Murder Mysteries: The White Whale of Narrative Generation?

Markus Eger
University of Costa Rica
San José, Costa Rica
markus.eger@ucr.ac.cr

Abstract

The field of narrative generation appears to have a particular fascination with the genre of murder mysteries, perhaps because the genre's defining characteristics include a clear, and predefined structure, a reliance on logic, and an immense popularity. However, while there have been several attempts to generate a satisfactory mystery story, most such efforts never made it beyond an initial prototype or idea. In this article we investigate what may draw researchers towards this genre, and then disseminate why the very properties that make murder mysteries such an appealing target also make it a particularly challenging domain. For the apparent rigid scaffolding structure provided by the genre, we discuss its flexibility; for the apparent hard grounding in logic we illuminate the deficiency of such; and for the seemingly clear communication maxims prescribed by written "rules" for the genre we explore which ones must be broken. Our goal is to show the actual scale of the difficulty of the problem, and which challenges have yet to be addressed. While it may seem that our focus is overly narrow, we will also explore the implications for narrative generation more broadly.

Introduction

> **Sherlock**: I'm a consulting detective. The only one in the world. I invented the job.
>
> Sherlock (TV Series), Episode 1.1, *A Study in Pink* (2010)

Murder mysteries are a mainstay of popular entertainment, perhaps lead by the fame of Sir Arthur Conan Doyle's character Sherlock Holmes, who has been named the "most portrayed literary human character in film and TV" [1]. In the last 10 years alone, Sherlock Holmes has been portrayed by such renowned actors as Will Ferrell (*Homes & Watson* (2018)), Robert Downey Jr. (*A Game Of Shadows* (2011)) or Sir Ian McKellen (*Mr. Holmes* (2015)) on the big screen, (sometimes very) roughly based on Doyle's original work. There have also been several TV series that adapted the character and placed him (or her, in one instance) in modern times in London (*Sherlock* by the BBC (2010-17)), New York (*Elementary* by CBS (2012-2019)), or Tokyo (*Miss Sherlock* by HBO Asia and Hulu Japan (2018)). Apart from this one, enormously popular, character, there is also a number of other detectives in literature, cinema and TV, including Agatha Christie's Miss Marple and Hercule Poirot, Rita Mae Brown's (and Sneaky Pie Brown's) Harry Haristeen and Mrs. Murphy, Carolyn Keene's (a collective pseudonym) Nancy Drew, or CBS's Jessica Fletcher (on *Murder She Wrote* (1984-1996)), among many, many others. Several video games based around the idea of detective fiction also exist, such as an entire series, consisting of 8 main games and 4 casual games, featuring Sherlock Holmes developed by Frogware, Disco Elysium by ZA/UM, or L. A. Noire by Team Bondi. It is therefore fair to say, that, as a genre, detective fiction spans an incredibly wide variety of settings, authors, styles, and media. This ubiquity also makes it an attractive target for research on narrative generation, with several approaches, differing almost as wildly as the stories themselves, having been proposed.

In this article, we want to disseminate the intrinsic qualities of detective fiction which make it such an alluring goal for AI-based generation, but at the same time highlight why this goal has eluded us so far, despite our best efforts as a community. We will start by providing a more precise characterization for a subset of detective stories, called "whodunits", which will serve as our main guiding example through our discussion of challenges encountered in their generation. We will discuss the appeal, and apparent "simplicity" (in a computational sense) of these stories, as well as why this simplicity is illusory in nature, and which research problems it actually leads to. While our main focus are the "classic" whodunit-style mysteries, we will also explain how many of our observations apply not just to detective fiction in general, but other genres as well. Once we have covered our viewpoint, we will also provide a review of existing approaches, and where they fall in addressing the challenges we present, and their status (if known), before concluding with an outlook on the vast potential for future work.

