IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 11, 2021, accepted March 1, 2021, date of publication March 8, 2021, date of current version March 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064242

Improving the Fitness Function of an Evolutionary Suspense Generator Through Sentiment Analysis

PABLO DELATORRE[1], CARLOS LEÓN[2], AND ALBERTO SALGUERO HIDALGO[1]
[1]Department of Computer Science, University of Cádiz, 11001 Cádiz, Spain
[2]Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, 28040 Madrid, Spain

Corresponding author: Pablo Delatorre (pablo.delatorre@uca.es)

This work was supported in part by the Andalusian Government under the University of Cadiz Programme for Researching and Innovation in Education under Grant SOL-201500054211-TRA, in part by the CANTOR Project under Grant PID2019-108927RB-I00, in part by the Spanish Ministry of Science and Innovation, and in part by the FEI INVITAR-IA Project of the University Complutense of Madrid under Grant FEI-EU-17-23.

ABSTRACT The perception of suspense in stories is affected not only by general literary aspects like narrative structure and linguistic features, but also by anticipation and evocation of feelings like aversion, disgust or empathy. As such, it is possible to alter the feeling of suspense by modifying components of a story that convey these feelings to the audience. Based on a previous straightforward model of suspense adaptation, this paper describes the design, implementation and evaluation of a computational system that adapts narrative scenes for conveying a specific user-defined amount of suspense. The system is designed to address the impact of different types of emotional components on the reader. The relative weighted suspense of these components is computed with a regression model based on a sentiment analysis tool, and used as a fitness function in an evolutionary algorithm. This new function is able to identify the different weights on the prediction of suspense in aspects like outcome, decorative elements, or threat's appearance. The results indicate that this approach represents a significant improvement over the previous existing approach.

INDEX TERMS Automatic story generation, genetic programming, predictive model, sentiment analysis, suspense.

I. INTRODUCTION

One common approach to automatic generation of stories is to generate content and evaluate the results against a particular function. This approach is of application in several generative schemes, namely state-space search or evolutionary approaches. The literature shows several examples of this family of generation techniques, and implementations of this method are common in Computational Creativity and Computational Narrative. Metrics for evaluating the effect of stories include coherence, causality, or character affinity [1]–[4].

For instance, the automatic storytelling system *BRUTUS* generates the sequence of events with a ruled-based system that produces a set of potential consequences for the characters [5]. *TALE-SPIN* [6] and *Fabulist* [7] tackle story generation as a planning process, taking a collection of characters with their corresponding intentions and goals, and generate a sequence of states and actions. Other systems like *Charade* [8] perform by simulating the evolution of a relationship between two characters, using their mutual affinities. *PropperWryter* creates Russian folktales following to Propp's formalization [9].

In particular, automatic story generation systems specifically addressing suspense usually follow the strategy of modeling the reader's expectation about the threatening outcome. They commonly assume that adversity will evoke a set of emotions that are related to the feeling of suspense. Broadly speaking, the goal of common suspense-based story generators is to compute sequences of states that lead to events that can be interpreted by readers as dangerous or difficult to overcome for the main characters [10]–[12].

For instance, the automatic story generation system *MEXICA* uses a cognitive model of the writing process that employs abductive reasoning to guarantee the tension and coherence of inter-character relationships [13]; *IDtension* generates "obstacle" states for the characters and selects the state in which the obstacle is the most difficult one to overcome [14]. Similarly, *Suspenser* generates the plot states with all the possible plans and selects the best one

The associate editor coordinating the review of this manuscript and approving it for publication was Hisao Ishibuchi.

based on the potential failures for the characters to overcome [15]. *Dramatis* tries to determine escape plans that the reader will consider most likely for the protagonist, but avoiding any condition of the narrative arc that contemplates a direct valid path between the current state and an undesired outcome [16], [17].

Whereas the techniques are diverse and provide useful insight towards their specific objectives, no story generation system explicitly addresses cognitive suspense beyond the implementation of conflict-resolution. This is usually done through relations, or, more commonly, the chance of eliminating escape alternatives for the protagonist [7], [18], [19]. Although these systems compute events according to a goal-planning process, the emotional impact of the different elements included in the plot are barely taken into account to choose the best states.

This emotional impact of the elements of a scene is known to play a fundamental role in the perception of suspense: a knife and a mask, worn by a criminal, may evidence the intention of a rougher murder than gunfire does; a disfigured threat may lead the audience to predict a higher danger. Natural events during a stormy night may evoke a greater feeling of suspense than a clear afternoon does, and the presence of ropes on a table may lead to foretell an abduction, even if these aspects do not finally play any role in the plot.

This kind of information influences the perceived intensity of the potential outcome, which is linked to the perceived suspense, as described in the literature [20]–[29]. Additionally, the discourse itself, understood as the specific order in which the events are narrated, can also trigger different emotional and affective responses in the audience [30], [31]. These ideas support the necessity to address the complexities of suspense taking into account the cognitive particularities of the phenomenon. As such, the computational analysis of suspense cannot be exclusively on the sequence of events.

This cognitive approach to the study of suspense has been previously studied and implemented as a component-based system [32]. This system performed text adaptation of suspenseful scenes in which the decorative elements (those present in the scene, but not playing any major role in the main plot) were modified to control the conveyed suspense.

The system is able to produce story modifications that do not modify the plot, but produce an output text whose perceived suspense is closer to a user-defined value. The original proposal uses the ANEW corpus [33], a list of 1034 terms along with their respective emotional values. These emotional values, or *dimensions* are the *valence* (or pleasure, ranging from *unpleasant* to *pleasant*), the *arousal* (or intensity of the emotion, ranging from *calm* to *excited*), and the *dominance* (or degree of control over the stimulus, ranging from *out of control* to *in control*). In particular, the corpus used by the system is a customized version: based on the original ANEW corpus, a new corpus of affective norms contextualized in a setting of fictional suspense was created. This corpus was called S-ANEW [34].

TABLE 1. Example with annotated terms of the S-ANEW corpus. As the term "cold" illustrates, a term can be associated to different annotations. Suspense scores range from 1 to 9.

term	suspense	annotation / type
cold	4.43	weather:personality:feeling
death	7.02	outcome
garden	3.30	place
hope	3.85	feeling
poster	2.12	decorator
pretty	2.45	appearance
shotgun	6.56	resource
sunset	5.26	daytime

Table 1 shows some examples of annotated terms of this corpus.

The emotional impact of the terms, such as those exemplified in Table 1, was studied in detail, and their effects of the final perception of suspense were quantified [35]–[39]. In [40], this information was used to compute a preliminary, original predictive model, and to implement an evolutionary algorithm. However, the system did not differentiate between distinct types of elements: Despite the fact of a potential *outcome* may have a different impact on suspense than a *decorative element* in the scene, the original fitness function just used the mean of suspense evoked by the different terms regardless of their type. The reason for using the averaging operator was that the literature does not provide quantitative difference among types of elements.

The results from the predictive algorithm yielded a medium-high correlation between the predicted suspense and the evaluation with human subjects ($r = 0.748$, $p < 0.05$). However, the possibility of a better correlation leads to hypothesize that focusing on the specific influence of the type of emotional components may led to a better performance than assuming the simple average. This is the objective of the present study.

