Check for updates

Alternate realities in interactive digital narratives – understanding and improving design and prosocial effects through empirical methods

Hartmut Koenitz[1] · Christian Roth[2] · Elisa Mekler[3]

Abstract

Interactive digital narratives (IDNs) create alternate realities using both fictional and non-fiction material. The interactive aspect of IDN turns audiences into participants and enables the exploration of alternative perspectives and actions within a single artifact. Such multifaceted representations make IDN a vehicle for representing complex issues, a crucial capability at a time when the limits of traditional narrative media to adequately represent complex issues such as climate change become apparent. Conversely, properly evaluated, generalized knowledge about how exactly IDNs engage and influence us and what this means for the design of such works is still scant and thus this topic needs scholarly attention. In this overview paper, we discuss the potential of IDN, but also the difficulties of realizing this potential in terms of design and of verifying the effectiveness through empirical research methods. The potential of IDN as dynamic, participatory, and encyclopedic artifacts can be clearly expressed, yet the same cannot be said when it comes to the design and especially the evaluation of intended prosocial effects, the topic this paper is focused on. We start by identifying the problem of IDN design resulting from a combination of the lack of generalized knowledge and formal professional training. Then, we discuss the challenge of measuring the effectiveness of IDN design for prosocial effects and report on several case studies. In this context, we discuss methodological issues and advocate for best practices. Finally, we consider future steps in addressing the continuing challenge of evaluating IDNs.

1 Introduction

Interactive digital narrative (IDN) is an umbrella term for manifestations that combine narrative with interactivity, including interactive documentaries/movies, narrative-focused video games, journalistic interactives, extended reality application and installation pieces

---

Extended author information available on the last page of the article

[1]. The main characteristic of IDN is that the audience becomes interactors, who can influence the progression, perspective, outcome, or content of a narrative through their actions. IDNs have a particular application in creating multimedia alternate realities where 'alternate' would not only be understood as differing from real life, but also as providing alternate perspectives within the same artifact, either as the result of pre-designed alternative views or by means of choices leading to different experiences and outcomes. An example of the former type is the Emmy award-winning interactive documentary *The Last Hijack Interactive* [2], which features different perspectives on piracy in Somalia (the ship's captain, the pirate, the ship captain's wife, negotiators). An example of the latter is *Detroit Become Human* [3], a narrative-focused video game, where the interactor controls the fate of several characters within a fictional setting in which androids reach consciousness and demand to be recognized as living beings. These two examples show that IDN can provide such alternate realities both in fictional and non-fiction contexts and with different purposes (education, public information, and entertainment). Indeed, IDN can be described as a thriving space. There are commercially and critically successful video games like *Dear Esther* [4], *Gone Home* [5], *The Stanley Parable* [6], *Firewatch* [7], *Oxenfree* [8], *The Last of Us* (1 + 2) [9, 10], *Neo Cab* [11], *Mutazione* [12] and *Unpacking* [13], award-winning interactive documentaries such as *Fort McMoney* [14]—on the effects of the oil industry on a small community turned 'boom town', the already mentioned *The Last Hijack Interactive* [2], *The Industry* [15]—on the illegal drugs industry in the Netherlands, and thought-provoking VR pieces such as *Clouds over Sidra* [16]—on life in a Syrian refugee camp), *A Breathtaking Journey* [17] on the experience of refugees smuggled into the EU in the back of a truck facing the danger of suffocation, and *Goliath* [18] on the experience of psychosis.

What these examples share is the ability to not only transport audiences to a different place in a similar way earlier mediated expressions such as novels and movies have done, but also to actively explore these alternative spaces. This means to influence perspective, progression and/or outcome and the ability to act in roles different from the ones in real life. Audiences can explore planets, but also face the hardship of living in a refugee camp or act as mayor faced with the manifold challenges of urban planning. With the ability to act and affect changes comes a considerable potential for understanding and learning about difficult political or personal situations, including the global contemporary issues of global warming, migration, and dis/misinformation, as well as personal mental health. Indeed, this potential is the foundation for the INDCOR COST action, an EU-funded network (https://indcor.eu) for capacity building which encompasses more than 200 researchers from 40 countries. The network, in which the authors have leading positions, is focused on laying the foundation for a more widespread application of IDN as a standard output for media production houses and news organizations, but also as an important tool in education. Towards that aim, the action has identified several salient aspects in need of coordinated development, including shared concepts and vocabulary, design conventions, evaluation methods and a better understanding of the societal impact.

In this article the authors focus on the connection between design and evaluation and describe approaches and challenges for understanding and improving design methods by means of evaluation. We have previously published on a related topic in the form of a conference paper aiming at verifying design intentions [19] in video games and aspects of that paper are included in this longer article. This article expands on the original topic and instead is focused on discussing the evaluation of prosocial effects of IDNs.

In this context, our contribution is four-fold: 1) increase awareness for the need to improve existing practices in evaluating IDN artifacts, 2) point out the benefits of evaluation for designers, 3) provide guidance to implement best practices in evaluating, and 4)

offer a starting point for developing improved evaluation methods. Our overarching aim is to engage the IDN-related community and install evaluation as a shared responsibility with considerable benefits for both design and research.

2 IDN and alternate realities

Before we continue with our topic of design and evaluation, we would like to provide a framing for IDN in general and in relation to alternate realities. Narrative has been described has having an important function in self-construction [20] as well as constructing and understanding general reality [21]. Conversely, Barthes and Duisit have described [22] narrative as a function of diverse manifestations, including painting and music, an understanding later reinforced by the 'cognitive turn' in narratology, which takes narrative as a cognitive frame for "mentally projected worlds" as David Herman has put it in the seminal book on cognitive narratology "Story Logic" [23]. Taken together, these developments clearly indicate that narrative is not restricted to manifestations traditionally understood as carrying narrative content such as print literature and film. This foundational realization has implications both on the conceptual level and for the design of narrative experiences. In parallel to the development of narratological perspectives, artists and scholars have identified the potential of the digital medium for narrative expressions, e.g. for interactive drama [24] and started to describe possibilities and theoretical categories [25, 26], most prominently Janet Murray in her influential volume *Hamlet on the Holodeck* [27]. Murry defines affordances (procedural, participatory, spatial, encyclopedic) as well as aesthetic qualities (immersion, agency and transformation).

On the basis of both developments (cognitive narratology as well as the investigations into the digital medium for narrative expressions) Koenitz has defined IDN as a "[…] narrative expression in various forms, implemented as a multimodal computational system with optional analog elements and experienced through a participatory process in which interactors have a non-trivial influence on progress, perspective, content, and/or outcome." [1] Furthermore, he has also developed a specific conceptual framework for IDN, the *SPP Model* (system, process, product) [1], taking further inspiration from Roy Ascott's theoretical perspective on cybernetic art [28, 29]. Koenitz describes IDN as a specific form of narrative, distinct in important ways from the more traditional forms of print literature and film and thus requiring particular theoretical approaches. In this article, we take Koenitz' SPP model as the theoretical foundation. In this framework, the *system* contains the *protostory*, the space of potential narratives that can be instantiated by interactors in an interactive *process* and result in narrative *products*. The narratives in an IDN system can be distinguished from earlier forms as dynamic and unfinished, requiring instantiation and thus participatory co-authorship by interactors to be fully realized. While IDN *products* could be indistinguishable from earlier forms in some cases, they are the result of the two earlier states (*system* and *process*) and should be regarded as output. Influenced by Eladhari's work on retellings [30], Koenitz has more recently implemented a distinction between objective (recorded) product and subjective (retold) product [1].

Roth et al. [31] have extended Koenitz' earlier version of the SPP model [32] to include a perspective on the interactor's internal processes. They describe it as a "double hermeneutic" where the interactor interprets and reflects on two aspects, (1) the narrative trace so far and (2) the possibilities for interaction. This perspective implements Karhulahti's concept of the same name for games [33] inspired by Giddens' original concept from sociology [34]. More recently, Koenitz [1] developed this notion further by adding the dimension

of replay as a third hermeneutic circle which represents the reflection of prior traversals of an IDN artifact. Koenitz therefore understands the double hermeneutic as characteristic for the first encounter with an IDN, which changes to a triple hermeneutic in all subsequent encounters.

