A Selective Summary of *Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence*

Nikhil Singh* MIT Media Lab
Guillermo Bernal* MIT Media Lab
Daria Savchenko* Harvard University
Elena L. Glassman Harvard University

Abstract

While developing a story, novices and published writers alike have had to look outside themselves for inspiration. Language models have recently been able to generate text fluently, producing new stochastic narratives upon request. However, effectively integrating such capabilities with human cognitive faculties and creative processes remains challenging. We propose to investigate this integration with a multimodal writing support interface that offers writing suggestions textually, visually, and aurally. We conduct an extensive study that combines elicitation of prior expectations before writing, observation and semi-structured interviews during writing, and outcome evaluations after writing. Our results illustrate individual and situational variation in machine-in-the-loop writing approaches, suggestion acceptance, and ways the system is helpful. Centrally, we report how participants perform *integrative leaps*, by which they do cognitive work to integrate suggestions of varying semantic relevance into their developing stories. We interpret these findings, offering modeling and design recommendations for future creative writing support technologies.

1 Introduction

Much remains unexplored about how emerging methods in AI, machine learning, and natural language processing might influence creative writing, in part due to the ambiguity and variability of human writing processes. These processes go beyond the linear projection from idea to a full text; research shows how planning narratives, translating ideas into visible textual material, and reviewing are all happening and interacting throughout the process rather than simple sequential stages (Nold, 1981; Flower and Hayes, 1981). However, this is a very familiar process for humans when communicating through writing; as every writer knows, having good ideas does not automatically produce a good text progression. The need for that "good idea" to be anchored and developed so that the reader can be invested takes a great deal of effort. In today's world, language generation models like GPT-2 (Radford et al., 2019), GPT-3 (Brown et al., 2020), and new ones coming down the line are typically silent on the inner processes of negotiation and decision that a human writer is working through. Additionally, contributions from these systems might take forms to influence writing other than text; writers are able to engage multiple perceptual channels through their work: they may activate multisensory imagination through evocative imagery, invoking auditory and olfactory phenomena, and other forms of sensory description.

We investigate how participants engage with a multimodal writing support system that bridges generated writing suggestions with multimedia retrieval to produce concept representations simultaneously in sight, sound, and language. We pair this interface with an extensive study that combines surveys, interaction, and semi-structured interviews during observed, think-aloud writing sessions. We examine and report in detail how participants receive, consider, and integrate suggestions from an intelligent tool into their writing. We explore prominent axes of individual and situational variation in these integrative behaviors, noting the different kinds of "leaps" participants make to understand suggestions and make the necessary compositional decisions to incorporate new information contained in them, ranging from copying and pasting to re-writing core aspects of their entire story.

In summary, our findings suggest that participants perform different kinds of *integrative leaps*, involving cognitive work to make suggestions useful to their writing. We interpret these and make commensurate design recommendations for future creative writing support tools. This paper is a selective summary of Singh, Bernal, Savchenko, and Glassman, 2022, focused on integrative leaps.

2 Related Work

2.1 Writing Support

Our central focus is the *process* of writing, and what this involves internally as it relates to interpreting and integrating incoming suggestions. On this topic, Flower and Hayes (Flower and Hayes, 1981) describe what they term a cognitive process theory of writing. They model several components as part of this: the task environment includes text produced upto a given point, as well as the rhetorical problem at hand, and the writing process(es) involve planning (generating ideas, organizing them, and setting goals), translating (transforming ideas into visible text), and reviewing (evaluating and revising). Our study examines how suggestions impact some of these kinds of processes (e.g. planning, translating, and reviewing).

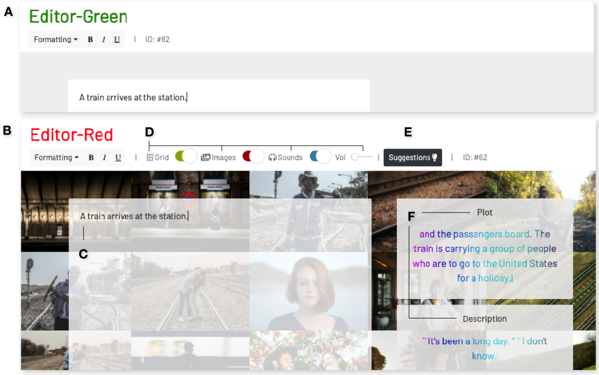
2.2 Interpretive Perspective

We approach our observation of participants' interaction through the lens of *interpretation*, which, as a concept, has been used in a number of papers in HCI (Sengers and Gaver, 2006; Lamb et al., 2018; Nake, 1994; Bardzell and Bardzell, 2016).

The interpretive perspective we maintain in this work is informed by an aspiration in anthropology to make visible the alignments of factors of interaction that might otherwise go unnoticed due to common-sense understanding. Building upon the dichotomy of social theory concepts of understanding as causal explanation (*erklären*) versus understanding as interpretation (*verstehen*), we specifically follow Max Weber's distinction (Weber, 1949) between *explanation* that captures the causal sequence of actions and *understanding* that attends to the meaning of those actions. Our research aims to analyze the interaction of the person with the AI system from the perspective of the latter, i.e., "meaning"—the meaning of actions from the point of view of the participants, who organically construct meaning in the process of engaging with complex systems. As such, interpretation in this research is a form of understanding that makes it possible to discern the meaning production that occurs within the interaction between the human and the AI system.

2.3 Explanatory Models of AI

We use the term "explanatory models" to refer to the super-set of two kinds of conceptual representations of computational systems, commonly referred to as "mental models" (Bansal et al., 2019) and "folk theories" (Eslami et al., 2016; DeVito et al., 2017) respectively. Human-AI researchers often use the concept “mental model of AI," informed by psychology and cognitive science. These are considered important for success in human-machine (and -human) collaboration, and while they offer insight into cognitive representations of a system's operation developed through experience, intuitive theories about the world structure cognition (Gelman and Legare, 2011). Folk theories are expectations based on some experience, but are not necessarily systematically checked (Rip, 2006). In this paper's results, we focus on suggestion integration and these explanatory models implicitly structure our investigative approach. We also capture explanatory models in greater detail through surveys, as detailed in our full paper.


