Let's Make Games Together: Explainability in Mixed-initiative Co-creative Game Design

Solange Margarido, Penousal Machado, Licínio Roque, Pedro Martins
University of Coimbra, CISUC, Department of Informatics Engineering
Coimbra, Portugal
{solange,machado,lir,pjmm}@dei.uc.pt

***Abstract*—There has been growing development of co-creative systems for game design, where both humans and computers work as colleagues, proactively contributing with creative input. However, the collaborative process is still not as seamless as in human-human co-creativity. A key element still underdeveloped in these approaches is the communication between the human and the machine, which can be facilitated by providing the computational agent with explanatory capabilities. Based on principles of explainability for co-creative systems from previous literature, we propose a framework of explainability specifically applied to mixed-initiative scenarios in game design. We illustrate the applications of the framework by suggesting possible solutions adapted to different use cases of existing approaches and, additionally, of our own proposed approach.**

***Index Terms*—explainability, mixed-initiative, game design, computational creativity, co-creativity**

I. Introduction

Mixed-initiative has been a topic of increasing relevance in game design. To enhance creative processes, researchers have a growing concern in evolving the interaction between humans and computational frameworks towards the manifestation of initiative and proactivity from both agents [1]. This kind of interaction is of particular interest in the game design field, as human-computer co-creativity can be very beneficial for the design process. Mixed-initiative approaches applied to game design have the potential of addressing the issues of lack of computer initiative in support tools and computational dominance in procedural content generation, by finding common ground between these systems.

Several mixed-initiative approaches to game design have been introduced over the last few years. These approaches address several types of game content at different complexities. Some examples include level design [1]–[8], the creation of visual game assets [9], the design of game mechanics [10], [11], narrative development [12], [13] and even the creation of of fully working games [14], [15]. Furthermore, there are a variety of techniques being used depending on the goals and intended functionality. Procedural content generation, recommendation systems and reinforcement learning are some of the techniques employed in the current state of the art. Despite the diversity of recent works, these approaches still fail to promote seamless collaboration between human and computer. Several factors surround this matter. In most cases, human and computer are still performing separated steps at a time. Therefore, not learning much from each other. Closely related to this, there is a lack of communication between the agents involved. If the human and the computer cannot understand the motivation behind each other's interventions and discuss them, it will cause friction in their collaboration. As such, explainability is an important feature for co-creative systems. Humans and computers need to exchange and explain their ideas so they can in fact work as colleagues. As human-computer co-creativity takes on an increasingly significant role in the scope of game design, we believe it to be necessary to address the issue of lack of explainability in the light of mixed-initiative approaches to game design.

In this paper, we draw from principles of Explainable Computational Creativity [16] and Explainable AI for Designers [17] and propose a framework of explainability in the context of mixed-initiative co-creative game design. We focus on identifying gaps in the communication and explanatory capabilities of some existing systems and suggest possible features for improvement with basis on the ideas presented in our framework.

Section II reviews the current state of mixed-initiative co-creativity and what are the prospects for future approaches. The works serving as guidelines for our discussion of explainability in the game design domain are also overviewed. Section III describes principles of explainability in the light of mixed-initiative co-creative game design. In Section IV, we present three use cases, two of existing approach and one we propose, from which we illustrate possible applications of the framework's principles. We finish by drawing some conclusions and expectations for future work.

II. Related Work

Davis [18] introduced the term *human-computer co-creativity* to describe a scenario in which the computer is presented as a collaborator in the creative process. Davis emphasizes that the computer is not merely led by predefined scripts, but instead adapts its interventions according to the user's input. As in human-human co-creativity, the interactions between the agents allow them to be mutually influenced by each other's contributions during the collaborative process. The terms *mixed-initiative*, *mixed-initiative co-creativity* and *human-computer co-creativity* can be used interchangeably, although authors may have different considerations about some particularities (e.g. if the human and the computer must be equal partners or can have different levels of responsibility in the creative process). Nevertheless, when considering an intrinsically creative context, as is the case of game design, the interaction between humans and machines must be seen as an extent of co-creativity.

Until recently, literature on human-computer co-creativity tendentially took on a human-centered perspective, focused on supporting human creativity. Assuming a more computer-centered perspective, Kantosalo and Toivonen [19] propose a scenario that allows the human and the machine to be more equal partners. They introduce alternating co-creativity, in which the human and the computational agent take turns in creating a creative artifact. Usually starting from scratch, the goal is to achieve a final result through collaboration that satisfies both partners. This is a chance for the human agent to get new inspirations, whereas the computer gets the resources for transformational creativity (i.e. the system changes its own conceptual space to admit new concepts [20]). This scenario requires high capabilities from the computational agent, as it must be able to identify, generate and evaluate concepts. Additionally, if a computational agent can perform transformational creativity to handle conflicts of disagreement and adapt to the new situation, then it can participate in *symmetric alternating* co-creativity.

