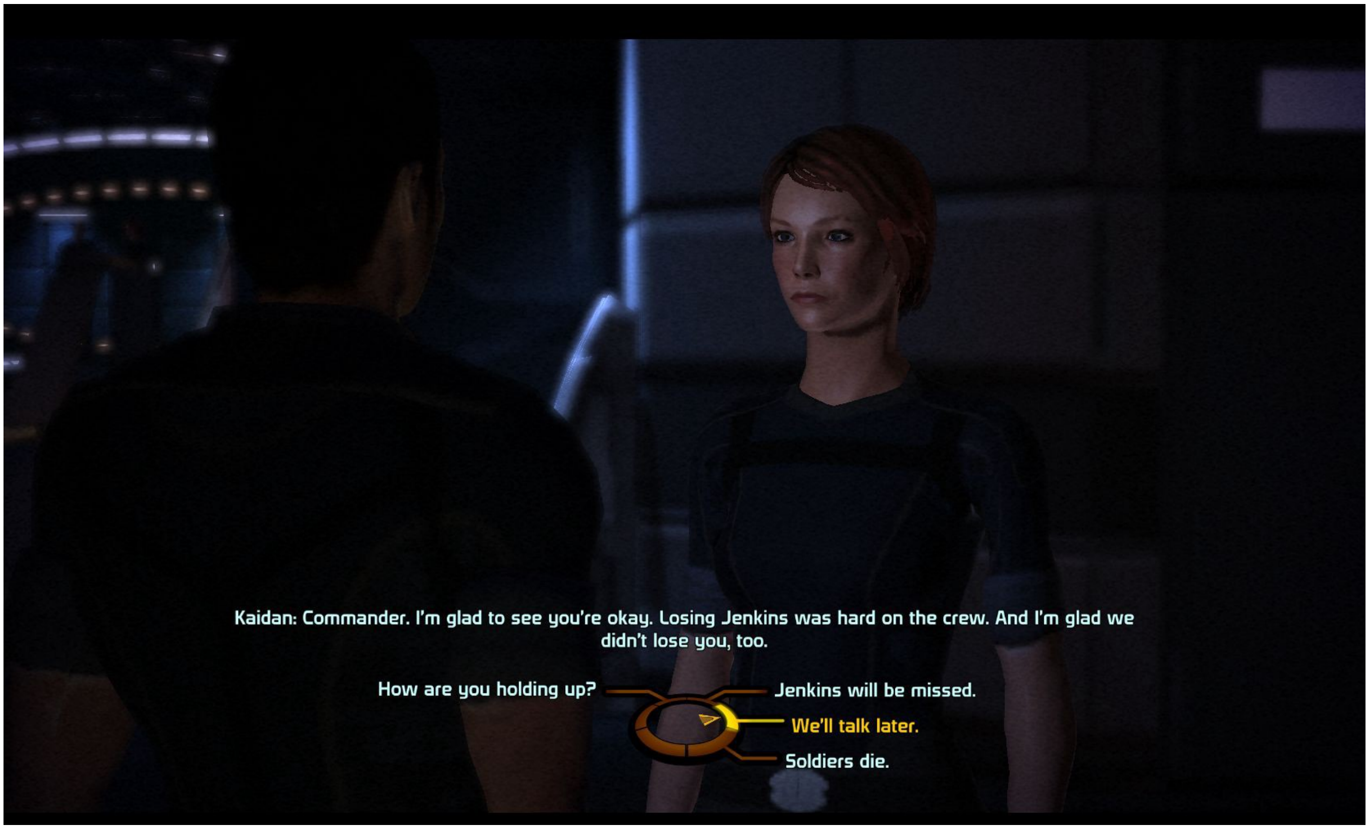
Intentional Collapse: Plausibly Human Randomized Text

Aaron A. Reed
Nov 7, 2019 · 8 min read

Here's an interesting problem: say you're writing a video game with a conversation tree, with many moments where the conversation might continue in one of several different ways. Normally, the decision of which path to take is driven by the player: they choose an option — presumably the one they find most interesting — and the line of dialogue that responds to it is duly triggered. **But what if the *system,* not the player, had to pick the most interesting way for the conversation to continue?** What strategies might it use to ensure the ongoing dialogue feels coherent and intentional, rather than purely random?



