Towards Computational Co-Creation in Modding Communities

Kazjon Grace and Mary Lou Maher
{k.grace,m.maher}@uncc.edu
University of North Carolina at Charlotte

Abstract

Large online repositories of player-generated game content are a popular component of modern interactive digital entertainment. These add-ons are typically referred to as "mods", and the communities of co-creators that coalesce around them "modding communities". We propose a method to augment these communities with computational agents that encourage creative submissions, which we define as those that are both unexpected and valuable. We propose to combine player experience modelling, computational creativity evaluation and mixed-initiative co-creation to drive modding communities towards more fun and diverse gaming experiences.

Introduction

Game modding communities can be considered creative communities of practice (Lave and Wenger 1991), in that they revolve around participation in a task requiring expertise. They involve mutual engagement, function as a joint enterprise, and make use of a shared repertoire of skills and tools. We propose to apply computational creativity techniques to these communities, with the goal of fostering participation in the community, diversity in contributions, and greater fun and engagement in the game being modded. Modding communities provide a large dataset of creative contributions, a community of amateur creators both expert and novice, and a well-defined creative task. These attributes make them a perfect experimental domain for computational co-creation.

Co-creation (Prahalad and Ramaswamy 2004) is the joint creation of products or services by originating creator(s) and a community of participants (traditionally, but not exclusively a company and their customers). The user base helps create value by sharing, remixing, engaging with and constructing new elements. In games design this can most commonly be found in modding communities, where motivated players create, remix and share game elements not part of the out-of-the-box product. We propose to augment co-creation platforms with intelligent agents that identify and promote creative contributions, critique and make suggestions to human contributors, and hopefully are eventually incorporated into modding communities as computationally creative contributors in their own right.

Co-creation involving both AI and humans, “mixed-initiative co-creation” (Yannakakis, Liapis, and Alexopoulos 2014) has been previously applied to game development tasks – where creativity support tools foster game designers’ creativity during level design. We broaden the scope of this approach to a community of amateurs rather than a single game designer – benefitting from the quantity and diversity of contributions contained therein. The community approach also emphasises the role of computational creativity as a companion to peer feedback, augmenting rather than replacing it. Combining a diverse group of agents (each with individual histories and therefore preferences) with a diverse group of amateur designers will allow us to study the effect of mixed-initiative co-creation on motivation, creativity and behaviour at a social level, an area of active interest to both the game design (Sotamaa 2010) and computational creativity (Saunders 2012) research communities.

To develop an agent model for evaluating the creativity of contributions to modding communities we build on our existing work in computationally evaluating the creativity of products (Maher and Fisher 2012; Grace et al. 2014a; 2014b; Grace and Maher 2014) and on the Player Experience Modelling (PEM) technique of Pedersen et al. (2009; 2010). Our framework for creativity suggests that creative artefacts are unexpected and useful, where both of those terms have domain-specific multi-faceted definitions. We adopt PEM for the usefulness dimension in our model, and define several aspects of unexpectedness below. We then describe a set of ways that computational co-creation agents could interact with the community. In this exploration we constrain our computational co-creation approach to level design tasks, but maintain that it could theoretically be generalised to all kinds of community contributions, from in-game items to total conversion mods.

Background

Computationally evaluating creativity

What makes game content – player-generated or otherwise – creative? There is no universally accepted operationalisable definition for creativity, and it has been suggested that what is “creative” should be considered a property of the so-
ciety that created it, rather than some universally objectifiable metric (Csikszentmihalyi 2009). Taylor (1988) presents dozens of definitions from the literature, noting the distinction between creative people, creative products, and creative processes. Newell, Shaw, and Simon (1959) define creative problem solving as occurring where one or more of the following conditions are met: the result has novelty and value for the solver or their culture, the thinking is unconventional, the thinking requires high motivation and persistence, and being ill-defined or requiring significant problem formulation. This definition has coalesced into the dyad of novelty and value, which is the closest thing to a broadly adopted definition of creativity that can be found in the literature (Taylor 1988). Critics of the novelty/value approach to evaluating creativity judge most implementations of novelty insufficient to capture the complexity and subtlety of the way observers react to creative products (Maher and Fisher 2012). In this work we adopt a variant of this dyad, in which novelty is expressed as unexpectedness, and value is expressed as usefulness. We explore our motivation for this below.