It is our expectation that future human-computer co-creativity converges towards an approach close to *symmetric alternating* co-creativity. However, to achieve seamless collaboration, the alternating should not be too rigid. That is, the turns taken by each intervener (human and computer) to construct and modify the creative artifact should not be too prolonged and independent of each other. Ideally, the system's response to the changes made should be almost instantaneous and provide explainability. After all, in human-human co-creativity, creative collaborators are not limited to alternating tasks. Instead, they are involved in a discussion facilitated by their communication capabilities. This characteristic allows ideas to emerge more fluidly and promotes a greater understanding of the motivations and intentions among the parties involved. To illustrate this approach, let us take the Sentient Sketchbook [21] as an example. As the user sketches a map, the system dynamically provides suggestions. The framework does not wait for the user to complete an entire design and then submit it, to only intervene with optimizations or alternatives later. This is the factor that makes the Sentient Sketchbook such an interesting co-creative system. Still, there is something missing. We believe this dynamic alternating can be fostered by granting more communicative and explanatory capabilities to the computational agent, allowing it to justify its contributions and be understood.

A. Overview of Explainability in Mixed-initiative Co-creativity

One main aspect that still distinguishes traditional co-creativity from mixed-initiative co-creativity is communication. As such, in order to achieve seamless collaboration between human and machine, we must facilitate their underlying communication. After all, communication is key in any kind of collaboration. Participants involved in a collaborative process are expected to explain why they made a certain decision, what are their intentions, what motivates their idea, compare their ideas to previous works, etc.

Llano et al. [16] introduced the concept of Explainable Computational Creativity (XCC) as a way to enable a two-way communication between computational creative systems and their human users, in order to promote the discussion of the ideas and decisions to improve the collaborative and creative process. For a better understanding of how XCC can be applied, the authors provide an example of a scenario where an Advertising Executive is working with an XCC system in order to design an advert for a toothpaste. Four design principles for XCC systems are presented:

- *Mental models*: representations of crucial elements of the creative environment that help conceptualize and understand how things work (e.g. the system can adjust its mental model with the information that the user does not like ideas with connections that are not obvious).
- *Long-term memory*: the capacity of storing and accessing details of past experiences (e.g. draw back from a previous experience of adding a tagline to support an advert concept).
- *Argumentation*: the process of reasoning about and supporting creative contributions (e.g. the system explaining that its idea is somewhat repulsive and providing example of how other repulsive adverts have been effective).
- *Exposing the creative process*: exposing the details of the processes and decisions within the underlying operation of a system (e.g. through an interface with a visual graph representation of the system's knowledge base).

The interactions that are conveyed with basis on these principles depend on the domain, the stakeholders and the stage of the creative process. Additionally, the communication can be carried out in its broadest sense, not only through the linguistic form.

Another proposal is presented by Zhu et al. [17]. They define a subfield of Explainable Artificial Intelligence (XAI) specifically focused on game designers: Explainable AI for Designers (XAID). This approach is centered on the specific needs and tasks of game designers in order to facilitate their co-creation with AI systems. Zhu et al. mapped the XAID space along three axes, each encompassing its own spectrum to adapt the approach to the technical properties of the system and the needs of the human designer. The *spectrum of explainability* ranges from explanations that provide an understanding of the underlying operation of the AI system, such as what was the sequence of actions and why the system made certain decisions, to observations that offer insight into the correlation between given inputs and the resulted outputs. The *spectrum of initiative* relates to the level of the system's interventions and, consequentially, the type and complexity of the explanations it may provide. This spectrum can range from a low level of initiative, as in the case of the system taking on a more supportive role, to a high level of initiative when it is expected to act more as a colleague. Lastly, the *spectrum* *of domain overlap* is associated with how much the tasks performed by the human and the computer overlap, which can be seen as the needed degree of co-creativity. This spectrum ranges from a scenario of a human and an AI system working on the same task to a scenario in which each agent is working on a different task that only slightly affects the other one.

Both of these works serve as guidelines for a framework of explainability for co-creative game design.

III. The Framework

Along the three axes of the XAID space, in this work we focus the *spectrum of initiative* on the highest level of autonomy and initiative, where the system takes the role of a colleague in the co-creative process. Regarding the *spectrum of domain overlap* we consider that in the vast majority of cases there is a large overlap between the tasks carried out by the agents involved. Only in this way can there be a true collaboration that allows an exchange of ideas and the agents to influence each other on the task at hand. We contemplate explanations and observations of the *spectrum of explainability*, as both are essential and must be applied depending on several factors, such as the profile of the co-creators and the game design domain.

Considering the XCC design principles, we provide an overview of how these can be applied in the domain of game design.