IDN are therefore conceptualized as dynamic systems that enable its audiences to influence the content and/or sequencing of narrative experiences in an interactive process, resulting in objective (recorded) or subjected (retold) products. In terms of concrete manifestation, IDN is a cross-cutting perspective encompassing all manifestations in which interactivity and narrative are connected in a non-trivial way. Today, the most prominent forms of IDNs are narrative-focused computer games and interactive documentaries, with XR applications gaining momentum in recent years.

In terms of applications, IDN has the potential to become an important tool for the representation of complex topics [35], enabling increased understanding and improved public discourse by means of presenting malleable alternate realities. More concretely, this potential stems from a number of specific abilities of IDN:

> IDN turn audiences into participants, who make decisions and see the respective consequences from them, gaining a systemic understanding.
> IDN works can contain vast amounts of information (Murray's "encyclopedic" affordance [27]) which in the case of complex issues translates into the ability to represent multiple, even competing, perspectives within a single artifact. In addition, it means that topics can be personalized and that IDN can provide a narrative interface to big data.
> IDN invite repeat engagement – replay – which enable audiences to reconsider decisions and explore different perspectives, furthering the systemic understanding.
> IDN can contain live data and thus can be kept up-to-date.

Overall, IDN combines the power of narrative as a long-established means of communication and knowledge transfer with the systemic properties of computational media. Every engagement with an IDN work creates a linear experience as a result of the audiences' choices, comprehensible as a narrative, while a larger systemic representation is in place which holds additional information.

The question is how to realize the potential of IDN for prosocial purposes on a broader scale – and how to verify positive effects. The answer to this question must come from a combination of design and evaluation – design to create captivating and compelling experiences as well as evaluation to verify whether the desired effects occur. As we have explained in the introduction, many good examples exist, however, these are still "one-off" examples and what is lacking in terms of design are established workflows and design conventions, especially in contrast to newspaper journalism or TV news and feature production. The remainder of this paper will address these issues, first considering design and then evaluation.

3 Realizing the potential of IDN: the challenge of effective IDN design

We have previously discussed IDN design [19, 36, 37] and offer a summary here in order to orient readers. The basic issue with the design of IDN works is that designers cannot rely on methods or conventions established for non-interactive works. When audiences have agency, creators no longer have the same control as do novel authors or film directors. Consequently, many conventions that are effective when creators determine when and

what audiences read, see or hear (such as framing of shots in a movie or the placement of a section in novel) are no longer usable when the audience has decision-making power and can control their pace, view, perspective as well as their goal. As Janet Murray points out, "Design processes are often stalled by […] unproductive attempts to apply legacy conventions to new digital frameworks" [38]. IDN design therefore is a part of "inventing a medium," (the title of Murray's book) and this effort requires "inventing something for which there is no standard model, like word processing in the age of the typewriter, or video games in the age of pinball." (ibid) With the advent of more advanced generative methods and co-creative opportunities, traditional narrative conventions become even more out of place. What we mean here are algorithmic methods e.g. for real-time generation of landscapes or for procedural generation of narrative, but also user-generated content, e.g. in Sandbox-type environments, where audience members can build structures and change landscapes.

This design problem can be described as the compound result of both the lack of formal IDN design knowledge and the lack of formal training. Unfortunately, even though the design of interactive digital narratives can be traced back more than five decades now, to the 1960s, with works like *Eliza* [39] or Grime's story generator [40], a shared body of design conventions for IDN has not emerged yet. When it comes to formal training, specialized programs are rare (less than ten worldwide at this time of writing[1]) and in many educational programs in Game Design, interactive narrative aspects are the topic of only a single course, which is insufficient for a deeper understanding. Established interactive narrative designers are for the most part self-trained and their respective knowledge is individual, uses private vocabulary and thus is not easily transferable, a state that has been described as the "Babylonian Confusion" of the IDN field [41].

We can already detect the formation of IDN design conventions on a pragmatic level, e.g. delayed consequences in many narrative games by Telltale, epitomized in the notification "Clementine will remember that" – in *The Walking Dead Game* [42]. The announcement of a delayed consequence is one way of making player agency more tangible. Players are reminded that their actions influence later points in the narrative, which can influence the further progression and outcome of the story.

Yet, while design conventions emerge over time, conceptually differing opinions exist in terms of categories and level of abstraction. This issue is visible in efforts to collect design patterns for video games by Kreimeier [43] and Björk/Holopainen [44]. As we have pointed out earlier these efforts are incompatible due to the use of different categories as we explained for the case of 'Paper Rock Scissors': "Kreimeier's patterns have the descriptors Name, Problem, Solution, Consequences, Example, Björk/Holopainen instead use Name, Description, Consequences, Using the Pattern and Relations" [19]. In addition, the high level of abstraction of both sets of design patterns creates a conceptual void between concrete examples and the descriptions.

In addition, several important questions have not yet found widely accepted answers. For example, when it comes to video games, the relationship between narrative design and other design methods, supposedly the ones that are focused on mechanics and rules, is unclear.

---

[1] BAs in "Game Writing"—University of Skövde, "Computer Games Design: Story Development "—University of the East of London, "Games Design and Creative Writing"—Brunel University, "Game Development: Writing"—Falmouth University, "Game Writing: B.A. Game Design Art "—Cogswell University (a further BA program at Bangor University is "not currently offered"), MA in "Digital and Interactive Storytelling" (focused on non-fiction)—University of Westminster.

While several influential books – which are widely used in education – have argued for a separation [45–47], others have argued for an integrated approach, e.g. Dubbelman [48]. This current state of affairs can be described as the lack of shared terminology combined with the lack of an accessible body of knowledge. The first issue is something the EU COST Action INDCOR is attempting to solve with its effort in creating an encyclopedia [49]. The second issue tempts narrative designers into falling back on design methods established in earlier media such as print literature and film—Murray's already mentioned "unproductive attempts to apply legacy conventions" [38]. This approach (e.g. [50]) also creates a problem by making IDN a derivative form. And here lies the danger: as long as IDN design relies on the methods of the novel or the movie, it will invite unfavorable comparisons to the original. Indeed, as Ian Bogost reminds us, "Video games are better without stories. Film, television, and literature all tell them better" [51]. Bogost's perspective betrays a widespread misunderstanding, as the purpose of IDN design is to create experiences that apply the specific potential of an interactive expression and not to convey stories in the manner of traditional media. IDN design does not tell, it provides opportunities to decide and experience. Bogost is therefore both right and wrong – correct in the assessment that conventional media are better in conveying conventional narratives – and incorrect in misunderstanding the purpose of IDN as a means to convey narratives in the same way traditional media dies. IDNs are different narratives that take advantage of procedurality and participation.

The creation of such interactive narrative experiences is challenging, as new design methods have to be invented and successfully implemented. In addition, there is the danger Murray alerts us about [38] in failing to exploit the expressive potential of interactive media and thus creating unsatisfying products. Conversely, Murray regards the "invention and refinement" [38] of design building blocks as a focus area for research in digital media as an expressive practice. Our approach to the problem is threefold: first by clearly distinguishing different levels of design categories in the form of higher-level design principles and lower-level building blocks/conventions [37]. Secondly, by using empirical methods to verify the effectiveness of particular design methods [19] and finally by collecting and promoting a shared body of design building blocks [52].