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1] https://www.radiotimes.com/news/2012-05-17/sherlock-holmes-awarded-guinness-world-record/

Murder Mystery Story Generation

> There are two kinds of people who sit around all day thinking about killing people...mystery writers and serial killers. I'm the kind that pays better.
>
> Richard Castle

Before we dive into what makes detective stories so challenging, we first want to define more precisely the types of stories we are (mainly) discussing, and what makes them so appealing for research purposes. Note that we talk about murder mysteries, rather than detective stories in general, which would be a much a broader category which encompasses more character-focused narratives, such as the True Detective TV series. Murder mysteries, as the name implies, are instead centered around a mystery, in the sense of a big unknown, that will be unraveled over the course of the story, and this mystery revolves around a murder (or multiple murders, as the case may be). The two main sub-types of these stories are called the "whodunit", in which the murderer is not revealed until the end, and the "howcatchem", in which the audience knows the murderer from the beginning and observes how the detective arrives at their conclusion while knowing more than them, the latter being popularized by the TV series Columbo. We will mainly focus on the former, for which S.S. Van Dine (the author-pseudonym of art critic Willard Huntington Wright), has collected a list of properties with which each "proper" member of this story type ought to comply (Van Dine 1928). A similar list has been compiled by Ronald Knox, another author of such stories (Knox 1929). Both lists agree that the general structure a murder mystery should follow starts with a (partial) description of the crime committed ("No lesser crime than murder will suffice.", according to Van Dine), followed by a "rational and scientific" (Van Dine) investigation based on observations by the detective, which has to feature the guilty party early on in the story, but they "must not be anyone whose thoughts the reader has been allowed to know" (Knox), and after which the detective identifies the guilty party "by logical deductions" (Van Dine). Additionally, both authors agree that the investigation process with all its clues must be laid bare to the reader, with Van Dine stating, as his first rule no less, that "The reader must have equal opportunity with the detective for solving the mystery. All clues must be plainly stated and described." Indeed, the appeal of these mystery stories is the "I see"-moment at the end, when the solution is presented, and the reader can review the entire story in their head to verify that they "could have" solved it, if only they had paid more attention and drawn the proper conclusions.

We hope that this description already provides some hint what makes murder mysteries such an alluring target for story generation. In the long history of story generation, most approaches were based, some perhaps more loosely than others, on some form of logic, and here we have an entire (popular) genre of stories that are, at first glance, directly grounded in logic! It only seems logical (no pun intended) to use the same reasoning that is used by the story generator internally as the mechanism by which the detective performs their investigation. The process of generating a murder mystery story becomes, in a way, isomorphic to finding a murderer in a virtual world governed by the logic du jour: Solving one of these problems automatically solves the other. We could, so it might be believed, exploit this isomorphism by introducing logical operators that provide the detective with clues, and which also facilitate the story goal of telling a story following the aforementioned structure. The mere presence of this structure adds an additional (perceived) ease to the task, as we already know which parts our story must possess, and what has to happen in each of them. However, as the title of this article may already indicate, the task of generating murder mysteries is a lot more involved than this cursory analysis would suggest. In the next section we will deconstruct, piece by piece, just what makes this so challenging.

Challenges

> **Miss Marple**: "One does not like to make definite assertions unless one has a little more definite knowledge."
>
> Agatha Christie, *Nemesis* (1971)

Fixed as the structure of murder mysteries might seem, it actually leaves a lot of room for variation. In this section we will present our thoughts on different aspects of this variation. First, while the defining characteristic of murder mystery story is its fixed structure, what makes each individual such story worth the audience's while is what makes it different from the other mysteries out there, and this novelty is not just limited to giving different names to the characters or permuting characteristics between them. Second, while logic ostensibly plays a large role in how a murderer is caught, the actual details are not always so strictly logical. And third, perhaps the key ingredient tieing all of the genre together, is to convince the audience that they "could have" identified the culprit themselves, by providing an exposition of all relevant facts, while actually hiding said facts in plain sight. In addition to these three main aspects, we will also briefly present an expanded view, on how these same challenges are applicable to narratives and their generation more broadly.