A. Mental Models

As representations of how things work in a co-creative environment, mental models encompass different types of elements. These elements can be related to team aspects, domain aspects and interpersonal aspects, as described by Llano et al. [16]. The type of elements that can be represented by the co-creative systems involved in a game design process is quite transversal to other domains of co-creativity. Still, we can think of specific elements that should be represented in the mental model of the system when working on certain types of game content. For instance, in level design, it is essential for the mental model to have representations of the game rules so that the computational agent's suggestions for the level structure are valid. In the same context, it is necessary for the mental model to include information about the global and particular goals of each level, as well as the expectations and preferences of the human co-creator: is it intended for it to be a starting level and therefore easier or even in a tutorial format? Or a more advanced level is intended, in which the complexity and difficulty should be greater? Do we want the player to be able to easily predict the characteristics of the level and how to overcome it? Or do we want to add some unexpected and surprising elements for the player? Do we want the player to learn some new mechanics, and do we need to think about how to introduce it? If we consider another scenario in which the design task itself is to create the rules and objectives of the game, other elements are essential to register in the mental model. What do we intend to convey with this game? Do we want the game to be more strategy, skill, or exploration-based? Is it targeted at casual gamers or hardcore gamers? Is there any message or moral lesson we want to convey?

Such information may be previously defined before starting the co-creative process or can be decided in discussion with the other co-creator. Both the human and the computational agent can present their own motivations and expectations and bring them into the conversation throughout the collaboration. Here the system should update its mental model with the new information it acquires over time to be able to adapt and improve its future interventions. Mental models allow the system to learn and understand the characteristics of the human designer, their interactions, the environment, and the creative domain, which provides it with the resources to explain and defend its contributions.

B. Long-term Memory

As humans, we are beings in constant learning and evolution. We keep acquiring more and more knowledge and capabilities through experience. Such is possible because we are endowed with long-term memory that allows us to draw on lessons learned from previous experiences. This ability is valuable for the creative process as it enables us to have richer interventions, justify our contributions, avoid mistakes or repetitive content, among other purposes. For a computational agent to collaborate with a human agent as an equal, it must also be equipped with this competence.

In a game design process, the system's use of memory and the type of information that is stored does not differ much from other types of co-creative processes. In general, it is important to save current and previous experiences of co-creation to take advantage of this knowledge in future collaborations. That includes storing the choices and reactions of the human designer so the system can converge to (or challenge) their preferences. The computer must also store new concepts that it learns in order to become increasingly capable. For instance, in a character design approach, the system can store concepts that the human co-creator has related (e.g., "knight" is related to "courageous") and that the system has not yet made this correlation. In addition, artifacts that have been successfully created (depending on the type of game content being tackled, it can be maps, visual assets, game mechanics, among others) should be stored to know what is the expectation for a final artifact, but also to avoid creating future content that is too similar. Remembering these types of information improves the creative capacity of the computational agent.

C. Argumentation

When it comes to explainability in a co-creative system, the main aspect that comes to mind is that the system is, in fact, capable of explaining. In fruitful collaborations, the agents discuss each others contributions, providing explanation of the process and motivations behind their ideas. It is through argumentation that participants can make themselves understood and play an active role in the co-creative process, increasing the value of their interventions.

The type of argument to be used will always depend on several factors, including the domain of game design in which one is working, the stage of the creative process, the intentions of the explanation (e.g., clarifying, convincing, contesting), and the profile of the human co-creator. Let us think about the kind of discussion that can arise for different types of game content. When suggesting a certain design for a game level, the computational agent must explain in what way that level is interesting and how it reaches the expected degree of difficulty. When creating a character, the human designer may want to understand why traits that have been suggested by the system are appropriate for the type of character being created. Collaborating on narrative development, we can imagine several questions that agents should be able to clarify: How does this sequence of events contribute to enriching the game narrative? What influence will it have on the final outcome? What kind of emotion do they intend to convey to the player? Fear? Sadness? Empathy? Surprise? Excitement? Furthermore, in several scenarios, such as in the creation of game dynamics or rules, the co-creative system must justify how a particular suggestion meets the vision of the game. This explanation should be done considering various aspects, namely the game genre and the concepts and values intended to convey through the game.

Not only to clarify or support their contributions, but a computational agent can also use argumentation to challenge or champion an idea of a human agent. As an example, the system may contest an idea it deems unoriginal: "I think your idea of a platform game with a plumber as the main character will not be well received, as it is the same concept as a widely known game. Unless we add some differentiation point to it, assuming it as a spin-off or a parody of the original game, I would recommend that we think of another idea."

There are several ways and purposes for a co-creative system to use argumentation. However, it is important to remember that explainability must be carried out in two directions. That is, the computer must make itself understood by the human designer, but it must also understand them. As such, ideally, the system must have some way of being capable of receiving arguments from the human designer and registering them in a structure that is comprehensible.

Naturally, the principle of argumentation raises questions about how communication can be mediated to enable this discussion and exchange of ideas between the parties, a topic that we will address ahead.