3.1 Establishing a design convention through evaluation

As a concrete example of the authors' approach of testing the effect of design decisions, this section presents two studies comparing different designs. Researchers need to develop their own prototypes or collaborate with designers. For our investigation we teamed up with the Dutch indie game company *Wispfire* which created the narrative adventure game Herald [53]. Wispfire won a grant to conduct 'artistic research for story-driven games' with the goal to increase the sense of player agency of their interactive narrative. Together with the authors of this paper, a research project was set up as an iterative process, basing study designs on the outcome of previous playtests. The developers wanted to give players an indication of the direction of the fictional revolution in the game. In this way, the designers hoped to provide players with an indication in which direction their choices led them. Initial reflections and design brainstorms led to the idea that player actions influence an in-game scale with the two factions of the game, colony and empire, on either side (see Fig. 1). The user interface (UI) in IDN is often a crucial factor for making players aware of their agency, to motivate them to act and to communicate the impact of their choices.

Within this study, our project group hypothesized that visual feedback in the UI gives players a stronger sense of perceived agency and invites them to consider their choices

Fig. 1 Implementation of the scale in the Herald evaluation prototype

more deeply. 55 participants played the prototype and filled out an extended version of Roth's measurement toolbox directly after. The goal of the study was to explore how well participants notice and interpret a scale at the top of the screen in the game. Results of the qualitative part of the study indicated that the meaning of the bar was unclear to players, resulting in widely varying statements. Players were unable to properly explain the direct and indirect consequences of moving the arrow on the bar. Based on these findings, a new study was designed, omitting the scale and introducing a gameplay element that visualizes player choices in the form of journal drawings.

To be able to compare the findings between both studies, the same scenes from the Herald game were used again with a similar sample. After each meaningful scene, player character Devan takes a moment to draw outcomes in his journal, thus creating a visual story of the player's progression. Blank parts indicate the choices the player didn't make (see Fig. 2).

As a result of the journal drawing addition, we hypothesized the following effects:

Players get an idea of all the possible ways in which playthroughs can differ.
Players can keep track of their progression and will be able to associate their earlier choices with later outcomes.
Journal drawings give players a sense of accomplishment.
Drawings show that every action has its consequences.

Fig. 2 Implementation of the journal drawing

Participants got the same questionnaire as in study 2, minus the questions regarding the scale and plus questions focusing on the drawings. Twenty-seven participants played the game and filled out the questionnaire directly after. Rating statements on a 7-point Likert scale, players did not perceive the journal drawings as interrupting the game. They felt a sense of achievement when unlocking new drawings and saw the drawings as clearly depicting their narrative choices.

We found that the sense of achievement coming from the drawings correlates with perceived global agency ($r = 0.388*$). Finally, we were interested in the different effects of the study 1 gameplay element "the scale" in comparison to the study 2 element "the journal drawings". Comparing both design approaches with the goal of the perception of agency, the implementation of journal drawings turned out to be more effective. All participants understood the meaning whereas we observed considerable struggles with the interpretation of the in-game scale. While the in-game scale element had not been recognized and completely understood by every participant, the journal drawings were not overlooked and clearly understood.

T-tests comparing the user experience dimensions of the two studies revealed that both local and global agency ratings were significantly higher for the prototype featuring the journal drawings. This result indicates that the drawings are a welcome way to make narrative agency more tangible with game flow and perceived autonomy being significantly higher in the study 2 condition. The drawings provided clear feedback on how the story progressed based on the deliberate player choices. From these results we can thus understand journal drawings as a candidate design convention for the design principle "agency feedback". Continuous use and verification would turn it into a design convention.

Different narrative mechanics such as the ones used in this example can be evaluated with the measurement tool to see their impact on the 12 user experience dimensions of Roth's toolbox. However, it is important to realize that narrative mechanics are interacting within a complex system. Evaluation of interactions between several variables requires intricate study designs and larger numbers of participants to possess the required statistical power.

4 Evaluating pro-social effects

This section focuses on the second topic of this paper, the question of evaluating the effectiveness of a particular work in terms of prosocial effects. The authors see this type of evaluation as related to the evaluation of design discussed earlier [19, 37] and in some cases, the two aspects might be closely related, e.g. when a particular design building block is meant to make audiences reflect on a decision or perspective. However, the difference can still be clearly expressed in terms of granularity, scope and temporal duration, as we take the question of evaluating the effectiveness to be about the cumulative effect of all design decisions taken together as a lasting and potentially delayed effect on interactors' perspective and behavior. The question of effectiveness can be reformulated thusly: 'What do interactors take away from the experience?'.

While considerable effort has been put into answering the question whether violence in video games spills over into the real world (it seems for the most part it does not, but transfers can happen on different levels, e.g. influencing dreams, see [54]), we still know little about prosocial effects. In 2018, van Riet et al. [55] observed the following:

> As Peng, Lee, and Heeter (2010) remarked, "Dozens of [...] games for social change have been developed and played by millions of people, [but] few empirical study ... has been conducted to evaluate how effective these games are" (pp. 723–724). Unfortunately, this observation is still true 8 years after it was made" (ibid)

Sadly, we do not see much progress in the last five years either. What we observed in our practice are five trends and neither of them fully address this problem. First, many studies in forms of IDN such as video games and related forms fail to meet the standards established in disciplines like psychology or learning sciences. We will describe many of these issues in the next section, including small sample sizes and incomplete execution (e.g. power not calculated).

Second, much effort is still focused on the direct experience of players (see [56] for an overview), but not on what audiences take away from the experience. Thus, our interest goes beyond user experience (UX) topics such as usability. Good UX and Interaction Design (IX) are prerequisites for impactful experiences with immediate and longer-term effects.

Third – and this issue is connected to the second one – there is a lack of longitudinal studies focusing on longer-term effects. This is a problem which needs to be addressed, because some initial evidence exists which points to delayed effects of interactive experiences, e.g. [57], where a stronger effect in the attitude towards homeless people was detected after three weeks.

Fourth, the considerable body of research which is concerned with the evaluation of interactive forms for educational purposes (see [58] for an overview), does not fully address our issue either. These efforts focus on clearly defined learning outcomes, most often in the setting of conventional educational institutions such as schools or universities. In contrast, what we are concerned with are effects on people outside an educational context. Instead of clearly defined learning outcomes, there are more loosely understood goals, similar to the intended effect of a newspaper article: to inform, to contextualize, to enable understanding and thus provide the basis for democratic discourse and prosocial behavior in democratic societies.

Finally, another problematic development is the focus on empathy in both design and evaluation, taking this quality as a decisive indication of success while ignoring a growing body of work which explains the considerable limitations of the concept [59, 60]. Indeed, this might be the underlying issue which negatively affects van t Riet et al.'s results [55] – that the game they evaluate is focused on creating an emotional reaction that can be evoked as well or better by traditional forms like documentaries.

Taken together, the above outlined five issues describe the reason why evidence for the effectiveness of interactive multimedia experiences as tools for improving democratic discourse and for affecting positive societal change is still scant. Our discussion also explains why there is no simple solution. To move forward, we need to gain a better understanding of what we are looking for and to overcome the 'empathy fallacy'. The question we need to answer is the following 'What qualities of IDN can create prosocial effects, with what design building blocks can we create these qualities and how can we reliably evaluate them?'.

4.1 Evaluation methods: foundational considerations

In this section, we first consider scientific methodologies toward the evaluation of the effects of IDN on its audiences. Both qualitative and quantitative methods have their uses and consequently, we recommend mixed-method approaches, using explicit, subjective (interviews, questionnaires) and implicit, objective data (physiological measurements, statistics from artifacts). Qualitative research provides an important starting point by identifying abstract design approaches and concrete implementations that can be verified in a next step within quantitative studies on larger samples.

Qualitative methods (content and design analysis, thematic analysis) can, for example, catalog specific interfaces and design properties [61], or reveal emotional qualities of specific design tropes [62]. Phenomenological approaches as well as auto-ethnographic methods (self-reflection and writing) deliver starting points for the investigation of effects. These initial approaches do not require a large number of participants to be insightful. Focus group interviews can be conducted with different target groups, with group sizes as small as four. Analyzing user reviews of IDNs, for example narrative games (e.g. on Steam, metacritic), can point out successes and flaws [63].