A
Editor-Green
Formatting ▾ B I U | ID: #62
B
Editor-Red
D
E
Formatting ▾ B I U | Grid Images Sounds Vol | Suggestions | ID: #62
C
F
Plot
Description
A train arrives at the station.

A train arrives at the station.

and the passengers board. The train is carrying a group cf people who are to go to the United States for a holiday.

"It's been a long day." "I don't know.

Figure 1: Our experimental writing interfaces.

3 System Prototype

Our experimental prototype consists of two writing interfaces: **Editor-Green**, a minimal "blank page" tool, and **Editor-Red**, our augmented multimodal tool. To minimize cognitive bias when conducting our user study, we chose to give names to the editors that would seem roughly equivalent. The system also contains a server that runs language models, as well as a real-time database to track inputs, responses from the server, and interactions, e.g., interface settings. Fig. 1 shows both interfaces, including an active multimodal response in **(B)** with images and sounds. Fig. 2 shows the underlying data flow through the system architecture that makes these interfaces possible.

We fine-tuned the same language model on two different datasets, producing two final models. The base model is a medium-sized GPT-2 architecture with pre-trained weights obtained from huggingface[1]. The first experimental model is fine-tuned on a corpus of movie summaries (Bamman et al., 2013), which we observe tend to contain high-level plot components and event sequences. As such, we label suggestions arising from this model as "Plot" suggestions. The second is fine-tuned on a writing prompts dataset (Fan et al., 2018), which features prompts and story responses taken from a prominent online forum for amateur fiction. Following the observation by Fast et al. that amateur fiction "tends to be explicit about both scene-setting and emotion, with a higher density of adjective descriptors" (Fast et al., 2016) as well as our own review of this dataset and the fine-tuned model, we label this second experimental model's outputs as "Description" suggestions.

4 Study

4.1 Participants

Participants were recruited through large department and living group mailing lists at R1 universities, including one social sciences department and several Computer Science-adjacent lists, as well as a post on Reddit. As a pre-condition, applicants filled out a form confirming that they were fluent in English and at least 18 years old. We maintained a balanced pool of participants who identified as native and non-native English speakers, and with and without Computer Science backgrounds.

Twenty-seven participants completed the writing tasks. Data from four had to be excluded due to firewall-related issues, mid-session server problems, and unwillingness to complete the task as instructed. All participants reported having at least a high school diploma. Participants' ages ranged from 18 to 45, with 48% of participants in the range of 18-22. 65% of participants reported that English was their first language. When asked "Do you struggle with writing?", 78% of the participants responded affirmatively.

4.2 Study Structure

After consenting, participants were given a short ($\leq$5m) overview of the study procedure, followed by a 10-minute introductory survey. They then completed the first writing task (20m), with either **Editor-Green** or **Editor-Red** depending on their (order-counterbalanced) group assignment. They were instructed to write a story using one of the following prompts: *The phone began to ring* or *A train arrives at the station* (alternating prompts between groups to control for the effect of the prompt). Both prompts were designed to be short, somewhat vague, and contain the beginning of some action (phone call and train arrival). Participants then completed a second writing task (20m) with the other editor. After each writing task, participants completed the corresponding follow-up survey, i.e., for **Editor-Green** (<5m) or **Editor-Red** (10m). Finally, all participants completed a survey comparing the two writing experiences, plus a demographics/background section. The overall duration was about 75 minutes, with a $25 Amazon gift card as compensation.

Two researchers separately conducted study sessions via Zoom videoconferencing. The sessions, including screen-sharing (except when answering surveys), were recorded with permission, and the researchers took notes throughout. While writing, participants were explicitly encouraged to comment and react aloud as they wrote, processed information, and responded to incoming suggestions and media. The interviewers periodically prompted participants to communicate about their thought processes and experiences.

4.3 Observation and Thick Description

Participants commented on how they wrote outside vs. within the study, explained their ideation process, their judgments of the system's suggestions, how they were making decisions to incorporate suggestions or not, and gave reasons. This enabled us to produce "thick description" (Geertz, 1973). Observing this interaction allowed us to capture the reasoning of participants for incorporating or ignoring suggestions, and also glean how participants make sense of their interaction with the system and their strategy on structuring this interaction to support their writing.

4.4 Data Analysis

We analyzed (1) logged texts, suggestions, and interactions, (2) transcripts of think-aloud writing sessions, (3) interviewers' notes, and (4) survey responses. One coder inductively coded the data in two rounds, followed by rearranging codes and turning in-vivo codes either into new themes, or adding them to existing codes. At this point, a second researcher did their round of coding and partially re-coded the data. The two coders dis-

[1] https://huggingface.co/

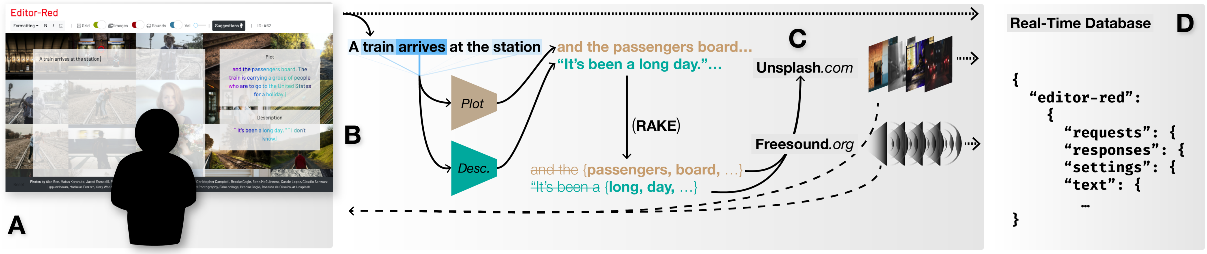

Editor-Red
Formatting
Grid
Images
Sounds
Vol
Suggestions
ID: #62
A train arrives at the station|
Plot
and the passengers board. The train is carrying a group of people who are to go to the United States for a holiday.|
Description
" It's been a long day. " " I don't know.|
A
B
A train arrives at the station
Plot
Desc.
and the passengers board...
"It's been a long day."...
(RAKE)
~~and the~~ {passengers, board, ...}
~~"It's been a~~ {long, day, ...}
C
Unsplash.com
Freesound.org
Real-Time Database
D
{
"editor-red":
{
"requests": {
"responses": {
"settings": {
"text": {
...
}

Figure 2: Flow of data through our system.

cussed and reached agreement on the codes. A third round of coding by a third researcher was done to align and streamline all the codes. In this paper, we focus on results from (2) and (3), although our interpretation is informed by (4), and (1) helped us reconstruct writing sessions' contents to carefully examine them later on.

