Interactive Narrative Design
Beyond the Secret Art Status:
A Method to Verify Design Conventions for Interactive Narrative
Hartmut Koenitz
Christian Roth
Teun Dubbelman
Noam Knoller
HKU UNIVERSITY OF THE ARTS, UTRECHT

ABSTRACT
In recent years, game narrative has emerged as an area for novel game concepts and as a strategy to distinguish a particular title. However, innovation in this area comes primarily from indie companies and individual efforts by noted designers. There is a lack of trained specialists ready to produce interactive narrative experiences. Many existing practitioners are self-trained and often rely on intuition in their design practice. A key element missing from the effort towards a more sustained development and improved professional training is a set of design conventions that fulfill a role comparable to cinematic conventions like continuity editing or montage. Therefore, our research focuses on identifying, verifying and collecting such design strategies. We describe an empirical method to verify candidate design conventions through the evaluation of user reaction to A/B prototypes, which improves upon the trial-and-error process of old.

KEYWORDS
design conventions; game design; interactive narrative design; pedagogy; education; user experience measurement.

RESUMO
Nos últimos anos, a narrativa nos jogos eletrónicos surgiu como uma área para novos conceitos de jogo e como estratégia para caraterizar obra específicas. No entanto, a inovação nesta área vem sobretudo de empresas independentes e de esforços individuais de designers reconhecidos. Faltam especialistas treinados capazes produzir experiências narrativas interativas. Muitos dos praticantes são autodidatas e confiam na intuição nas suas práticas de design. Um elemento-chave que falta ao esforço de desenvolvimento sustentado e a uma melhor formação profissional é um conjunto de convenções de design que cumpram um papel comparável às convenções cinematográficas, como edição ou montagem de continuidade. Esta investigação centra-se na identificação, verificação e recolha dessas estratégias de design. Descrevemos um método empírico para verificar as convenções de design propostas através da avaliação da reação do utilizador aos protótipos A/B, o que melhora os processos anteriores baseados em tentativa e erro.

PALAVRAS-CHAVE
convenções de design; design de jogos; design de narrativas interativas; pedagogia; educação; medição da experiência do utilizador.
A decade after the narratology vs. ludology debate, game narrative has emerged as an area for novel game concepts and as a strategy to distinguish a particular title. The critical and economic success of Telltale, with titles such as *The Walking Dead* (Telltale Games, 2012), *The Wolf Amongst us* (Telltale Games, 2013) or *Minecraft: Story Mode* (Telltale Games, 2015), but also of independent productions from *Save the Date* (Cornell, 2013) to *The Stanley Parable* (Wreden, 2011) and *Firewatch* (Campo Santo, 2016) is testimony to this development. At the same time, this demonstrable interest by audiences is not met by a wider body of specialized knowledge and training in narrative-focused game design.

Game design programs might include a course in game narrative, but such an effort barely scratches the surface of a complex topic. At the time of this writing, no degree program in interactive narrative design exists. Many existing practitioners are self-trained and often rely on intuition and experimental approaches in their design practice. We find innovation to be coming from noted designers like David Cage (*Fahrenheit/Indigo Prophecy, Heavy Rain* (Quantic Dream, 2010)) and midsize game companies like Telltale (which was much smaller when it invented the now famous “telltale formula”). The latter’s case is instructive, as the company slowly built up, in an iterative process of continuous refinement, to the narrative design formula that had made *The Walking Dead* such an outstanding success. However, there are already signs that the ‘Telltale formula’ has become stale, since — in the eyes of some critics and players — later titles provide new content, but only minor additions to the overall design approach (Sinha, 2016; S. Wright, 2015). This lack of development comes as no surprise, since narrative-focused game design is a double secret art. First, it is practiced only by relatively few ‘initiated’ self-trained practitioners. Second, where more knowledge has been accumulated in the commercial space, it is treated as a closely-guarded company secret. Clearly, knowledge about narrative game design exists (as Dubbelman (2016) reminds us), and it is disseminated at places like the GDC (Game Developer Conference) narrative summit. However, as we have identified earlier, there is a lack of professional training (Koenitz 2014) further compounded by the fact that much of the available knowledge is either too abstract (e.g. high-level concepts like Murray’s “Scripting the interactor” (Murray, 1997)) or too specific (valid only for particular works as discussed in post
mortems at the GDC narrative summit) (Koenitz, 2015). This state of affairs is an impediment to overall development in this area of game design and related fields such as interactive documentaries, journalistic ‘interactives’ and installation pieces. What is missing is an established body of design conventions for game narrative that fulfill a similar role to our more developed understanding of theatrical or cinematic conventions.