The rest of the sections detail this study, as follows: Section II introduces the implementation of the evolutionary algorithm and the original fitness function to be improved; Section III describes the method and generation of a set of scenes by combining different narrative elements in clusters; Section IV shows how the predictive models of the affective values of each cluster from the automatic semantic analysis of the scenes were computed; Section V describes the implementation of the new fitness function according to the weights of the model; and Section VI describes the evaluation of the new fitness by the evaluation of human subjects. Finally, Section VII and Section VIII respectively discuss and summarize the results.

II. IMPLEMENTATION OF THE ORIGINAL FITNESS FUNCTION

Following, the implementation of the evolutionary algorithm and the original fitness function to be improved are described.

The evolutionary algorithm was implemented with the Java Genetic Algorithms Package (JGAP), and performs like a Simple Genetic Algorithm (SGA) [41]. The probability of mutation was set to 0.5, and the crossover probability was set to 0.4. It follows a generation scheme in which each candidate term for the narrative is a gene to be evolved. The system was designed to follow the aforementioned hypothesis that the emotional effect of certain elements differs from others. For instance, the weather conditions could have less impact in suspense than the different descriptors of the threat's physical appearance.

In order to model this, genes are *clustered* together according to the kind of information they convey (annotations, in the S-ANEW corpus). Therefore, specific subsets of terms (g) are grouped into clusters ($U_i = \{g_{i,1}..g_{i,n_i}\}$). For instance, terms representing places are included in a unique cluster $U_p = \{g_{p,1}..g_{p,n_p}\}$, where any $g_{p,j}, j \in [1, n_p]$ would be a term classified as a *place* in the S-ANEW corpus, and n_p is the number of places appearing in the story. Likewise, another cluster $U_o = \{g_{o,1}..g_{o,n_o}\}$ could contain instances of terms representing expected outcomes (terms annotated as *outcome* in the corpus), and n_o is the number of outcomes appearing in the story (for instance, one per character).

All genes belonging to cluster U_i will be assigned a particular weight ($w_i \in [0.0, 1.0]$) in the predictive model. This way, the suspense Φ for each cluster is computed as $\Phi_i = (w_i/n_i) \times \sum_{j=1}^{n_i} G_{i,j}$, where n_i is the number of genes of the cluster U_i, and $G_{i,j}$ is the suspense for the narrative element $g_{i,j}$, extracted from the S-ANEW corpus. The fitness function was computed as the absolute value of the subtraction of the mean of all Φ_i and the input desired final suspense of the generated plot.

In the original work [40], this fitness function assigned the same weight to each cluster; that is, $w_i = w_j = 1, \forall\, U_i, U_j$. This is because no previous study in the literature provides quantitative values for these weights, as previously indicated. Therefore, although the algorithm yielded a medium-high correlation between predicted suspense and evaluation with human subjects ($r = 0.748$, $p < 0.05$), the idea that the emotional impact of the elements of each cluster U_i is different is analyzed in this study. This leads to a new design of the fitness function that tries to improve the prediction accuracy of the expected suspense.

III. METHODOLOGY TO IMPROVE THE FITNESS FUNCTION

The present research sets off to improve the performance of the fitness function of the evolutionary algorithm described in Section II. In the original approach, the hypothesis that different kinds of terms have different impact on the overall perception of suspense was not evidenced. Gaining insight on this hypothesis can improve the suspense generation process since the algorithms will be able to control the expected suspense generation in more detail.

For this purpose, using humans to evaluate suspense is not feasible because of methodological issues. First, there could be a lack of context in which subjects do not assign the appropriate emotional values to terms because they are not really feeling the suspense of a real narrative, which would invalidate the result. Second, the cost of the process would be extremely high: The S-ANEW set is composed by 1034 terms, and the necessary cross-comparison together with the number of different samples per element would be intractable.

For this paper and in order to overcome this limitation, the multilingual sentiment analysis tool Lexalytics' Semantria [42]–[47] was used to analyze and evaluate the quantitative impact of a set of 486 suspenseful short scenes generated from three different original texts. The terms of each scene were obtained by combining different narrative elements in nine different clusters.

From this set of base stories, a regression model between the clusters' suspense and the sentiment analysis was computed.

A. GENERATION OF THE PRELIMINARY SET OF TEMPLATES

The study was conducted using three different templates of short narratives. The templates' placeholders are used by the evolutionary algorithm to include the elements to be evaluated. Once generated, the resulting scenes were used both for the semantic analysis and for testing against human subjects with the new fitness function, as detailed later.

The narratives chosen for the research fulfill a number of properties:

- All the scenes finish in a climax situation of high narrative tension according to the classic Freytag's dramatic structure [48]. Otherwise, scenes that resolve the narrative tension could potentially produce a stress relief that could influence the assessment of the perceived suspense [49].
- In order to provide coverage for the target set of stories, it must be ensured that certain families of terms always appear in the templates. Specifically, all the templates must include the next clusters:
  - *place* (p) where the action unfolds
  - *daytime* (e) and *weather* (h)
  - *decorative* (d) elements in the scene
  - *personality* (s), *appearance* (a), *resource* (r), and *feeling* (f) of characters
  - *expected outcome* (o)

  Clusters a, r and f must contain two genes: one for the main character (a_m, r_m, f_m) and another one for the threat (a_t, r_t, f_t). This means that nine clusters and total of twelve genes will compose the chromosome and the training features for the regression model.

A strong focus has been put on creating sufficiently different narrative templates. In order to get it, basic strategies to conceive suspense scenes were studied [50]–[55]. After the analysis, the type of scenes found which were different in terms of narrative techniques that were considered not to overlap were three. As a result, three templates were finally generated:

1) "Pursued", where a character is stalked and, later, pursued.

TABLE 2. Number of words, number of sentences and Flesch readability index (RI) per narrative template.

scene	words	sentences	RI	order											
Pursued	126	14	33.34	h	e	p	s	a_m	f_m	r_m	d	a_t	r_t	f_t	o
The call	145	12	45.21	r_m	d	p	a_t	a_m	s	f_t	f_m	r_t	o	h	e
Wake up	232	24	60.64	p	d	h	e	f_m	o	r_m	s	f_t	r_t	a_m	a_t

TABLE 3. Placeholders for template substitution in the short narrative scenes. The placeholders describe term family and syntactic concordance.