Kaidan: Commander. I'm glad to see you're okay. Losing Jenkins was hard on the crew. And I'm glad we didn't lose you, too.

How are you holding up?

Jenkins will be missed.

We'll talk later.

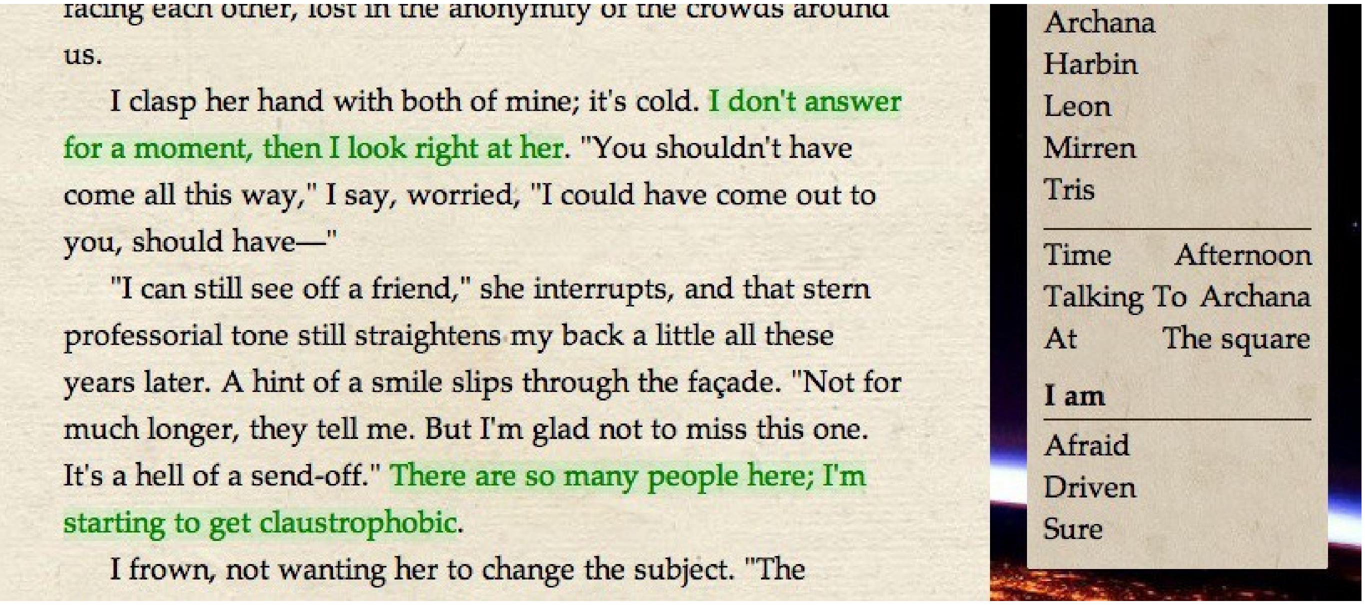
Soldiers die.

The player makes a choice in Mass Effect which controls which conversation text they see next; the conversation system itself doesn't have to worry about which choice would be best.

My upcoming novel *Subcutanean* is a book that varies its text for each new printing, choosing at hundreds of decision points from a set of possible alternatives that range in scope from single words to entire scenes. No two copies are ever quite the same. (I've written previously about why you would do this and the format I used to write it.) The code that renders a new book has a module called *Collapser* in charge of turning the master text with variants into a single rendered book, extending a metaphor of "quantum text" that exists in many states simultaneously until it's "collapsed" down to a single observed outcome. While I at first envisioned *Collapser* as a mere randomizer, choosing one option from each set of variant texts as it encountered them, as the project moved forward it became clear there were more interesting ways to make those decisions.

For one thing, *Collapser* needs to help ensure that each novel seems internally coherent: though springing from a quantum soup of possibilities, each instantiation needs to tell a single, consistent version of the story. In regular fiction writing, it's rare to make a significant change that doesn't have ripple effects elsewhere: likewise, while *Subcutanean* has many minor details that simply vary in-place, if choosing one piece of text might affect any other text in the book, that decision must be stored in a variable. The variable itself then becomes what's randomly set, and we consult it at future variable-gated decision points to know which variant to print. This is used in *Subcutanean* for everything from which major revelation to use in a climactic scene to minor details like the brand name of a grappling hook.

