

Narrative Gold Digging: Sifting and Narrating Stories from a Procedural Simulation

Jonathan Lessard
Concordia University
Montreal, Quebec, Canada
jonathan.lessard@concordia.ca

Samuel Paré-Chouinard
Concordia University
Montreal, Quebec, Canada
spchouinard@gmail.com

Stephen Friedrich
Concordia University
Montreal, Quebec, Canada
stephen.friedrich@mail.concordia.ca

Leila Kosseim
Concordia University
Montreal, Quebec, Canada
leila.kosseim@concordia.ca

Abstract

This paper reports on an experiment in formalizing and implementing “story-sifting” heuristics to automatically identify likely “interesting stories” from a large mass of events transpiring from the *Chroniqueur* procedural social simulation. Following preliminary experimentation with various concepts inspired by narratology, we proceeded to formally evaluate “unexpectedness” and “dramatic situations” as key heuristics to sample six stories from a simulation run and rendered each as two short stories, one written by a human author and the other by a LLM (in order to help distinguish the effect of narrative from story). These were scored by 24 participants for narrative appreciation. Inconclusive results about heuristics making any significant difference prompted self-reflection on what makes emergent narratives interesting (in a distinct way from traditional fiction), pointing both at “story effects” (some selected event chains constituting better story material) and “narrative effect” (better narrativization being able to cope with poor story material).

CCS Concepts

• **Applied computing** → Arts and humanities; Media arts; • **Computing methodologies** → Modeling and simulation; Simulation evaluation.

Keywords

Story sifting, Emergent Narrative, Narratology, Social Simulation

ACM Reference Format:

Jonathan Lessard, Stephen Friedrich, Samuel Paré-Chouinard, and Leila Kosseim. 2026. Narrative Gold Digging: Sifting and Narrating Stories from a Procedural Simulation. In *Foundations of Digital Games (FDG '26)*, August 10–13, 2026, Copenhagen, Denmark. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3815598.3815687>

1 Introduction

In his 2018 PhD thesis on emergent narrative, James Ryan [15] proposed the “overgenerate and curate” approach in which potentially interesting stories are identified after the fact amongst the

mass of events produced by a social (agent-based) simulation. He named this process “story-sifting”, evoking the image of the gold prospector sieving large quantities of sand (banal simulated events) looking for valuable nuggets. Once identified, these outstanding events could be “narrativized”, that is, rendered discursively or visually for an audience to appreciate. This contribution was mostly theoretical, and the technicalities of its actualization were left for future research.

The *Chroniqueur* project was, amongst other aims, an attempt to fully realize this pipeline: a complex social simulation generates a large quantity of coherent, potentially narrative events; story-sifting heuristics are then used to search this data for potentially “tellable” [1] causal chains; which can ultimately be narrativized as readable short-stories.

While we cannot claim in any way to have “solved” this set of very complex problems at the threshold of computation and narratology, we have gathered insights (both positive and negative) that we hope might be useful to fellow designers and developers. We will briefly present the *Chroniqueur* procedural social simulation system, our experiments with narratology-inspired story-sifting heuristics, and the results of a user evaluation test of sifted stories in the form of narrativized short stories. The unsatisfying results lead to a critical re-examination of our assumptions, highlighting some interesting lessons to be learned from these efforts.

2 Relevant Work

James Ryan [15] first uses the term “story sifter” to describe a system by which meaningful stories are identified amongst the heap of events taking place in simulated worlds. This theoretical proposition sparked significant research in the past years which mostly focused on developing tools to allow users to “story sift” using specialized queries [8, 9] sketched dramatic arcs [10], or visual interfaces [6, 11, 17].

Fewer attempts were made to devise “absolute” story sifting methods independent from specific user intent. Researchers Méndez and Gervás experimented with ChatGPT, but this could only process a small number of events at a time and the authors were not convinced by the results [4]. Méndez and Gervás more recently developed an evolutionary approach to assess narrative interest of simulational logs [3]. While promising, the approach would not in the current state scale to the very large event-space of project *Chroniqueur*.