code	description	examples	
`{:feature$variable}`	Replaces the placeholder by a narrative element annotated in the corpus as *feature* which will be internally referred as *variable*. The transformation can happen anywhere in the template scene, and it is needed for the rest of the transformations using this variable.	`{:outcome$o}`	Replaces this code by an outcome, internally called *o*; for instance, "tortured".
		`{:daytime$e}`	Replaces this code by a daytime, internally called *e*; for instance, "sunset".
`{?gender$variable}`	Replaces this code by the name of a person with the specified gender (or random if it is not provided), which be internally referred as *variable*. The transformation can happen anywhere in the template scene, and it is needed the rest of the transformations using this variable.	`{?female$main}`	Replaces this code by a female name, internally called *main*; for instance, "Sarah".
		`{?male$partner}`	Replaces this code by a male name, internally called *partner*; for instance, "Peter".
		`{?$threat}`	Replaces this code by a random name, either male or female, internally called *threat*; for instance, "Helen" or "John".
`{$variable}`	Replaces this code by the term internally referred as *variable*.	`{$o}`	Replaces this code by the value of *o*; in the previous examples, "tortured".
		`{$partner}`	Replaces this code by the value of *partner*; in the previous examples, "Peter".
`{$variable->hesheit}`	Replaces this code by the corresponding personal pronoun of the value of *variable*, when applicable.	`{$partner->hesheit}`	Replaces this code by the personal pronoun of *partner*; in the previous examples, "he".
`{$variable->himherit}`	Replaces this code by the corresponding object pronoun of the value of *variable*, when applicable.	`{$main->himherit}`	Replaces this code by the object pronoun of *main*; in the previous examples, "her".
`{$variable->hisherits}`	Replaces this code by the corresponding possessive adjective of the value of *variable*, when applicable.	`{$main->hisherits}`	Replaces this code by the possessive adjective of *main*; in the previous examples, "her".
`{$variable->hishers}`	Replaces this code by the corresponding possessive pronoun of the value of *variable*, when applicable.	`{$partner->hishers}`	Replaces this code by the possessive pronoun of *partner*; in the previous examples, "his".

2) "The call", where a character is involved in a "countdown" situation, in which has to take a decision.
3) "Wake up", where a character is confined in an unknown place.

Likewise, while templates were conceived as short scenes the number of words, number of sentences and Flesch readability index (RI) [56], and order of the elements were different. These differences are shown in Table 2.

B. GENERATION OF THE SCENES

Variable replacement in the templates was carried out with a simple parser. The evolutionary algorithm replaces the placeholders in the templates by term values belonging to the specific family defined by such placeholders (*places* or *outcomes*, for example) [40].

Table 3 details the placeholder variables used in the templates. It can be seen in the table that the placeholders include not only the family of suitable emotional terms, but also specific replacement properties.

The texts to be evaluated were generated for each narrative template. In order to do this, two different scenes per desired suspense value in the range $[1.0, 1.1, 1.2, \cdots, 9.0]$ were generated per template (that is, all values from 1.0 to 9.0, with a step of 0.1). In total, 162 different instances per template. For the first generation, population size and offspring were set to 100, and the maximum number of iterations was set to 200. For the second generation, population size and offspring were set to 20, and the maximum number of iterations was set to 30. Two iterations were performed because a high value produces less variability on the terms (and a slower generation), but a higher accuracy with respect to the input suspense. On the other hand, a lower number of iterations tend to produce results that are less similar to the input suspense, but present a higher variability between the gene values (and, therefore, between scenes). From the 162 instances for each three templates, 486 short scenes were produced. These were individually analyzed by the sentiment analysis tool.

Three examples of scenes are shown in Listings 1, 2 and 3. Each example corresponds to one of the three input templates. The instantiation of the placeholders related to suspense terms are underlined in the texts for facilitating their identification. Input (desired) suspense (Φ_0), suspense computed using the baseline fitness function (Φ), sentiment analysis evaluation (S), and the suspense value of each underlined term of these examples (along with its order in the text) are shown in Table 4.

TABLE 4. Parameter values for the example scenes 1, 2 and 3: input (desired) suspense (Φ_0), computed suspense using the baseline fitness function (Φ), sentiment analysis evaluation (S), suspense value (*susp*) of each included term (*term*), and order in the story (*ord*). Additionally, population size and offspring (P_0), and maximum iterations (I_{max}) are included.

	Pursued			The call			Wake up		
	P_0=100, I_{max}=200			P_0=100, I_{max}=200			P_0=20, I_{max}=30		
	Φ_0=4.60, Φ=4.57			Φ_0=3.70, Φ=3.70			Φ_0=6.10, Φ=5.61		
	S=-0.111 (neg.)			S=-0.113 (neg.)			S=-0.336 (neg.)		
g	**ord**	**term**	**susp**	**ord**	**term**	**susp**	**ord**	**term**	**susp**
d	8	lamp	2.88	2	crown	2.01	2	scissors	5.28
e	2	twilight	5.05	12	daylight	2.57	4	sunset	5.26
h	1	frigid	5.04	11	breezy	2.46	3	cold	4.43
o	12	killed	7.52	10	punished	5.85	6	tortured	7.20
p	3	alley	5.24	3	windmill	3.19	1	jail	5.87
s	4	timid	3.35	6	curious	3.97	8	hostile	3.63
a_m	5	muscular	3.32	5	blond	2.10	11	sickness	5.80
f_m	6	hopeful	3.69	8	ridiculous	4.21	5	helpless	7.10
r_m	7	grenade	5.50	1	hammer	4.71	7	rifle	6.32
a_t	9	sick	4.92	4	strong	3.31	12	deformed	5.41
f_t	11	enjoying	4.01	7	happy	4.72	9	traumatized	6.94
r_t	10	dagger	4.47	9	rock	3.76	10	pistol	6.27

> It was a frigid twilight. Charles walked through the alley. He was timid and physically muscular. Lately, he felt hopeful. Suddenly, Charles had the feeling that something was following him. He reached into his backpack, from which he took a grenade. Charles kept it in his pocket, thinking it could serve as a deterrent, just in case. He hadn't finished thinking about it, when a figure appeared out of nowhere, next to a lamp. It seemed like a man. His appearance was sick. He carried a dagger, and Charles realized. Then, enjoying, the man shouted to Charles: "You deserve to be killed!". Charles started to run through the alley, but he stumbled and fell down. His pursuer was closer and closer.

LISTING 1. Example scene automatically adapted from the "Pursued" template.

> "Drop that and raise your hands". Robert said these words as he pointed to the man with his hammer. The man was called Michael, and he sat next at a table on which there was a crown. Under the light, in that windmill, he looked strong, but Robert didn't care. Robert was blond, and he was also curious. He stared at Michael, who seemed to be happy. Robert took the opportunity to check his own feelings, and concluded he felt ridiculous. Be that as it may, Michael still had that mobile phone in his hands, his finger ready to press the screen. The fact that, in another place, a rock hit a person depended on that call. Even if the call did not come, the person would still be punished by one of the Michael's henchmen. On that breezy daylight, Robert wondered how he should act.

LISTING 2. Example scene automatically adapted from the "The call" template.

IV. COMPUTING THE PREDICTIVE MODELS

All the scenes were analyzed by the sentiment tool. According to the output value, all of them yielded a negative sentiment value. Additionally and as expected, the correlation found between the input suspense (Φ_0) and the suspense

> When Paul opened his eyes, he noticed that he was tied to a rickety chair in the center of a jail. Some scissors were in the walls. In front of them, there was what must have been a small window. It was a cold sunset. Paul felt helpless. He wondered how long he had been unconscious. Suddenly, Paul realized he was not alone. There was another person there, right behind his back, and just at that moment, a female's voice spoke close to his ear: "Time to pay off old debts". Paul recognized that voice. Meredith! It could not be possible. After so long... "Time to be tortured, Paul!". From behind, her captor passed her hand before his eyes to show him a rifle. Paul took advantage of the moment to move quickly. He threw his head back and hit the other. Then, he jumped aside as high as he could and turned to fall backwards, breaking the old chair and being free. Quickly, he approached Meredith, who was still dazed, and pulled the rifle out of her hands. Paul wasn't sure how he had done it so fast, but he felt that he was returning to his personality: just hostile. Recovered, Meredith looked at him, traumatized. She put her hand in her pocket and pulled out a pistol. There they were, facing each other. Paul was sickness, Meredith was deformed.