5 Results

To provide a more granular exposition of the suggestion integration patterns, we detail a collection of *integrative leaps*. These leaps describe how participants alter the meaning and structure of their narratives while integrating suggestions. Our data on suggestion integration contains 47 instances of integrative leaps from 19 (out of 23) participants, when the observing researcher identified a moment that the participant engaged with and actively incorporated suggestions from the system. Participants often explicitly commented on their integration process in addition to our observations and analysis. P6, P8, P16, and P22 did not appear to incorporate **Editor-Red**'s suggestions in any identifiable way.

The integrative leaps can be analyzed along a number of axes, summarized in Table 1. First, we consider the "edit" distance (e.g., lexical, semantic, etc.) between the suggestion as presented to the user and as incorporated into the story. We characterize these as *direct integration* ($N = 30$), e.g., verbatim or restructured verbatim for a textual suggestion or a textual analogue of the object or idea represented in a visual or auditory suggestion (Figure 3), or *indirect integration* ($N = 17$), where participants' explanations highlighted modifications they made in the process of suggestion incorporation (Figure 4).

Second, for both direct and indirect integration, we look at how incorporated suggestions relate to global aspects of their story's direction and most prominent elements. When participants used suggestions to explore new lines of narration, we call it *exploratory integration* ($N = 28$), in contrast to taking suggestions to continue with their chosen narrative by adding more details, which we call *confirmatory integration* ($N = 19$). This forms the horizontal axis of Figures 3 and 4.

Finally, we attend to the role suggestions play in creative problem solving during both direct and indirect integration. Do they simply solve a localized problem by "closing" some aspect of the narrative in a necessary, analytical, or expected way? For example, naming a character that has already been described, or explaining why a character went from place A to place B if both of those events have been established. Or do they "open" up options to consider, resulting in abstract, novel, or unexpected events, patterns, or directions? We describe these as *convergent integration* ($N = 31$) and *divergent integration* ($N = 16$), the vertical axis of Figures 3 and 4.

Further examples of these leaps are given in Appendix B.

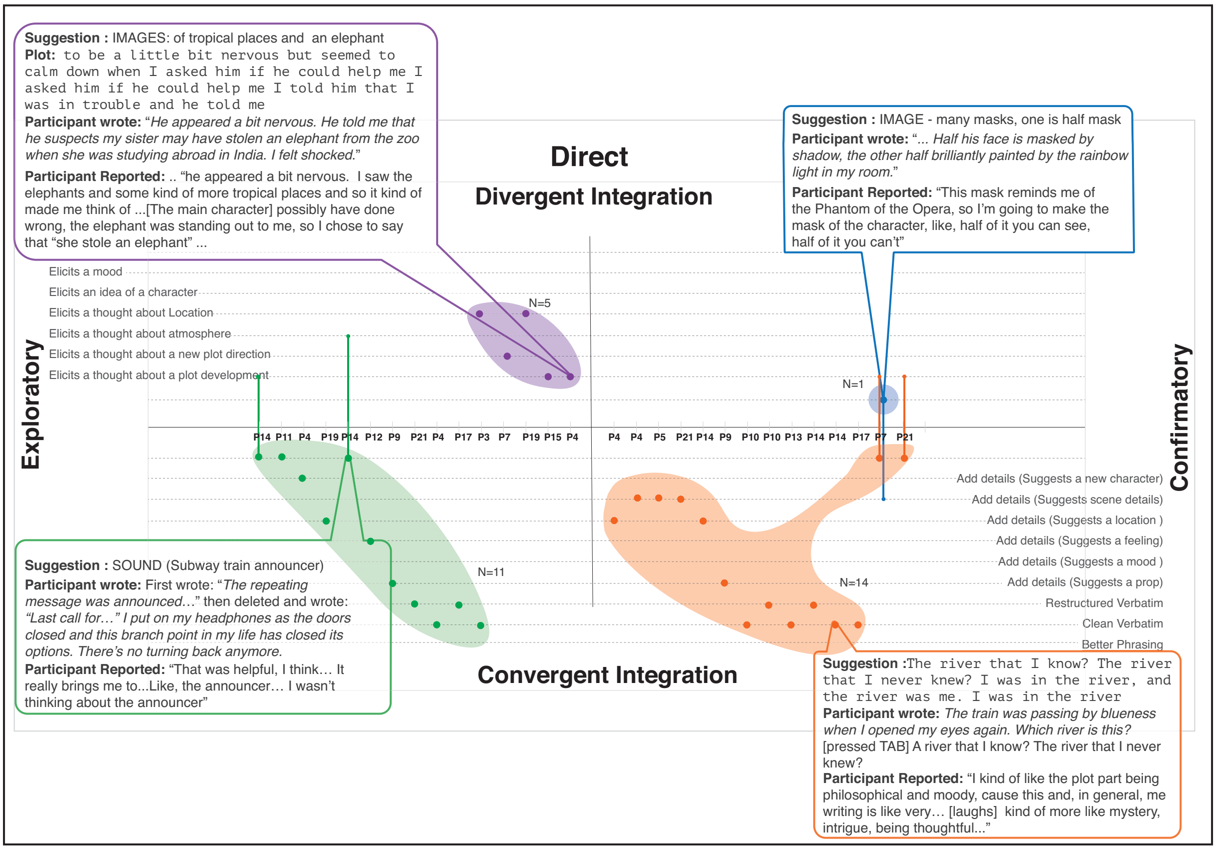
Based on Figures 3 and 4, we can see that participants generally made more *direct* leaps than *indirect* leaps; most direct leaps were also *convergent* (though there are several exceptions, as with *exploratory-divergent*), and *indirect* leaps were slightly biased toward *divergent* integrations. The following are some examples of these distinct types of integrative leaps:

Leap 1. Input (summary): "...There is a matter we have to attend to first before we will let anyone be checked in," said the officer calmly.
Suggestion: `I had been waiting for this moment for years.`
Integration: The train was already late and now this; who knows how long before I get on board?! I can't be late... maybe if I start now, I can drive over to... no, no, no. I'll never make it that way.