For quantitative evaluation, two main experimental setups exist: within-subject and between-subject. The within-subject setup exposes participants to different test conditions in sequence. To avoid possible sequential effects, the order of the conditions should be varied. In a between-subject setup, participants are randomly assigned to different test conditions. Data representing the user experience is usually acquired by means of questionnaires, often via validated scales measuring different experience dimensions, such as flow, presence, and enjoyment. The advantage of this study design is that it can be easily administered. Also, asking participants directly about their experience has high reliability and validity. Furthermore, administering a questionnaire after exposure is non-intrusive as it does not interrupt the participant's experience. For example, in their experiment comparing interactive vs. non-interactive narratives, Steinemann et al. [64] administered questionnaires to assess narrative engagement, identification, as well as enjoyment and appreciation.

That said, self-report questionnaires have come under increasing scrutiny. For example, the Game Experience Questionnaire (GEQ [65]), a scale intended to measure 7 distinct experiential dimensions, saw widespread use as a validated measure, despite the absence of a published validation study [66]. However, independent validation studies have identified issues regarding reliability, and could not replicate the stated 7-dimension factor structure, indicating limited validity [66]. These results showcase the need for more transparent reporting practices (e.g., whether a questionnaire has been statistically validated), as well as careful selection and validation of questionnaires to evaluate design conventions.

However, even validated post-hoc measurements come with disadvantages, chiefly the lack of information on temporal variations of the user experience. Yet, these are relevant for research in the effects of interactive narrative design, because good narratives, by their very nature, feature different pacing and thus elicit a range of affective responses over time. Indeed, regarding video game design, Pagulayan, et al. [67] remind us that the success of a play environment is determined by the process of playing, not its outcome. Post hoc questionnaires can ask participants to assess their experiences during gameplay, yet these experiences might be hard to recall in a precise manner. This is especially problematic for experiences lasting longer than half an hour. In addition, participants might go through phases of different experiences during exposure. For designers, it is often crucial to identify "unattractive" sequences. Schønau-Fog tries to address this issue by means of interrupting the gameplay for interactor feedback [68]. However, this kind of intrusive measurements can severely disrupt the experience, which is especially problematic when the researcher wants to obtain data about the effects of an experience, which is no longer the same if interrupted.

A promising approach is in 'diegetic measurements,' by which we mean a further hiding of the scientific aspect of the study and more seamless integration into the virtual world of the narrative game experience, e.g. by having players file a report from within the experience. Additionally, the tracking of interactor behavior and actions can be used for content analysis [69]. Physiological measurements during game experience are also a promising route [66], yet they pose additional problems by creating an abundance of data that can be difficult to interpret.

For our own studies we use Roth's measurement toolbox [70], which addresses a range of relevant user experiences in the context of narrative games with validated, distinct scales in post-hoc questionnaires. This framework enables the measurement of IDN user experience dimensions on a quantitative level and has been aligned with Murray's experiential qualities of agency, immersion, and transformation [71]. The measurement toolbox allows us to evaluate the IDN user experience on 12 dimensions, grouped under Murray's experiential qualities of agency, immersion, and transformation (see Fig. 3). For the purpose of understanding the overall effect of an IDN experience, transformation is of particular interest.

To experimentally verify the effects of a particular design we devise a predominantly quantitative approach that connects the creation of prototypes with the evaluation of user experience. We apply this approach to test the effectiveness of design strategies by comparing prototypes that differ only in one specific characteristic (A/B comparison) or one specific characteristic in different magnitudes (A/B/C/…).

As an example for the latter, we describe a study design to investigate the effect of different onboarding scenarios of the IDN *Breaking Points* [72] on the perceived player agency. We created three intro texts, increasingly Scripting the Interactor (StI) on their narrative impact (effectance). We differentiate between local effectance (showing direct impact of player choices) and global effectance (showing impact of choices in later scenes / delayed consequences).
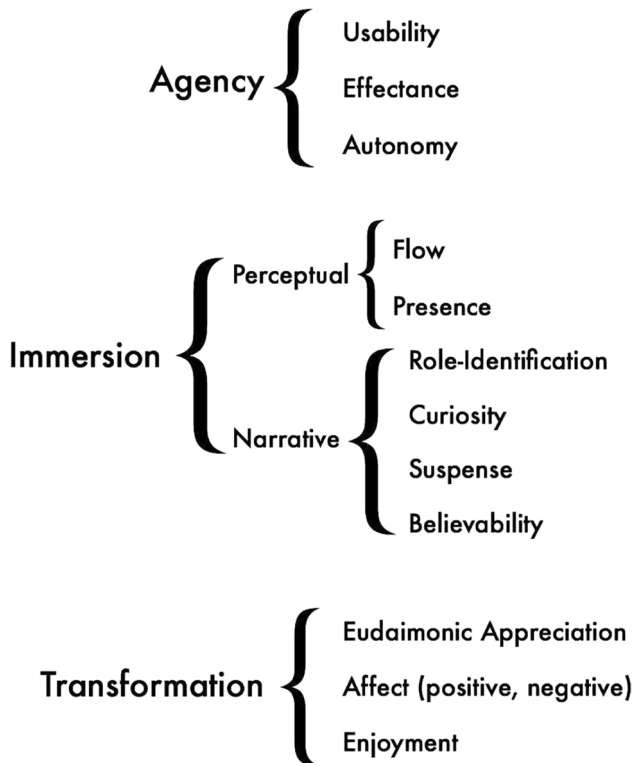

Fig. 3 User experience dimensions [70] mapped to Murray's taxonomy [27]

A—Basic text, no scripting on effectance (No direct StI).
B—Basic text plus specific scripting on effectance (Precise StI).
C—Basic text plus scripting on potential global effectance (Strong StI).
We posited the following theory-driven hypotheses:

H1: Priming the expectation of agency (local and global effectance) affects the user experience (intro A versus B and C)
H2: Strong StI (C) leads to a higher perception of global effectance than No StI (A) and Precise StI (B)

Participants were recruited via email and randomly distributed to the three online experiment conditions (A, B, C) of the narrative game. After interacting with the artifact for a variable time (usually 10 to 20 min), participants were guided to the online questionnaire based on Roth's measurement toolbox, presenting statements that must be agreed or disagreed to, on a 5-point Likert scale (quantitative measure). Additionally, participants were asked to freely write about positive and negative aspects of their experience (qualitative measure).

The most relevant insight from this study was that while we accepted H1, we had to reject H2. Players' perception of global agency didn't increase when they were being primed via the strong intro script. Interestingly, we saw an opposite effect: Exaggerated claims during the priming can significantly lower player identification with their interactor character compared to descriptions that were validated by the experience itself [73].

4.2 Best practices for user studies

Empirical methods can give us insights into the effects of an IDN experience. For user studies to be effective, several best practices should be followed. In this regard, Robinson et al. [74] offer crucial advice. In particular, we advocate to always start with a small pilot study to test the overall setup and see how participants react. This requires 5 to 10 participants per test condition. This might seem like an extra step, yet, in fact, it is less costly in terms of effort and time to make adjustments at this stage. In an exploratory study, we recommend researchers compute descriptive statistics, e.g., means, median, standard deviation, and visualize confidence intervals for each group.However, frequentist statistical approaches, such as t-tests or ANOVA, are not recommended for exploratory studies with small sample sizes as the pilot study mainly serves to test the study design itself.

In particular, researchers need to pay attention to the relationship between sample size and effect size. As we have argued before [19] effect size allows to move beyond the simplistic, 'Does it work or not?' to the far more sophisticated, 'How well does it work in a range of contexts?'" [75] (also see [76]). To compare effect sizes between studies with different participant sizes, one can use a specific unit that exists for this purpose, *Cohen's d*. [77]. The problem with low sample size is with regard to the power of the test.

Therefore, we strongly recommend calculating the power of a statistical test beforehand, given a particular effect size and sample size. For this purpose, one can use G*Power [78], a software package that can calculate the power of a statistical test, given a particular effect size and sample size.