D. Exposing the Creative Process

Sometimes, the explanations provided by the system may not be enough for the co-creator to understand its ideas and the creative process behind them. To fill these gaps, we can implement tactics to make its internal operation, interventions, and way of communicating more interpretable for a human co-creator. Such can be done by schematizing the underlying processes simply and clearly, highlighting the steps, metrics, inspirations, logical associations, and other relevant aspects. One possibility is the use of concept maps, in which connections are made to illustrate the relationships between concepts. In a collaborative process for character design where one seeks to define physical and personal traits, the option of visualizing this type of diagram would help to understand the logic behind the suggestions of the computational agent. For example, the visual schema can display that the concept "bully" relates to "aggressive" which consequently relates to "angry facial expression". In this way, the correlation between the initial concept of "bully" and the suggestion of the physical attribute of "angry facial expression" becomes more evident.

The exposure of the creative process can also be done dynamically during the collaboration, without the human user directly requesting it. It can be similar to one of the features of the Sentient Sketchbook's level design approach, where the system shows some metrics like "resource safety", "safe area", and "exploration", exposing what parameters its suggestions can improve. The exposure of metrics, inspirations, motivations, and reasoning behind the contributions of the computational agent can, in fact, be done in several ways other than schematically. During a collaborative process of scenery creation for a game, where visual aesthetics will be the most significant point, the system can show what its references are, whether photographs of real landscapes or art style from other existing games.

It would also be interesting for the computer to convey how it feels about an idea. Following on the previous example of a context of scenery creation, the system may have some method of expressing its emotions towards the artifact being created. The landscape can make it feel amazed, cozy or, on the other hand, embittered, scared. The work of McCormack et al. [22] is an example of how a system can expose its emotions during a co-creative process, an approach that also allows to humanize the computer and make it a more reliable participant and equated with the human co-creator.

E. Forms of Communication

Both the works on XCC and on XAID mention the possibility for different forms of communication. Zhu et al. [17] state that communication on mixed-initiative co-creative systems in game design is mostly done through visual feedback (e.g. by showing a list of properties that make a certain suggestion of an artifact desirable). In contrast, there is a lack of exploration in the use of natural language in these approaches. The use of natural language can be very useful not only to argue the value of an idea or artifact, but also to explain the creative process behind its generation. Llano et al. [16] emphasize that the type of communication must be adapted to the creative domain in which it is inserted.

In the realm of game design, the choice of a communicative medium can also be very variable and depends on the co-creative context and the type and complexity of explainability that is intended. In many cases, visual communication will be the most suitable, whether through diagrams, illustrations, animations, text. We are talking about contexts such as level design, game environment design (e.g. streets, pathways, vegetation), creature/character design, etc. In others contexts,

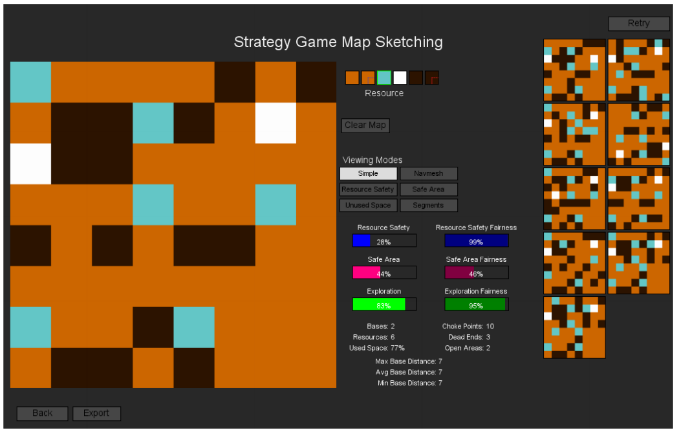

Fig. 1. Sentient Sketchbook. Left - the sketch editor; right - the automatically generated map suggestions; middle - the tile palette, the map display menu and an overview of the map's fitness dimensions and metrics [1].

as in the case of designing sound for a game, it makes sense that explainability can also take advantage of sound communication. For example, a system that collaborates in the creation of ambient sound can justify its interventions through sound communication, taking the opportunity to reproduce sounds that have been used as reference. In a narrative design context, it would be appropriate for the communication to also resort to natural language processing to promote understanding between the co-creative agents.

IV. Use Cases

In this section we present examples of how the framework principles can be applied for three use cases. We discuss how both the human and the computer can exert explainability during the co-creative process, focusing on the forms of communication that may be most appropriate for each case.