Koestler (1964) and Boden (1990) offer definitions of creativity based on the creative process rather than the resulting artefacts. These focus on the mental context of the creator – their understanding of the domain of the creative problem, which Koestler calls a “matrix” and Boden a “conceptual space”. Koestler’s matrices are “patterns of ordered behaviour governed by fixed rules”, while Boden’s spaces are dimensioned by the “organising principles that unify and give structure to a given domain of thinking”. Boden describes the pinnacle of creative thinking as the transformation of the space of what is possible to include artefacts that were previously outside that space, while Koestler describes creativity as the blending, or “bisociation” of elements from two distinct mental frames of reference. The transformational notions of the creative process postulated by Boden and Koestler are similar to those of unconventionality and reformulation raised by Newell. Each involves the relaxation of constraints and the adoption of elements – of process or product – considered outside the “norm”. Like the novelty/value dyad, transformational creativity has been criticised by some computational creativity researchers, mostly for its difficulty to operationalise (Wiggins 2006).

We have proposed unexpectedness as a bridge between the transformativity-based definitions of creativity and the novelty/value dyad (Grace and Maher 2014). We argue that adopting an observer-centric viewpoint for evaluating creative products allows us to encapsulate the range of ways people relate to creative things – transforming their understanding, disrupting domains, challenging social norms. In this approach an artefact is judged against the observer’s history of experiences with that kind of artefact, rather than against an objective set of artefacts. This parallels the user modelling approach taken in PEM.

We operationalise “novelty” to unexpectedness, allowing each artefact to be compared against a set of domain-appropriate expectations formed by an agent observer (constructed as an approximation of the expectations of a human player). This incorporates the ideas of reformulation, transformation and bisociation. We express “value” as usefulness: the degree to which an artefact is effective at the purpose its observer desires it for, whatever that may be.

### Computational creativity in games

The use of methods from computational creativity to evaluate procedurally generated game content has been increasingly explored in the last few years. These approaches are different to the majority of the flourishing field of Procedural Content Generation (PCG) as they emphasise – to a greater or lesser degree – the novelty dimension of what they generate.

The constrained novelty search approach (Liapis, Yannakakis, and Togelius 2013; 2014; Yannakakis, Liapis, and Alexopoulos 2014) is the most explicit incorporation of divergence in PCG evaluation of which we are aware. Novelty search is an evolutionary computing technique in which fitness is granted to individuals based on their deviation from norms established by comparison to the current and previous populations (Lehman and Stanley 2011). Constrained novelty search applies this to optimisation problems with clear divisions between feasible and infeasible solutions – such as level design. In this approach levels are evolved for both playability and novelty – by comparison against an archive of past novel designs – using a variety of constraint and novelty promotion methods, which are then compared based on their diversity. Many of the techniques in this work are applicable to our own, and we intend to build on them by replacing the “novelty archive” and its distance-based metric with a set of predictive models that constitute the expectations the system has about new designs. We have argued in previous work that the use of likelihood – according to a diverse set of expectation models – to model the unusualness of a new creative artefact is more effective than relying on distance-based comparisons (Grace et al. 2014b).

Togelius and Schmidhuber (2008) present a method for evolving game rulesets using the latter’s model of artificial curiosity (Schmidhuber 2010). Game rulesets and agents that play those games are co-evolved, and rulesets which are neither too easy nor too hard to learn to play well are rewarded. This is based on the notion that both games requiring no skill and games that are nearly impossible are not interesting. This relates to the notion of unexpectedness pursued in our work – in the work of Togelius and Schmidhuber there is a predictability sweet-spot, producing games that are neither too predictable nor too unexpected, while in our work games that are maximally unexpected while also being useful (playable, challenging, balanced, fun, etc) are preferred.

Other research in encouraging diversity and creativity among game content is in the area of NPC interestingness (Yannakakis and Hallam 2004; Yannakakis 2005). Simple models of player behaviour are used to bootstrap interesting AI behaviours through off-line learning which are then further evolved through on-line learning with real players. These models are also based on the “not too hard, not too
easy” strategy, but additionally explicitly reward opponents for behaving diversely. While we do not focus on NPC behaviour in our model, it is notable that diversity and interestingness — what we would call creativity metrics — have been applied elsewhere in automated game design tasks.