Variety

> **Jessica Fletcher**: "Of course it's the murder weapon! Whoever heard of framing anyone with the wrong gun?!"
>
> Murder She Wrote (TV Series), Episode 1.19, *Armed Response* (1978)

As we have established above, murder mysteries follow a relatively rigid structure, which gives them a rather procedural appearance. However, upon closer examination, the draw of many of these stories is that even within this fixed framework, there is a great deal of variety. The TV shows Elementary and Monk each had an episode in which the murderer turned out to be a person currently in a coma, but even with this very specific scenario the exact executions differed widely: In one, the person was in a temporary

coma, induced by her co-conspirator who was also her doctor and woken up at night to commit the crime, while the other relied on a delayed trigger mechanism, and the perpetrator semi-intentionally put themselves into a coma as an alibi. Further analysis of stories reveals that the method, motive and circumstances of each murder are highly specific to each episode. In one, the detective entrenches themselves in the world of ancient Asian artifacts and the Chinese circus (Sherlock, Episode 1.2), while another has them investigate the ins and outs of film-making (Murder She Wrote, Episode 8.12), or seek the help of a mathematician while uncovering the mysteries surrounding the Millennium Prize Problem of P vs. NP (Elementary, Episode 2.2). The original Sherlock Holmes Stories by Doyle also always involve a significant backstory, depending on which social sphere connects the murderer (and/or other suspects) and the victim. For a generative approach, this means that, while the general structure of the story is relatively fixed, this structure will have to be molded to a wide variety of different environments. Of course, it would be possible to generate 50 murders set aboard a moving train, maybe even with slight variations in the motive, and relationship between the characters, but we argue that such a generative process does not really address the stated problem. To actually solve this problem, a story generation system would need to have a large knowledge base with different possible scenarios. There is also the matter of surprise: Murder methods in contemporary mystery fiction are quite esoteric (and, often completely impractical), from using magnets to attract bullets (Elementary, Episode 3.1), to aforementioned delayed trigger based on gluing packages to the inside of mail boxes (Monk, Episode 2.7), to a hallucigenic drug that makes the victim try to suck their child's blood (Miss Sherlock, Episode 1.4). If a domain theory only has actions for "shoot" and "stab", or similar, such stories become impossible to express.

In summary, one main draw of murder mysteries is their variety, rather than their rigidity, and this variety requires an equally varied domain theory, or the capability to perform reasoning about potential actions afforded by a world. On the other hand, since our detective performs their reasoning in a cold and logical manner, we will want this reasoning capabilities anyway. However, in the next section we will deconstruct this perceived notion of logic, and expose its flaws.

Logical Conclusions

> **Col. Ross**: "Is there any other point to which you would wish to draw my attention?"
> **Sherlock Holmes**: "To the curious incident of the dog in the night-time."
> **Col. Ross**: "The dog did nothing in the night-time."
> **Sherlock Holmes**: "That was the curious incident."
>
> Sir Arthur Conan Doyle, *Short Story: The Adventure of Silver Blaze* (1892)

One particularly noteworthy aspect of murder mysteries is the disparity of knowledge between the different characters, and how this disparity is overcome. Indeed, one could view a murder mystery as the process to equalize knowledge between all main characters (except perhaps the victim, given its post-mortem status). The two main mechanisms employed by the author are those of conveying information, and a logical inference procedure on this information. Typically, the detective is presented as possessing the most advanced such inference procedure and then, in turn, conveys their deductions to the authorities, their side-kick, and/or the murderer. As practitioners of AI we are no strangers to logical inference engines, and with a suitable formalism we can also describe the necessary knowledge gathering actions, like encountering clues and questioning suspects. We also have means to reason about the beliefs about beliefs, representing what may be true, what others believe to be true, or what can not possible be taken as factual by the detective or other characters (Van Ditmarsch, van Der Hoek, and Kooi 2007). Similarly, we may have the murderer attempt to mislead the detective, because they believe the detective to be on the wrong trail, when they are actually being deceived by the detective themself.

The challenge with this viewpoint manifests itself in two related ways. First, while researchers may have a certain affinity for logic, most of the actual audience of a story may not have the same predisposition towards such deductions. Indeed, many "logical" inferences performed by actual human beings follow intuition more than strict rules. Sure, this inference process may not be sound or complete, but the biological cost-benefit analysis did not work out towards evolving perfectly rational brains. However, and this brings us to the second part of the challenge, this also gives authors more leeway with their stories. A common theme in murder mysteries is the detective confronting the murderer with how they deduced their guilt. In TV shows this is invariably followed by an exasperated or defeated look on part of the murderer, to remove all doubts that the real culprit was found, for example when the detective informs them that the lack of the dog's barking unmistakenly identifies them as the guilty party, even when there may be dozens of other plausible explanations for the canine's behavior. This relieves the author of the burden of coming up with a completely watertight proof on behalf of the detective, as long as the inference process is laid out with little enough room for ambiguity for the audience to believe that the murderer believes it. While there are various ways to achieve this, a particularly popular one involves an in-depth analysis of a convoluted or novel scenario by the detective that seemingly incorporates all gathered information (connecting to the previous point of the variety and novelty of the scenarios), and a straight up appeal to "once you have eliminated the impossible" without actually doing so. An actual court of law may require proof "beyond reasonable doubt", but in a narrative a proof "beyond casual doubt" appears to suffice.