A. Sentient Sketchbook

The Sentient Sketchbook [1], [21] is a well-known mixed-initiative approach to the design of game levels. This framework uses map sketches as low-fidelity representations of levels, allowing users to create and edit sketches (Fig. 1). As the designer interacts with the tool, while sketching an idea, the interface gives real-time suggestions that are evolved permutations of the designer's current sketch. The suggested alternative map designs are achieved through a genetic search algorithm that either evolves to optimize defined fitness functions or to create visually diverse maps. At any time, the user can select a suggestion to replace their current sketch. When the collaborative sketching process is over, the computer can convert the map into a fully playable 2D or 3D level.

Through an evaluation of the Sentient Sketchbook, Yannakakis et al. [1] identified that, in some design sessions, human participants did not use any of the computational agent's suggestions. One of the reasons given by participants for the little usefulness of the suggestions was that these were not aligned with the aspects that the user was focusing on at the moment. Although the system has the game rules and what makes a playable and balanced level in its mental model, it has no record of the preferences and motivations of the human co-creator, nor how he interacts and what options he takes throughout the process. Updating its mental model with this information could tackle this issue so the system can make contributions more conscious and tailored to the creative partner. Another point to consider is that Sentient Sketchbook targets a specific style of strategy games but could more easily adapt to other game genres if the computer was able to adjust its mental model to accept new rules or different metrics. During the co-creative process, it may happen that the human agent manifests a different goal than what is defined by the system or suggests new features (such as other types of tiles) that can be added. By being able to adjust elements that were already predefined in its mental model, the system acquires the capability to perform transformational creativity so that eventually they can reach a result that is pleasant for both creators while also increasing the creative possibilities and potential for collaboration with the co-creative system.

Most of the users involved in the Sentient Sketchbook evaluation tended towards symmetry to ensure that the level was balanced between opponents. The computational agent's interventions, on the other hand, challenge this belief by suggesting maps that manage to provide game balance despite being visibly asymmetrical. It turns out that the preconceptions of the human designer are difficult to break, and in most cases these suggestions are rejected. It is in these types of situations that the argumentative power of the system could have a major impact on the course of the creative process. The system already makes a small effort to possibly convince its creative partner by showing the values of several parameters of each suggestion compared to the current sketch at hand. Here the user can verify that, effectively, asymmetrical levels can maintain or even improve the playability and balance properties. Even so, this is not enough for the human creator to be willing to give in. If the system could further support its ideas, it might succeed in breaking the habits of the human creator and moving towards more novel and unexpected solutions.

While the user analyzes a suggestion, one of the possible strategies would be for the computer to show, in addition to the metrics, a short description of its creative process and the characteristics of that level. An example description could be: "In this level, I intended to provide more difficulty for players by reducing the number of resources available and the ease of access to them. I want to try a scenario where players have to make riskier decisions. The map structure allows the level of risk to be balanced between opponents.". Another option of a more visual nature would be to show a schematic or animation of the suggested level's gameplay, illustrating possible strategies that players can take and possible outcomes.

It is noteworthy that, from the little that the tool tries to expose its creative process, it focuses on how the suggestions can optimize the different properties of the level. However, part of the computational agent's suggestions is focused only on maximizing novelty without necessarily offering optimization.

In these cases, there is no explanation of the value of these suggestions, as it does not lie in improving the level quality but rather in serving as inspiration to create more visually diverse levels. As a result, users rarely choose to select suggestions solely focused on visual novelty. Therefore, it would benefit the collaboration for the system to justify how each suggestion is interesting. Whether through a textual or more visual approach, it is important for the system to mention that a suggestion can, for instance, offer a different gameplay from other previously created levels and force players to consider new strategies, explaining how so.

To create some disruption in the creative process, the computational agent can also intervene more actively at certain times. Its intervention would be timely if the user constantly maintained the same behavior pattern throughout the creative process (e.g., always drawing symmetrical levels or always using the same number of resources). After a determined time of following the same behavior pattern, the computational partner could indicate (e.g., through a text pop-up) that the created levels are becoming repetitive, and it can lead to the players losing interest. Simultaneously, it should highlight suggestions that can help stray further from the ingrained habit the human creator has been following.

B. *Germinate*

Kreminski et al. [14] presented "a mixed-initiative casual creator for rhetorical games". This framework extends Gemini — an abstract game generator — but with a more approachable user interface for casual users. This approach is focused on starting from high-level rhetorical goals and then working on details to meet those goals. The framework's interface allows the user to specify a number of properties that they would like to be present in the generated games. These properties can be defined through four types of cards: entity cards — game objects with a name and graphical representation (an emoji) that may respond to interactions with the player or other entities; resource cards — describing quantitative values that will relate to the game's goals; relationship cards — mapping of relationships between entities and/or resources; trigger cards — describing the outcome of specified trigger events. Gemini generates a batch of games based on these properties which are then shown in the Germinate's interface. The generated games present a set of created rules, goals, behaviors, and interactions that match the expressed intent of the human creator. An example of use could be a user trying to make a game about depression: they could define a "Friend" entity and a "Depression" and "Confidence" resources (plus the relationships and triggers they want to showcase). This example is shown in Figure 2. The user is free to try the various generated games and see if there are any results that they find interesting, then be able to make modifications to the properties to reflect a better understanding of their intent.