P5 was writing a slow-paced descriptive story using the prompt "A train arrives at the station."

Axis/Label	Description	N
Direct	(Almost) verbatim text or textual analogue of image/sound	30
Indirect	Modified incorporation, or inspires another idea	17
Exporatory	Explore new lines of narration	28
Confirmatory	Continue with existing narrative by adding more details	19
Divergent	Abstract novel, or unexpected events, patterns, or directions	16
Convergent	Necessary, analytical, or expected	31

Table 1: A summary of our labels for integrative leaps, through our coding process. We note the label (within each axis), a brief description of how the suggestions impact or are integrated into the story, and the count (out of 47 total instances that we identified).



Suggestion : IMAGES: of tropical places and an elephant
Plot: to be a little bit nervous but seemed to calm down when I asked him if he could help me I asked him if he could help me I told him that I was in trouble and he told me
Participant wrote: "*He appeared a bit nervous. He told me that he suspects my sister may have stolen an elephant from the zoo when she was studying abroad in India. I felt shocked.*"
Participant Reported: .. "he appeared a bit nervous. I saw the elephants and some kind of more tropical places and so it kind of made me think of ...[The main character] possibly have done wrong, the elephant was standing out to me, so I chose to say that "she stole an elephant" ...

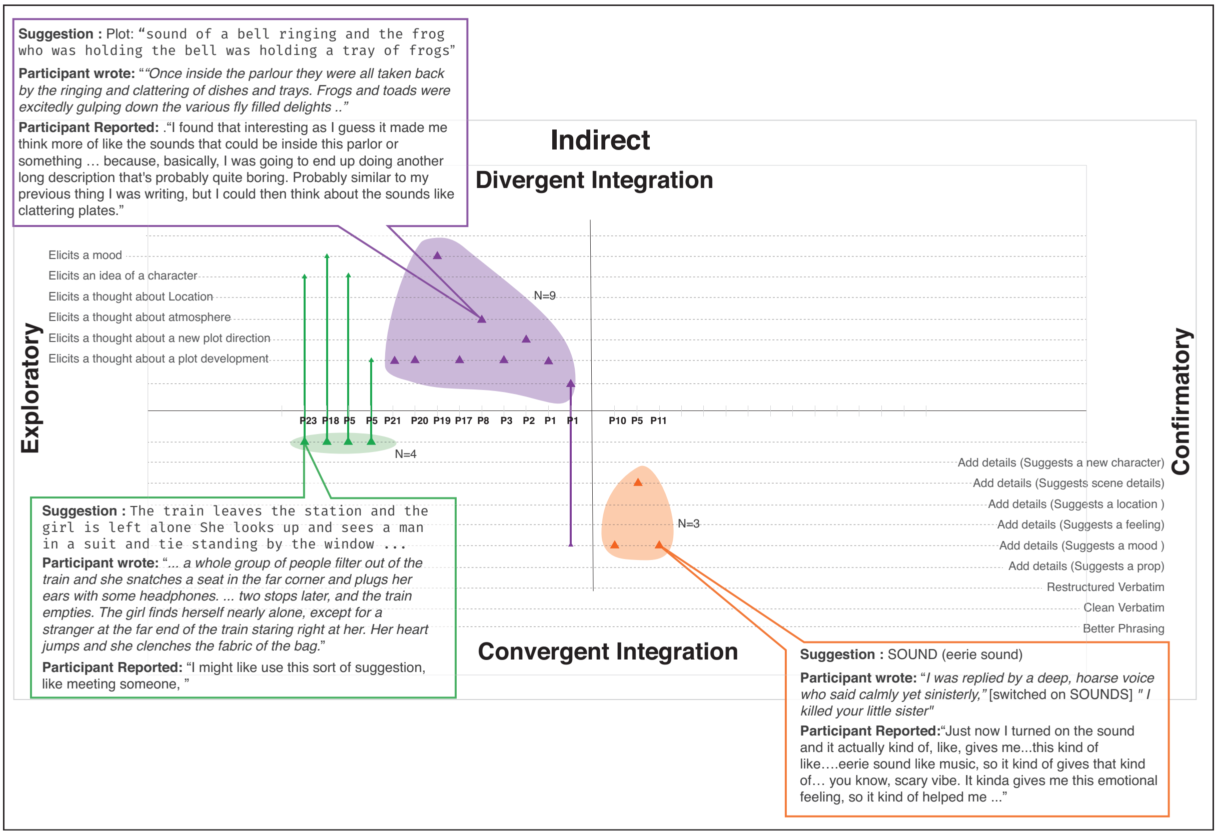
Suggestion : IMAGE - many masks, one is half mask
Participant wrote: "*... Half his face is masked by shadow, the other half brilliantly painted by the rainbow light in my room.*"
Participant Reported: "This mask reminds me of the Phantom of the Opera, so I'm going to make the mask of the character, like, half of it you can see, half of it you can't"

Suggestion : SOUND (Subway train announcer)
Participant wrote: First wrote: "*The repeating message was announced…*" then deleted and wrote: *"Last call for…" I put on my headphones as the doors closed and this branch point in my life has closed its options. There's no turning back anymore.*
Participant Reported: "That was helpful, I think… It really brings me to...Like, the announcer… I wasn't thinking about the announcer"

Suggestion : The river that I know? The river that I never knew? I was in the river, and the river was me. I was in the river
Participant wrote: *The train was passing by blueness when I opened my eyes again. Which river is this?* [pressed TAB] A river that I know? The river that I never knew?
Participant Reported: "I kind of like the plot part being philosophical and moody, cause this and, in general, me writing is like very… [laughs] kind of more like mystery, intrigue, being thoughtful..."