Using this method can alert us to the danger of underpowered study designs [79] using small sizes of 25–30 participants per test condition, which can lead to Type I and Type II errors. Type I error, resulting in false positives stem from increase in bias and the likelihood of inflated effects based on chance. Type II errors happen due to the fact that small

sample sizes lack the statistical power to find significant effects even though a genuine effect exists in the population. It is crucial to understand that significance tests of p-value depend not only on the size of the effect but also on sample size. Significant results could exist but when the effect is rather small, they are only detectable in adequately large samples [80]. Since statistical significance does not by itself include information about the size of an effect, we encourage fellow scholars to always report effect sizes (e.g. Cohen's measures[2]), which measure the strength of a result and do not depend on sample size. For a more comprehensive (statistical) explanation, see the paper by Vornhagen et al. [81].

4.3 Additional factors and further investigation

An important determining factor for users' experiences is their prior engagement and frequency of contact with video games or other IDN experiences such as interactive documentaries. This pre-existing knowledge has a direct impact on the experience through genre expectations and familiarity with controls. A good example is the established convention of the keyboard's WASD keys as character movement controls in 3D environments with the computer mouse or trackpad as camera and aiming input. This convention has developed over time—initially, the standard inputs in 3D games were the arrow keys and camera control via the mouse was not available.

For example, a group of students working with us found the controls of *Firewatch* [7] as being too difficult to figure out for non-gamers in the time allotted for a study and thus decided to instead choose a different work, *Life is Strange* [82]. An unequal distribution in regard to computer literacy between experiment groups might influence the test results. To overcome this issue in studies with between-subject design and different experiment conditions, participants need to be randomly assigned to the conditions while monitoring the balance of computer literacy distribution, measured beforehand via self-report scales on experience and usage. Before any further analysis, mean values of computer literacy should be compared to guarantee no significant attribution differences between the groups.

Another factor that might limit the external validity of user experience studies is the use of student-produced IDN prototypes, which usually cannot compete with commercial games produced over the course of several years by large teams in professional production studios. Finally, play sessions are often rather short (5 to 20 min) and therefore might not represent typical experiences, which are often designed to take hours to complete. However, study setups with very long test sessions come with their own limitations: participants might lose focus over time, and post-hoc measurements are less reliable as participants will only remember parts of the experience.

When conducting user experience research, it is crucial to test with all target groups in mind. Samples should represent the population or a specific subgroup. However, user studies are often conducted with participants that are easily available, mainly students enrolled at the same university, instead of a more inclusive group needed for a representative outcome. More concretely, the use, for example, of game design students as subjects might not create a valid representation of the population and thus can result in limited external validity. Similarly, the validity of a study can also be limited due to societal and cultural factors. An overall design which might be effective in Western Europe might not work well

---

[2] Different statistical tests require specific effect size measures. For the comparison of group mean values, Cohen's d became a frequently used measure to estimate effect sizes and required sample sizes.

in Southeast Asia. Exploring these external effects requires more sophisticated and wide-ranging testing. In this vein, the call for decolonizing interactive experiences [83–86] is also a call for culturally aware and cross-culturally accessible design.

When it comes to results, the first question should be: do the results make sense? In this section, we offered some potential explanations for discrepancies. In general, replications of studies are needed to support concrete findings. Exact replications of an experiment need to operationalize both dependent and independent variables in exactly the same way as the original study (cf. [87]). At the same time, variations of studies are needed, as the analysis of effects need to be verified with different target audiences.

4.4 Case studies

In this section, we describe two case studies on the topic of prosocial behavior to demonstrate different scientific evaluation approaches. The first study represents an exploratory approach combining quantitative measures (questionnaire) and qualitative measures (focus group discussion). The second study applies a quantitative confirmatory design with a large sample allowing to compute Structural Equation Modeling (SEM) [88]. SEM is a statistical technique used to analyze complex relationships among variables. It incorporates both observed and latent (unobserved) variables, allowing researchers to test and estimate the strength and direction of causal relationships between variables. SEM provides a comprehensive framework for assessing theoretical models and is applied to fields such as social sciences, psychology, and economics.

4.4.1 Angstfabriek

This explorative user experience study is an example for interactive narratives in the form of participatory theatrical installations. The *Angstfabriek* (Dutch for fear factory) is an educational installation (Fig. 4) in the form of a complete building and lets visitors experience fearmongering and the related safety industry, with the goal of eliciting reflection, insight and discussion. The question Roth investigated in this case was whether an interactive narrative installation is able to offer a transformative learning experience.

Roth [89] evaluated the potentially transformative user experience via a focus group interview and a pilot user experience study ($N=32$). The research revealed the importance of sufficient

Fig. 4 Entrance of the Angstfabriek installation

scripting of visitors regarding their role and agency, highlighting the conceptual connection between interactive digital narrative design and interactive theater design. This case study showcases the value of an exploratory qualitative approach by means of observation and a focus-group interview, followed by a survey consisting of the user experience dimensions from the aforementioned measurement toolbox, combining qualitative and quantitative measurements.

Onboarding is an important factor when designing roleplaying experiences. Previous research asserts that for successful perspective-taking in this context, participants need to receive sufficient background information and scripting [90]. In the case of Angstfabriek, participants are first primed about positive aspects of fear by different means (scanning for fears, VR simulation with fear inducing topics, talking with the director) using a narrative motif of "fear keeps us safe". However, just when visitors think they would leave, they are approached by an actor playing a dissident worker who wants to reveal what is going on behind the scenes. To support this change in perspective, participants are then invited to wear work jackets to investigate the factory undercover, thus changing their role from visitor to undercover employee. Their first task is the production of fear-inducing media messages. Roth found that 12 of the 32 visitors sabotaged this task by not complying with the assignment during the fear media creation. However, the majority were not aware of their changed role with increased agency to interrupt the process. This agency can range from subtle resistance to more extreme measures, like pulling the plug of the different machines, completely stopping the production. No one from this sample or the focus group made use of more disruptive actions by sabotaging the factory itself and it is striking that the first priming was so effective that more than half of the visitors stayed compliant. An explanation for this behavior comes from both the survey (regarding role-believability) and the focus group interviews, which indicate that the role of whistleblower remained underdeveloped and mainly determined by the visitors' imagination. Similarly, reflection on the experience was not fostered by design and it was up to the visitors to meet and discuss their experience afterwards.

This evaluation of *Angstfabriek* helped to identify deficiencies in the design of the experience, in particular in when it comes scripting interactors for the role change of dissident employee. This issue undermined the transformative potential of the experience. The chosen evaluation approach, featuring a focus group of experts and post-hoc questionnaires for a small sample of visitors delivered initial insights, forming the basis for controlled experiments, like the study design investigating the influence of different scripts described in 4.1.

In terms of design conventions, we identified the need for a particular design convention related to role-change, as it is evident that the existing "scripting of the interactor" [27] by means of the dissident worker's speech and the change of clothing was insufficient. We speculate that a more drastic intervention is needed, an "opposite role change intervention" convention, for example by making visitors directly the target of fear mongering before the role change.