This work is licensed under a Creative Commons Attribution 4.0 International License. *FDG '26, Copenhagen, Denmark*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2495-4/2026/08
<https://doi.org/10.1145/3815598.3815687>

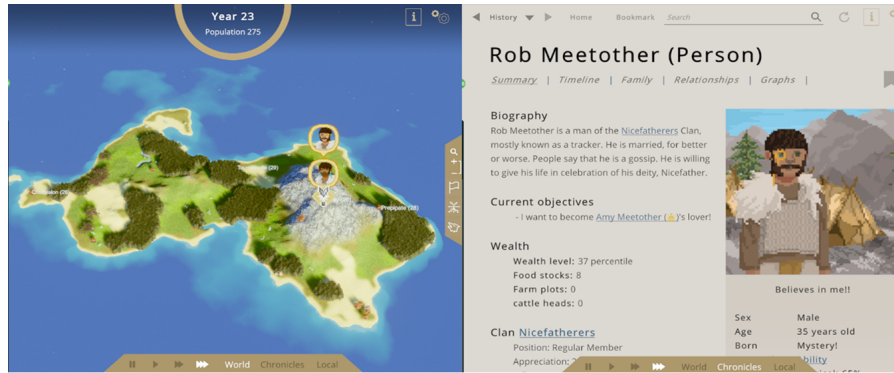


Figure 1: : Screenshots from *Chroniqueur*

With few exceptions, story-sifting research has primarily evaluated algorithmic relevance and interface usability rather than reader-perceived narrative interest. The main methodological precedent to the research presented here is Kreminski *et al*'s 2022 paper "Select the Unexpected: A Statistical Heuristic for Story Sifting" [7]. While the story-sifting framework is still a user-defined pattern-based search, the authors propose ranking the results according to "unexpectedness" as a heuristic of narrative potential. They asked users to rank selected event logs manually to see whether their choice would confirm the prediction that less probable items would be more interesting (it did to some extent). We are also interested in testing "absolute" heuristics (including unexpectedness) and have actual readers evaluate the outcomes. However, rather than ranking system log events, we opted to render these events as full narratives for a more natural evaluation.

Beyond theoretical and methodological precedents, it should be considered that every fictional social simulation has its own specificities (both in design and content) making each experiment hard to generalize. This is probably a domain where similar experiments need to be done in the context of different "worlds" in order to see clearer patterns and good practices.

3 The Chroniqueur system

The *Chroniqueur* system was a research-creation project developed from 2020-2024 by the LabLabLab research lab to investigate the design of complex social simulations with the specific intent of affording rich emergent narrative. It was inspired by James Ryan's work [15] as well as landmark games such as *Dwarf Fortress*, *The Sims* or *Crusader Kings*. It led to scientific publications [11, 13] and was presented in multiple venues as creative work.

While *Chroniqueur* does many things, at its center is an agent-based social simulation that attempted to procedurally model the neolithic revolution: as scattered groups of hunter-gatherers progressively "fill up" a finite procedurally generated island, the increased scarcity of resources available leads to the progressive adoption of work-intensive animal and crop domestication, social stratification and urbanization. While the system was never "neolithic complete" so to say, it did reasonably implement what Sahlins termed "stone age economics" [16]. As we had observed that a "means of living" simulation did not consistently lead to narratively rich events (in

fact, rarely!), we also implemented the possibility for "dramatic situations" to occur, taking inspiration from playwright George Polti [14].

Relevant to this paper is that the *Chroniqueur* system can procedurally simulate hundreds of years of "human-like" history; recording along the way all occurring events, the agents involved, and their effects.

4 Story-Sifting Heuristics

In their 2022 paper, Kreminski *et al.* took inspiration from narratology to devise their "select the unexpected" heuristic for narrative potential [7]. We took a similar approach and considered various narratological concepts which could inspire absolute story sifting heuristics.

4.1 Eventfulness

How do we choose what to say when someone asks "how was your day"? Our days are filled with events and actions, some ours, many witnessed, and yet we quickly mentally "sift" through this to latch on to things worth telling: the slipping on a banana bit will make an anecdote, the routine of putting on socks, not. What are our search heuristics? Narratology has some leads through the concepts of "tellability" [1]: what is the quality of something worth telling, or "eventfulness" [5]: what makes an Event (with a capital "e") the stuff of stories. The latter is particularly interesting as our simulation (like many others) is event driven. Procedures that would allow "scoring" events in terms of "eventfulness" would hypothetically bring us closer to tellability.

A model that stood out for us was Schmid's criteria of eventfulness [18] as they represented concrete factors to attempt evaluating our events. We have attempted to implement each of these, with various levels of success.

4.1.1 Relevance. According to Schmid: "Eventfulness increases in conjunction with the degree to which the change of state is felt to be an essential part of the narrative world in which it occurs" [18]. This first criterion is the most problematic in searching for stories at world-scale as there is no narrative premise against which to evaluate relevance. We considered we could use this criterion to do situated story sifting, that is finding events relevant to a given character (for example, the player-character) by scoring higher

events concerning themselves, friends or relatives. While this is a workable approach for systems with clear protagonists around whom to center the narrative universe, for this experiment we were interested in “absolute” story sifting rather than relative.