Interestingly, a look back at the early years of video games shows that narrative game design is not exactly a new idea. Pioneering text-based games like Adventure (Crowther, 1976) lead to the creation of the IF genre including successful narrative-focused commercial games like Zork (Lebling, 1980; Lebling, Blank, & Anderson, 1979) and Planetfall (Infocom, 1983). However, the rejection of narrative as an avenue for video games in the narratology vs. ludology debate (Aarseth, 2001; 2004; Eskelinen, 2001; Frasca, 1999; Juul, 1999; 2001), in concert with a disregard for the specific needs of interactive narrative design (e.g. Jesse Schell’s rant against specific approaches in Fullerton’s Game Design Workshop (Fullerton, Swain, & Hoffman, 2008)) has resulted in a scarcity of efforts to analyze and develop narrative game design.

There is also a connection between the minor status of educational efforts in game narrative and the lack of established interactive narrative design conventions. Education depends on the availability of teaching materials. Our work is therefore intended to have impact in both education and practical application by existing practitioners. We describe a method to verify narrative design conventions using empirical methods. Instead of the trial-and-error process of old, our method evaluates user reaction to A/B prototypes using an established toolkit measuring immersion, agency and transformation through a granular set of 13 psychological dimensions (Roth and Koenitz, 2016; Roth and Vermeulen, 2013; Roth, Klimmt, Vermeulen, and Vorderer, 2011; Roth, Vorderer, Klimmt, and Vermeulen, 2010).

In the remainder of the paper, we discuss the notion of design conventions and related work before describing the measurement toolbox we use and the process for verification. Finally, we explain the concrete setup and draw the connection to education.

II. DESIGN CONVENTIONS

The concept of guidelines for the creation of narrative artifacts might be traced back to early guides on literary composition. The introduction of conventions is a process whereby authors or designers experiment with formal patterns that can carry cultural meaning, audiences indicate whether these patterns are indeed perceived as intended, and through iteration (Derrida, 1988) they become conventional. In cinema, for example, many conventions now taught in film schools — for example continuity editing in Hollywood cinema (Thompson and Bordwell, 2012) — have historically evolved over a period of several decades of...
trial and error. Therefore, it might seem that conventions cannot be imposed on the field. They become accepted design conventions once both authors and audiences have appropriated them, and they become implicit cultural knowledge. However, some conventions have been the result of experimental setups and quasi-academic investigations, for example by pioneering Russian filmmaker Lev Kuleshov, who devised a method to establish the effects of montage on viewers (Kuleshov, 1974). Kuleshov interjected the same shot of an expressionless actor between different images (e.g. of food or a funeral) and was able to demonstrate that audiences perceived the actor's expression differently depending on associations originating from the context provided by the shots that surrounded them. These experiments provided the basis of the Soviet montage technique of film editing, as theorized by Sergey Eisenstein (first in lectures, then from 1937 in written publications (Eisenstein, 2010)).

Scholarly methods in user experience research (Calvillo Gámez, Cairns, and Cox, 2009; Dow, 2007; Hassenzahl, Diefenbach, and Göritz, 2010; Roth, Vermuelen, Vorderer, and Klimmt, 2012a; P. Wright and McCarthy, 2008; Zammitto, Mirza-Babaei, Livingston, Kobayashi, and Nacke, 2014) now provide the means to analyze the effects of particular design approaches, and thus to identify and evaluate candidate conventions empirically, without the need for extensive trial and error. In other words, design conventions can be evaluated through their impact on user experiences and this is the area we focus on.

III. RELATED WORK

Comparative methods for the investigation of user experiences in interactive narrative are well established. Dow et al. (Dow, Mehta, Harmon, MacIntyre, and Mateas, 2007) examined user experiences across different versions of the interactive drama Façade. Participants used an augmented reality version (in which a see-through display projects the characters into the physical recreation of the apartment), a desktop version with speech communication, and a desktop version using a keyboard input. Through interviews and observations, the authors found that immersive augmented reality can increase perceived presence.