On the surface, this may actually seem to make the problem easier, since we could now use "some" less involved reasoning system, and still arrive at a reasonable deduction. However, it is less clear what this reasoning system would actually be. Formal logic has the advantage of being more readily implementable for use in a computational system, whereas the inference process we are looking for would have to be able to pass as intuition. Put another way: We may be

able to automatically construct a proof for a special case of the Erdős discrepancy problem, but generating an intuitive explanation, or even just an overview, of said proof remains an open problem [2]. Note that the explanations and shortcuts in deductions we described here happen endogenously in the story world. Conveying them to an audience is another issue entirely, which also brings with it its own set of challenges, as described in the next section.

Discourse Generation

> "Ah!" Poirot shook his forefinger so fiercely at me that I quailed before it. "Beware! Peril to the detective who says: 'It is so small — it does not matter. It will not agree. I will forget it.' That way lies confusion! Everything matters."
>
> Agatha Christie, *The Mysterious Affair at Styles* (1920)

At this point we may consider all we have learned so far and conclude that generating a murder mystery *story* is already a very challenging undertaking, so we may as well attempt to tackle that before we even think about how we would convey it to an audience, a problem which is often called *discourse generation* in the literature. However, in this section we will describe how the two are intrinsically linked, and we would argue that one can not be done without at least considering the other. First, let us consider the case of shortcuts in logical reasoning described above. As we have claimed, the detective will make use of them as to not overburden the complexity of the story. Now let us consider the case where the detective actually performs a thorough, and complete logical inference to determine the culprit, and we then use a separate process, the discourse generator, to determine how to simplify its telling to make it palatable to an audience. We may consider this a reasonable approach, but all we have does is shift the problem to another part of our system. Regardless of where we address this challenge, we must have a model of how a human audience could be convinced of the validity of the "proof".

Before our detective can even perform any reasoning, though, they must gather evidence or clues, which form the basis of these inferences, and as S. S. Van Dine stated "All clues must be plainly stated and described", directly alluding to the process of generating an appropriate discourse. However, what it means for a clue to be "plainly" stated is a question of definition, and in typical mystery narratives, whether they are presented in written or audio-visual form, clues are, while not obscured, often presented in passing, and only at the end will the detective actually draw the audience's attention to a particular piece of previously communicated information. In the very first episode of Elementary, for example, the key insight leading to the identification of the murderer was encountering a cell phone in a bag of rice, where it was put to dry. The camera work in this episode made every effort to plausibly show all necessary pieces, without necessarily drawing attention to them. First, a washing machine is shown to have a boot print on it, which was verbally attributed to anger (Quote: "Mixed his colors with his whites? Who knows."), then the camera slowly pans to a cell phone charger, and continues to pan through the rest of the apartment, including the pantry, while cutting back to the detective to convey that they are observing the place. During the pan over the pantry, the bag of rice is clearly visible, and the same camera shot is later used in a flashback, but it is no more highlighted in this moment than the jars of pickles or the dish soap. Other discourse techniques that fall in the same category include textual descriptions of characters, which clearly mention a relevant fact, but also a host of irrelevant ones, or testimony by a witness, which - almost naturally - contains many details that do not actually pertain to the case at hand.