LISTING 3. Example scene automatically adapted from the "Wake up" template.

after computing the terms based on the original fitness (Φ) is almost total ($r = 0.997$, $p < 0.000$).

However, the correlation between the computed suspense Φ and the evaluation of the semantic analysis (S) was $r = -0.721$, $p < 0.000$. Again, this medium correlation leads to hypothesize that focusing on the specific influence of clusters separately, with different weights, can led to a better performance than assuming the simple average.

In order to do this, a regression between the sentiment analysis score S (dependent variable) and each one of the nine cluster independently (independent variables) was calculated. The goal was to find the values of α and β for each cluster U_i

that maximizes the fit described in Equation 1.

$$S = \beta_0 + \sum_i \beta_i U_i^{\alpha_i}, \quad \alpha_i \in \mathbb{Z},\ \beta_i \in \mathbb{R},$$
$$i \in \{d, e, h, o, p, s, a_{m,t}, f_{m,t}, r_{m,t}\} \quad (1)$$

To check that the clusters had a different influence depending on the plot and the way in which the information is presented to the reader, three different models were computed, one for each template. The best accuracy was obtained with $\alpha_i = 1, \forall i$. Other non-linear options were tried for the three scene templates, but none gave a significantly better result. The three models were different and produced a higher signification and weight in distinct clusters.

Additionally, an ANOVA test between the models was carried out. In all cases, p was less than 0.05. This led to hypothesize that a specific predictive model must be considered for each type of scene in order to improve the effectiveness of the fitness function.

V. IMPLEMENTATION OF THE NEW FITNESS FUNCTION

Considering the impact of the genes of each cluster in the model of each scene, the implementation of the new fitness function was based on penalty criteria. The general idea is that the more suspense weight an element has, the more penalizes the fitness function if it is far from the desired suspense. That is, each variable penalized the fitness value according to its weight (w_i) and the distance with the desired suspense (Φ_0). Thus, for a set of genes $g_{i,j} \in U_i$ with mean value $\overline{G_i}$ in the cluster and a weight $w_i \in [0, 1] \subset \mathbb{R}$, the penalty criteria follows the formula $P(i) = w_i \times |\Phi_0 - \overline{G_i}|$. Thus, the fitness value (Z) would follow Equation 2, where Φ_{max} is the maximum value of suspense in the range.

$$Z = \Phi_{max} - \sum_i P(i) \quad (2)$$

The impact factor of each variable in the model was computed under the general assumption that the value of the independent constant β_0 was related with the effect of both the structure and the fixed elements of the scene, not with the added terms. This hypothesis makes it possible to partially avoid the impact of the independent variable in the fitness function, which in any case would only provide a constant linear value not related to any variable. In order to get the factors, the weight w_i is proportional to the impact of the variable in its respective model, so the lower this impact is, the less the effect in the penalization $P(i)$ and, hence, proportionally the less impact in the fitness value Z.

Table 5 shows the values of β_i, their signification in the model, and the penalty weight w_i per each type of scene. Each weight w_i is computed by interpolating the corresponding model's parameter β_i to a normalized [0.0, 1.0] scale.

Figure 1 shows how the regression models distribute the weights of the narrative elements for each scene. For example and according to this distribution, most suspense weights in the scene "The call" fall on the personality of the protagonist. In fact, the feelings of the characters seem to determine an important part of the level of suspense in all the stories. Regarding the empathy [57], this effect is consistent with relevant literature [58]–[61]. Similarly, the appearance of the characters (the threat, mainly) seems to significantly influence the suspense generated. However, other expected aspects, such as the outcome or the resources of the threat, do not seem to have had an important impact on the scenes.

TABLE 5. Values of β_i, signification, and computed weight $w_i \in [0, 1] \subset \mathbb{R}$ in the regression models, one per type of scene.

	Pursued		The call		Wake up	
gene	β	w	β	w	β	w
d	-.007	.026	.007	.034	.008*	.066
e	.034 *	.111	-.003	.013	-.011*	.096
h	.006	.021	.001	.005	-.002	.017
o	-.001	.004	-.016*	.082	-.005	.043
p	-.045 *	.148	-.007	.036	.006	.051
s	-.004	.014	-.061**	.319	-.025**	.217
a_m	-.049 *	.161	-.018*	.092	-.007	.057
f_m	-.054 **	.178	-.036*	.183	-.006	.051
r_m	-.019	.064	.005	.023	-.004	.036
a_t	-.044 *	.144	-.025*	.130	-.021**	.182
f_t	-.035 *	.114	-.015*	.076	-.019*	.163
r_t	-.004	.015	.002	.008	-.003	.022

*: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.005

VI. EVALUATING THE RESULTS WITH THE NEW FITNESS FUNCTION

A. METHOD

A total of $N = 72$ participants, 43 males (59.72%) and 29 females (40.27%), voluntary participated in this study. Subjects ranged in age from 19 to 32 years old ($M = 21.62$, $SD = 2.01$). The evaluation was carried out online.

$N = 72$ scenes per each type of scene ("Pursued", "The call", and "Wake up") were generated. For each template, half the stories were generation with the baseline fitness function and, the other half, with the new fitness function. They were randomly assigned to the participants so that they would receive one story per template (3 stories in total). In total, 216 stories were evaluated.

Each registered participant was contacted by e-mail with an Excel spreadsheet containing four sheets. The first sheet contained the instructions. The other three sheets contained, each one, one of the three scenes assigned to the participant. The sheets were randomly shuffled for avoiding the sequence effect. The scene text was located on the leftmost, topmost cell $(1, A)$. The cell was dimensioned so that the whole content would visibly fit in the cell. The next cell to the right $(1, B)$ contained a fully visible drop-down list with values from 1 to 9 for rating the story. The rest of the sheet was empty.

According to the instructions, the participants had to rate the amount of suspense they felt, reporting the corresponding value in the drop-down list. Participants were asked to fill in and send the completed file in within two weeks. These instructions and the use of non-presential surveys were similar to those provided in related studies [34], [62]–[64]. After the time period for sending back the questionnaires expired, all participants had submitted a completely filled-in spreadsheet.