Aylett et al. (Aylett, Louchart, Dias, Paiva, and Vala, 2005) conducted a small-scale (N = 11) user test with the emergent Interactive Storytelling system FearNot!, and collected children’s responses on a short set of evaluation items. They investigated whether conversations were perceived as interesting and ‘felt real,’ as well as how autonomous characters seemed to respond to user input (e.g., whether they seemed to be listening to user advice). In comparison to a scripted version of the same setting — featuring non-autonomous characters — a larger-scaled study (Hall et al., 2004) showed that character responsiveness in the emergent software was experienced as less real, less interesting, and less responsive.
to user input. The authors concluded that the conversations in the scripted version were perceived as more coherent than the conversations with autonomous characters, which sometimes seemed unresponsive to user advice.

These earlier studies have been useful in optimizing system parameters and creating more effective links between the goals of a given interactive narrative and user requirements. However, the measurements applied in these studies do not enable the systematic testing and comparison of design conventions. Roth’s approach (Roth et al., 2010) overcomes this limitation and enables us to empirically compare candidates for design conventions. Based on Entertainment Theory (Bryant and Vorderer, 2013), Roth’s measurement toolbox covers a wide range of user experience dimensions (effectance, autonomy, usability, suspense, curiosity, presence, identification, character believability, flow, eudaimonic appreciation, enjoyment, positive and negative affect). We have recently mapped these dimensions (Roth and Koenitz, 2016) to Murray’s more broadly understood concepts (Murray, 1997): agency, immersion, and transformation.

This measurement toolbox has been applied within a range of different studies (Roth et al., 2010; Roth and Vermeulen, 2013; Roth et al., 2012b) and proved to be useful for the evaluation of interactive digital narrative (IDN) user experiences. For example, a prototype-based user study compared a turn-based dialogue design to a real-time approach (Endrass, Klimmt, Mehlmann, André, and Roth, 2011). Findings suggest that participants value real-time dialog more highly in terms of perceived autonomy and elicited curiosity. Another study (Roth et al., 2012b) evaluated the effects of different player modes (embodied actor mode versus disembodied ghost mode) on the user experience (sense of immersion vs. control). The results confirmed the assumption that control-related experiences such as autonomy, effectance, and flow, ranked higher in disembodied ghost mode, resulting in a higher sense of accomplishment and user satisfaction. Another study investigated the effect of auditory feedback on user enjoyment of an interactive surround video. In this A/B comparison, participants watched two variants of an interactive movie on the virtual reality hardware Oculus Rift, either with or without sound feedback as rewarding indicator for successful interactions. The authors had hypothesized that auditory feedback would increase the perception of effectance and had to reject this assumption (Vosmeer, Roth, and Schouten, 2015). Instead, interactors rated presence significantly lower as the sounds hampered their feeling of immersion. These studies show the validity of user experience testing when considering design decisions.

IV. VERIFYING DESIGN CONVENTIONS

In our approach towards the verification of design conventions, we apply experimental designs and methods from psychological research. The design of our ex-
Experimental setup is based on earlier experience with IDN user experience evaluation (Roth and Vermeulen, 2013), with prototype-based user studies (Endrass et al., 2011) and with the A/B comparison approach (Roth et al., 2012b).

Design conventions can be verified through their impact on the user experience. There are two broad approaches (Bernhaupt, 2010). First, there is the evaluation of hardware technology and the software itself, in terms of user interface usability, which investigates general system responsiveness and usability. Second, there is an approach which focuses on the user experience as the overall appreciation and enjoyment of the experience delivered by the system, e.g. regarding the presentation of the story world, the characters, and the interactions. In respect to the latter, our specific approach is grounded in Entertainment Theory (Bryant and Vorderer, 2013; Vorderer and Reinecke, 2015; Vorderer, Klimmt, and Ritterfeld, 2004). In this context, ‘entertainment’ — in line with psychological research and the communication sciences — is used to mean the evaluation and experience from the audience’s perspective. ‘Entertainment’ as we use it, is thus not the opposite of ‘artistic value,’ but inclusive of it.