4.4.2 How I became homeless

Finally, we outline an example of a confirmatory study to investigate whether an interactive news post can increase prosocial behavior. Based on theoretical considerations around the processes and outcomes of interactive narratives [91], Steinemann et al. [64] hypothesized that a player-directed version of an online news post recounting the experiences of a single parent becoming unexpectedly homeless, would encourage players to donate significantly more than the non-interactive text. Specifically, the authors hypothesized that interactivity would increase prosocial behavior via increasing identification, responsibility (i.e., agency) and appreciation (see Fig. 5).
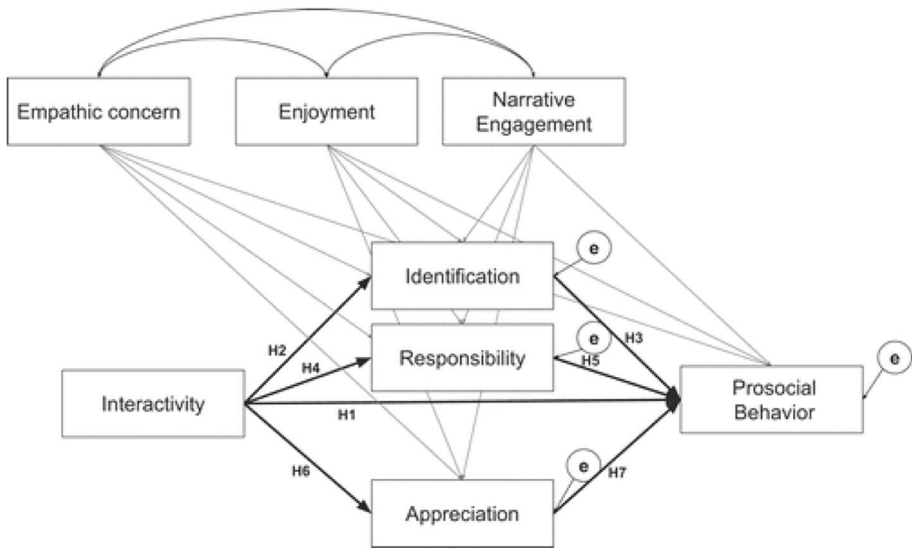
Fig. 5 A model of the expected processes between interactivity and prosocial behavior. Lines in bold indicate hypotheses-relevant pathways. From Steinemann et al. [64]

Based on a statistical power analysis, over 600 participants were recruited and randomly assigned to engage with either the interactive or non-interactive narrative. Afterwards, they were asked to rate their experience and indicate how much of their study compensation they wished to donate to a charity supporting homeless people. Given the complexity of the expected processes between interactivity and user behavior, the authors opted to analyze the data via structural equation modeling rather than ANOVAs. Against their expectations, interactivity did not increase donations directly. Nevertheless, the hypothesized processes were partially confirmed: Interactivity increased participants' sense of responsibility for the narrative outcome, whereas increased appreciation was linked to a significantly higher percentage of money donated. These findings emphasize the need for confirmatory research to verify theoretical claims and question established (yet often unchallenged) assumptions. The result also further draws attention to the need of developing and verifying design conventions, as interactivity needs to be understood as a "raw power," which requires skilled and targeted application to be effective for a particular purpose.

To summarize, in Section 4, we considered issues and challenges in evaluating prosocial effects of IDN artifacts. We first identified a lack of evidence for prosocial effects which we related to issues in the design and execution of studies. We then proposed strategies for improved study design and execution and drew attention to further issues and possible solutions. Finally, we described several case studies displaying best practices.

5 Future work

We are keenly aware of the challenges involved in proper evaluation due to both the required effort in setting up user studies and the involved costs, both in terms of time spent on the effort and in recruiting participants. However, the benefits in terms of much improved user

experience and satisfaction of clients by means of verified effects make this effort worthwhile. As an important step in this direction, we suggest increasing collaborations between media creators and scholars. The latter can contribute expertise in evaluation and also bring in participants, often students for credit points, while the former provide expertise from the practice and the important corrective of the academic through real-world application. As a concrete measure, we propose to create a global participant pool for online studies to further test the influence of different design approaches and evaluation methods. Here, scholarly and professional organization should take leading roles and work together, for example ARDIN (Association for Research in Digital Interactive Narrative. http://ardin.online) and IGDA (International Game Developers Association. https://igda.org).

The rigorous scientific approach we describe allow researchers to conduct meta-analyses, which are quantitative reviews analyzing the results of multiple scientific studies investigating the same research question. In the context of IDN design research, a meta-analysis could for example investigate a series of studies regarding the same design convention candidate. As every method has its advantages and disadvantages, we recommend a combination of established research methods to reach a more balanced perspective. Multi-method analysis (MMA) [92] is an approach which proposes to combine several methods at the same time. A practical application of MMA is in the form of MMAJams, which describes an effort of several researchers coming together to analyze an artifact in a concentrated and time-limited format, similar to video game jams, which are time-limited competitions for the creation of video games. The advantage of this approach is that it can produce results in a matter of hours.

While MMA can mitigate the limitations of particular methods, there is still a need for increased scrutiny of the methods frequently applied for evaluation of IDN design artifacts such as Flow theory [93] or Self Determination Theory [94, 95] in line with recent criticism [96, 97] in the wake of the replication crisis [98, 99]. Given the well-argued criticism, we see the need for further evaluation method development, with a particular emphasis on application also for designers and on rigorous replication.

6 Conclusion

In this paper, we have discussed interactive digital narratives as works that create multimedia alternate realities even within a single artifact and can thus be used to represent multi-faceted, complex issues. However, there are at least two salient issues that prevent a more widespread application of IDN – the lack of a shared body of design knowledge and the low number of empirical studies verifying the intended prosocial effects. Our paper addresses these issues by discussing evaluation methods as a means to identify design conventions and to verify prosocial effects. Our aim is to 1) increase awareness for the need to improve existing practices in evaluating IDN artifacts, 2) point out the benefits of evaluation for designers, 3) provide guidance to implement best practices in evaluating, and 4) offer a starting point for developing improved evaluation methods. In this way, we aim to engage the IDN-related community more in evaluation and evaluation method development. Furthermore, we hope that our methodological approach to makes the expressive and prosocial potential of IDN—and with it the aspect of alternate realities—more accessible to media organizations and individual creators.

Data availability The datasets generated during and/or analysed during the current study are available in a GoogleDrive repository: https://drive.google.com/drive/folders/1dDhcDuBGLVwjxaQQ1mqb5-CaMYfz6DIA?usp=sharing

Declarations

Conflicts of interests/Competing interests. The authors have no relevant financial or non-financial interests to disclose.

References

1. Koenitz H (2023) Understanding Interactive Digital Narrative: Immersive Expressions for a Complex Time. Routledge, London and New York https://doi.org/10.4324/9781003106425
2. Submarine Channel (2014) The Last Hijack Interactive. https://lasthijack.submarinechannel.com
3. Quantic Dream (2018) Detroit: Become Human. https://www.quanticdream.com/en/detroit-become-human
4. The Chinese Room (2008) Dear Esther. https://www.thechineseroom.co.uk/games/dear-esther
5. The Fullbright Company (2013) Gone Home. https://gonehome.com
6. Wreden, D (2013) The Stanley Parable. https://www.stanleyparable.com
7. Campo Santo (2016) Firewatch. http://www.firewatchgame.com
8. Night School Studio (2016) Oxenfree. https://nightschoolstudio.com/oxenfree/
9. Naughty Dog (2014) The Last of Us. https://www.playstation.com/en-us/games/the-last-of-us-part-i/
10. Naughty Dog (2020) The Last of Us 2. https://www.playstation.com/en-us/games/the-last-of-us-part-ii-remastered/
11. Chance Agency (2019) Neo Cab. https://neocabgame.com
12. Die Gute Fabrik (2019) Mutazione. http://mutazionegame.com
13. Witch Beam (2021) Unpacking. https://www.unpackinggame.com
14. David Dufresne (2013) Fort McMoney. National Film Board of Canada. https://www.nfb.ca/interactive/fort_mcmoney_en
15. Submarine Channel (2018) The Industry. https://theindustryinteractive.com/home
16. Arora G, Pousman B, Milk C (2015) Clouds over Sidra. https://www.unicefusa.org/stories/clouds-over-sidra-award-winning-virtual-reality-experience/29675. Accessed 10 Jan 2024
17. Kors, MJL, Ferri, G, Van der Spek, ED, Ketel, C, Schouten, BAM (2016) A breathtaking journey. On the design of an empathy-arousing mixed-reality game. In: CHI PLAY. pp. 91–104. ACM, New York, NY, USA https://doi.org/10.1145/2967934.2968110
18. Anagram (2021) Goliath. https://weareanagram.co.uk/project/goliath
19. Koenitz, H, Roth, C, Knoller, N, Dubbelman, T (2018) Clementine will remember that – methods to establish design conventions for video game narrative. In: Proceedings of the 2018 DiGRA international conference the game is the message, Torino, Italy
20. Ricoeur P (1991) Narrative Identity. Philosophy Today 35:73–81. https://doi.org/10.5840/philtoday199135136
21. Bruner J (1991) The narrative construction of reality. Crit Inq 18:1–21. https://doi.org/10.1086/448619