Togelius et al. (2010) describe a system for evolving a Pareto front of strategy game maps along the axes of playability, fairness, skill differentiation, and interestingness. This closely parallels the objective function in our work: playability, fairness, and skill differentiation are aspects of usefulness, while interestingness is a kind of novelty. The authors develop fitness functions that are specific to the design of good Starcraft maps, but represent general principles — balance, replayability, strategic depth — that apply more broadly. These metrics also demonstrate the acceptance of the observer-centric evaluation view among the PCG evaluation community, a fact that distinguishes it from the broader computational creativity research community. The interestingness metric in Togelius et al. is an exception: it is a within-map diversity promoter, and does not compare that map to previously-generated ones in any way. Other automatic terrain generation research has investigated the diversity of output produced, but not explicitly modelled novelty/unexpectedness in their objective function (Frade, de Vega, and Cotta 2010a; 2010b). Our work will build on these approaches in two ways: by incorporating observer-centric expectations to promote the diversity of solutions, and by incorporating these evaluative models into a community containing both player- and AI-generated content.

ANGELINA (Cook and Colton 2011; Cook, Colton, and Gow 2012; 2014) combines ruleset generation, NPC generation, level design and other aspects into a complete model of automated game design. The system, which has been through several iterations and now produces 3D games in the Unity engine, co-evolves game components to produce a complete game. ANGELINA’s fitness functions incorporate observer-centric routines where fake players are generated and conduct playthroughs using simple static rulesets. While it does not explicitly incorporate a novelty objective, ANGELINA is at the far end of the spectrum of intelligent game design support tools, operating largely without human designer intervention. We propose a tighter coupling between human and AI: computational critique, suggestion and contribution to a community of amateur human game designers.

**Computationally evaluating creativity in player-contributed level design**

We develop a framework for how a computational agent embedded in an online modding community — a society of amateur game designers working in a constrained domain — could use our creativity evaluation techniques to productively contribute. We base this framework on the assumption that high quality mod content is not just fun and interesting to play, but is creative in its domain — unexpected and novel given the community and the underlying game. This assumption is born out by observations of innovation among modding communities, where the periodic emergence of new game styles and new kinds of content are considered a sign of a healthy and provocative community. It is this diversity and continuous evolution that we intend our framework for computational co-creation to stimulate. While we believe this approach is general enough to apply to a variety of genres and communities, any given implementation of it must focus on a specific domain in order to permit the design of appropriate representations and metrics. For the purpose of this paper we discuss level design contributions in general, but we note that any instantiation of this framework would have a more narrowly defined focus.

**Unexpectedness in player-contributed level design**

Grace et al. (2014a) identify four kinds of expectation that are relevant to evaluating a creative product, each of which we relate to evaluating level design:

1. **Categorical expectations** about a design given a classification either inferred by the observer or attributed by an external source.
2. **Trend expectations** about a design given domain trends (the change in designs over time).
3. **Relational expectations** about a design given an aspect of the design.
4. **Comprehensiveness expectations** that the observers knowledge of the domain will suffice to describe new designs.

The simplest categorical expectations are analogous to distance-based novelty metrics — the expectation is that artefacts in the domain will continue to resemble past experiences. For example, consider the expectation (learned from experiencing past maps) that maps contain a certain amount of resources, rendering particularly high or low resource concentrations highly unlikely. The agent would learn a probability distribution over the expected values of the level design’s variables, and rate new observations by their unexpectedness. This concept can be applied to sub-categories in the domain — perhaps “desert” maps have different concentrations of resources, or perhaps maps tagged “quick game” by players tend to have shorter routes to enemy start locations. Unexpectedness of this kind may represent a new direction for the community as a whole or a redefinition of a tag or sub-genre.

Trend expectations capture changes in the community’s tastes and preferences over time. In modding communities these trends can be caused by imitators of a novel discovery (consider the proliferation of hero defence maps for Warcraft III after the style was established), by an officially released game expansion changing the space of possibilities, or by the emerging metagame of a competitive strategy game. Unexpectedness of this kind may represent a contribution that bucks domain trends — and which, if sufficiently influential, may lead to the emergence of a new trend itself.

Relational expectations involve predictions about how features of game contributions vary together — what the presence or absence of one feature or range of a variable, says about the presence or value of other variables/features. For example, consider the presence of a pattern of narrow, blind
corners in a strategy game map. This may be found, on inference from the set of known maps, to be highly correlated with the availability of resources or weapons that encourage close-ranged fighting. Or consider that smaller maps may be predictive of a shorter average play time (as relational expectations can be built using both performative as well as descriptive design variables). Unexpectedness of this kind may represent a novel play-type; an experience that challenges player expectations and forces them to think outside-the-box to succeed.