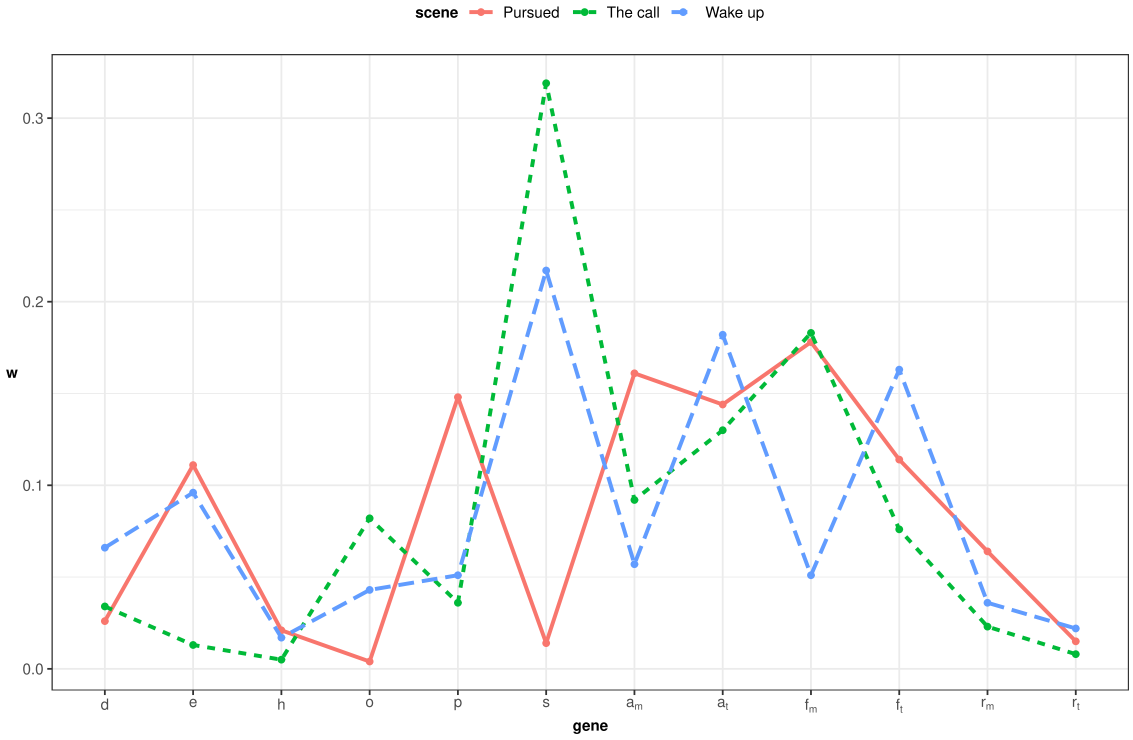

FIGURE 1. **Distribution of the weight (*w*) of different genes per scene.**

B. RESULTS OF THE NEW FITNESS FUNCTION

After the data was processed, the analysis of the correlation between reported and computed suspense were conducted. The objective was to compare the performances of the baseline and new fitness functions for each type of scene. The correlation results are described next.

- For "Pursued", the computed suspense using the mean as the basis for the baseline fitness function reported a correlation of $r = 0.686, p < 0.000$ with the suspense reported by the subjects. Instead, the new weighted fitness function based on the semantic analysis reported a significantly higher correlation of $r = 0.898$, $p < 0.000$. No significant differences between genders or age ranges were found.
- For the scene "The call", the computed suspense using the baseline fitness function reported a correlation of $r = 0.801, p < 0.000$ with the suspense reported by the subjects, and the new fitness function reported a similar but higher correlation of $r = 0.872, p < 0.000$. No significant differences between genders or age ranges were found.
- Finally, for the scene "Wake up", the computed suspense using the baseline fitness function reported a correlation of $r = 0.755, p < 0.000$ with the suspense reported by the subject, and the new fitness reported also a higher correlation of $r = 0.824, p < 0.000$. No significant differences between genders or age ranges were found.

Results show that the new fitness function, which addresses the differences in relative weight between clusters in the prediction, performs better than the baseline fitness function. The performance improvement can be observed in the three scene templates. Additionally, it supports the initial hypothesis that the emotional impact of the elements of each cluster U_i influences suspense differently, although the way in which it does depend on the particular semantic affection of each type of story.

VII. DISCUSSION

This section discusses some design decisions and potential limitations of this study.

Firstly, the design and the textual realization of the narrative templates have been carried out by the authors. Whereas it can be argued that this might add a certain amount of bias in the research, the granularity of the research methodology forces a strict control on the content. The text of the narrative templates and the values for the placeholders must match, which means that it is necessary to provide coverage for the whole list of possible values. This control is virtually impossible to obtain without designing the narrative templates with it in mind. Although using existing suspenseful narratives could have been an option, finding the right passages would have been a very costly task, and it would have also introduced potential human assessment bias. It is therefore assumed that the chosen option is currently the most reasonable alternative.

Secondly, not all parameters β_i in the model were statistically significant for computing the weights w. Although all clusters were included without taking significance into account, it would have been reasonably to consider that the value of a gene that has no significance in the regression model would also have no significance in the fitness penalty. However, the exclusion of some genes would have implied that they were not used by the fitness function, meaning that the had no impact on the overall interpretation of the scene and could have been simply randomized. Although this situation may initially seem appropriate for the generative system, two potential drawbacks are highlighted: a) as reported below, the accuracy of the semantic analysis tools is good, but still not perfect, which means that the results could be hiding the effect of some elements that actually have impact in the plot; and b) in any case, and as aforementioned, weights w_i are proportional to the effect of the variable, so the lower this impact is, the less the effect in the penalization $P(i)$ and, hence, proportionally the less impact in the resulting fitness value Z. In this respect, the risk of excluding any gene is avoided: the lack of significance of some variables in the model does not indicate that they cannot be included, but that they will not significantly change the result in the regression model.

Thirdly, the research methodology includes the use of a semantic parser, Lexalytics' Semantria. The parsing process has an analysis accuracy that is relatively slow and not perfect (as it is the case with any currently available sentiment analysis tool), so it is not possible to assume that the extraction of the emotional effect is fully correct in all contexts [65]–[67]. Additionally, the context is known to affect the emotional interpretation of terms, as previously studied [34], [37]. There exist initiatives that try to schematize or represent semantic and lexematic analysis of suspense-related genres, but the results are inconclusive and are not available as implementations [68], [69]. This does not mean that relying on semantic analysis tools implies a lack of precision (in fact, it has been observed that it increases precision). It must however be consider that the parsing does not fully cover all scenarios, and that the reader's interpretation of the context also plays an important role. Consequently, the present study may not rule out additional effects of specific instances of certain suspense texts. An in-depth, exhaustive analysis of a broader set of narratives is part of the future work.

On a separate issue, Section VI describes a research methodology in which participants, after having enrolled in the experiment, were contacted and queried by e-mail. The absence of a face-to-face supervision makes it difficult to control the sessions or duration of the task per participant, which may affect the task performance [70]. However, a number of experiments gathering affective ratings obtained the answers through similar techniques such as remote web surveys or spreadsheets [63], [64], [71], [72]. These methods are used by several authors due to the advantages in terms of resources and time [73]–[75], and, nowadays, because of the current pandemic situation. In any case, relevant literature points out very small differences between face-to-face and remote questionnaires [75]–[79]. The improvement on availability of the chosen approach makes up for the small potential differences.

Finally, the generative algorithm used for selecting the narrative elements is effective but simple. In this regard, a study of alternatives different from the current SGA strategy must be addressed. This topic will also be carried out in future contributions.