We would argue that being able to accuractly produce discourse of this form is essential for a good mystery story, as one of the stated goals is to make the audience believe that they could have solved the case if they had just paid more attention. In truth, though, this requires a delicate balancing act. If facts are stated too plainly, the mystery loses its, well, mystery, but if they are hidden too well, the audience will feel like they did not have a fair chance at uncovering the truth. There is a lot of subtelty involved in how to make information appear plain in retrospect. In the rice bag example from above, the rice is stored in the bottom-most rack of the pantry that is shown, which means, while the camera actually moves over all other contents of the pantry starting at the top, it comes to rest ever-so-briefly when showing the rice before cutting to the next shot. This short moment of rest is also all that is used in the later flashback-moment, to explicitly remove all other stimuli and draw attention to the rice. However, this actually just serves to convince the viewer that they "already knew" about the rice. For our purposes, we can conclude that, in order to convey facts "plainly and clearly", we first need a better understanding of the cognitive processes involved in absorbing information, or, rather, how to convince an audience that they could (and "should") have absorbed information.

At this point, we hope to have convinced the reader that the generation of murder mysteries is not only challenging, but actually rather less grounded in pure logic than it appears at first glance. Note that we, by no means, and despite the somewhat snarky title of this paper, intend to dissuade anyone from pursuing this line of research. Instead, our goal is to illuminate the many challenges that lie on the way. If anything, we want to emphasize that each of these problems is a worthwhile endeavor to address. All of the challenges we have addressed so far also have applications to narrative generation more broadly, as we will discuss in the next section.

Other Genres

> **Lt. Columbo**: Oh, one more thing. Gee, I almost forgot what I came here to ask you.
>
> Columbo (TV Series), Episode 7.2, *Murder Under Glass* (1978)

Murder mysteries, and the "whodunit" variety in particular,

[2] https://phys.org/news/2014-02-math-proof-large-humans.html

are creative expressions molded into a rather rigid structure, like discussed. However, these same creative expressions occur in narratives of many different kinds, and similar challenges must therefore be overcome in their generation. As noted, variety is what makes murder mysteries actually interesting, and a similar argument could be made for other stories as well. Of course, this is not exactly a new thought, with such measures as expressive range and other metrics (Summerville 2018), or narrative diversity (Amos-Binks, Roberts, and Young 2016) being applied to measure how distinct the produced stories are, or techniques such as novelty pruning used to generate a more diverse set of stories (Farrell and Ware 2016).

However, the other aspects we discussed also constitute worthwhile additions to narratives in general. Catching a murderer is just one manifestation of the phenomenon of a surprise reveal, a very commonly used narrative trope [3]. And exactly as in a murder mystery the reveal has to be believed by the audience, which often means that they feel they could have known, or at the very least predicted it as a possibility (Interestingly enough, research suggest the audience may even enjoy the story more if they knew about the reveal beforehand (Leavitt and Christenfeld 2013)). To provide such an experience requires the same level as logical "rigor" and discourse prowess to convey to the audience exactly the right amount of information. As an example, consider the Sixth Sense, in which Bruce Willis plays a psychologist treating a child that sees dead people. At the end of the movie it is revealed that Bruce Willis' character himself was dead the entire time and the child was the only one that could see him. An actual, purely logical analysis of the events of the movie would reveal several glaring flaws (such as him not realizing that no one else can see him for weeks), but the storytelling and camera work show only enough information to provide the appearance of a normal "life", while also semi-revealing enough flaws in the hypothesis that he is still alive to allow the audience to believe the ending presented by the author.

Related Work

> **Jessica Fletcher**: You know, if I read one more paragraph tonight, this manuscript is gonna start looking like one big typo.
>
> Murder She Wrote (TV Series), Episode 4.11, *Doom With a View* (1987)

The popularity of procedurally generating murder mysteries can be traced to one of the very first story generation systems ever devised, Sheldon Klein's Automatic Novel Writer (Klein et al. 1971; 1973). This system used a Monte Carlo Simulation to stochastically produce a murder mystery using one of four possible motives and produced a 2100 word story text in English as its output. While the actual program appears to have been lost to time, the system is reported as being impressive for its time, especially the discourse generator, which utilized deep structure networks to reason about the semantics of the produced sentences to provide fine-grained control over what to reveal to an audience. While this provided a solution to, e.g., tell the audience that a murder occurred, but not who committed it, it required hand-writing such semantic rules for every possible action that could be recounted. Most advances in the field have been made more recently, though. *Crystal Island*, for example, is a learning-game framed as a mystery story where the player/student has to solve a science mystery (Rowe et al. 2009). As the goal of the game is to teach a particular concept, this alleviates the need for (overly) great variety, but the game still incorporates many techniques that are highly relevant, such as a model of the player's knowledge. On the other hand, as the primary objective is not entertainment, the model the system employs for modeling what the player already knows, and how different facts can be connected to derive new information is more rigid than would be necessary for a more leisure-oriented application, and there is also little interest to obscure the exposition of knowledge. In fact, we would argue that a learning environment has almost the opposite goal of a classical murder mystery, as we *want* the student to figure out the solution by themselves, whereas in a murder mystery we want the audience to be surprised, and just feel like they "could have" found the solution themselves.