4.1.2 Improbability. Schmid predicts that “Eventfulness increases in proportion to the extent to which a change in state deviates from the doxa of the narrative (i.e. what is expected from the narrative world)” [18]. This is probably the most obvious and seemingly objective criterion (the paper’s discussion will revisit this naïve assumption) that unusual, unexpected, and outstanding events tend to be more interesting. This heuristic was also tested by Kreminski [7] with positive results as well as Dimulescu *et al.* [2].

Computationally, this can be seen as an “outlier” detection problem with multiple potential solutions. Because of our interest in stories (and not just events), we chose to search for improbable causally related event chains. In *Chroniqueur*, some events can cause other events, and these chains are remembered in an implementation of what Ryan calls “causal bookkeeping” [15]. We model the whole event space as Markov chains and then can find the most improbable sequences.

4.1.3 Irreversibility. It makes sense that “[e]ventfulness increases with the irreversibility of the new condition which arises from the change of state” [18]. A marriage is more difficult to walk out of than a date and therefore is significantly more eventful narratively. Although theoretically simple, we could never find a satisfying general way to search for this. Because of the greatly varying object and character states, there are very few statistical records of states being exactly restored after a given event (which could point to reversibility). We would have to track very specific variables or resort to hard coding some events on an irreversibility scale. We predicted this would only yield back the events we had already identified and so judged it was not worthwhile for the moment. We are sure others will find smarter ways to implement this.

4.1.4 Non-iterativity. “Repeated transformations, even if they are both relevant and unpredictable, represent at best a low level of eventfulness” [18]. With this factor we have again the global history problem. For example: if a marriage is a low-iterativity event for an individual, it is a very common event at the social level. In effect, every wedding is probably a tellable story but that doesn’t help us picking up wedding stories in the world at large. We tried scoring events by their “commonness” in the world event space and then sort chains according to their most uncommon event. However this yielded results not very different from “improbability” and so was not kept for the user test.

4.1.5 Effect/Persistence. Schmid states that “... [Eventfulness] increases with the consequences for the thought and action of the affected subject in the framework of the narrated world” [18]. Along these lines, we hypothesized that events changing more properties in the world than others would likely be more eventful. Again, here the issue was that there are so many variables with different scales that it is complicated to tally significant measures. For example, the birth of a child is just a unit change in a person’s “children” list while one can be given 20 “food units” and lose “30 health units” but none of these things are comparable in terms of subjective effect. Like with non-iterativity, attempting to unify these variables with a

single weighted “key” would similarly introduce the problem of simply producing what the researchers ask for. We tried higher level approximations for example counting how many events can trace their cause in a single event (making it “high effect”) or how many characters were impacted along the causal chain. Each of these approaches yielded the issue of favoring always the same classes of events. Clan migrations, for example, involve many people and may trigger multiple effects down the line. Indeed, they represent tellable events but if the heuristic always returns a migration then it boils down to a “migration” search. We have also dropped this heuristic for user evaluation.

4.2 Dramatic situations

During the development of the *Chroniqueur* social simulation, we drew inspiration from the notion of dramatic situations to inform the design of events (or combinations thereof) with high narrative potential (Lessard *et al.* 2022b). It made sense to evaluate whether the presence of these dramatic situations within a cause chain did in fact increase its tellability. Ten of Polti’s 36 dramatic situations [14] can be searched for in the event space:

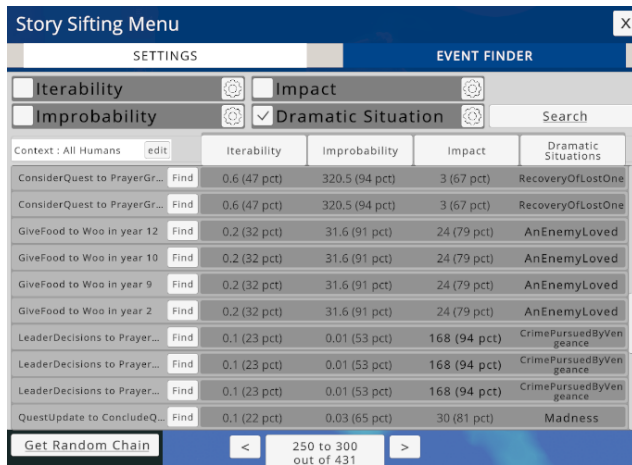
1. Supplication: Any “Plea” event
2. Crime pursued by vengeance: Any successful “Vendetta” or “Find Thief” quest
3. Vengeance taken for kin upon kin: Any successful “Vendetta” quest between close family members
4. Revolt: Any successful “Chiefdom Independence” war event
5. Enmity of kin: Any “Vendetta” or “Find Thief” quest with a close family member as the Person of interest
6. Rivalry of Kin: A “Find Thief” quest with a close family member, a “Become Chief” quest in a clan whose chief is close family member
7. Murderous Adultery: A lethal “Jealous Violence” event
8. Madness: A “Breakdown” event that leads to a divorce or a duel
9. An enemy loved: A “Fall in Love” event towards a person in an enemy clan (negative clan appreciation)
10. Recovery of a lost one: A successful “Find Person” quest