22. Barthes R, Duisit L (1975) An introduction to the structural analysis of narrative. New Literary History 6:237. https://doi.org/10.2307/468419
23. Herman D (2002) Story logic. U of Nebraska Press, Lincoln, NE
24. Laurel B (1986) Toward the design of a computer-based interactive fantasy system. http://books.google.de/books?id=sNMTAQAAIAAJ&q=intitle:Toward+the+design+of+a+computer+based+interactive+fantasy+system&dq=intitle:Toward+the+design+of+a+computer+based+interactive+fantasy+system&hl=&cd=1&source=gbs_api. Accessed 10 Jan 2024
25. Joyce, M (1995) Of two minds. University of Michigan Press https://doi.org/10.3998/mpub.10599
26. Jennings P (1996) Narrative structures for new media. Leonardo 29:345–350
27. Murray JH (1997) Hamlet on the holodeck: the future of narrative in cyberspace. Free Press, New York
28. Ascott, R (1964) The construction of change. Cambridge Opinion
29. Ascott R (1968) The cybernetic stance: My process and purpose. Leonardo 1:105. https://doi.org/10.2307/1571947
30. Eladhari, MP (2018) Re-tellings: The fourth layer of narrative as an instrument for critique. In: Interactive storytelling: 11th international conference for interactive digital storytelling, ICIDS 2018. pp. 65–78. Springer Berlin Heidelberg, Cham. https://doi.org/10.1007/978-3-030-04028-4_5
31. Roth, C, van Nuenen, T, Koenitz, H (2018) Ludonarrative hermeneutics: A way out and the narrative paradox. In: Rouse, R., Koenitz, H., and Haahr, M. (eds.) Interactive storytelling: 11th international conference for interactive digital storytelling, ICIDS 2018. pp. 93–106. The 3rd International Conference for Interactive Digital Storytelling, Cham
32. Koenitz H (2015) Towards a specific theory of interactive digital narrative. In: Koenitz H, Ferri G, Haahr M, Sezen TI (eds) Interactive digital narrative. Routledge, New York, pp 91–105
33. Karhulahti, V-M (2012) Double fine adventure and the double hermeneutic videogame. In: Foundations of digital games. pp. 19–26. ACM Press, New York, New York, USA https://doi.org/10.1145/2367616.2367619
34. Giddens, A (1987) Social theory and modern sociology. Stanford University Press
35. Koenitz, H, Barbara, J, Eladhari, MP (2023) Interactive digital narrative (IDN)—new ways to represent complexity and facilitate digitally empowered citizens. New Review of Hypermedia and Multimedia. 0, 1–21. https://doi.org/10.1080/13614568.2023.2181503
36. Koenitz, H (2015) Design approaches for interactive digital narrative. In: Schoenau-Fog, H., Bruni, L.E., Louchart, S., and Baceviciute, S. (eds.) Interactive storytelling 8th international conference on interactive digital storytelling, ICIDS 2015. Springer, Cham https://doi.org/10.1007/978-3-319-27036-4_5
37. Koenitz, H, Roth, C, Dubbelman, T, Knoller, N (2017) What is a convention in interactive narrative design? In: Nunes, N., Oakley, I., and Nisi, V. (eds.) Interactive storytelling: 10th international conference on interactive digital storytelling, ICIDS 2017. pp. 295–298. Springer International, Cham
38. Murray JH (2011) Inventing the medium : principles of interaction design as a cultural practice. MIT Press, Cambridge, Mass
39. Weizenbaum J (1966) Eliza — a computer program for the study of natural language communication between man and machine. Commun ACM 9:36–45. https://doi.org/10.1145/365153.365168
40. Nunes, N, Oakley, I, Nisi, V eds (2017) Grimes' fairy tales: A 1960s story generator. Springer International Publishing, Cham https://doi.org/10.1007/978-3-319-71027-3_8
41. Koenitz, H, Eladhari, MP (2019) Challenges of IDN research and teaching. In: Cardona-Rivera, R.E., Sullivan, A., and Young, R.M. (eds.) Interactive storytelling: 12th international conference on interactive digital storytelling, ICIDS 2019. pp. 26–39. Springer Nature, Cham https://doi.org/10.1007/978-3-030-33894-7_4
42. Telltale Games (2012) The walking dead. https://www.telltale.com/
43. Kreimeier B (n.d.) The case for game design patterns. http://www.gamasutra.com/features/20020313/kreimeier_01.htm. Accessed 10 Jan 2024
44. Björk, S, Holopainen, J (2004) Patterns in game design (game development series). Charles River Media
45. Juul, J (2005) Half-real. MIT Press
46. Schell J (2008) The art of game design: a book of lenses. Elsevier/Morgan Kaufmann Amsterdam, Boston
47. Fullerton T, Swain C, Hoffman SS (2008) Game design workshop. Morgan Kaufmann, Burlington, MA
48. Dubbelman, T (2016) Narrative game mechanics. In: Nack, F. and Gordon, A.S. (eds.) Interactive storytelling 9th international conference on interactive digital storytelling, ICIDS 2016. pp. 39–50. Springer International Publishing https://doi.org/10.1007/978-3-319-48279-8_4
49. Koenitz, H, Eladhari, MP, Louchart, S, Nack, F, Roth, C, Mekler, E, Makai, PK (2022) A Shared Vocabulary for Interactive Digital Narrative (IDN): An Encyclopedia Project. In: Bostan, B. (ed.)

Games and Narrative: Theory and Practice. pp. 207–229. Springer International Publishing, Cham https://doi.org/10.1007/978-3-030-81538-7_14

50. LeRay L (n.d.) Developing A cat's manor in saudi arabia for a western audience. https://www.gamasutra.com/view/news/285904/Developing_A_Cats_Manor_in_Saudi_Arabia_for_a_Western_Audience.php. Accessed 10 Jan 2024

51. Bogost I (n.d.) Video games are better without stories. https://www.theatlantic.com/technology/archive/2017/04/video-games-stories/524148/?utm_source=atlfb. Accessed 10 Jan 2024

52. Koenitz H, Roth C, Dubbelman T (2018) Engaging the community in collecting interactive narrative design conventions. CHI PLAY. https://doi.org/10.1145/3270316.3271533

53. van den Schilden R (2017) Herald: an interactive period drama. http://heraldgame.com

54. Roth, C (2019) Video Game Violence from the Perspective of Cognitive Psychology. Role Identification and Role Distancing in A WAY OUT. In: Alvarez Igarzábal, F., Debus, M.S., and Maughan, C.L. (eds.) Violence | Perception | Video Games. pp. 53–62. transcript Verlag https://doi.org/10.1515/9783839450512-006

55. van t Riet J, Meeuwes AC, van der Voorden L, Jansz J (2018) Investigating the effects of a persuasive digital game on immersion, identification, and willingness to help. Basic Appl Soc Psychol 40:180–194

56. Daoudi I (2022) Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda. Educ Inf Technol. https://doi.org/10.1007/s10639-022-11087-4

57. Ruggiero D (2015) The effect of a persuasive social impact game on affective learning and attitude. Comput Hum Behav 45:213–221. https://doi.org/10.1016/j.chb.2014.11.062

58. Vlachopoulos D, Makri A (2017) The effect of games and simulations on higher education: a systematic literature review. Int J Educ Technol Higher Educ 14:22. https://doi.org/10.1186/s41239-017-0062-1

59. Bloom, P (2018) Against empathy. Vintage

60. Rouse, R (2021) Against the instrumentalization of empathy: Immersive technologies and social change. In: Augmented and mixed reality for communities. pp. 3–19. , Boca Raton, FL

61. Toups, ZO, Lalone, N, Alharthi, SA, Sharma, H (2019) Making maps available for play: Analyzing the design of game cartography interfaces. ACM Trans Comput-Hum Interact 26 https://doi.org/10.1145/3336144