Figure 3: Diagram of participants' exploratory/confirmatory and divergent/convergent *direct* integrative leaps.

At some point, the protagonist was stopped by an officer and told that the train would not be boarding as there were some issues. P5 requested a suggestion and one of the suggestions was "I had been waiting for this moment for years." The participant wrote: "The train was already late and now this; who knows how long before I get on board?! I can't be late… maybe if I start now, I can drive over to… no, no, no. I'll never make it that way." To the interviewer who ran the session, there was no obvious connection between the suggestion and what the participant subsequently wrote. However, P5 explained that the suggestion "I had been waiting for this moment for years" made them think "more of a frustration for the train being late" and they imagined that there was something that the character was supposed to get to on time in another city. So this idea was translated into making the character impatient. We label this *indirect* (waiting for years to frustration and impatience), *exploratory* (switches from describing scene and events to narrating internal dialogue about the character's feelings) and *convergent* (an expected reaction to the situation that describes the effect of the train's lateness).


Indirect
Divergent Integration
Convergent Integration
Exploratory
Confirmatory
Elicits a mood
Elicits an idea of a character
Elicits a thought about Location
Elicits a thought about atmosphere
Elicits a thought about a new plot direction
Elicits a thought about a plot development
Add details (Suggests a new character)
Add details (Suggests scene details)
Add details (Suggests a location)
Add details (Suggests a feeling)
Add details (Suggests a mood)
Add details (Suggests a prop)
Restructured Verbatim
Clean Verbatim
Better Phrasing
P23 P18 P5 P5 P21 P20 P19 P17 P8 P3 P2 P1 P1 P10 P5 P11
N=9
N=4
N=3

Suggestion : Plot: “sound of a bell ringing and the frog who was holding the bell was holding a tray of frogs”

Participant wrote: “*“Once inside the parlour they were all taken back by the ringing and clattering of dishes and trays. Frogs and toads were excitedly gulping down the various fly filled delights ..”*

Participant Reported: .“I found that interesting as I guess it made me think more of like the sounds that could be inside this parlor or something … because, basically, I was going to end up doing another long description that's probably quite boring. Probably similar to my previous thing I was writing, but I could then think about the sounds like clattering plates.”

Suggestion : The train leaves the station and the girl is left alone She looks up and sees a man in a suit and tie standing by the window ...

Participant wrote: “*... a whole group of people filter out of the train and she snatches a seat in the far corner and plugs her ears with some headphones. ... two stops later, and the train empties. The girl finds herself nearly alone, except for a stranger at the far end of the train staring right at her. Her heart jumps and she clenches the fabric of the bag.*”

Participant Reported: “I might like use this sort of suggestion, like meeting someone, ”

Suggestion : SOUND (eerie sound)

Participant wrote: “*I was replied by a deep, hoarse voice who said calmly yet sinisterly,”* [switched on SOUNDS] *" I killed your little sister"*

Participant Reported:“Just now I turned on the sound and it actually kind of, like, gives me...this kind of like….eerie sound like music, so it kind of gives that kind of… you know, scary vibe. It kinda gives me this emotional feeling, so it kind of helped me ...”

Figure 4: Diagram of exploratory/confirmatory and divergent/convergent *indirect* integrative leaps made by the participants.

Leap 2. Input (summary):. . . the detective met me at the door. He appeared
Suggestion: { , ,...} + `to be a little` **`bit nervous`** `but seemed to calm down when I asked him...`
Integration: He appeared a **bit nervous**. He told me that he suspects my sister may have stolen an **elephant** from **the zoo** when she was studying abroad in India. I felt shocked.

P4, following the prompt "The phone began to ring," was developing a story about a police detective who asked the narrator to come to the police station because their sister was in trouble. P4 felt unsure as to how to continue and what it could be that the detective could have been accusing their sister of. This participant was really perplexed with what in their previous writing could have prompted the subsequent suggestions involving zoos, animals, and tropical places (these were in the retrieved images) but still decided to go ahead and integrate the suggestions into their story. P4 explained their reasoning in integrating the system's suggestion: “I don't know why these images popped up and how they are related to what I wrote before. But I saw the elephants and some kind of more tropical places and so ...I was thinking what could she possibly have done wrong that she could be in trouble and so the elephant was standing out to me, so I chose to say that 'she stole an elephant.'" The participant concluded their story by writing, in an attempt to rationalize and make sense of the elephant's role:

> “ I knew my sister loved animals, especially larger ones, but I never would have expected this. Where would she have left it? I had so many questions. I asked if I could talk to my sister. "Did you steal an elephant??" "I don't know what he's talking about. I've never seen it before." ”

In this example, image suggestions (e.g. elephants) and text suggestions ("nervous") were verbally expressed in the person's writing, so we call this integration *direct*. Also this suggestion elicits a thought about a plot development (*exploratory*), and opens up new questions for the story (*divergent*) rather than closes any existing ones.

Leap 3. Input (summary):. . . We were neighbors growing up, so I was pretty close with her sister too.
Suggestion: 🔊 1; 2 (crowds)
Integration: In my mad dash to get to the hospital, I forgot that the 4th of July **parade** was happening today just blocks down. . .

In P21's story, they were describing a character driving to the hospital and the system gave auditory suggestions that P21 described as chanting and explained: "There is chanting happening, it makes me think she got into traffic because there's a protest happening, ...or a parade." So P21 wrote: "In my mad dash to get to the hospital, I forgot that the 4th of July parade was happening today just blocks down from the hospital. I'm stuck at an intersection where the parade is passing by..." In this example, sound suggestions prompted the participant to think about what could have caused the traffic, so we call this integration *indirect*. The integration of this suggestion also significantly altered the course of the plot (*exploratory*) creating new avenues of the story development (*divergent*).