Comprehensiveness expectations model the “impact” that observing a new design has on the computational agent’s understanding of the domain. These are closely related to the notion of transformational creativity – should an agent be confident in its conceptual understanding of a domain then an observation that forces a restructuring of that understanding is transformative. This rewards designs that are not just novel, but affect how other designs are categorised. Consider the example of a contributed map that combines two previously-separate play styles, such as tower defence and hero arena maps. This hybrid might force a re-examination – in both the minds of players and the computational agent – of how such genres are defined. As a result many maps previously thought to belong to one or the other couple be re-categorised, possibly to a third, new genre.

Of note in constructing each kind of expectation is the nature of games as an interactive medium, where the experience emerges from the interaction between player and artefact, rather than being solely a property of the artefact itself. This represents a departure from previous expectation-based approaches to creativity evaluation, which have focussed on either sequentially experienced artefacts like music (Pearce and Wiggins 2012) or holistically experienced artefacts like mobile phones (Grace et al. 2014a).

We are investigating three approaches to addressing the evaluation of interactive experiences: learning from recorded player data, approximation through static analysis of game content, and simulation using AI substitutes. The player data approach has the advantage of accurately reflecting the player experience, but the disadvantage of not permitting diversity between agents (as all would have access to the same player data and they would not truly be “playing” themselves). The approximation approach has the advantage of simplicity, but would require carefully constructed and validated game-specific models of what expectations arose from what features. The simulation approach has the advantage of producing large amounts of usable, diverse data, but the disadvantage of significantly restricting what kinds of games and content may be modelled.

Each of these kinds of expectation could be implemented in many ways, over many representations, over many kinds of contributed game content. We intend to start with level design as it is both one of the most well-researched PCG domains and one for which well-populated online repositories exist among mod communities.

Usefulness in player-contributed level design
The usefulness of game content in an online modding community can be assessed three ways: by the popularity and/or ratings assigned to it by the community, by inferring preferences from player behaviour, or by algorithmic evaluation or simulation. Popularity metrics (which we call “social evaluation”) are the most representative of the community’s desires, although it should be noted that not all good content needs to be popular to the masses, it may be highly-well regarded among a minority sub-community. One disadvantage of social measures of usefulness is that they can only be calculated after a design has been publicly released and some time has passed. Another disadvantage is the significant social pressures that may confound public evaluation – the influence of gatekeepers, peer pressure, social connectedness, and so on.

PEM approaches incorporate the latter two kinds of usefulness measure: metrics inferred from user behaviour or designed by researchers to reflect it. These methods range from simple statistical properties of levels through to complete simulations of game play-throughs. PEM typically involves detailed observation of multi play-throughs by a single player to extract their personal preferences. By contrast our approach will involve more abstract population-level analysis, necessitating server logs, uploadable replays or some similar access to large quantities of player data. The Space Syntax approach to spatial analysis, which originated in architecture and urban planning and involves quantitative analysis of visibility, connectedness and flow, is one form of population-level analysis that could be used to extend PEM to a community scale. Data about how players have interacted with contributed maps is not currently available in the majority of participatory gaming communities, but efforts to collect such data are beginning to emerge, particularly on large digital distribution platforms like Steam. Further work is needed to be done to establish how accurately such analyses approximate aspects of the player experience like balance, replayability, and strategic depth on a population scale.

Kinds of agent contributions to a modding community
The central component of our framework is that agents in a mixed-initiative community, both computational and human, contribute game content. We propose that the computational agents interact with the community in six ways: public individual feedback, private individual feedback, summative collation, suggested inspiration, social recommending, and remix. The first five of these involve justified evaluation and critique of contributions authored by others (see Charnley, Pease, and Colton (2012) for a discussion of framing and justification in computational creativity), while the sixth involves the computational agent doing PCG and contributing itself – using the work of others as a starting point. In detail, the six kinds of contribution are:

- **Public individual feedback** occurs when an agent observes a design that performs particularly well along one or more of its creativity dimensions and leaves a public comment saying so. This would consist of a justified evaluation: what is liked and why. We restrict public, unsolicited feedback to positive judgements to avoid alienating contributors, many of which may be novice creators. This restric-
tion may be relaxed upon evaluation of the agents’ critical abilities and user responses to them.