This might seem painfully obvious and straightforward, but I think authors of procedural text have a tendency to under-use variables, because they take extra work to set up than inline variations. As a result, output texts tend to have fewer moments of recurring consistency, and feel less like texts written by human authors. Years ago my generative text experiment *Almost Goodbye* inserted random "satellite sentences" into the writing of a hand-authored story, to customize details (like a scene's setting) for decisions the player had made. For instance, if in one playthrough a conversation takes place in a diner, the system would insert location-specific asides like "A group clinked wine glasses at a nearby table, laughing raucously." Just a couple little moments like that are enough to make it feel like a scene has been hand-written for that specific place. Mentioning something once is nice, but weaving in mentions across multiple moments and scenes is worldbuilding.



facing each other, lost in the anonymity of the crowds around us.

I clasp her hand with both of mine; it's cold. I don't answer for a moment, then I look right at her. "You shouldn't have come all this way," I say, worried, "I could have come out to you, should have—"

"I can still see off a friend," she interrupts, and that stern professorial tone still straightens my back a little all these years later. A hint of a smile slips through the façade. "Not for much longer, they tell me. But I'm glad not to miss this one. It's a hell of a send-off." There are so many people here; I'm starting to get claustrophobic.

I frown, not wanting her to change the subject. "The

Archana
Harbin
Leon
Mirren
Tris

Time Afternoon
Talking To Archana
At The square

I am

Afraid
Driven
Sure

Inserting contextual satellite sentences into a story in the author's "Almost Goodbye" (2012)

Second, you don't always need to make decisions about what text to print "just in time," i.e. right before you print it. In fact, if you make the big decisions right from the beginning, that lets you be much more clever in how you set those moments up before they arrive: again, just like in writing regular fiction. We call this "foreshadowing" when it works on a symbolic or metaphorical level, and, more prosaically, "laying pipe" when it involves setting up plot beats that later pay off.

One of my favorite examples of pipe-laying is in the movie *Aliens,* when Ripley, rescued after drifting in deep space for a century, finds her skills as a pilot are out of date. She can only get work as a dock loader operating industrial robots. This is a great bit of world-building in the moment, and sets up the next plot beat where Paul Reiser's weaselly character can use her precarious finances as leverage, but its primary job is actually laying pipe for something that happens all the way at the very end of the film: justifying how Ripley has the means to meet the alien queen on the battlefield as an equal.

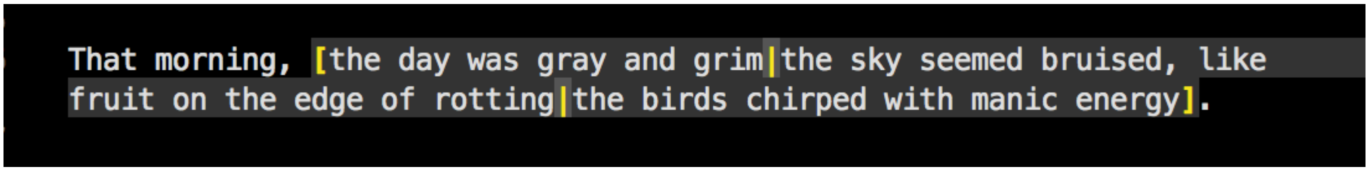
Still from Aliens (1986) by 20th Century Fox, or maybe Disney now I guess? Huh.

If, like James Cameron, a generative text system knows in advance how an upcoming pivotal moment's going to play out, it can do the work to set that moment up in earlier scenes. *Subcutanean* has many moments that lay pipe and foreshadow moments that come much later. The very first page of the book, for instance, can involve one of three different stories depending on which particular pivotal phone conversation will appear all the way up in Chapter 10. Each of these stories is a writerly response to the problem of priming the reader to appreciate the impact of that much later conversation when it arrives. To be clear, *Collapser* isn't doing any reasoning itself over those bits of text — it just knows if Variant 129 was selected there, Variant 37 has to come here — but that basic capability is critical.