6 Discussion

Several participants rejected suggestions for a perceived lack of coherence or relevance to their developing texts, which comports with prior work on language model assisted writing (Calderwood et al., 2020; Clark et al., 2018). Building on this, we have also shown that several others in our study did not see this as an obstacle to working with the system and in some cases appreciated less immediately semantically relevant suggestions and were even able to incorporate ideas from less linguistically coherent suggestions. Our expectation from observing participants is that this has primarily to do with a difference in participants' approach to and needs during creative writing. As such, the relevance of suggestions may not be a simple variable to always aim toward maximizing; rather, the optimal level of relevance might vary by writer. Sometimes, it might also vary depending on other circumstances; for example, some participants noted that less relevant suggestions likely required more time to integrate, and that they might do so given additional time to write. This may also be reflected in the fact that on average, participants wrote less text in **Editor-Red** than in **Editor-Green**, though we note this is also related to other aspects of the interaction in our study, e.g., novelty of the interface, talking more while using **Editor-Red**, etc.).

The ambiguity in assessing relevance extended to the multimodal concept representations; even when not used directly, their contribution to the environment might vary with relevance. For example P8, who didn't visibly incorporate any suggestions, noted they were "impressed that the sound suggestions seemed to pick up on the creepy, suspenseful tone of the story right away, and it could be helpful if the image suggestions followed the tone more closely" as compared with P5 who wrote that the "sound wasn't directly influencing my ideas but having background noise was relaxing."

Balancing relevance with variety is likely to be important in making suggestions useful to participants, in our assessment. Participants especially noted the homogeneity of images: "I mentioned a phone and the grid overlay just shoved several iterations of smartphones, it would be nice if it could show different types of telephones" (P20). This also extended to demographic factors: "there's just a bunch of white guys staring at me and I don't know why" (P2) and "they are all images of straight blonde Caucasian women" (P5). We noted that these instances were not directly related to query material, indicating that these might reflect broad biases in available images.

Technical approaches to generative modeling and information retrieval to support creative processes should, in our view, consider individual and situational variation in relevance and variety. Modeling this is likely non-trivial and raises questions such as: what is relevant when and to whom? When are precise, logical suggestions needed, and when are surprising, unusual suggestions needed? The integrative leaps we have reported on suggest the practical challenges in automatically inferring this trade-off, or even reducing it to a simple, one-dimensional control. A helpful source of information in our case is the writers; finding channels for writers to communicate their personal stylistic and contextual narrative needs to both interfaces and the underlying models, for example in natural language or by providing examples, may help these systems robustly support creative expression by being flexible and allowing users to clearly and naturally communicate their needs and intentions.

6.1 Sources of Support

A wide range of participants' comments highlight that the system acted as a support tool in diverse ways beyond directly offering useful suggestions. Those participants who actively integrated the system's suggestions admitted that **Editor-Red** was structuring their process of writing. For instance, P1 admitted that they found themselves at a certain point "writing *for* the suggestions," seeing **Editor-Red** as "a form of motivation to continue writing" in order to get better suggestions. P3 commented that **Editor-Red** helped them "keep going" and "continue along" with their writing when they otherwise would have stopped.

In the "blank page" writing with **Editor-Green**, 10 participants out of 23 visibly relied on cultural (books, TV shows, music videos) and personal (memories, personal experiences, and immediate surroundings, e.g., describing what one can see from the window) references. For example, P8 writing in **Editor-Green** with the prompt "A train arrives at the station," explained they were thinking about "the train station and Anna Karenina, kind of thing." P9 writing in **Editor-Green** with the prompt "The phone began to ring" explains that "the phone" made them think about a landline, a landline made them think about a hotel, and that, in turn, made them think about the last trip they had when they were staying in a hotel, which prompted a subsequent description they made in **Editor-Green** (post-**Editor-Red** writing). In **Editor-Red**, 5 of these 10 did not visibly use any cultural or personal references in their writing.

6.2 Dynamics of suggestion integration

Ideas for writing often came from participants' readiness to do cognitive work in extending, adjusting, and altering suggestions and/or prior text to better suit the combination of text they had written and either any thoughts in their mind about how to proceed (*confirmatory*) or ideas about altering the narrative to lead in a new direction (*exploratory*). One constraint we observed is the possibility of an easy transition. This, in turn, is individually and contextually varying. Those participants whom we identified as willing to cooperate with **Editor-Red** and incorporate its suggestions, did not seem to mind suggestions being "absurd," "crazy," and "out there." These suggestions sometimes led to considerable changes to the subsequent and prior narratives; participants made decisive creative moves when they were willing to engage in this way.

The transition towards a suggestion that is unexpected and/or unrelated to the input text is dependent on the readiness and motivation of a user to the requisite cognitive and/or emotional work toward a meaningful synthesis of elements. These observations align with Freiman's characterization of the writer's drafting process, involving a "state of unknowing", a "kind of faith" that something will emerge from the drafting, and ultimately how "something that perhaps lacked cohesion or structure now becomes more concrete or coherent in the making of the text" (Freiman, 2015). Freiman suggests this happens by the writer making cognitive, affective, linguistic, and other creative decisions through a series of drafts and changes.

How are distant suggestions able to be meaningfully integrated into users' existing narratives? Integrating **Editor-Red** suggestions sometimes involves a considerable amount of cognitive reorganization of narrative information, in the sense of reorganizing what one already knows (e.g. Piaget's equilibration (Piaget, 1985)) or, in this case, has already written. One possible mechanism for this is self-explanation, which is an attempt to make sense of new information by explaining it to oneself (Chi, 2000). Here, self-explanation may provide an inferential process to reorganize the narrative by finding possible connections and associations, similarity, extracting abstract properties, or making referential links (for example, as we described earlier with P4 having the precondition of a crime, seeing an elephant that seems irrelevant, and explaining the presence of the elephant by making it the object of the crime involved). Other possible mechanisms for combining distant concepts have also been described in prior literature, such as causal reasoning (Kunda et al., 1990), comparison and construction (Wisniewski, 1997), conceptual integration or "blending" (Turner and Fauconnier, 1999; Dancygier, 2006), and satisfying constraints like diagnosticity, plausibility, and informativeness (Costello and Keane, 2000).