4.3 The Story-Sifting Interface

Figure 2 displays the story-sifting interface that was developed for *Chroniqueur*. It has access to all the causal chains of events to have ever happened in the world up to the current point and can search them against specific criteria as described previously. Results can also be sorted according to specific criteria when multiple ones were selected. The first column gives the user a sense of the chain by summarizing it as the beginning to end events, for example the first item here is a chain that begins with a “Consider Quest” event and ends with “Prayer Granted”. The “Find” button next to it will display all actual similar chains with the possibility to see all the detailed events that compose them (with links to an interactive encyclopedia of the world: [11]).

5 Evaluating heuristics through narrative

Rather than test in the *Chroniqueur* platform directly, we instead used the system to generate stories which would then be narrativized and presented those narrativizations to our participants.

Figure 2: *Chroniqueur's* Story Sifting Interface

We opted for this approach for two reasons. First, abstracting the simulational elements out of the story through prose retellings allowed us to test narrativization and story-sifting separately. Our focus was on the effectiveness of story sifting algorithms first and foremost, and not whether *Chroniqueur* itself faithfully conveyed the results of those algorithms to the player. Second, obscuring the procedural nature of the narrative allowed us to focus on the construction of the narratives as such, divorced from artifacts of their procedural generation. Procedurally generated text in *Chroniqueur* has its own writing conventions generated by a templated grammar system [12], whose idiosyncrasies might distract from the underlying narrative. By doing so we were knowingly averting what James Ryan referred to as the "Aesthetics of the uncanny" [15] even if this can be seen as a loss.

Of course, this also introduces a separate issue - namely, that the varying quality of human writing in the narrativization might obscure the underlying narrative the same way that *Chroniqueur* generated text might. In order to mitigate this, we decided to narrativize each set of events twice, once by a human author (always the same) and once using GPT4 given the same prompt, and evaluating the results of these narrativizations both separately and in aggregate.

Moving forward we formulated the following hypotheses:

1. Story-sifting heuristics would yield more interesting stories than a randomly picked one, and the more heuristics a story validates the better it will be.
2. Human-authored stories would tend to be more enjoyable than GPT-authored stories (i.e. there is an authorship effect to narrative interest).

5.1 Story Generation and Testing Protocol

To generate the stories, a world was generated in *Chroniqueur* and allowed to run without human intervention until a sufficient corpus of game events satisfying our search criteria had built up. The world 'Preparipate' selected after a period of 61 in-game years had generated 228K unique events featured in 27K event chains (story candidates). Using our story-sifting tool (see Figure 2), the

Chain of events from Accident to Woo in year 24

Abigail Singleworld's Accident in Year 24:

I was taking a leisurely stroll on a mountain path when suddenly I stepped into a pit covered by leaves and I lost 2 health.

Abigail Singleworld's Breakdown in Year 24:

I think I just had a big breakdown. I lashed out at Glen Backs because, if I didn't, I think I would have gone mad...

...

Abigail Singleworld's Breakdown in Year 24:

I had a breakdown. I grew unhealthily obsessed with Glen Backs.

Abigail Singleworld's Woo in Year 24:

Glen Backs blushed as he listened to the poetry I composed in his name.

Figure 3: Example of event log provided for story generation

game's events were parsed for Dramatic Situations and two, "Crime Pursued by Vengeance" and "Supplication" were chosen for their average rate of occurrence - neither too rare nor too ubiquitous - and tendency to naturally consist of multiple narrative beats. Each of the Dramatic Situations was sampled twice: one sample being a random story fitting the Dramatic Situation and the other being the most-improbable story fitting the Dramatic Situation. Two additional stories were identified, one being the most-improbable sequence of events that nevertheless did not adhere to any Dramatic Situation and the other being a random, unremarkable event.

The chains of events of each story and key information on the involved people were extracted. These were each then given to a human author and to GPT4 with the following instructions:

Please produce a short story of ~500 words using the following characters and plot events. Feel free to invent story elements, including action and dialogue. You may change small details, but do not deviate from the story outline as I have given it to you. Please write the story in an entertaining, engaging tone, like mainstream pop fiction or pop history. Please also come up with an appropriate, clever title for the story. The key plot events are as follows.

This would be followed by the list of events and *dramatis personae* as generated by *Chroniqueur*; see Figure 3 for example.