VIII. CONCLUSIONS AND FUTURE WORK

This paper has described an evolutionary algorithm which, given a narrative template, adapts the placeholders for producing a version that generates a suspense close to a desired suspense. In particular, one of the main contributions of the paper is the methodology for the improvement of the fitness function. This improvement has been achieved by refining the calculation of the specific families of terms (clusters) instead of assuming a simple average, and the use of a semantic parser to train the predictive model. Based on the same plot, different values for the placeholders produce different perceived amounts of suspense. The empirical results evidence a significant improvement over the baseline fitness function, shown as a higher correlation between the input and the perceived suspense than with the original, simpler fitness function.

Moreover, an analysis of the results shows that there are elements that seem to have an impact on several stories, such as the personality and emotions of the characters, and the appearance of the threat. Other aspects that could be determined as important, such as the outcome, do not seem to have much influence in the scenes used. Likewise, it is observed that the importance of each narrative element varies according to the scene and, consequently, these require an independent semantic analysis and weighting. Thus, it seems that there is no universal formula that can equally compute the impact of the terms for the whole spectrum of suspense scenes. This issue, on the other hand, is common in natural language processing and sentiment analysis given the mental processes, the representation and meaning, and the context itself [80]. The cognitive effects of any text are variable depending on how the text is structured and the context in which it is presented. In fact and in addition to the elements of the story, different affective responses can be evoked simply by manipulating the order in which a story's events are narrated [30], [31].

In any case, this study has been carried out with three scenes and with a maximum of two genes per cluster. A study with more scenes and, particularly, with a greater number of elements and not restricted to the use of a certain number of types of narrative elements, and the use of other semantic tools could help to observe additional patterns. The current results lead to hypothesize that a general simple formula does not provide coverage for all the cases of narrative and suspense. Instead, a range of family-specific models for different aspects of suspense and plot seems to be the approach that could yield the best overall performance.

ETHICAL STATEMENT

The study reported on this paper were carried out in accordance with the recommendations of national and international ethics guidelines, *Código Deontológico del Psicólogo* and American Psychological Association. The study did not present any invasive procedure, and it did not carry any risk to the participants' mental or physical health, thus not requiring ethics approval according to the Spanish law BOE 14/2007 and the ethical guidelines of authors' institutions. All subjects participated voluntarily and gave written informed consent in accordance with the Declaration of Helsinki. They were free to leave the experiment at any time. There was no compensation for participating in the evaluations.

REFERENCES

[1] K. Wang, V. Q. Bui, and H. A. Abbass, "Evolving stories: Tree adjoining grammar guided genetic programming for complex plot generation," in *Proc. Asia–Pacific Conf. Simulated Evol. Learn. (SEAL)*. Berlin, Germany: Springer, 2010, pp. 135–145.

[2] C. León and P. Gervás, "Creativity in story generation from the ground up: Non-deterministic simulation driven by narrative," in *Proc. 5th Int. Conf. Comput. Creativity (ICCC)*, Jun. 2014, pp. 201–210.

[3] E. Concepción, P. Gervás, and G. Méndez, "An API-based approach to co-creation in automatic storytelling," in *Proc. 6th Int. Workshop Comput. Creativity, Concept Invention, General Intell. (C GI)*, 2017, pp. 1–13.

[4] L. Yao, N. Peng, R. Weischedel, K. Knight, D. Zhao, and R. Yan, "Plan-and-write: Towards better automatic storytelling," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 7378–7385.

[5] S. Bringsjord and D. A. Ferrucci, *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, A Storytelling Machine*. New York, NY, USA: Taylor & Francis, 1999.

[6] J. R. Meehan, "Tale-spin, an interactive program that writes stories," in *Proc. 5th Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 77, 1977, pp. 91–98.

[7] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, no. 1, pp. 217–268, Sep. 2010.

[8] G. Méndez, P. Gervás, and C. León, "A model of character affinity for agent-based story generation," in *Proc. 9th Int. Conf. Knowl., Inf. Creativity Support Syst.*, vol. 11, 2014, p. 2014.

[9] E. Concepción, P. Gervás, and G. Méndez, "A common model for representing stories in automatic storytelling," in *Proc. 6th Int. Workshop Comput. Creativity, Concept Invention, General Intell. (C GI)*, 2017, pp. 1–12.

[10] M. S. R. Pérez ý Pérez, "MEXICA: A computer model of a cognitive account of creative writing," *J. Exp. Theor. Artif. Intell.*, vol. 13, no. 2, pp. 119–139, Apr. 2001.

[11] N. Szilas, "IDtension: A narrative engine for interactive drama," in *Proc. Technol. Interact. Digit. Storytelling Entertainment (TIDSE)*, vol. 3, 2003, pp. 187–203.

[12] B. O'Neill and M. Riedl, "Dramatis: A computational model of suspense," in *Proc. 28th Conf. Artif. Intell. (AAAI)*, vol. 2, 2014, pp. 944–950.

[13] P. Gervás and C. León, "The need for multi-aspectual representation of narratives in modelling their creative process," in *Proc. 5th Workshop Comput. Models Narrative*, in OASIcs-OpenAccess Series in Informatics, vol. 41, E. G. Bruneau, M. A. Finlayson, and J. C. Meister, Eds. QC, Canada, Jul. 2014, pp. 61–76.

[14] N. Szilas, "A computational model of an intelligent narrator for interactive narratives," *Appl. Artif. Intell.*, vol. 21, no. 8, pp. 753–801, Sep. 2007.

[15] Y.-G. Cheong and R. M. Young, "A computational model of narrative generation for suspense," *Assoc. Advancement Artif. Intell. J.*, vol. 21, pp. 1906–1907, Sep. 2006.

[16] B. O'Neill, "A computational model of suspense for the augmentation of intelligent story generation," Ph.D. dissertation, School Interact. Comput., Georgia Inst. Technol., Atlanta, GA, USA, 2013.

[17] R. A. Doust, "A domain-independent model of suspense in narrative," Ph.D. dissertation, Dept. Math., Comput. Technol., Open Univ., Milton Keynes, U.K., 2015.

[18] R. E. Cardona-Rivera, B. A. Cassell, S. G. Ware, and R. M. Young, "Indexter: A computational model of the event-indexing situation model for characterizing narratives," in *Proc. Workshop Comput. Models Narrative at Lang. Resour. Eval. Conf.*, 2012, pp. 32–41.

[19] B.-C. Bae and R. M. Young, "A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach," in *Interactive Storytelling*. Berlin, Germany: Springer, 2008, pp. 156–167.

[20] M. Hron, "Torture goes pop!" *Peace Rev.*, vol. 20, no. 1, pp. 22–30, Mar. 2008.

[21] P. Delatorre, C. León, A. Salguero, M. Palomo-Duarte, and P. Gervás, "Outcome inference based on threat resources in suspenseful scenes," in *Proc. 5th AISB Symp. Comput. Creativity*, no. 5. Liverpool, U.K.: Society With AI, 2018, pp. 18–25.

[22] A. Thain, *Bodies in Suspense: Time and Affect in Cinema*. Minneapolis, MI, USA: Univ. of Minnesota Press, 2017.

