Evolution and Human Behavior 47 (2026) 106813

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Vulnerability and the computational logic of fear: insights from the horror genre

Edgar Dubourg [a], Coltan Scrivner [b,*]

[a] Institut Jean Nicod, Département d'études cognitives, École normale supérieure, Université PSL 29 Rue d'Ulm, 75005 Paris, France
[b] Department of Psychology, Arizona State University, 950 S McAllister Ave, Tempe, AZ 85281, United States

ARTICLE INFO

Keywords:
Fear
Horror
Fiction
Emotion
Formidability

ABSTRACT

Fear is a universal feature of storytelling, yet the structural conditions that make fictional threats compelling remain poorly understood. Here, we propose the *Protagonist Vulnerability Index* (PVI), an evolutionarily grounded computational approach to explain why some narratives evoke stronger fear responses than others. PVI quantifies protagonist vulnerability by assessing the imbalance in formidability between protagonists and antagonists and the risk of attack faced by the protagonist. Across 691 films, higher PVI values predicted classification as horror, the presence of fear-related keywords in non-horror films, and stronger physiological fear responses indexed by heart rate. Linking film preferences to psychological and demographic data from more than 3.5 million individuals on Facebook, we found that preference for high-PVI films was associated with lower agreeableness, conscientiousness, and extraversion, and with higher openness. Openness moderated the negative association between neuroticism and engagement with fear-related content, indicating that curiosity can counteract threat avoidance in anxious individuals. These findings clarify the structural and psychological conditions that activate evolved threat-management systems. The results show how horror operates as a narrative simulation of extreme formidability asymmetry, and provide a framework for predicting, and potentially engineering, fear in fiction.

1. Introduction

1.1. Formidability and the nature of fear

The concept of formidability, derived from Resource Holding Potential in the animal behavior literature (Parker, 1974) refers to the perceived capacity of an agent to inflict harm (Sell et al., 2009). Across species, size and strength (including muscle mass and natural weaponry) are the strongest predictors of high formidability and conflict success (Archer, 1988; Emlen, 2008; Fessler et al., 2012; Sell et al., 2009). In addition to size and strength, behavioral tendencies such as motivation, persistence, and aggression also influence conflict success and perceived formidability (Barlow et al., 1986; Elwood, 2009; Hofmann & Schildberger, 2001; Hurd, 2006; Scrivner et al., 2020). Formidability can also be augmented with social, technological, or situational advantages that affect one's ability to inflict or withstand harm (e.g., see Fessler et al., 2012; Fessler & Holbrook, 2014).

Accordingly, animals, including humans, have evolved threat-assessment systems that use cues of formidability to regulate conflict avoidance (Durkee et al., 2018; Parker, 1974; Sell et al., 2009). When the perceived balance of formidability is skewed against the perceiver, the asymmetry evokes a sense of vulnerability: a computational appraisal of poor survival odds in an asymmetric conflict. Fear functions as the emotional output of that appraisal, coordinating defensive attention, arousal, and avoidance (Armfield, 2006; Mobbs et al., 2009). Thus, perceiving a more formidable adversary activates fear not simply because the adversary is powerful, but because the perceiver computes their own limited capacity to predict, control, or escape harm.

1.2. The computational structure of horror

Philosophers and psychologists have described horror as phenomenologically distinct from fear. Carroll (1990) argued that horror is produced by encounters with entities that violate ordinary ontological categories. Empirically, horror is more often elicited by abnormal or "schema-incongruent" harm than by ordinary danger (Taylor & Uchida, 2022). Large-scale mapping of emotional experience also places horror adjacent to, but distinct from, fear. In a study of 27 emotion categories

* Corresponding author at: Department of Psychology, Arizona State University, 950 S McAllister, Ave, Tempe, AZ 85281, USA
E-mail address: cscrivner1@gmail.com (C. Scrivner).

https://doi.org/10.1016/j.evolhumbehav.2025.106813
Received 2 July 2025; Received in revised form 8 November 2025; Accepted 8 December 2025
Available online 11 December 2025

derived from self-reports across thousands of stimuli, Cowen and Keltner (2017) identified horror as a distinct category lying along the semantic gradient between fear and disgust.