62. Bopp, JA, Müller, LJ, Aeschbach, LF, Opwis, K (2019) Exploring emotional attachment to game characters. In: CHIPlay https://doi.org/10.1145/3311350.3347169

63. Cole, T, Cairns, P, Gillies, M (2015) Emotional and functional challenge in core and avant-garde games. In: the 2015 annual symposium. pp. 121–126. ACM Press, New York, New York, USA https://doi.org/10.1145/2793107.2793147

64. Steinemann ST, Iten GH, Opwis K, Forde SF, Frasseck L, Mekler ED (2017) Interactive narratives affecting social change. J Media Psychol 29:54–66. https://doi.org/10.1027/1864-1105/a000211

65. IJsselsteijn, W, Poels, K, de Kort, Y (2008) The Game Experience Questionnaire: Development of a self-report measure to assess player experiences of digital games. TU Eindhoven

66. Law, E, Brühlmann, F, Mekler, E (2018) Systematic review and validation of the game experience questionnaire (geq)-implications for citation and reporting practice. CHIPlay 2020 https://doi.org/10.1145/3242671.3242683

67. Pagulayan, R, Keeker, K, Romero, R, Wixon, D, Fuller, T (2009) User-centered design in games. In: Human–Computer interaction handbook. pp. 741–759. CRC Press https://doi.org/10.1201/9781410615862.ch37

68. Schønau-Fog H (2012) Sure, I would like to continue. Bull Sci Technol Soc 32:405–412. https://doi.org/10.1177/0270467612469068

69. Roth, C, Vermeulen, I (2013) Breaching Interactive Storytelling's Implicit Agreement: A Content Analysis of Façade User Behaviors. In: Koenitz, H., Sezen, T.I., Ferri, G., Haahr, M., Sezen, D., and Çatak, G. (eds.) Interactive Storytelling. pp. 168–173. Springer International Publishing, Cham https://doi.org/10.1007/978-3-319-02756-2_20

70. Roth C (2016) Experiencing interactive storytelling. PhD Thesis, VU University Amsterdam. https://research.vu.nl/en/publications/experiencing-interactive-storytelling. Accessed 10 Jan 2024

71. Roth, C, Koenitz, H (2016) Evaluating the user experience of interactive digital narrative. In: the 1st international workshop. pp. 31–36. ACM Press, New York, New York, USA https://doi.org/10.1145/2983298.2983302

72. Koenitz, H, Sezen, TI, Sezen, D (2013) Breaking Points A Continuously Developing Interactive Digital Narrative. In: Koenitz, H., Sezen, T.I., Ferri, G., Haahr, M., Sezen, D., and Çatak, G. (eds.) Interactive Storytelling. pp. 107–113. Springer International Publishing, Cham

73. Roth, C, Koenitz, H (2017) Towards creating a body of evidence-based interactive digital narrative design knowledge: Approaches and challenges. In: AltMM'17. , Mountain View, CA https://doi.org/10.1145/3132361.3133942

74. Robinson, R, Murray, J, Isbister, K (2018) You're Giving Me Mixed Signals! In: the 2018 CHI conference. pp. 1–6. ACM, New York, NY, USA https://doi.org/10.1145/3170427.3188469

75. Coe, R (2002) It's the effect size, stupid: what effect size is and why it is important

76. Sullivan GM, Feinn R (2012) Using effect Size—or why the PValue is not enough. J Grad Med Educ 4:279–282. https://doi.org/10.4300/JGME-D-12-00156.1

77. Cohen J (1992) A power primer. Psychol Bull 112:155–159. https://doi.org/10.1037/0033-2909.112.1.155

78. Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191. https://doi.org/10.3758/BF03193146

79. Bakker M, van Dijk A, Wicherts JM (2012) The rules of the game called psychological science. Perspect Psychol Sci 7:543–554. https://doi.org/10.1177/1745691612459060

80. Cohen J (1995) The earth is round (p ¡ .05): Rejoinder. Am Psychol 50:1103–1103. https://doi.org/10.1037/0003-066X.50.12.1103

81. Vornhagen, JB, Tyack, A, Mekler, E (2020) Statistical significance testing at chi play: Challenges and opportunities for more transparency. CHIPlay 2020 https://doi.org/10.1145/3410404.3414229

82. Dontnod Entertainment (2015) Life is strange. https://dont-nod.com/en/games/life-is-strange/

83. Mukherjee S (2018) Playing Subaltern: Video Games and Postcolonialism. Games and Culture 13:504–520. https://doi.org/10.1177/1555412015627258

84. de Paula, B (2021) Decolonizing Game Literacy. In: Global Citizenship for Adult Education. pp. 325–333. Routledge https://doi.org/10.4324/9781003050421-41

85. Trammell A (2022) Decolonizing play. Crit Stud Media Commun 39:239–246. https://doi.org/10.1080/15295036.2022.2080844

86. Silva, C, Reyes, MC, Koenitz, H (2022) Towards a Decolonial Framework for IDN. In: Vosmeer, M. and Holloway-Attaway, L. (eds.) Interactive Storytelling. pp. 193–205. Springer International Publishing, Cham https://doi.org/10.1007/978-3-031-22298-6_12

87. Stroebe W, Strack F (2014) The alleged crisis and the illusion of exact replication. Perspect Psychol Sci 9:59–71. https://doi.org/10.1177/1745691613514450

88. Kline, RB (2023) Principles and practice of structural equation modeling. Guilford publications

89. Roth, C (2019) The 'Angstfabriek' experience: Factoring fear into transformative interactive narrative design. In: Interactive storytelling. pp. 101–114. Springer, Cham https://doi.org/10.1007/978-3-030-33894-7_11

90. Downey MT (1995) Doing history in a fifth-grade classroom: Perspective taking and historical thinking. Presented at the American Educational Research Association (AERA) Annual Meeting, San Francisco

91. Green MC, Jenkins KM (2014) Interactive narratives: Processes and outcomes in user-directed stories. J Commun 64:479–500. https://doi.org/10.1111/jcom.12093

92. Eladhari MP, Koenitz H (2020) MMAJams-Multi-Method analysis of games in research and education. In: DiGRA '20 – Proceedings of the 2020 DiGRA International Conference: Play Everywhere

93. Csikszentmihalyi M (1990) Flow: the psychology of optimal experience. Harper & Row, New York

94. Deci, EL (1984) The Psychology of Self-Determination. Free Press

95. Deci, E (1992) The relation of interest to the motivation of behavior: A self-determination theory perspective. In: Renninger, K.A., Hidi, S., and Krapp, A. (eds.) The Role of Interest in Learning and Development. L. Erlbaum Associates

96. Tyack, A, Mekler, ED (2020) Self-Determination Theory in HCI Games Research: Current Uses and Open Questions. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–22. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3313831.3376723

97. Deterding, S, Cutting, J (2023) Objective Difficulty-Skill Balance Impacts Perceived Balance but Not Behaviour: A Test of Flow and Self-Determination Theory Predictions. Proc. ACM Hum.-Comput. Interact. 7, 419:1179–419:1205 https://doi.org/10.1145/3611065

98. Shrout PE, Rodgers JL (2018) Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. Annu Rev Psychol 69:487–510. https://doi.org/10.1146/annurev-psych-122216-011845

99. Echtler, F, Häußler, M (2018) Open Source, Open Science, and the Replication Crisis in HCI. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–8. Association for Computing Machinery, New York, NY, USA https://doi.org/10.1145/3170427.3188395

Authors and Affiliations

Hartmut Koenitz[1] · **Christian Roth**[2] · **Elisa Mekler**[3]

✉ Hartmut Koenitz
hartmut.koenitz@sh.se

Christian Roth
christian.roth@hku.nl

Elisa Mekler
elme@itu.dk

1   Södertörn University, Huddinge, Sweden

2   HKU University of the Arts Utrecht, Utrecht, Netherlands

3   IT-University of Copenhagen, Copenhagen, Denmark