An interesting side effect of editing *Subcutanean*'s quantum text was discovering which major moments *didn't* require adjusting any previous text when alternate versions were added. Sometimes that revealed interesting things about the story structure — I hadn't realized that certain beats were effectively standalone moments that might have gone anywhere — but sometimes it pointed out a weakness in my writing. In some cases it helped me realize I hadn't properly foreshadowed or laid pipe for a big revelation, and needed to go find a few places in earlier chapters to properly signpost what was coming.

. . .

But most of *Subcutanean*'s variant texts weren't big moments or recurring details: they were just alternate ways a particular bit of the story could be told. Here's a hypothetical but representative example:



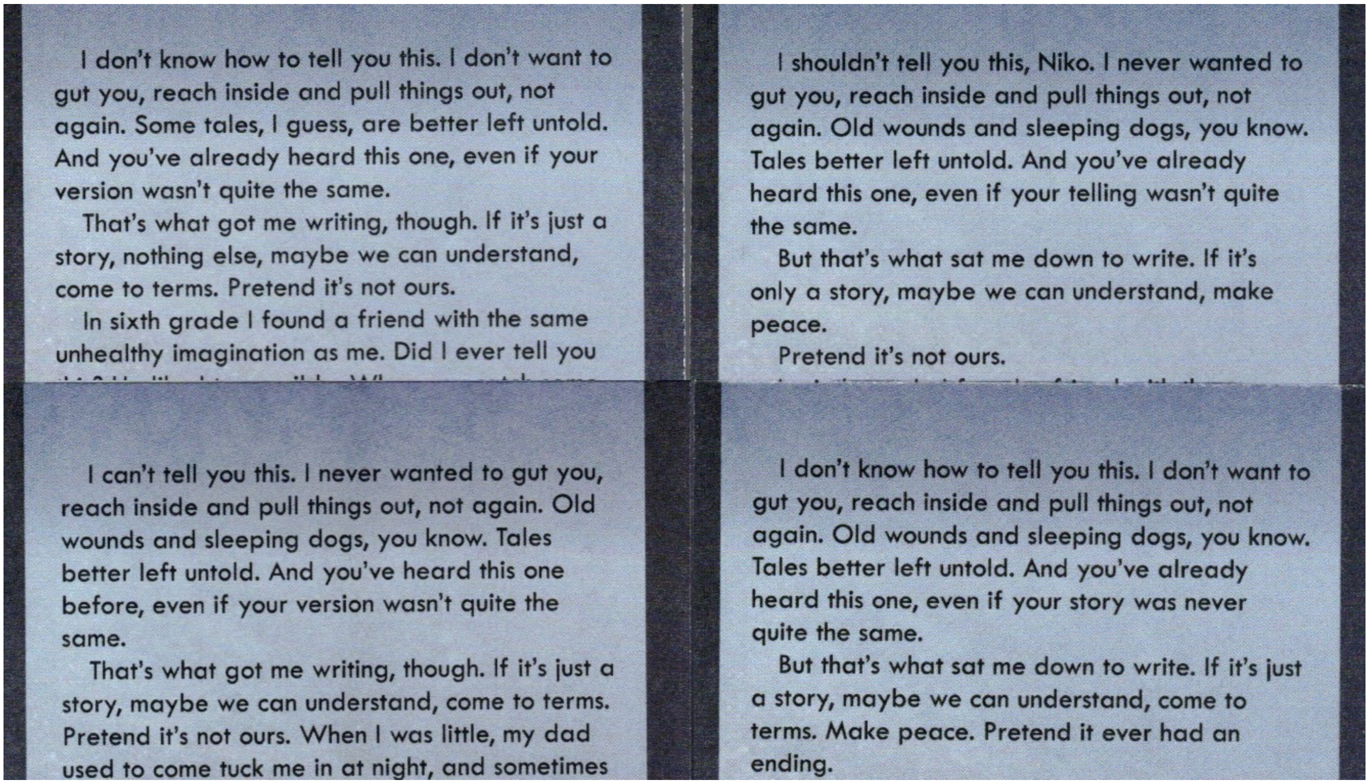
```
That morning, [the day was gray and grim|the sky seemed bruised, like
fruit on the edge of rotting|the birds chirped with manic energy].
```

I struggled for a while with whether it made sense to write these kinds of alternatives at all. Maybe it was just a bad habit I'd picked up on earlier procedural text projects: writing variant texts just because I could. A random decision between these variations seemed to serve little purpose. But then I started to wonder what would happen if *Collapser* could reason about these decisions, too?