- **Private individual feedback** occurs when a user requests that an agent provide a review of one of its submissions, resulting in a comprehensive report over all the dimensions of creativity the agent is able to measure. This report would be issued privately with an option for the user to publicise the results if they desire. These judgements – particularly the extremely high or low scores – would be accompanied by justifying critique.

- **Summative collation** occurs when an agent ranks and publishes a list of contributions thought to be particularly creative along one or more dimensions of evaluation. These lists are equivalent to “favourites” or “playlists”, and will accompany a description of why the agent believes the dimension being exhibited is of value to the community. Making these lists agent-centric rather than centralised (and thus subjective to each agents experiences) helps communicate agent diversity to users and avoid the perception of a single omniscient computational arbiter.

- **Suggested inspiration** occurs when an agent observes a contribution for which a previous contribution exists that is similar but better. In this case the agent will make a suggestion that the user explore the inspiration(s) and give a description of what about them the agent likes. This can be phrased in the form “I like that you are doing X, check out works a, b and c for more examples of that!” The intention here is to increase the diversity (and hopefully, eventually, the creativity) of experience of human contributors.

- **Social recommending** is similar to suggested inspiration but recommends that a particular user or group of users rather than a particular work. This resembles traditional reciprocal recommendation approaches (Pizzato et al. 2010), except that matches will be chosen based on the likelihood that the contributions of each user will be stimulative to the other.

- **Remix** is the most complex of the kinds of interaction between the computational agents and the co-creation community. Remix occurs when an agent generates and submits a new contribution based on existing work. The agent will accompany its creation with a justification of what it found inspiring in its chosen stimuli, and what aspect it was trying to improve on. Remix could eventually lead to authorship in cases where the agent’s contribution significantly diverges from its source material(s). This is the ultimate goal of our approach: computational agents that participate in communities of human creators as equals, building on and stimulating the creativity of other agents, human or AI.

### Requirements for computational creativity evaluation in gaming communities

The framework presented here demands certain requirements from the domain to which it is applied. In this paper we have framed it as generally as possible, but now attach several caveats to the community of modders, the game being modded, and the contributions being made to it:

- The **community** in which mixed-initiative co-creation is to be applied must be of sufficient scale, with at least hundreds of contributions and thousands of players so that machine learning approaches are feasible. It should also have an existing tradition of (human-driven) critique, ensuring that contributors intend to receive evaluations and criticisms of their work. It should contain diverse users with a range of skill levels in game design, such as is found in communities of amateurs. And finally it should be of sufficient receptiveness to the idea of machine-driven critique, a point on which user-centred research (such as Wizard of Oz experiments) may be necessary to properly explore.

- The **game** for which mods are being solicited, evaluated and remixed can belong to a variety of genres, so long as there are defined in-game goals and a coherence of higher-level strategies to facilitate simulation (for example an open-ended building game like Minecraft is probably not appropriate, but a constrained building game like a bridge simulator may be). The game must also facilitate the uploading of player telemetry (for analysis) and be able to be played competently by an AI (for simulation).

- The **contributions** of modders which are to be computationally critiqued must facilitate computational representation: maps or levels are appropriate as they can be represented spatially or geometrically, items or units are appropriate as they tend to have quantifiable in-game attributes, but art assets or other aesthetic contributions are likely not appropriate.

### Discussion

In this position paper we have outlined how the unexpectedness and usefulness approach to computational creativity evaluation could be applied to mixed-initiative co-creation among a modding community. While we have presented our approach generally, agnostic of the game genre, community type or contribution kinds in question, we note that any individual instantiation of this approach must restrict each of those factors in order to make creativity evaluation feasible. We have discussed the kinds of unexpectedness and usefulness that could be implemented in the domain of level design, building on our own framework for the former and on Player Experience Modelling for the latter.

We are in the process of porting our machine learning models from the domain of product design (see Grace et al. (2014b)) to game design, developing representational techniques and kinds of unexpectedness that are appropriate to this new domain. We believe that the idea of computational co-creation in online game design communities, implemented using notions like those presented here, is a promising new domain for experimental game AI research, and one that can help productively bring computational game creativity to a broader audience.
References


Sotamaa, O. 2010. When the game is not enough: Motivations and practices among computer game modding culture. Games and Culture.