We argue that the phenomenology of horror arises when a threat is appraised as capable of inflicting severe harm due to overwhelming formidability. This framework provides a deeper computational basis for previous conceptualizations of horror. When the imbalance in formidability between an individual and a threat becomes extreme, the appraisal exceeds the parameters of ordinary threat schemas. Harm posed by the antagonist appears severe and violates expectations about what an agent should be capable of inflicting on another. This pattern of appraisal produces the characteristic phenomenology of horror: fear mixed with helplessness.

1.3. The cognitive architecture of horror stories

By empathizing with characters in stories, audiences can simulate the experiences of those characters, leading to the elicitation of relevant emotions for the situation (Grodal, 2009; Oatley, 1999). When audiences engage with narratives in which the protagonist is vulnerable to a more formidable antagonist, they tend to empathize with that character and may simulate the character's vulnerability as if it were their own, resulting in a fearful experience (Lynch, 2018). In some cases, audiences may simulate the experience and fear *for* the vulnerable protagonist (Smith, 1995; Tan, 1996), engaging not through direct identification but through sympathetic concern (Singh, 2021). These instances become clear when audience members in a horror film shout words of warning at the screen, such as, "don't go in there!" and "he's behind you!" In fact, fans of horror films appear to be particularly good at perspective taking (Scrivner, 2024), which may make the vulnerability of the protagonist feel even more salient in scenes of danger.

Fear and feelings of vulnerability don't emerge only from the formidability of the antagonist, but also from the space in which they are encountered. In nature, animals navigate through a world of perceived risk that shifts with both the presence of predators and the hostility of the surrounding environment, what ecologists call the "landscape of fear" (Laundré et al., 2010). A prey animal's vigilance, movement, and arousal all depend on the immediacy of the threat and perceived danger of the environment (Mobbs et al., 2020). The horror genre operates by manipulating the same parameters. A monster's presence increases perceived threat, while the setting (e.g., dark, claustrophobic, or unpredictable) amplifies environmental hostility. Together they create a fictional landscape of fear: an imagined ecology in which the viewer, like prey, must navigate uncertainty, assess danger, and anticipate attack.

Taken together, we propose that the defining structure of the horror genre centers around a vulnerable protagonist who is threatened by a more formidable antagonist. The protagonist's extreme vulnerability creates conditions for sustained fear and tension, the affective hallmarks of horror. Consequently, narratives in which the antagonist holds a clear advantage in formidability and poses a credible risk of attack should evoke greater fear in audiences.

1.4. Predictions

Our general hypothesis is that the frightening nature of horror stories stems primarily from the fact that their character structure successfully triggers phylogenetically old neural circuits that mediate threat avoidance and predator evasion. These mechanisms take as input both the threat posed by an antagonist (i.e., their formidability) and the protagonist's ability to counter or escape that threat (i.e., their own formidability). If this hypothesis is correct, then the following predictions should hold:

P1. Horror narratives will be more likely than non-horror narratives to feature vulnerable protagonists who possess low relative formidability compared to their antagonists.

P2. Films featuring vulnerable protagonists (formidability imbalance favoring the antagonist) will be more frightening.

2. General method

2.1. Overview

To systematically assess the formidability imbalance in horror films, we use an automated annotation method powered by GPT. Traditional manual coding methods struggle with consistency and scalability when applied to large, diverse datasets. LLMs like GPT offer a reliable alternative by synthesizing and standardizing information across different sources (in the case of movies, texts like scripts, reviews, comments, or summaries) on which they have been trained. Importantly, in this study the model was not provided with any external text input such as summaries or scripts; instead, it received only the title and year of release for each film and drew on its latent, generalized knowledge of each film acquired during pretraining. Recent studies have shown that GPT can match or surpass human raters in cultural annotation tasks (Bongini et al., 2023; Gilardi et al., 2023; Pei et al., 2023; Rathje et al., 2023). Its zero-shot learning capability allows it to perform annotation tasks without retraining (Bongini et al., 2023; Ding et al., 2023; Kuzman et al., 2023; Pei et al., 2023). Additionally, Bongini et al. (2023) highlight GPT's ability to incorporate new cultural knowledge without fine-tuning, making it highly suitable for this study. Previous studies have successfully used LLMs to annotate cultural products, including video games (Dubourg & Chambon, 2025), literary works (Dubourg, Safan, et al., 2025), and movies and novels from around the world (Dubourg, Thouzeau, et al., 2025).