This yielded 12 short stories, each with a GPT4-written and a human-written version, divided across six categories:

1. Most-Improbable Crime Pursued by Vengeance
2. Random Crime Pursued by Vengeance
3. Most-Improbable Supplication
4. Random Supplication
5. Most Improbable, No Dramatic Situation
6. Random, No Dramatic Situation

These twelve stories¹ were then split across two questionnaires, with each questionnaire containing a story from each of the six categories - three that were human-authored, and three that were

¹ Accessible here: <https://www.lablablab.net/StorySifting/SiftedStories.pdf>

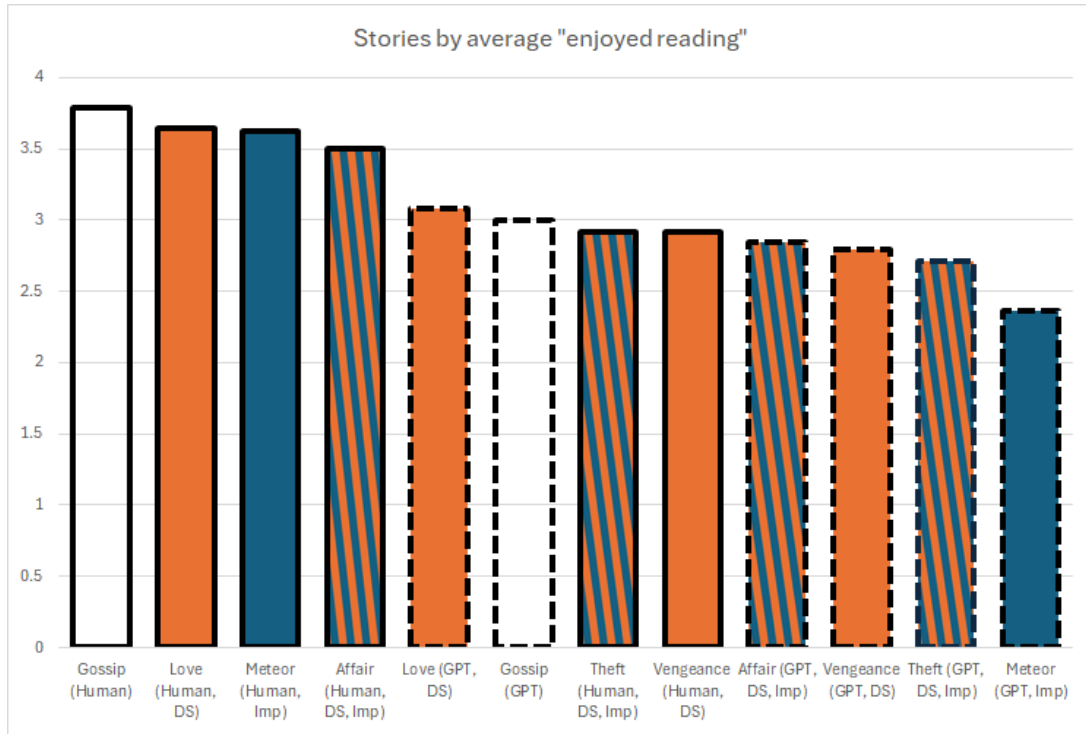


Figure 4: Average enjoyment per story

generated with GPT4. Participants were asked to score each story on a 5-point Likert scale across six criteria:

1. I found this story interesting
2. I found that the events of this story made sense
3. I enjoyed reading this story
4. I felt the characters in this story were well-rounded and believable
5. I found this story surprising
6. I could easily follow this story.

Participants in this survey were recruited from among faculty and students, both undergraduate and graduate, of the Fine Arts, History, and English departments at Concordia University. They were asked to report on a Likert scale their familiarity with fiction in general, their familiarity with structural elements of fiction specifically, and their interest in emergent narrative games.

Qualitative data was also collected through a space for additional comments. The resultant survey data was analyzed and responses both within and between stories were compared. 24 participants read the stories and filled the surveys, 12 for each set of questionnaires.

6 Results and Discussion

To put it simply, the results of this user evaluation do not show that our story heuristics have any significant relationship with reading enjoyment. Ironically, the narrative that stood out as the clear favorite was the control one, being neither particularly improbable nor featuring any dramatic situations (see “Gossip” in Figure 4).

Given the limited variety of narratives evaluated and small sample size, it would be premature to discredit the approach altogether. There was only one narrative evaluated with no heuristics and it was picked randomly rather than testing more contrasted extremes (very probable, for example, rather than simply not highly improbable). We should also note that as a social simulation *Chroniqueur* has its own idiosyncrasies that might make it less favorable to story searching than other systems, and that the heuristics could have been modeled in many other ways. However the results do point us towards some interesting preliminary observations.