We live in a glorious time for procedural text: anyone can make a Twitter bot, and tools abound for scraping, processing, understanding, and generating textual content. The TextBlob Python library, for instance, offers one-line sentiment analysis that will tell you if a sentence is generally positive ("the birds chirped…") or negative ("the day was gray and grim"). As I cast around for a more interesting way to choose between textual variants, I started thinking you might be able to use tools like this to create unique but consistent narrators for each telling of *Subcutanean*.

So say when *Collapser* starts to render a new book, it sets a variable *NarratorOptimist* to either true or false. Each time it has to make a decision about a set of alternate texts, it runs each one through TextBlob's sentiment analysis, and weights each option accordingly. Suddenly you have a narrator who will consistently adopt a particular style (say, preferring pessimistic utterances like "The day was gray and grim") as it assembles the book.

Subcutanean has a number of these narrator variables. Some narrators are more verbose or prefer bigger words than others; others prefer to say things as simply as possible. Some prefer more subjective language: they would rather use a bruised sky metaphor over an objective statement about its color. Some narrators enjoy alliteration ("gray and grim"). Some would prefer to paraphrase dialogue rather than quote it directly (by simply disfavoring variants that include quotation marks). What each *Collaps*ing ends up with is a unique set of narratorial preferences that together determine how texts get selected in that particular rendering. I'd found a reason to write all of those minor variations, after all: in aggregate, they added another layer of consistency to each generated book. Even if on a more subtle level, they were helping keep each possible output novel consistent, just like the pipe-laying for big moments and variables for matching up minor details.



I don't know how to tell you this. I don't want to gut you, reach inside and pull things out, not again. Some tales, I guess, are better left untold. And you've already heard this one, even if your version wasn't quite the same.

That's what got me writing, though. If it's just a story, nothing else, maybe we can understand, come to terms. Pretend it's not ours.

In sixth grade I found a friend with the same unhealthy imagination as me. Did I ever tell you

I shouldn't tell you this, Niko. I never wanted to gut you, reach inside and pull things out, not again. Old wounds and sleeping dogs, you know. Tales better left untold. And you've already heard this one, even if your telling wasn't quite the same.

But that's what sat me down to write. If it's only a story, maybe we can understand, make peace.

Pretend it's not ours.

I can't tell you this. I never wanted to gut you, reach inside and pull things out, not again. Old wounds and sleeping dogs, you know. Tales better left untold. And you've heard this one before, even if your version wasn't quite the same.

That's what got me writing, though. If it's just a story, maybe we can understand, come to terms. Pretend it's not ours. When I was little, my dad used to come tuck me in at night, and sometimes

I don't know how to tell you this. I don't want to gut you, reach inside and pull things out, not again. Old wounds and sleeping dogs, you know. Tales better left untold. And you've already heard this one, even if your story was never quite the same.