[23] K. Custers and J. Van den Bulck, "The association between soap opera and music video viewing and fear of crime in adolescents: Exploring a mediated fear model," *Commun. Res.*, vol. 44, no. 1, pp. 96–116, Feb. 2017. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/483704

[24] C. Hale, "Fear of crime: A review of the literature," *Int. Rev. Victimol.*, vol. 4, no. 2, pp. 79–150, Jan. 1996.

[25] J. Jackson, "A psychological perspective on vulnerability in the fear of crime," *Psychol., Crime Law*, vol. 15, no. 4, pp. 365–390, May 2009.

[26] J. Jackson, "Revisiting risk sensitivity in the fear of crime," *J. Res. Crime Delinquency*, vol. 48, no. 4, pp. 513–537, Nov. 2011.

[27] M. Killias, "Different measures of vulnerability in their relation to different dimensions of fear of crime," *Brit. J. Criminol.*, vol. 40, no. 3, pp. 437–450, Jun. 2000.

[28] A. O'Donovan, G. J. Devilly, and R. M. Rapee, "Antecedents to women's fear of rape," *Behav. Change*, vol. 24, no. 3, pp. 135–145, Aug. 2007.

[29] F. W. Winkel, "Fear of crime and criminal victimization: Testing a theory of psychological incapacitation of the 'stressor' based on downward comparison processes," *Brit. J. Criminol.*, vol. 38, no. 3, pp. 473–484, Jan. 1998.

[30] H. Hoeken and M. van Vliet, "Suspense, curiosity, and surprise: How discourse structure influences the affective and cognitive processing of a story," *Poetics*, vol. 27, no. 4, pp. 277–286, May 2000.

[31] W. F. Brewer and K. Ohtsuka, "Story structure, characterization, just world organization, and reader affect in American and Hungarian short stories," *Poetics*, vol. 17, nos. 4–5, pp. 395–415, Oct. 1988.

[32] P. Delatorre, B. Arfè, P. Gervás, and M. Palomo-Duarte, "A component-based architecture for suspense modelling," in *Proc. 3rd Int. Symp. Comput. Creativity (CC AISB)*, 2016, pp. 32–39. [Online]. Available: http://hdl.handle.net/10498/18328

[33] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," Center Res. Psychophysiol., Univ. Florida, Gainesville, FL, USA, Tech. Rep. C-1, 1999.

[34] P. Delatorre, A. Salguero, C. León, and A. Tapscott, "The impact of context on affective norms: A case of study with suspense," *Frontiers Psychol.*, vol. 10, p. 1988, Aug. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.01988

[35] P. Delatorre, A. Berns, M. Palomo-Duarte, P. Gervás, and F. Madueño, "Diseno de un juego serio basado en el suspense," in *Proc. 2nd Congreso de la Sociedad Española para las Ciencias del Videojuego, CEUR Workshop*, vol. 1394, D. Camacho, M. A. Gómez-Martín, and P. A. González-Calero, Eds. Barcelona, Spain: CEUR-WS.org, Jun. 2015, pp. 102–111. [Online]. Available: http://hdl.handle.net/10498/18005

[36] P. Delatorre, M. Palomo-Duarte, and P. Gervás, "Formalising suspense from immersive environments," in *Proc. 3rd Congreso de la Sociedad Española para las Ciencias del Videojuego (CoSECiVi)*, Jun. 2016, pp. 25–36.

[37] P. Delatorre, C. León, P. Gervás, and M. Palomo-Duarte, "A computational model of the cognitive impact of decorative elements on the perception of suspense," *Connection Sci.*, vol. 29, no. 4, pp. 295–331, Oct. 2017, doi: 10.1080/09540091.2017.1345856.

[38] P. Delatorre, C. León, A. Salguero, C. Mateo-Gil, and P. Gervás, "Impact of interactivity on information management for suspense in storytelling," in *Proc. AISB Annu. Conv.*, J. P. J. Bryson and M. De Vos, Eds. Liverpool, U.K.: Society With AI, 2017, pp. 110–116. [Online]. Available: http://aisb2017.cs.bath.ac.uk/conference-edition-proceedings.pdf

[39] P. Delatorre, C. León, A. Salguero, M. Palomo-Duarte, and P. Gervás, "Confronting a paradox: A new perspective of the impact of uncertainty in suspense," *Frontiers Psychol.*, vol. 9, p. 1392, Aug. 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01392

[40] P. Delatorre, C. León, A. G. Salguero, and A. Tapscott, "Predicting the effects of suspenseful outcome for automatic storytelling," *Knowl.-Based Syst.*, vol. 209, Dec. 2020, Art. no. 106450.

[41] M. D. Vose, *The Simple Genetic Algorithm: Foundations and Theory* (A Bradford Book). Cambridge, MA, USA: MIT Press, 1999.

[42] R. Dale, "NLP meets the cloud," *Natural Lang. Eng.*, vol. 21, no. 4, pp. 653–659, Aug. 2015.

[43] I. S. Abeywardena, "Public opinion on OER and MOOC: A sentiment analysis of Twitter data," in *Proc. Int. Conf. Open Flexible Educ.*, 2014, pp. 296–307.

[44] M.-J. Lee, T.-R. Lee, S.-J. Lee, J.-S. Jang, and E. J. Kim, "Machine learning-based data mining method for sentiment analysis of the Sewol ferry Disaster's effect on social stress," *Frontiers Psychiatry*, vol. 11, Dec. 2020, Art. no. 505673.