While we asked various questions, this paper focuses on understanding what factors influenced the general reading enjoyment of participants. Figure 4 summarizes these results by comparing all stories in terms of average “enjoyed reading”. Bars with solid outlines were human authored while those with dotted outlines were written by GPT4. The absence of colour means that no heuristics were present in the story, orange signals the presence of at least one dramatic situation, blue stories were highly improbable, and striped bars represent a mix of both. The bars are labeled with the story identifier, each is shown twice as it was narrativized both by a human and GPT. For example, we can see that the “Meteor” story was improbable but did not display any dramatic situations, its human version ranked 3rd in appreciation while the GPT4 version ranked 12th.

In the following sections we invoke both the survey results as well as our designer/developer insights to raise questions and hypotheses about this story sifting approach.

6.1 Problems of Improbability

6.1.1 Irrelevant improbability. As we were preparing our experiment we stumbled upon many issues with improbability as a heuristic. For example, in order to have noticeable outliers, we needed to let the simulation go longer, generating a larger quantity of events. However, longer simulations led to the emergence of extreme outliers that defied narrative explanations. In particular, the search for causal chains meant that longer chains were typically rarer and at some point the heuristics mostly returned very long chains of events that did not necessarily amount to much narrative interest. In other words, the seemingly objective criterion of “improbable” required significant adjustments for it to return “good” results: varied but not too weird.

Related to this observation is the fact that statistically rare chains of events are not necessarily narratively meaningful. For example, the *Starstruck* and *When Stars Collide* stories (“Meteor” stories on Figure 4) are based on the following improbable causal chain:

```
FindFood -> HuntingAndGathering -> Accident ->
GetHurt -> Die -> Funeral -> GroupInteraction ->
SocialInteraction -> LifeChangingConversation.
```

In short, Butch Singleworld was traveling great distances to find food for his family (the grammar-generated text specified that he was specifically looking for frogs!) when he was struck by a meteorite and died. During the funeral ceremony an acquaintance named Jayda had a conversation with a certain Melody which led her to become attracted to rich people².

This is a convincingly improbable sequence but from a narrative standpoint, the finale simply seems irrelevant and puzzled many readers. How is Butch’s accident story related to Jayda’s dating considerations? The causality is real in the *Chroniqueur* world but a large part of story construction is in the pruning of events unrelated to what is constituted as the story.

6.1.2 Diegetic Improbability vs Narrative Improbability. This test’s most appreciated story, which also scored third as most surprising, was founded on the following chain: GiveFood -> SocialInteraction -> ShareInformation. In brief: while clan leader Bifrons was giving a food contribution to tribe chief Violette, the latter decided to confide in him that she was having an affair with a certain Viné.

This sequence is not remarkably improbable in the *Chroniqueur* world but it is quite unusual as a story. Why did Violette choose to confess to this person and in this circumstance? Things are also left in uncertainty: what will Bifrons do with this information? What are we led to understand? The story has a titillating aura of mystery.

In other words, world-improbability does not equate narrative-improbability. This observation is coherent with Schmid’s prediction that “improbable” should be considered in terms of the local narrative itself (whatever has led to the event tested for improbability) and not in terms of the genre [18] and, we can maybe extrapolate: the *world*. This complicates story-sifting by suggesting a need for some awareness of readers’ narrative expectations when it comes to possible future events.

²In *Chroniqueur*, social interaction events have a rare chance of causing a “life changing conversation event” which provokes the random transfer of a personality trait from one character to another. Affinity traits like “being attracted to the rich” had been added to complicate romantic stories and hopefully lead to more interesting stories.

6.1.3 Dramatic Situations and “Cookie-Cutter” Heuristics. A few qualitative comments from readers pointed towards the repetition of certain plot points, notably “affairs”. We know this type of event to be too prevalent already in the world space, in spite of multiple attempts to tone it down, which is a problem separate from story sifting. However, there is also the warning that if you look for a “love triangle” dramatic situation, you will not be surprised to find one. However, the dramatic situation literature is also convincing in its argument that while some of these situations are endemic in literature, they are endlessly refreshed by the local details of the story [14]. Who is involved in the love triangle? What events led to it? What are the outcomes?

We can hypothesize that for dramatic situations to work as a heuristic, the simulated world should be ripe enough with the details that might make each situation unique. In a previous paper on the use of dramatic situations in *Chroniqueur*, we showed how three randomly picked instances of the “Plea” dramatic situation could turn out very differently [13]. The difference, however, is that we allowed authors to dig into the details of the simulations to retrieve as much information as needed to flesh out these stories. In this experiment, we standardized the information available for each story experiment to support a more automated pipeline which led to much more abstracted causal chains. We have again another complication towards automated story-sifting and narrativization which is that each story candidate might need very different background information to be evaluated and narrated.