Our method (see Dubourg, Thouzeau, et al., 2024, for the method) leverages GPT's ability to rate protagonist's and antagonist's formidability across structured scales. This allows for the computation of key variables for testing whether horror stories systematically feature an imbalanced power dynamic. Following Dubourg, Valentin, and Baumard (2024), we finalized and pre-registered the prompts after conducting a series of prompt engineering iterations and pilot tests on a subset of movies (see Supplementary Materials for the full prompts and scales).

2.2. Annotated variables

To ensure precise and consistent ratings, we had multiple prompts to assess formidability. As formidability include the capacity to inflict costs on others (offense), the capacity to defend oneself from harm (defense), and the motivation to pursue or persist in these efforts (motivation & persistence), we used three different prompts to score both the protagonist's and the antagonist's formidability. Each prompt was used two times: once for *Protagonist Formidability* (P) and once for *Antagonist Formidability* (A).

Then, we assessed the presence and prominence of the threat throughout the movie. The degree to which the antagonist is present and actively threatening varies across films. A horror film with a highly formidable antagonist may not be frightening if the threat is rarely encountered, whereas a constant, looming threat can amplify fear even if the antagonist is not overwhelmingly powerful. To capture this, we define *Threat Presence* (T) as the extent to which the antagonist or their influence is felt throughout the movie, from completely absent to an overwhelming, constant menace.

Next, we assessed the level of danger posed by the environment itself, independent of any specific antagonist or character actions. Some environments inherently decrease the odds of the protagonist's survival through harsh conditions, lethal hazards, relatively few areas of refuge, or areas for ambush predators to hide. Thus, *Environmental Hostility* (E) captures how dangerous the setting is for the protagonist.

2.3. Protagonist vulnerability index (PVI)

To assess protagonist vulnerability, we developed the Protagonist Vulnerability Index (PVI) formula, computed by combining character-level and situational features annotated by GPT.

For both the protagonist (P) and the antagonist (A) formidability scores, we aggregated sub-components (offense, defense, and motivation/persistence) using a root mean square (RMS), which gives greater weight to extreme values and reflects the idea that a single overwhelming trait can determine the character's effectiveness in conflict. We then incorporated our two situational variables, Threat Presence (T) and Environmental Hostility (E). Because E is intended to scale the overall threat level, we transformed its original 1–10 range into a scale from 0.5 to 1.5. This rescaled value is then used as a multiplier in the final formula: when the environment is relatively safe (values below 1), it dampens the threat; when the environment is highly hostile (values above 1), it amplifies it.

We then devised a formula using these variables that calculates protagonist vulnerability. The Protagonist Vulnerability Index (PVI) was calculated with the formula $(A/P) \times E \times T$, capturing the idea that threat management systems are triggered by an imbalance in power favoring the antagonist, and that this can be augmented by an unfavorable environment and ever-present antagonist. We standardized our PVI score (z-score) to facilitate interpretation and comparability across analyses. Because the distribution of PVI was right-skewed, we applied a logarithmic transformation to improve normality. Prior to this transformation, we shifted all values to ensure they were strictly positive by subtracting the minimum and adding 1. The resulting log-transformed score was used in subsequent analyses.

To evaluate the robustness of our results, we recomputed the Protagonist Vulnerability Index (PVI) across 36 alternative formulations, systematically varying every reasonable analytic choice involved in its computation. Each formulation was defined by a combination of four dimensions: (1) Aggregation method: antagonist and protagonist formidability scores (offense, defense, motivation) were combined either using the root-mean-square (RMS) or the arithmetic mean. (2) Inclusion of additional vulnerability factors; we tested whether to include threat presence (T) and/or environmental hostility (E) in addition to the protagonist/antagonist ratio, resulting in four possibilities: no additional factors, T only, E only, or both T and E. (3) Rescaling of environmental hostility (E): when included, E was either left on its original 1–10 scale or rescaled to a smaller range (0.5–1.5 or 1–2) to reduce its influence. (4) Score transformation: after computing the index, we applied three alternative transformations: no transformation, log-shift (to reduce skew), or z-score standardization (for normalization). The full factorial combination of these parameters produced 36 unique variants. Each version was computed for all films in the dataset and used for our three main predictions: (1) predicting horror vs non-horror films (Study 1), (2) among non-horror films, predicting those with vs without fear-related keywords (Study 2), (3) among horror films, predicting average heart rate (Study 3).