[45] D. G. A. Rodrigues, "Hotel online reviews: What influences the sentiment and rating? A sentiment analysis on Lisbon hotels," M.S. thesis, Dept. Inf. Syst. Technol. Manage., NOVA Inf. Manage. School (NIMS), Lisboa, Portugal, 2020.
[46] J. Kim, G. Naylor, E. Sivadas, and V. Sugumaran, "The unrealized value of incentivized eWOM recommendations," *Marketing Lett.*, vol. 27, no. 3, pp. 411–421, Sep. 2016.
[47] A. M. Alkalbani, A. M. Ghamry, F. K. Hussain, and O. K. Hussain, "Sentiment analysis and classification for software as a service reviews," in *Proc. IEEE 30th Int. Conf. Adv. Inf. Netw. Appl. (AINA)*. Crans-Montana, Switzerland: IEEE, Mar. 2016, pp. 53–58.
[48] G. Freytag, *Freytag's Technique of the Drama: An Exposition of Dramatic Composition and Art*. Glenview, IL, USA: Scott Foresman, 1894.
[49] M. Sternberg, *Expositional Modes and Temporal Ordering in Fiction*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1978.
[50] H. J. Wulff, "Suspense and the influence of cataphora on viewers' expectations," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1996, pp. 1–17.
[51] P. Highsmith, *Plotting and Writing Suspense Fiction*. New York, NY, USA: St. Martin's Griffin, 1990.
[52] T. J. Connelly, *Cinema of Confinement*. Evanston, IL, USA: Northwestern Univ. Press, 2019.
[53] S. Smith, *Hitchcock: Suspense, Humour and Tone*. London, U.K.: Bloomsbury, 2019.
[54] N. E. Batty, "The art of suspense: Rushdie's 1001 (mid-) nights," *ARIEL A, Rev. Int. English Literature*, vol. 18, no. 3, pp. 1–17, 1987.
[55] A. M. Alaa and M. Van Der Schaar, "Balancing suspense and surprise: Timely decision making with endogenous information acquisition," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 2910–2918.
[56] P. B. Bafna and J. R. Saini, "On readability metrics of goal statements of universities and brand-promoting lexicons for industries," in *Data Management, Analytics and Innovation*, vol. 1. Singapore: Springer, 2020, p. 63.
[57] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll, "Learning by feeling: Evoking empathy with synthetic characters," *Appl. Artif. Intell.*, vol. 19, nos. 3–4, pp. 235–266, Mar. 2005.
[58] D. Beecher, "Suspense," *Philosophy Literature*, vol. 31, no. 2, pp. 255–279, 2007.
[59] K. Dijkstra, R. A. Zwaan, A. C. Graesser, and J. P. Magliano, "Character and reader emotions in literary texts," *Poetics*, vol. 23, nos. 1–2, pp. 139–157, Jan. 1995.
[60] D. Zillmann, *The Psychology of Suspense in Dramatic Exposition*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 1996, pp. 199–231.
[61] S. Keen, "A theory of narrative empathy," *Narrative*, vol. 14, no. 3, pp. 207–236, 2006.
[62] T. M. Eilola and J. Havelka, "Affective norms for 210 British English and Finnish nouns," *Behav. Res. Methods*, vol. 42, no. 1, pp. 134–140, Feb. 2010.
[63] A. Moors, J. De Houwer, D. Hermans, S. Wanmaker, K. van Schie, A.-L. Van Harmelen, M. De Schryver, J. De Winne, and M. Brysbaert, "Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words," *Behav. Res. Methods*, vol. 45, no. 1, pp. 169–177, Mar. 2013.
[64] M. Montefinese, E. Ambrosini, B. Fairfield, and N. Mammarella, "The adaptation of the affective norms for English words (ANEW) for Italian," *Behav. Res. Methods*, vol. 46, no. 3, pp. 887–903, Sep. 2014.
[65] D. Georgiou, A. MacFarlane, and T. Russell-Rose, "Extracting sentiment from healthcare survey data: An evaluation of sentiment analysis tools," in *Proc. Sci. Inf. Conf. (SAI)*, Jul. 2015, pp. 352–361.
[66] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Syst. Appl.*, vol. 77, pp. 236–246, Jul. 2017.
[67] S. Gao, J. Hao, and Y. Fu, "The application and comparison of Web services for sentiment analysis in tourism," in *Proc. 12th Int. Conf. Service Syst. Service Manage. (ICSSSM)*, Jun. 2015, pp. 1–6.
[68] M. J. F. Guijarro, "A functional lexematic analysis of a horror short story," *Cuadernos de investigación filológica*, vol. 23, pp. 99–114, Jan. 1997.
[69] S. H. Lee, J. Cui, and J. W. Kim, "Sentiment analysis on movie review through building modified sentiment dictionary by movie genre," *J. Intell. Inf. Syst.*, vol. 22, no. 2, pp. 97–113, Jun. 2016.
[70] I. Gasquet, B. Falissard, and P. Ravaud, "Impact of reminders and method of questionnaire distribution on patient response to mail-back satisfaction survey," *J. Clin. Epidemiol.*, vol. 54, no. 11, pp. 1174–1180, Nov. 2001.
[71] A. P. Soares, M. Comesaña, A. P. Pinheiro, A. Simões, and C. S. Frade, "The adaptation of the affective norms for English words (ANEW) for European Portuguese," *Behav. Res. Methods*, vol. 44, no. 1, pp. 256–269, Mar. 2012.
[72] M. Guasch, P. Ferré, and I. Fraga, "Spanish norms for affective and lexico-semantic variables for 1,400 words," *Behav. Res. Methods*, vol. 48, no. 4, pp. 1358–1369, Dec. 2016, doi: 10.3758/s13428-015-0684-y.
[73] U.-D. Reips, "The Web experiment method: Advantages, disadvantages, and solutions," in *Psychological Experiments on the Internet*, M. H. Birnbaum, Ed. Amsterdam, The Netherlands: Elsevier, 2000, pp. 89–117.
[74] A. Risso, "Metodología de investigación de las www," in *Metodología de las Ciencias del comportamiento, Asociación Española de Metodología de las Ciencias del Comportamiento*, La Rioja, Ed. Murcia, Spain: Univ. of Murcia, Dialnet, 2002, pp. 487–491.
[75] T. Buchanan and J. L. Smith, "Using the Internet for psychological research: Personality testing on the World Wide Web," *Brit. J. Psychol.*, vol. 90, no. 1, pp. 125–144, Feb. 1999.
[76] M. A. Smith and B. Leigh, "Virtual subjects: Using the Internet as an alternative source of subjects and research environment," *Behav. Res. Methods, Instrum., Comput.*, vol. 29, no. 4, pp. 496–505, Dec. 1997.
[77] J. M. Stanton, "An empirical assessment of data collection using the Internet," *Personnel Psychol.*, vol. 51, no. 3, pp. 709–725, Sep. 1998.
[78] T. Buchanan, "Potential of the Internet for personality research," in *Psychological Experiments on the Internet*, M. H. Birnbaum and M. O. Birnbaum, Eds. Amsterdam, The Netherlands: Elsevier, 2000, pp. 121–140.
[79] K. A. Pasveer and J. H. Ellard, "The making of a personality inventory: Help from the WWW," *Behav. Res. Methods, Instrum., Comput.*, vol. 30, no. 2, pp. 309–313, Jun. 1998.
[80] G. G. Chowdhury, "Natural language processing," in *Fundamentals of Artificial Intelligence*. New Delhi, India: Springer, 2020, pp. 603–649.

PABLO DELATORRE received the M.Sc. and B.Sc. degrees in computer engineering from the Universidad de Sevilla and the Ph.D. degree from the Complutense University of Madrid (UCM). His research is focused in narrative and interactive suspense applied to automatic story generation, ontologies to the recognition of activities of daily living, and education. He combines this research activity with his responsibilities as a Teacher at the University of Cadiz. Beyond the academic profession, he worked ten years in the ICT sector, managing national and international corporate projects.

CARLOS LEÓN is currently an Associate Professor with the Complutense University of Madrid. His research focuses on the use of computational creativity techniques for automatic generation of content in videogames and interactive systems, with a strong focus on human cognition. He has published more than 60 articles in this field. He is a member of the Advisory Board of the Association for Computational Creativity.

ALBERTO SALGUERO HIDALGO received the B.Sc. and M.Sc. degrees in computer science from the University of Granada, Granada, Spain, in 2001 and 2004, respectively, and the Ph.D. degree in computer science from the University of Cádiz, Spain, in 2013. He has been an Interim Professor of computer science with the University of Cádiz, since 2010. His research interest includes ontologies and how they can be used to improve information systems. More specifically, for the past few years, he has been focused on applying ontologies to the recognition of activities of daily living.