The system does a proper job of trying to understand the co-creator's intentions and motivations, working at a rhetorical level where the user can express what they want to represent through relationships between entities and resources. How-

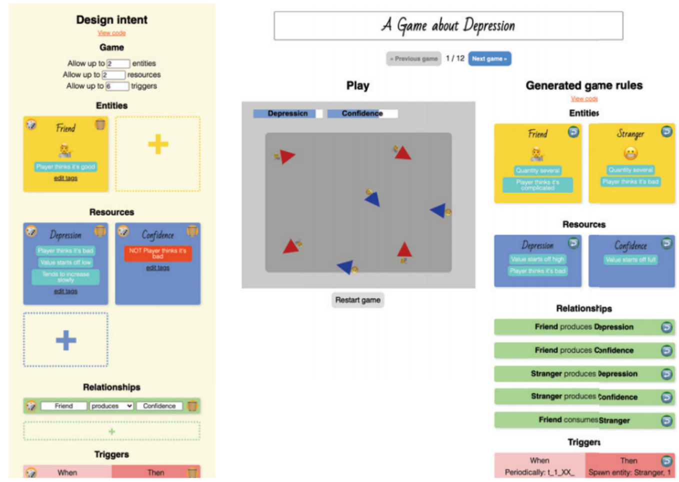

Fig. 2. Germinate. Left - properties specified by the user; center - a playable version of one of the generated games (a single game is shown at a time); right - the actual properties of the generated game [14].

ever, this scheme may not be sufficient for users to express themselves clearly. For instance, the human agent might want to convey the following message to the player: "Isolation and toxic relationships can exacerbate depression. If a person goes through therapy and is able to maintain only healthy relationships, they can gain the confidence to get out of that state of mind.". The user may be able to express this idea in several steps, defining, for example, the resource "Depression" as bad for the player and the entity "Friend" as good. Still, some concepts may be harder to express directly. One such case is the concept of isolation being bad and worsening depression. A possible interpretation could be setting the "Depression" resource to gradually increase over time and decrease it in the event of a collision with a "Friend" entity. However, formulating this logic may not be intuitive for the human creator. Another aspect is that the user cannot clearly indicate the objective he would like to set for the game, like so: "The goal is to completely eliminate depression and maximize confidence.". Thus, there is still room for improvement regarding the possibilities of representing the co-creator's intentions in the mental model of the system.

Regarding argumentation, the system tries to explain the generated games by displaying the several mechanics and interactions of a game in card format, in an identical scheme to the cards created by the user when defining their intent. This way, as the authors mention, the human creator can read the cards and better understand how the rules are structured, what is the logic behind the game mechanics and what message it is trying to convey. Even so, the rules, gameplay and goal of a generated game may not be clear to the user, which was verified in the results of the framework evaluation. To solve this issue, the computer should provide more direct explanations about the generated games. Specifically, it would be appropriate for it to have a description of the rhetorical meaning of the game. For instance, if the system generated a game with a "Book" entity that produces "Confidence", it should explain the intrinsic meaning of this relationship. Additionally, it should show a straightforward explanation of the rules and mechanics of the game and what objective the player has to achieve in order to win. This explanation could even be fostered by taking on a tutorial format, ideally accompanied by a visual demonstration of the gameplay.

It should be noted that the computational agent seeks to work with the intentions of the human agent but does not question them. Sometimes, users themselves may be unsure of what they want to convey with the game and create relationship and trigger cards that are inconsistent and may even be contradictory. There are cases where the system is not even capable of generating games from the user's intentions, because some incongruity in the requirements does not allow it. However, the system only indicates that it was unable to generate any games without mentioning the source of the problem. Here the system should intervene to point out contradictions of its co-creator. It can also encourage the user to rethink the meaning of what they have expressed in case it considers that the user's intentions are confused or uncreative, possibly suggesting other alternatives.

Another possibility to improve the explainability of the system is suggested by Kreminski et al. Since Germinate is built on top of Gemini, they consider the additional functionality of exposing the derivation trees of this framework. Gemini is internally composed of specific arguments that relate the various characteristics of the game to certain rhetorical meanings. Although exposing these derivations trees can allow the human creator to have a more detailed view of the system's internal operation and the logic behind its creative process, the computational agent should care to display these graphical representations in an understandable way. Especially as this system is designed for casual creators, we must take into account that their technical knowledge is often limited. A more simplified representation of these trees should be thought of in order for the message to get through without being too confusing and overwhelming for the user.