In the current research, we investigate both dimensions — the impact of design decisions in terms of both usability and overall appreciation. This kind of hybrid enjoyment is a highly complex experiential state with a variety of manifestations (e.g. exhilaration, suspense, and identification) and numerous determinants attached to both the system delivering the experience and the person perceiving the system. To evaluate whether and how a given IDN system elicits enjoyment and meaning in users, it is necessary to conceptualize in advance the kind of experiential qualities a system might deliver. Subsequently, experimental exposure studies with control groups aim to measure these experiences and thus help identify what elements determine the appreciation of interactive narrative systems. In this context, the aforementioned toolbox (Roth, 2016) provides the necessary conceptualization and granular measurements to empirically evaluate the effectiveness of design strategies.

V. IDENTIFYING CANDIDATE CONVENTIONS

We start the process by identifying candidates for verification as design conventions. We use existing scholarly publications as sources for conventions, but also structured interviews with practitioners, expert statements and analysis of artifacts. For example, Janet Murray, in her investigation of Weizenbaum’s early AI therapist simulation Eliza (Weizenbaum, 1966), identifies the — possibly unconscious — design approach of “scripting the interactor” (Murray, 1997) as a major factor why this early experiment was compelling. Another candidate design convention was identified by Kevin Bruner, CEO of Telltale Games, who
stated\(^1\) that an important aspect of his company’s success with narrative-focused games is in using “choice notifications” that alert players of potential later consequences. A further possible convention is “cross-session memory.” This candidate was gained through analysis of the game *Save The Date* (Cornell, 2013), in which the player experiences the consequences of decisions made in play sessions completed earlier.

VI. USER STUDY DESIGN

The design of our experimental setup is based on earlier experience with IDN user experience evaluation (Roth and Vermeulen, 2013), for example, with prototype-based user studies comparing two design alternatives (Endrass et al., 2011; Roth et al., 2012b). Therefore, we create prototype pairs — one applying a specific design strategy and another omitting it.

In this way, we enable the double-blind two-group post-test-only randomized experimental design, which is an established quantitative research approach for assessing cause-effect relationships (Campbell and Stanley, 1963; Shadish, Cook, and Campbell, 2002). It is one of the best research methods to evaluate the comparative effects of two treatments, in our case the experience of an interactive digital narrative in two versions, differing only in regards to one specific design convention candidate (experimental group featuring the convention candidate and control group exposed to a version without it). In the field of online user experience research, this experimental two-sample hypothesis testing approach to measuring the effects of design alternatives is called “split testing” or “A/B testing.”

Following best practices for these kinds of user studies, we assign participants randomly to the two groups. This procedure leads to a probabilistically equivalent distribution of participants to the two experimental conditions. In every evaluation, we take participants’ demographics, game preferences, previous experience with IND, and other possible mediators into account. In this way, we are able to control the equivalence of the groups and can be used to counterbalance a possible distribution bias when measured prior to the distribution.

An extended version of this user study design, the Randomized-Multigroup Design (Bordens and Abbott, 2002), introduces one or more additional levels of the independent variable. In our case these would be variants of a design convention candidate. This extended setup would allow for the differentiation of different measurement ‘levels,’ e.g. of agency (a lot influence on a narrative level vs. a bit influence vs. no influence). This would result in a A/B1/…/Bn testing setup. A downside of this setup is the requirement of larger samples, as each added condition multiplies the necessary amount of participants.

---

\(^1\) In a keynote at ICIDS 2016 conference, Institute for Creative Technologies, November 16, 2016.
VII. PRACTICAL SETUP

Our practical setup involves student projects with a duration of two to three months, with weekly supervision meetings and extensive project documentation. We use regular meetings to discuss suitable narrative design approaches, to play-test, and to check on progress and potential issues. In-between these meetings, students upload their prototypes, work plan (containing milestones) as well as their project diary so they can be evaluated online at any time (e.g. using Google Drive).

Every project starts with a brainstorming phase, resulting in a concept draft. Before using any digital software, a paper prototype of the interactive digital narrative is built and play-tested together with the student. By doing so, potential issues with interaction design, pacing, narrative structure and the use of variables can be identified and solved without spending time on rewriting an application. Once the paper prototype is deemed satisfactory, students recreate it using the ASAPS authoring tool (Koenitz and Chen, 2012), a software suite for the creation of 2D multiplatform interactive narratives.