A common theme in work on murder mystery generation is that there is a number of projects for which "initial" or "preliminary" reports are published, but which never seem to materialize completely. NOLIST, for example, provides "a first step towards developing games with a high degree of interaction and a coherent narrative" that uses Bayesian Networks to model the probability that the player beliefs certain facts about the murder (such as what the murder weapon might have been) (Bangsø et al. 2004). James Ryan's inspiring work on "Talk of the Town" introduced characters that could misremember and lie (Ryan et al. 2015), a very useful addition to belief manipulation techniques for mysteries, but the project was unfortunately abandoned as his focus shifted on other aspects of story generation (Ryan 2018). Jaschek et al. (2019) used linear logic as implemented by Chris Martens' Ceptre (Martens 2015) to produce what they call "puzzles" that could ostensibly be part of a murder mystery story, but that do not constitute complete narratives, nor display great variety. Mohr et al. presented a procedurally generated murder mystery game (Mohr, Eger, and Martens 2018) using Eger and Marten's implementation of a Dynamic Epistemic Logic (Eger and Martens 2017b). While the system generates a complete game experience, it is limited to one very specific environment, and only generates permutations of some basic properties, and no further work appears to have done on the project.

Perhaps the most complete murder mystery generator is by Barbosa et al., who use a plan-based tool implemented in SWI-Prolog that models communication, perception and reasoning actions and is even able to present the generated stories in a comic book style format (Barbosa et al. 2014). The stories generated by this tool are pretty impressive, and already constitute the general outline of murder mysteries. However, the action set is also on a relatively high level of abstraction, leading to permutations of "character X observed the murder or the motive and tells the detective about

[3] https://tvtropes.org/pmwiki/pmwiki.php/Main/PlotTwist

it". The system would even have mechanisms in place to enable characters to lie, but since the goal is to generate a complete story, this option is seldomly used, as it would make the entire story less "efficient". The detective also does not always seem to consider lies to be a possibility, and accepts facts at face-value whenever it serves to conclude the story.

At this point, we also find it important to note that we in no way hold the limitations of their systems, or the lack of further progress, against any of these authors. On the contrary, we believe this just aids to illustrate our point that the problem is much more challenging that it might appear at first glance, and we ourselves have presented yet another approach based on logic to model belief updates that can be added to the list of projects awaiting further work (Eger and Martens 2017a). It was this work exactly, and our investigation into solutions other authors may have come up with, that lead us to write the present article in the first place.

That is not to say that there is no work to address the challenges we discussed at all. Perhaps the most promising way to overcome many limitations due to our models' rigid logical structure is the incorporation of a human in the generation process, such as in Kreminski et al.'s Cozy Mystery Construction Kit (Kreminski et al. 2019), which was in turn inspired by Ben Samuel et al.'s *Bad News* (Samuel et al. 2016). By including a human, many of the problems we described can be delegated to them, including determining just how much logical rigor is needed, or how to best present the story. In their article, Kreminski et al. also highlight the possibility of putting more focus on character relationships and emotions, which are also hard to address. A different approach was taken by Barros et al., who crawl open data, such as Wikipedia articles or OpenStreetMap data to generate a diverse set of characters and locations (Barros et al. 2018). This greatly aids with having characters which are not merely random permutations of properties, and gives some scaffolding to the story-world, but the resulting game still basically requires acquiring all clues in order to strictly eliminate all possible subjects except for the murderer.