C. An Envisioned Approach to Character Design

For future work, we aim to develop a mixed-initiative approach to character design. Having analyzed the current state-of-the-art, we realized that most approaches target level design while other types of game content are overlooked. We specifically chose character design because there is still no mixed-initiative approach to it as far as we know. Furthermore, it has enough depth to allow a diversity of aspects to be tackled, such as visuals, animation, believability, attachment, or even narrative. The goal is for our framework to allow the human creator to begin the creative process with some initial idea of a character and, together with the computational partner, build its several characteristics from there. These characteristics range from its personal traits to the entire visual aspect, which the human and the computer work on until reaching a final sprite of the character.

The computational agent contributes by dynamically providing suggestions, especially at the visual level, which can be compared with the DesignScape approach [23] focused on layout design. However, in our case, the creative process starts with a vague idea and a blank canvas, progressing with the system suggesting more novel content that can be added until constructing a complete character. To give an example, the user can start by indicating that they want to create an "emo" character. From there, the system starts suggesting several elements that can be related, such as "dark clothes", "eyeliner", "teen", "sensitivity", and "depression". The user can select elements they agree with and add others that may not have been suggested by the computer. At the same time, the system also suggests visual elements to compose the character's appearance, such as hairstyles, shirts, pants, shoes, accessories, among others. The human designer can select the suggestions they like the most, and the system will adapt its following suggestions according to its partner's choices. The user also has the option to save suggestions they like but do not want to apply directly to the character, which further enables the system to understand their preferences. When both the human agent and the computational agent are satisfied, they can export the 2D sprite with the character's final look and save its list of traits.

Throughout Section III, we have already described several ways in which explainability can improve human-computer collaboration in the context of character design. Updating the mental model of the system according to the user's options and their interactions throughout the co-creative process is a way for the system to better understand the characteristics and preferences of its partner. We also discussed how storing past co-creation experiences helps the system improve its future interventions. This setup in which the user can select, save and add suggestions provides the necessary resources for the computer to register this information. Additionally, it is important to recap that all new suggestions that the human partner may add are concepts that the system can save to expand its knowledge base. In the example above of creating an "emo" character, the user can suggest the characteristic of "pale skin", describing a new relationship between concepts that the system can learn.

During the co-creative process, it may happen that the human designer does not understand why the computer has suggested certain traits for the type of character they are creating. If there is no explainability, it is expected that the user will simply ignore suggestions they do not understand. In these situations, it is essential for the computational agent to justify why a particular trait is suitable. One way of doing this is by showing a textual description of the suggestion when the user is hovering it. For instance, it can explain the "sensitivity" characteristic as follows: "The emo style originates from a musical genre that emphasizes emotional expression. It is more associated with the age group of teenagers, who tend to be more emotive.". When it comes to a character's physical attribute, such as "dark clothes", the system can also support its arguments with visual references such as photographs of people with emo style or images of characters from games, series, or movies. These explanations can help the human partner become more receptive to suggestions that initially do not seem fitting. The system can also expose its creative process through visual schemes such as concept maps. When analyzing this diagram, the user can see that “emo” is directly related to the concept of “emotional”, which in turn is related to the concept of “sensitivity”. In this way, the user can better understand the thought logic of its computational partner.

V. Conclusions

In this work we presented a framework of explainability applied to mixed-initiative co-creativity in game design. We drew from XAID [17] and XCC [16] principles and illustrated how these can be applied in various game design approaches, highlighting the different forms of communication that suit each context. We argue that current mixed-initiative approaches still fail to promote seamless collaboration between humans and computers, but equipping agents with explanatory capabilities is a step towards approaches that promote greater understanding among co-creators. This is particularly relevant in the field of game design, which is the subject of extensive research on mixed-initiative approaches to optimize processes and foster creativity.

We expect that our effort to illustrate the application of the principles through use cases can serve as guidelines for future work of mixed-initiative co-creative game design, where participants can understand the motivations and intentions of each other and successfully justify their contributions. We know that the implementation of explainability is challenging. Nevertheless, the gradual introduction of strategies such as simple textual descriptions that reveal part of the creative process, the presentation of visual references, or the use of concept maps can already have a high impact on the quality of collaboration.

Acknowledgment

This work is funded by the FCT – Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit – UIDB/00326/2020 or project code UIDP/00326/2020. The first author is funded by Foundation for Science and Technology (FCT), Portugal, under the grant 2021.07047.BD.

References

[1] G. N. Yannakakis, A. Liapis, and C. Alexopoulos, “Mixed-initiative co-creativity,” in *9th International Conference on the Foundations of Digital Games*. Foundations of Digital Games, 2014.

[2] G. Smith, J. Whitehead, and M. Mateas, “Tanagra: Reactive Planning and Constraint Solving for Mixed-Initiative Level Design,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 3, no. 3, pp. 201–215, sep 2011.

[3] A. Baldwin, S. Dahlskog, J. M. Font, and J. Holmberg, “Mixed-initiative procedural generation of dungeons using game design patterns,” in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, aug 2017, pp. 25–32.

