Value of Player Data*. Springer London. https://doi.org/10.1007/978-1-4471-4769-5

Seif El-Nasr, M., Gagné, A., Moura, D., & Aghabeigi, B. (2013). Visual Analytics Tools – A Lens into Player's Temporal Progression and Behavior. In M. Seif El-Nasr, A. Drachen, & A. Canossa (Eds.), *Game Analytics: Maximizing the Value of Player Data* (pp. 435–470). Springer. https://doi.org/10.1007/978-1-4471-4769-5_19

Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, *50*, 116–123. https://doi.org/10.1016/j.csi.2016.09.014

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & Education*, *46*(3), 223–234. https://doi.org/10.1016/j.compedu.2005.11.003

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, *3*(3), Article 3. https://doi.org/10.18608/jla.2016.33.3

Shu, L., & Liu, M. (2019). Student Engagement in Game-Based Learning: A Literature Review from 2008 to 2018. *Journal of Educational Multimedia and Hypermedia*, *28*(2), 193–215.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction*, *55*(2), 503-524.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, *137*(3), 421–442. https://doi.org/10.1037/a0022777

Squire, K. (2007). *Open-ended video games: A model for developing learning for the interactive age* (pp. 167-198). MacArthur Foundation Digital Media and Learning Initiative.

Squire, K. (2021). *Making Games for Impact*. MIT Press.

Su, Y., Backlund, P., & Engström, H. (2021). Comprehensive review and classification of game analytics. *Service Oriented Computing and Applications*, *15*(2), 141–156. https://doi.org/10.1007/s11761-020-00303-z

Sweet, S. J., & Rupp, A. A. (2012). Using the ECD Framework to Support Evidentiary Reasoning in the Context of a Simulation Study for Detecting Learner Differences in Epistemic Games. *Journal of Educational Data Mining*, *4*(1), 183–223.

Tchounikine, P. (2019). Learners' agency and CSCL technologies: Towards an emancipatory perspective. *International Journal of Computer-Supported Collaborative Learning*, *14*(2), 237–250. https://doi.org/10.1007/s11412-019-09302-5

Tissenbaum, M., Kumar, V., & Berland, M. (2016). *Modeling Visitor Behavior in a Game-Based Engineering Museum Exhibit with Hidden Markov Models*. International Educational Data Mining Society. https://eric.ed.gov/?id=ED592698

Torrente, J., del Blanco, Á., Marchiori, E. J., Moreno-Ger, P., & Fernández-Manjón, B. (2010). : Introducing educational games in the learning process. *IEEE EDUCON 2010 Conference*, 1121–1126. https://doi.org/10.1109/EDUCON.2010.5493056

Toussaint, R., & Lavergne, M.-H. (2005). *Problèmes complexes flous en environnement et pensée réflexive d'élèves du secondaire*. https://doi.org/10.4267/2042/8855

Trestini, M. (2019). L'Environnement Numérique d'Apprentissage inscrit dans le paradigme de la modélisation systémique de la complexité. *Information, Organisation, Connaissances*, *2*(1). https://doi.org/10.21494/ISTE.OP.2019.0381

Ushey, K., Allaire, J. J., Wickham, H., Ritchie, G., & RStudio. (2024). *rstudioapi: Safely Access the RStudio API* (0.16.0) [Computer software]. https://cran.r-project.org/web/packages/rstudioapi/index.html

Vicarious. (2024, July 17). https://dictionary.cambridge.org/fr/dictionnaire/anglais/vicarious

Vienneau, R. (with Internet Archive). (2004). *Apprentissage et enseignement: Théories et pratiques*. Montréal : Gaëtan Morin. http://archive.org/details/apprentissageete0000vien

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*(3), 229–243. https://doi.org/10.2190/FLHV-K4WA-WPVQ-H0YM

Wang, M., & Nunes, M. B. (2019). Matching serious games with museum's educational roles: Smart education in practice. *Interactive Technology and Smart Education*, *16*(4), 319–342. https://doi.org/10.1108/ITSE-03-2019-0013

Wang, S.-H., & Wang, H.-Y. (2017). Using an epistemic game to facilitate students' problem-solving: The case of hospitality management. *Technology, Pedagogy and Education*, *26*(3), 283–302. https://doi.org/10.1080/1475939X.2016.1234408

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Brand, T. van den, Posit, & PBC. (2024). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (3.5.1) [Computer software]. https://cloud.r-project.org/web/packages/ggplot2/index.html

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Software, P., & PBC. (2023). *dplyr: A Grammar of Data Manipulation* (1.1.4) [Computer software]. https://cran.r-project.org/web/packages/dplyr/index.html

Wickham, H., Henry, L., & RStudio. (2023). *purrr: Functional Programming Tools* (1.0.2) [Computer software]. https://cran.r-project.org/web/packages/purrr/index.html

Wickham, H., Vaughan, D., Girlich, M., Ushey, K., Software, P., & PBC. (2024). *tidyr: Tidy Messy Data* (1.3.1) [Computer software]. https://cran.r-project.org/web/packages/tidyr/index.html

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

XantaCross. (2011). *Diffference in matching between Euclidean and Dynamic Time Warping* [Graphic]. https://commons.wikimedia.org/wiki/File:Euclidean_vs_DTW.jpg

9 APPENDIX

9.1 *A priori* analysis to choose criteria of food system complexity transcribed by AL2049

An a priori analysis was performed before the quantitative analyses of digital traces. It involved a researcher involved in the whole design process of AL2049. We decomposed complexity into seven complexity criteria for food production systems (for details, see 'Complexity' chapter). We sought to explain how AL2049 game designers tried to transcribe complexity in the game and to pick the objective indicators that can be found in the game's traces. The objective indicators are thus measuring how learners tackle complexity, which was the prime pedagogical objective. The objectives of this a priori analysis were, first, to define the seven criteria of complexity (adapted from Ladyman et al. (2013)) through the prism of AL2049 gameplay and, second, to identify indicators of complexity in players actions and ultimately in the traces. The seven complexity criteria were: 1) Nonlinearity, 2) Feedback, 3) Spontaneous order, 4) Robustness and lack of central control, 5) Emergence, 6) Hierarchical organization, 7) Numerosity. For each criterion, we first described how the chosen complexity criterion was represented in AL2049 (e.g., "how nonlinearity is part of the game?"). Second, we linked it to the expected digital trace linked to the situation (e.g., "the number of changes in the 'Energy' variables") and an hypothesis was written involving the trace (e.g., "in Phase I, all types of energies are increasing during phase I but not necessarily in phase II"). Among the seven criteria, we chose to restrict the quantitative analyses to a lesser number of criteria. Our choice was restricted to what the game mechanic of energy allocation can give as observable indicators. We thus chose to focus on one criterion: nonlinearity.

Nonlinearity is shown in AL2049 gameplay through its ability to show the players long term underestimated links which are only made visible in phase II. During phase I, the game is straightforward and rather linear in essence, involving one simple goal. At the end of this first phase, when the two other gauges are now made visible, the game becomes non-linear: one action can lead to multiple consequences in multiple factors, some of which are thought to be unpredictable to the player.