There are also several techniques developed in narrative generation more broadly that have direct applications to detective stories, perhaps most prominently various models of belief for use by characters (Wadsley and Ryan 2013; Shirvani, Farrell, and Ware 2018; Thorne and Young 2017). While each of these models involves a rigorous logical backing, the sheer variety of them demonstrates that there is flexibility, and where there is flexibility there is an opportunity to introduce a perhaps less rigorous model to account for human nuances. This is particularly promising, as all of these models target the same general, plan-based approach of story generation. In order to then show a generated story in various media, there are, of course, also computational models for discourse generation, of which we particularly want to highlight the Bardic system (Barot et al. 2017), based on the Firebolt camera planning system (Thorne et al. 2019). This system allows the definition of low-level camera shots and what discourse is realized by each of them. While the Firebolt system itself does not impose any meaning on the camera shots or reason about what is conveyed by them, it constitutes a possible basis for a system that does.

Conclusion

> **Dr. House**: I'm sure this goes against everything you've been taught, but right and wrong do exist. Just because you don't know what the right answer is, maybe there's even no way you could know what the right answer is, doesn't make your answer right or even okay. It's much simpler than that. It's just plain wrong.
>
> House, MD (TV Series), Episode 1.21, *Three Stories* (2005)

In this article, we have provided three grand challenges that pertain to the generation of murder mystery stories: A diversity in the generated stories that goes beyond mere permutations of properties, a more human-oriented reasoning method, as well as the proper presentation of the information gathered by the detective. At the same time, we would like to acknowledge that story generation is a hard problem in general, and great advances have already been made in the field. The goal of our argument was therefore to change the perspective on murder mysteries from one of a simple and rigid story structure to a rather rich one that requires more sophistication than it appears at first glance.

After what may be perceived as a rather bleak look on what is a rather popular subject of study, we would be remiss to conclude on a more hopeful note. While, as we argued, murder mysteries, and related genres present many hard challenges, we do not think that the pursuit of these challenges can not be fruitful or worthwhile. On the contrary, and perhaps despite the title of this article, we believe that addressing this challenges is not only possible, but necessary to advance our field. And even though many of the problems we discussed are interrelated, there is room to address pieces of them at a time. Take, for example, the problem with discourse generation we presented. A complete murder mystery would need to take into account what is presented, and how subtle this presentation is done, in order to keep up the shroud of mystery. On the other hand, a more thorough investigation into attention and consciously and subconsciously perceived information can proceed outside the context of such stories. Indeed, there have been studies on the subject to some degree (MacLachlan and Logan 1993; Smith and Gevins 2004), that are missing a link back to procedural generation. Similarly, the logical (and illogical) reasoning processes of humans and the heuristics they use have been studied (Solomon 1992), but there has yet to be an operationalization of these models specifically for narrative generation. And the particular challenge of providing a wide variety of scenarios provides fertile ground for better domain engineering methods, or even automated construction of story domains (not only in the planning sense). In conclusion, we believe rather than seeing murder mysteries as a "low-hanging fruit", they are actually rather complex domains to generate stories for, and should serve as one of the longer-term goals for our discipline.

References

Amos-Binks, A.; Roberts, D. L.; and Young, R. M. 2016. Summarizing and comparing story plans. In *7th Workshop on Computational Models of Narrative (CMN 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Bangsø, O.; Jensen, O. G.; Jensen, F. V.; Andersen, P. B.; and Kocka, T. 2004. Non-linear interactive storytelling using object-oriented bayesian networks. In *Proceedings of the international conference on computer games: Artificial intelligence, design and education*.

Barbosa, S. D.; De Lima, E. S.; Furtado, A. L.; and Feijó, B. 2014. Generation and dramatization of detective stories. *SBC Journal on Interactive Systems* 5(2):39–52.

Barot, C.; Branon, M.; Cardona-Rivera, R. E.; Eger, M.; Glatz, M.; Green, N.; Mattice, J.; Potts, C. M.; Robertson, J.; Shukonobe, M.; et al. 2017. Bardic: Generating multimedia narrative reports for game logs. In *Working Notes of the AIIDE Workshop on Intelligent Narrative Technologies*.

Barros, G. A. B.; Green, M. C.; Liapis, A.; and Togelius, J. 2018. Who killed albert einstein? from open data to murder mystery games. *IEEE Transactions on Games* 11(1):79–89.