[4] S. P. Walton, A. A. M. Rahat, and J. Stovold, “Mixed-initiative procedural content generation using level design patterns and interactive evolutionary optimisation,” *arXiv preprint arXiv:2005.07478*, may 2020.

[5] D. Karavolos, A. Bouwer, and R. Bidarra, “Mixed-initiative design of game levels: integrating mission and space into level generation,” in *10th International Conference on the Foundations of Digital Games*. Foundations of Digital Games, 2015.

[6] M. Charity, A. Khalifa, and J. Togelius, “Baba is Y’all: Collaborative Mixed-Initiative Level Design,” in *2020 IEEE Conference on Games (CoG)*. IEEE, aug 2020, pp. 542–549. [Online]. Available: https://ieeexplore.ieee.org/document/9231807/

[7] E. Butler, A. M. Smith, Y.-E. Liu, and Z. Popovic, “A mixed-initiative tool for designing level progressions in games,” in *Proceedings of the 26th annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM, oct 2013.

[8] O. Delarosa, H. Dong, M. Ruan, A. Khalifa, and J. Togelius, “Mixed-Initiative Level Design with RL Brush,” in *Artificial Intelligence in Music, Sound, Art and Design. EvoMUSART 2021*. Springer International Publishing, 2021, pp. 412–426. [Online]. Available: http://link.springer.com/10.1007/978-3-030-72914-1_27

[9] A. Liapis, G. N. Yannakakis, and J. Togelius, “Adapting Models of Visual Aesthetics for Personalized Content Creation,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 3, pp. 213–228, sep 2012.

[10] T. Machado, D. Gopstein, A. Nealen, and J. Togelius, “Pitako - Recommending Game Design Elements in Cicero,” in *2019 IEEE Conference on Games (CoG)*. IEEE, aug 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8848081/

[11] T. Machado, D. Gopstein, A. Wang, O. Nov, A. Nealen, and J. Togelius, “Evaluation of a recommender system for assisting novice game designers,” in *Proceedings of the 15th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2019*, 2019.

[12] M. Kreminski, M. Dickinson, M. Mateas, and N. Wardrip-Fruin, “Why Are We Like This?: The AI Architecture of a Co-Creative Storytelling Game,” in *International Conference on the Foundations of Digital Games*. New York, NY, USA: ACM, sep 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3402942.3402953

[13] J. M. Thomas and R. M. Young, “Author in the Loop: Using Mixed-Initiative Planning to Improve Interactive Narrative,” in *Workshop on AI Planning for Computer Games and Synthetic Characters*, 2006.

[14] M. Kreminski, M. Dickinson, J. C. Osborn, A. Summerville, M. Mateas, and N. Wardrip-Fruin, “Germinate: A Mixed-Initiative Casual Creator for Rhetorical Games,” in *Proceedings of the Sixteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2020. [Online]. Available: https://ojs.aaai.org/index.php/AIIDE/article/view/7417

[15] M. J. Nelson, S. Colton, E. J. Powley, S. E. Gaudl, P. Ivey, R. Saunders, B. P. Ferrer, and M. Cook, “Mixed-initiative approaches to on-device mobile game design,” in *CHI’17 Workshop on Mixed-Initiative Creative Interfaces*, 2017. [Online]. Available: http://repository.falmouth.ac.uk/id/eprint/2568

[16] M. T. Llano, M. D’Inverno, M. Yee-King, J. McCormack, A. Ilsar, A. Pease, and S. Colton, “Explainable Computational Creativity,” in *11th International Conference on Computational Creativity*, 2020, pp. 334–341.

[17] J. Zhu, A. Liapis, S. Risi, R. Bidarra, and G. M. Youngblood, “Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation,” in *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, aug 2018, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/8490433/

[18] N. Davis, “Human-computer co-creativity: blending human and computational creativity,” in *Proceedings of the Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2013.

[19] A. Kantosalo and H. Toivonen, “Modes for creative human-computer collaboration: alternating and task-divided co-creativity,” in *Proceedings of the Seventh International Conference on Computational Creativity*, Paris, 2016.

[20] M. A. Boden, *The Creative Mind*. Routledge, feb 2004. [Online]. Available: https://www.taylorfrancis.com/books/9781134379583

[21] A. Liapis, G. N. Yannakakis, and J. Togelius, “Sentient sketchbook : computer-assisted game level authoring,” in *8th International Conference on the Foundations of Digital Games*. Chania: Foundations of Digital Games, 2013.

[22] J. McCormack, T. Gifford, P. Hutchings, M. T. Llano Rodriguez, M. Yee-King, and M. D’Inverno, “In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, may 2019, pp. 1–11.

[23] P. O’Donovan, A. Agarwala, and A. Hertzmann, “DesignScape: Design with Interactive Layout Suggestions,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, apr 2015, pp. 1221–1224.