VIII. ASAPS SOFTWARE

We use the ASAPS authoring tool (Figure 1) for three reasons: a) the focus on 2D removes the complexity of 3D production and model making and thus enables rapid production of prototypes; b) all elements of the artifact are easily accessible for analysis in a non-proprietary format; c) the software allows for optional user tracking, providing an opportunity for additional data collection and subsequent scholarly treatment.

![ASAPS Software architecture](image)

**Figure 1.** ASAPS Software architecture.
IX. EDUCATIONAL APPLICATION

One application domain for the results of this kind of work is in education, as an important part in the professional training of interactive narrative designers. However, this still leaves open the question of how to best disseminate this knowledge. How can we educate a wider group of interactive narrative designers — Janet Murray’s “cyberbards” (Murray, 1997) — so that they understand the demands and possibilities at the intersection of interaction and narration (Koenitz and Louchart, 2015; Spierling and Szilas, 2009)? The challenge lies in the difference (Mateas, 2010): interactive narrative design is fundamentally different from approaches in well-established narrative practices, like filmmaking or creative writing; it demands practitioners to be able to design and develop interactive systems aimed to create narrative experiences. Therefore, successful IDN design requires a broad understanding of interaction and engaging narration. One promising approach for the education of IDN authors lies in teaching best practices. This, as we mentioned above, is where the lack of a canon of design conventions becomes also a pedagogical challenge. Therefore, the identification of generalizable IDN design conventions (Koenitz, 2015; Murray, 2012) becomes a necessary precondition for a more formal educational approach to interactive narrative design.

X. PEDAGOGICAL IMPLEMENTATION

Once the material for teaching is available, the next question would be: how to embed the knowledge into a concrete pedagogical effort and effectively teach IDN design conventions?

The practice of IDN can be situated within the field of creative technologies; this interdisciplinary and transdisciplinary field within the creative industries is technology-driven, user-oriented, and includes creative practices like interactive documentaries, game development and interaction design. Pedagogical practices in the domain of creative technologies (CT) generally favor practice-based education, contextual authenticity and playful learning (Connor, Marks, and Walker, 2015). In short, CT pedagogy favors learning by doing in authentic scenarios. Essential skills, attitudes and knowledge are appropriated by learners in situations that recreate the conditions where specific skills, attitudes and knowledge become useful, meaningful and applicable (Maddrell, 2015).

Following this insight, IDN design conventions can best be taught when a learning situation does not simply offer the conventions to the learner, but incorporates them in the actual process of designing and developing IDN artifacts. In most disciplines in the domain of creative technologies, this process is iterative, and relies on cycles of ideation, prototyping and testing (Meinel, Leifer, and Plattner, 2011). Learners appropriate the conventions most effectively when
they are engaged in the activity of creating new ideas, building prototypes and performing user tests.

XI. FUTURE WORK

For now, our investigation focuses on universal IDN design conventions. However, design conventions that can be identified in one IDN artifact might not necessarily be applicable to others. We should therefore be careful not to approach IDN design conventions as essentially universal and at least allow that they might be context sensitive and thus work best within the contextual constraints of particular formats, from interactive video games to interactive documentaries. This hypothesis is supported by a comparison with non-interactive media: some narrative conventions in a particular cinematic genre (e.g. horror movies) are applicable to other genres (e.g. romantic comedies) or even other narrative media, while others are not. Likewise, only some design conventions for interactive video may be applicable to video games and vice versa. A next step in our investigation will be an attempt to identify conventions in specific domains and thus enable a distinction between general conventions and ones that are more specific.

Furthermore, we plan to extend the scope of our evaluation methods to larger field tests. By offering some of the prototypes online, we can reach a wider audience and thus further improve the quality of our results in evaluating IDN design convention candidates. Future experimental designs could also include multiple variations, and thus allow for more complex experimental designs.

XII. CONCLUSION

In this paper, we described an approach for identifying and verifying design conventions for video games narrative. We discussed the nature and emergence of conventions in traditional media, before presenting our methodology to verify conventions by applying the measurements of user experience dimensions to A/B prototypes. Finally, we considered educational applications of the results of our process and future work.

The identification and verification of potential design conventions is an important step towards a specific pedagogy and a more accessible professional and artistic practice — the end of the secret art and the beginning of a professional discipline of interactive narrative designers.

REFERENCES


© 2018 Hartmut Koenitz, Christian Roth, Teun Dubbelman, Noam Knoller. Licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 International (CC BY-NC-ND 4.0).