Eger, M., and Martens, C. 2017a. Character beliefs in story generation. In *Working Notes of the AIIDE Workshop on Intelligent Narrative Technologies*.

Eger, M., and Martens, C. 2017b. Practical specification of belief manipulation in games. In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Farrell, R., and Ware, S. G. 2016. Fast and diverse narrative planning through novelty pruning. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Jaschek, C.; Beckmann, T.; Garcia, J. A.; and Raffe, W. L. 2019. Mysterious murder-mcts-driven murder mystery generation. In *2019 IEEE Conference on Games (CoG)*, 1–8. IEEE.

Klein, S.; Oakley, J. D.; Suurballe, D. I.; and Ziesemer, R. A. 1971. A program for generating reports on the status and history of stochastically modifiable semantic models of arbitrary universes. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Klein, S.; Aeschlimann, J. F.; Balsiger, D. F.; Converse, S. L.; Foster, M.; Lao, R.; Oakley, J. D.; Smith, J.; et al. 1973. Automatic novel writing: A status report. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Knox, R. 1929. Ten rules for a good detective story. *Publishers' Weekly* 5:1739.

Kreminski, M.; Acharya, D.; Junius, N.; Oliver, E.; Compton, K.; Dickinson, M.; Focht, C.; Mason, S.; Mazeika, S.; and Wardrip-Fruin, N. 2019. Cozy mystery construction kit: Prototyping toward an ai-assisted collaborative storytelling mystery game. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 1–9.

Leavitt, J. D., and Christenfeld, N. J. 2013. The fluency of spoilers: Why giving away endings improves stories. *Scientific Study of Literature* 3(1):93–104.

MacLachlan, J., and Logan, M. 1993. Camera shot length in tv commercials and their memorability and persuasiveness. *Journal of Advertising Research* 33(2):57–62.

Martens, C. 2015. Ceptre: A language for modeling generative interactive systems. In *Eleventh artificial intelligence and interactive digital entertainment conference*.

Mohr, H.; Eger, M.; and Martens, C. 2018. Eliminating the impossible: A procedurally generated murder mystery. In *Working Notes of the AIIDE Workshop on Experimental AI in Games*.

Rowe, J.; Mott, B.; McQuiggan, S.; Robison, J.; Lee, S.; and Lester, J. 2009. Crystal island: A narrative-centered learning environment for eighth grade microbiology. In *Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK*, 11–20.

Ryan, J. O.; Summerville, A.; Mateas, M.; and Wardrip-Fruin, N. 2015. Toward characters who observe, tell, misremember, and lie. *Working Notes of the AIIDE Workshop on Experimental AI in Games* 2.

Ryan, J. 2018. *Curating simulated storyworlds*. Ph.D. Dissertation, UC Santa Cruz.

Samuel, B.; Ryan, J.; Summerville, A. J.; Mateas, M.; and Wardrip-Fruin, N. 2016. Bad news: An experiment in computationally assisted performance. In *International Conference on Interactive Digital Storytelling*, 108–120. Springer.

Shirvani, A.; Farrell, R.; and Ware, S. G. 2018. Combining intentionality and belief: Revisiting believable character plans. In *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Smith, M. E., and Gevins, A. 2004. Attention and brain activity while watching television: Components of viewer engagement. *Media Psychology* 6(3):285–305.

Solomon, M. 1992. Scientific rationality and human reasoning. *Philosophy of Science* 59(3):439–455.

Summerville, A. 2018. Expanding expressive range: Evaluation methodologies for procedural content generation. In *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Thorne, B. R., and Young, R. M. 2017. Generating stories that include failed actions by modeling false character beliefs. In *Working Notes of the AIIDE Workshop on Intelligent Narrative Technologies*.

Thorne, B. R.; Winer, D. R.; Barot, C.; and Young, R. M. 2019. Firebolt: A system for automated low-level cinematic narrative realization. In *International Conference on Interactive Digital Storytelling*, 333–342. Springer.

Van Dine, S. 1928. Twenty rules for writing detective stories.

Van Ditmarsch, H.; van Der Hoek, W.; and Kooi, B. 2007. *Dynamic epistemic logic*, volume 337. Springer Science & Business Media.

Wadsley, T., and Ryan, M. 2013. A belief-desire-intention model for narrative generation. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.