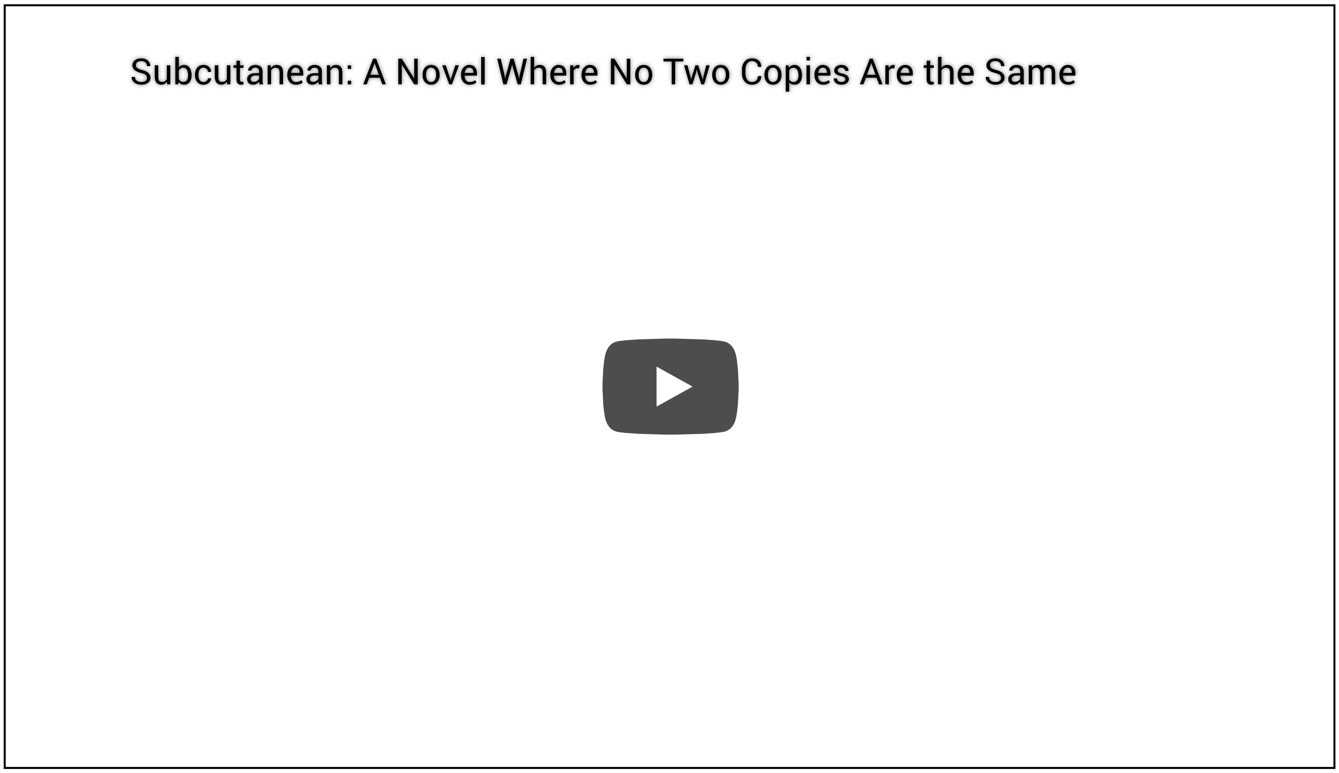
But that's what sat me down to write. If it's just a story, maybe we can understand, come to terms. Make peace. Pretend it ever had an ending.

The top of four promotional bookmarks made for Subcutanean with text from the start of the book; both small per-word variants and the start of two different opening stories (bottom of left two) are visible.

The narrator variable system has proven to be lovely in several ways. First, it's entirely automatic: I don't have to manually tag this text in any way, because each narrator variable has its own definitions for the kinds of text it likes. The narrator variables also proved a useful editing tool independent of the actual generation process. For instance, at first I had a narrator variable for preferring active versus passive voice. After a while, perhaps somewhat obviously, it became clear that the active voice was almost always better. But because I could flip a switch and say "generate me the version of this book that maximizes use of the passive voice," it was easy to find places where the passive version was weak and shouldn't appear in *any* version, and remove or replace it. Eventually I no longer needed this narrator and retired it: but it still served a useful purpose in giving me a new window onto the space of possible texts I was writing, and the ability to prune that space towards a version where all possible outputs had strong writing.

In the end, while the elevator pitch for *Subcutanean* states that its variant text is randomly shuffled, in actual fact it's rare for *Collapser* to make any decision entirely at random. Instead, each printing cuts more intentionally through the master text's large possibility space. This means there are a smaller number of possible outputs (though still a very large number): but now each one feels, hopefully, more consistently authored and more plausibly told. Each one is something closer to what a human would have written, rather than an obscure oddity from a dark corner of Borges' infinite library.

> ***Subcutanean is crowdfunding now:*** *find out how you can preorder your own unique copy. You can also follow the project on Twitter, Facebook, or Goodreads, or check out my previous design posts.*



Subcutanean: A Novel Where No Two Copies Are the Same

Books Procedural Text Randomness Generative Art Interactive Fiction