6.1.4 Narrative Effect. Since our experiment evaluates two different narratives for every story, we can gather some observations on the classical narrative/story (or fabula/syuzhet) distinction. The main takeaway, albeit unsurprising, is that evaluators generally preferred human authored stories in this specific context. Only two GPT4-written stories surpassed human ones in average “enjoyed reading”, but they only compete with the less preferred ones (we’ll discuss these cases later).

The more interesting case is that of the “meteor” story whose human version came in 3rd favorite while the GPT version ranks dead last. We see here a strong “narrativization” effect wherein the same events can be either considered very interesting in one rendition (3 readers gave it a 5/5 in “enjoyed reading”) or not at all in another (3 readers gave it 1/5). We have already summarized earlier the story of Butch’s accident leading to a life changing conversation between two relatives at his funeral and how the last episode was difficult to read as a relevant conclusion. GPT4’s narrative bridges the gap by evoking that funerals are propitious to introspection. While this is reasonable, it fails to cope with the absurdly irrelevant nature of said introspection: Jayda’s sudden preference for rich partners. The last line’s serious conclusive comment further highlights the rupture between the narrative’s tone and the events recounted: “even in the darkest times, there is a glimmer of hope and the potential for new beginnings”. In other words: yes, the frog-catcher died from a random meteor; but at least Jayda now wants to marry rich (!)

Confounded with the lack of continuity and silliness, the human author constructed a more narratively meaningful segue and opted to render the conversation in a subtle tongue-in-cheek mode that managed to make things work for the readers (see Figure 5). The

"All these people come to weep for Butch, but will they say the same things about me?" Jayda asked with a sigh. Melody thought about it for a moment.

"The village isn't mourning Butch the man, it's mourning Butch the provider. It's his wealth that people really grieve for," she said.

"But I could never hope to be as wealthy as he was. What, do you want me to spend all year out in the mud? We'd never see each other anymore."

Melody shrugged. "Well, then you could always marry rich."

Jayda nodded slowly. "Yes..." she said after a moment, the setting sun sparkling gold in her eyes. "Yes, I rather suppose I could."

Figure 5: Excerpt from human-authored “meteor” story

point here is not to underline GPT4’s failure but rather draw attention to the fact that good narrativization approaches could probably compensate for imperfect story sifting methods.

6.1.5 Story-Effect. We noted earlier that two GPT4-rendered stories ranked higher than a few human written ones namely the *gossip* and *love* stories. Interestingly, the most appreciated human narratives were derived from these same stories as well. This suggests that the pre-narrative story material might have had intrinsic narrative qualities, it didn’t need a *good* storyteller.

Inversely, the *theft* and *vengeance* stories yielded both the human author’s least appreciated narratives (scoring below GPT4-generated narratives) as well as GPT’s second to least preferred. The story material was so poor even the human author struggled to render it as passing narrative.

These observations allow some optimism towards future story sifting experiments. Beyond the obvious narrative-effect (that a good storyteller might transcend poorly sifted events), there is also evidence towards a story-effect: that good story material might shine *in spite* of mediocre narrativization. [19]

7 Conclusion: Simulation Narratives as Standalone Form

A harsh realization from this experiment was that the evaluators surely did not like these stories as much as we did. Why is that? Taking some distance, the evaluators are right of course. These stories, taken as is, are not great. The difference is that we are very familiar with the *Chroniqueur* world, a position that allows us to spot and appreciate all deviations from the norm, things that are unexpected and surprising within that world. Not unlike the jazz connoisseur who can relish in some fresh variation that non-jazz fans cannot distinguish from any generic “jazziness”. This leads us to the maybe obvious but not often discussed proposition that simulation-based emergent narratives should probably not be expected to be appreciated in abstraction of their idiosyncratic worlds.

It is probably a questionable underlying assumption behind the story-sifting research agenda that events transpiring from fictional simulations have the potential to be interesting “in absolute terms”. One of James Ryan’s ambitions at the time of his thesis [15] was the production of an automated podcast recounting interesting events that would have emerged from his simulation of a small American town. There are in fact very few precedents to support this ambition (justifying its nature as research, of course). There are a few (somewhat) famous *Dwarf Fortress* retellings that can spark the interest of an audience larger than the game’s players but we can still imagine that this narrative delight is predicated on some understanding of the *Dwarf Fortress* world. Steven Sych [19] highlighted that at the core of many player retellings is the witnessing of some weirdness or incongruity that would be identifiable only by those well-versed in the simulated worlds in which they happened.