Earlier work has illustrated how completely unrelated ideas and unusual word combinations can be evocative and productive for creative writing (Ward and Lawson, 2009; Card, 1990; Donaldson, 2008). In the case of causal reasoning, the surprisingness of combinations may provoke additional and exploratory processes and thereby the production of creative ideas (Kunda et al., 1990). Distant suggestions might also be useful by explicitly prompting more critical evaluations of written content, i.e. what Flower and Hayes call "evaluating" and "revising" (Flower and Hayes, 1981). We can model distant suggestions with such semantic difficulties as we observe as being useful inefficiencies which prompt critical evaluations of drafts and suggestions, metacognitive reflection about narrative development, and ultimately axes for more substantial narrative reorientation, where otherwise there would be no prompt or incentive to re-engage with and reconsider prior thoughts and writing. More work is needed to examine this possibility in detail.

6.3 Language modeling

More modeling power can result in increased coherence and relevance, especially for *convergent* integrations and as processed sequences (i.e. stories) get longer, if pretrained on appropriately large and diverse datasets. Fine-tuning for stylistic personalization may help with *confirmatory* integrations, and fine-tuning on creatively-oriented text may help several kinds of integrative leaps. In parallel, models with implicitly richer knowledge bases (Petroni et al., 2019) may also find interesting relations with aspects of users' writing, and assist them in performing contextually appropriate and creatively fulfilling integrations more *direct*ly.

However, larger models are typically slower, more difficult to fine-tune and host, and increasingly closed-source, expensive to obtain access to, and private. Additionally, we noted many instances in which the cognitive work done by participants was the operative force in making suggestions helpful and ultimately able to contribute to their writing, especially for *indirect*, *exploratory*, and *divergent* integrations. For these participants, writing styles, and situations, larger language models may not necessarily help much, but would incur costs in interactivity, which were already pointed out by some participants in our current prototype. In our case, suggestions typically took 3-5 seconds after requests (given that we were running two separate fine-tuned models, extracting keywords, running searches, etc.), depending on the length of the input text; larger models may take much longer (one estimate of GPT3-Davinci: 147WPM (Branwen, 2020)) and are challenging to host and serve interactive requests with due to the resources needed.

Some participants also indicated a desire to influence or control suggestions with prior information, e.g. high-level story goals, moods, feelings, and ideas. While relevance can already be expressed to language models at *sampling* time to some extent, through decoding parameters like *temperature*, the ability to semantically "steer" (Dathathri et al., 2019) relevance towards more fruitful integrations, rather than expressing it as a numerical value, might also better support diverse writers' diverse needs, as illustrated by the different types of leaps we detailed. Such steering can be explicitly enabled (Krause et al., 2020; Keskar et al., 2019; Lin and Riedl, 2021), for example, by conditional modeling, or, in the absence of specialized approaches, even discovered by so-called "prompt engineering" which has been successfully used by many for language-controlled visual art generation (Patashnik et al., 2021) with general-purpose vision+language models (Radford et al., 2021).

7 Conclusion

In this work, we reported on *integrative leaps*, by which participants integrate writing suggestions by performing cognitive work to make transitions possible, and discussed implications for creative writing support tools.

8 Acknowledgements

This material is based upon work supported by the NSF under Grant No. 2107391.

References

David Bamman, Brendan O'Connor, and Noah A Smith. 2013. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 352–361.

Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S Lasecki, Daniel S Weld, and Eric Horvitz. 2019. Beyond accuracy: The role of mental models in human-ai team performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 2–11.

Jeffrey Bardzell and Shaowen Bardzell. 2016. Humanistic hci. *Interactions*, 23(2):20–29.

Gwern Branwen. 2020. Gpt-3 creative fiction.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Alex Calderwood, Vivian Qiu, Katy Ilonka Gero, and Lydia B Chilton. 2020. How novelists use generative language models: An exploratory user study. In *HAI-GEN+ user2agent@ IUI*.

Orson Scott Card. 1990. *How to write science fiction and fantasy*. Writer's digest books Cincinnati, OH.

Michelene TH Chi. 2000. Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in instructional psychology*, 5:161–238.

Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces*, pages 329–340.

Fintan J Costello and Mark T Keane. 2000. Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24(2):299–349.

Barbara Dancygier. 2006. What can blending do for you? *Language and Literature*, 15(1):5–15.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2019. Plug and play language models: A simple approach to controlled text generation. *arXiv preprint arXiv:1912.02164*.

Michael A DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. " algorithms ruin everything" # riptwitter, folk theories, and resistance to algorithmic change in social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 3163–3174.

Stephen R Donaldson. 2008. *The gap into conflict: The real story*. CNIB.

Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First i" like" it, then i hide it: Folk theories of social feeds. In *Proceedings of the 2016 cHI conference on human factors in computing systems*, pages 2371–2382.

Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical neural story generation. *arXiv preprint arXiv:1805.04833*.

Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 4647–4657.

Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College composition and communication*, 32(4):365–387.

Marcelle Freiman. 2015. A 'cognitive turn'in creative writing–cognition, body and imagination. *New Writing*, 12(2):127–142.

Clifford Geertz. 1973. *The interpretation of cultures*, volume 5019. Basic books.

Susan A Gelman and Cristine H Legare. 2011. Concepts and folk theories. *Annual review of anthropology*, 40:379–398.

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

Ben Krause, Akhilesh Deepak Gotmare, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher, and Nazneen Fatema Rajani. 2020. Gedi: Generative discriminator guided sequence generation. *arXiv preprint arXiv:2009.06367*.

Ziva Kunda, Dale T Miller, and Theresa Claire. 1990. Combining social concepts: The role of causal reasoning. *Cognitive Science*, 14(4):551–577.

Carolyn Lamb, Daniel G Brown, and Charles LA Clarke. 2018. Evaluating computational creativity: An interdisciplinary tutorial. *ACM Computing Surveys (CSUR)*, 51(2):1–34.

Zhiyu Lin and Mark O Riedl. 2021. Plug-and-blend: A framework for plug-and-play controllable story generation with sketches. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 17, pages 58–65.

Frieder Nake. 1994. Human-computer interaction: signs and signals interfacing. *Languages of design*, 2(193-205).

Ellen W Nold. 1981. Revising. *Writing: the nature, development, and teaching of written communication*, 2:67–79.

Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. 2021. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2085–2094.

Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*.

Jean Piaget. 1985. *The equilibration of cognitive structures: The central problem of intellectual development*. University of Chicago press.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Arie Rip. 2006. Folk theories of nanotechnologists. *Science as culture*, 15(4):349–365.

Phoebe Sengers and Bill Gaver. 2006. Staying open to interpretation: engaging multiple meanings in design and evaluation. In *Proceedings of the 6th conference on Designing Interactive systems*, pages 99–108.

Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Trans. Comput.-Hum. Interact.* Just Accepted.

Mark Turner and Gilles Fauconnier. 1999. A mechanism of creativity. *Alternation*, 6(2):273–292.

Thomas B Ward and E Thomas Lawson. 2009. Creative cognition in science fiction and fantasy writing.

Max Weber. 1949. *Max Weber on the methodology of the social sciences*. Free Press.

Edward J Wisniewski. 1997. When concepts combine. *Psychonomic bulletin & review*, 4(2):167–183.

A Photo Credits

Photos in Fig. 1 and 2 by Alan Ren, Matus Karahuta, Javad Esmaeili, Cassie Lopez (1), Christopher Campbell, Brooke Cagle (1), Benn McGuinness, Cassie Lopez (2), Claudio Schwarz, purzlbaum, Matheus Ferrero, Cory Woodward, Tim Photoguy, 2 Bro's Media, Nature Uninterrupted Photography, Fabe collage, Brooke Cagle (2), Ronaldo de Oliveira at Unsplash.

B Integrative Leaps: Additional Examples

Leap 4. Input (summary): [Emotional dialogue, son is held captive...] ..."What?" She replied back. "Who are you talking about?" "It's them," he whimpered. "But I-I don't have anything to tell them. I don't have the information they're looking for."
Suggestion: `I'm just a` **`normal person`** `who is in a hurry to get home...`
Integration: She freezes. What is he talking about? This isn't making any sense... yes, she has an estranged relationship with her son, but **they are normal people.** "You're not making any sense." "It's **not normal. None of this is normal**" he responds shakily. She hears a scream and the phone cuts out.
Explanation:

> "...I'm just thinking about how to continue this story but I don't really have much... but the suggestion under *Plot* is giving me some... you know, "**I'm just a normal person**" line...I still don't have any sort of direction with the story... this feature seems to be good to help me, like, continue along, where otherwise I think I will just stop writing... "

P3, following the "The phone began to ring" prompt, was writing an intense story of a mother getting a phone call from her estranged son. Through a number of previous suggestion interactions, the participant wrote a story where the son on the phone call was in trouble, as some people were holding a gun to his head and demanding some information he didn't have. The next round of suggestions contained "I'm just a normal person who is in a hurry to get home." Following that, the participant wrote "She freezes. What is he talking about? This isn't making any sense... yes, she has an estranged relationship with her son, but they are normal people." As the participant explained, the phrase in the suggestion "I'm just a normal person" stood out to them and prompted them to develop it into the mother's inner thoughts trying to come to terms with the fact that her son and she herself are probably in big trouble. We labeled this example as *direct* (almost verbatim integration: **normal person** to **normal people**), *exploratory* (the participant did not have a clear idea of the narrative) and *convergent* (solving a local question of how the main character reacts to the news that her son is in trouble).

Leap 5. Input (summary): [Best friend phone call...] ..."I ran into your ex-boyfriend at the hospital". I was in shock. I hadn't seen him since 4 years ago when he left me to run away to Cuba with some new woman.
Suggestion: { , , ... }
Integration: "Wait why were you in the hospital?" I asked my friend. "My sister was in **car** crash. She's okay, but she broke a rib." I completely forgot about what she said about my ex being in the area, assuming it was hours ago, and rushed to the hospital. We were neighbors growing up, so I was pretty close with her sister too.
Explanation:

> "I'm seeing **cars**, so maybe she was in a **car** crash."

P21 was developing a story from the prompt "The phone began to ring" and was describing a call from the best friend of the main character. P21 wrote the first part of the dialogue ""Wait why were you in the hospital?" I asked my friend" and the subsequent round of the suggestions contained images with cars. The participant immediately took on the idea: "I'm seeing cars, so maybe he was in a car crash." and to continue the dialogue, P21 wrote: "My sister was in a car crash. She's okay, but she broke a rib." Since the suggestions helped to keep the writing going and did not prompt the participant into a new avenue of thought, as well as being a textual representation of a suggested visual object, this entry is labeled as *direct* (**images of cars** to **car crash**), *confirmatory* (reinforces the existing narrative), and *convergent* (closes a local question of why the person is in the hospital).

Leap 6. **Input (summary)**: [Tadpoles taking the train back home from Kindergarten. . .] . . . Once inside the parlour they were all taken back by the
Suggestion: `sound of a bell ringing and the frog who was holding the bell was holding a tray of frogs and he was holding a tray of tadpoles who were all waiting for the new tadpoles`
Integration: Once inside the parlour they were all taken back by the ringing and clattering of dishes and trays. Frogs and toads were excitedly gulping down the various fly filled delights inside. "Georgia! Barry! Tadette!" beamed Mr Willeker. "You all look so well!" Please take a look at the menu.
Explanation:

> " I found that interesting as I guess it made me think more of like the sounds that could be inside this parlor or something . . . because, basically, I was going to end up doing another long description that's probably quite boring. Probably similar to my previous thing I was writing, but I could then think about the sounds like clattering plates. "

Following the prompt "A train arrives at the station" P9 started writing a fantasy story about frogs waiting for their tadpoles to get back from Tadpole Kindergarten. Another round of suggestions read: "sound of a bell ringing and the frog who was holding the bell was holding a tray of frogs. . . " As the participant explained, the specific "sound of a bell ringing" in the suggestion made them think about sounds in general and what kind of sounds can be in the setting of their story. The participant wrote "Once inside the parlour they were all taken back by the ringing and clattering of dishes and trays. " Here, the participant took a concrete description of sound (sound of a bell ringing) and then made a shift from concrete description to the general concept of sound and made a decision about what kind of particular sound will be in their story ("clattering of dishes and trays").