This is both a problem and a relief to the story-sifting project. It is a problem because it points towards the necessity for would-be readers of emergent narratives to be familiar with their source simulated worlds. In the digital game domain, this is often the case. Players engage for hours with a game like *Dwarf Fortress* or *The Sims* before spotting and taking pleasure in emergent stories. However, this realization is also a relief because given this awareness of the worlds, heuristics like improbability might actually score much higher in reader appreciation as they would highlight interesting anomalies or coincidences. This is for further research.

Acknowledgments

Merci aux Fonds de Recherche du Québec, dossier 335420. <https://doi.org/10.69777/335420>.

References

- [1] Raphaël Baroni. 2014. Tellability. In *The Living Handbook of Narratology*. Peter Hühn, Jan Christoph Meister, John Pier, and Wolf Schmid (Eds.). Hamburg University. <https://www-archiv.fdm.uni-hamburg.de/lhn/node/30.html>
- [2] Andrei Dimulescu and Jean-Louis Dessalles. 2009. Understanding Narrative Interest: Some Evidence on the Role of Unexpectedness. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 31. 2922–2927.

- [3] Pablo Gervás and Gonzalo Méndez. 2024. Evolutionary Story Sifting over the Log of a Social Simulation. In *Artificial Life and Evolutionary Computation*. Springer, Cham, 381–393.
- [4] Gonzalo Méndez and Pablo Gervás. 2023. Using ChatGPT for Story Sifting in Narrative Generation. In *Proceedings of the 14th International Conference on Computational Creativity (ICCC '23)*. University of Waterloo, Waterloo, Canada.
- [5] Peter Hühn. 2013. Event and Eventfulness. In *The Living Handbook of Narratology*. Peter Hühn, Jan Christoph Meister, John Pier, and Wolf Schmid (Eds.). Hamburg University. <https://www-archiv.fdm.uni-hamburg.de/lhn/node/39.html>
- [6] Samuel Johnson-Bey and Michael Mateas. 2021. Centrifuge: A Visual Tool for Authoring Sifting Patterns for Character-Based Simulationist Storyworlds. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment Workshops*.
- [7] Max Kreminski, Matthew Dickinson, and Michael Mateas. 2022. Select the Unexpected: A Statistical Heuristic for Story Sifting. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment Workshops*.
- [8] Max Kreminski, Matthew Dickinson, and Noah Wardrip-Fruin. 2019. Felt: A Simple Story Sifter. In *Interactive Storytelling. ICIDS 2019. Lecture Notes in Computer Science*, Vol. 11869. Springer, Cham, 14–27.
- [9] Max Kreminski, Noah Wardrip-Fruin, and Michael Mateas. 2020. Toward Example-Driven Program Synthesis of Story Sifting Patterns. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment Workshops*.
- [10] Wilson Leong, Julie Porteous, and John Thangarajah. 2022. Automated Sifting of Stories from Simulated Storyworlds. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS '22)*. 4950–4956.
- [11] Jonathan Lessard and Antoine Beauchesne. 2022. Automatic Interactive Documentation for Emergent Story Discovery. In *Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22)*. Athens, Greece. ACM.
- [12] Jonathan Lessard and Quintin Kybartas. 2021. Blabbeur – An Accessible Text Generation Authoring System for Unity. In *Interactive Storytelling. ICIDS 2021. Lecture Notes in Computer Science*, Vol. 13138. Springer, Cham, 206–218.
- [13] Jonathan Lessard and Samuel Paré-Chouinard. 2022. Dramatic Situations for Emergent Narrative System Authorship. In *Interactive Storytelling. ICIDS 2022. Lecture Notes in Computer Science*, Vol. 13590. Springer, Cham, 258–270.
- [14] Georges Polti. 1912. *Les Trente-six Situations dramatiques*. Mercure de France, Paris, France.
- [15] James Ryan. 2018. *Curating Simulated Storyworlds*. Ph.D. Dissertation. University of California, Santa Cruz, CA.
- [16] Marshall Sahlins. 1974. *Stone Age Economics*. Aldine Transaction, Chicago, IL.
- [17] Ben Samuel, Adam Summerville, James Ryan, and Lauren England. 2021. A Quantified Analysis of Bad News for Story Sifting Interfaces. In *Interactive Storytelling. ICIDS 2021. Lecture Notes in Computer Science*, Vol. 13138. Springer, Cham, 142–156.
- [18] Wolf Schmid. 2003. Narrativity and Eventfulness. In *What Is Narratology? Tom Kindt and Hans-Harald Müller (Eds.). Walter de Gruyter, Berlin, Germany*, 17–33.
- [19] Steven Sych. 2020. When the Fourth Layer Meets the Fourth Wall: The Case for Critical Game Retellings. In *Interactive Storytelling. ICIDS 2020. Lecture Notes in Computer Science*, Vol. 12497. Springer, Cham, 203–211.