3. Study 1: predicting horror classification

3.1. Data

We annotated a total of 691 movies using our structured GPT-based method. These films were selectively drawn from the IMDb database. We retained only those movies for which we had access to personality and demographic data from Nave et al. (2020), which aggregates the average Big Five traits, age, and gender of Facebook users who "Liked" each movie (N = 3.5 million). The oldest movie dates back to 1939, while the most recent was released in 2017, with a median year of 2004. Genre diversity is high: the most common categories include drama (n = 337), comedy (n = 277), action (n = 178), romance (n = 153), crime (n = 118), and adventure (n = 112). Horror, the focus of the study, is also well represented (n = 73), along with other related genres such as thriller (n = 106), fantasy (n = 86), and science fiction (n = 67). Movies vary substantially in popularity and critical reception. Average IMDb ratings range from 1.9 to 9.3, with a mean of 6.94 and a median of 7.00. The number of user votes ranges from 533 to over 2.2 million, with a median of 198,469 and a mean of 281,354.

3.2. Analyses and results

3.2.1. Individual variables as predictors of genre classification

First, we wanted to evaluate the extent to which our four variables independently predict whether a movie belongs to the horror genre. We fitted a logistic regression model using protagonist formidability (P), antagonist formidability (A), threat presence (T), and environmental hostility (E) as predictors. The results were consistent with our theoretical expectations: horror films are more likely to involve weak protagonists (ß = −0.80, $p < 0.001$), strong antagonists (ß = 0.61, $p < 0.001$), pervasive threat presence (ß = 0.79, $p = 0.002$), and hostile environments (ß = 0.33, $p < 0.001$). To assess the specificity of these effects, we ran the same model across a range of other genres and plotted the resulting coefficients. As expected, horror stood out with the most pronounced configuration of low protagonist formidability, high antagonist formidability, frequent threat, and hostile environment. Other genres, by contrast, either showed weaker or reversed patterns—for instance, mystery movies also tended to feature strong antagonists and persistent threats, but they differed from horror films by involving significantly stronger protagonists and showing no particular pattern of environmental hostility (Fig. 1**.A**).

3.2.2. PVI scores as predictors of genre classification

To evaluate whether our Protagonist Vulnerability Index (PVI, described in Section 2.3) meaningfully varies across narrative genres, we computed PVI scores for all films in our sample and compared average scores between genre-defined groups. We focused on six major groups: films classified exclusively as horror, films classified as horror in combination with other genres, and four other genre categories: thriller, action, fantasy, and romance. The films we analyzed in these other four categories did not have horror listed as a secondary genre on IMDb.

We conducted pairwise comparisons using independent-samples t-tests between each of these genre groups to assess whether the mean PVI significantly differed across them. The results reveal a systematic gradient: films associated with the horror genre (especially exclusive horror) tend to exhibit higher PVI scores. More specifically, horror films were significantly higher in PVI than all other categories (all $p < 0.001$), with the sole exception of the comparison between exclusive and non-exclusive horror, which did not reach significance ($p = 0.14$; see Supplementary Materials for all pairwise comparisons). Romance films scored significantly lower than all other genres (all $p < 0.001$), followed by fantasy, action, and thriller. These findings support the idea that horror is structurally defined by extreme asymmetries between protagonists and antagonists in threatening environments, and that the PVI can discriminate genres along this narrative dimension (Fig. 1**.B**).

To address class imbalance (73/691 horror films) and guard against overfitting, we ran a 5-fold stratified cross-validated logistic regression predicting horror (vs. non-horror) from PVI and its components (P, A, T, E). We report class-balanced metrics. Performance was robust: AUC-ROC = 0.887 (SE = 0.011), AUC-PR = 0.971 (SE = 0.004), and balanced accuracy = 0.714 (SE = 0.025) across folds. These results confirm that PVI-based features discriminate horror out-of-sample well above a predict-all-non-horror baseline.

To test whether our findings depend on the specific way PVI is computed, we reran the horror vs. non-horror classification analyses using all 36 alternative PVI formulations described in Section 2.3. In every case, higher PVI significantly predicted horror classification (36/36 models; see **Table S1**), indicating that the link between protagonist vulnerability and horror genre membership is robust to changes in

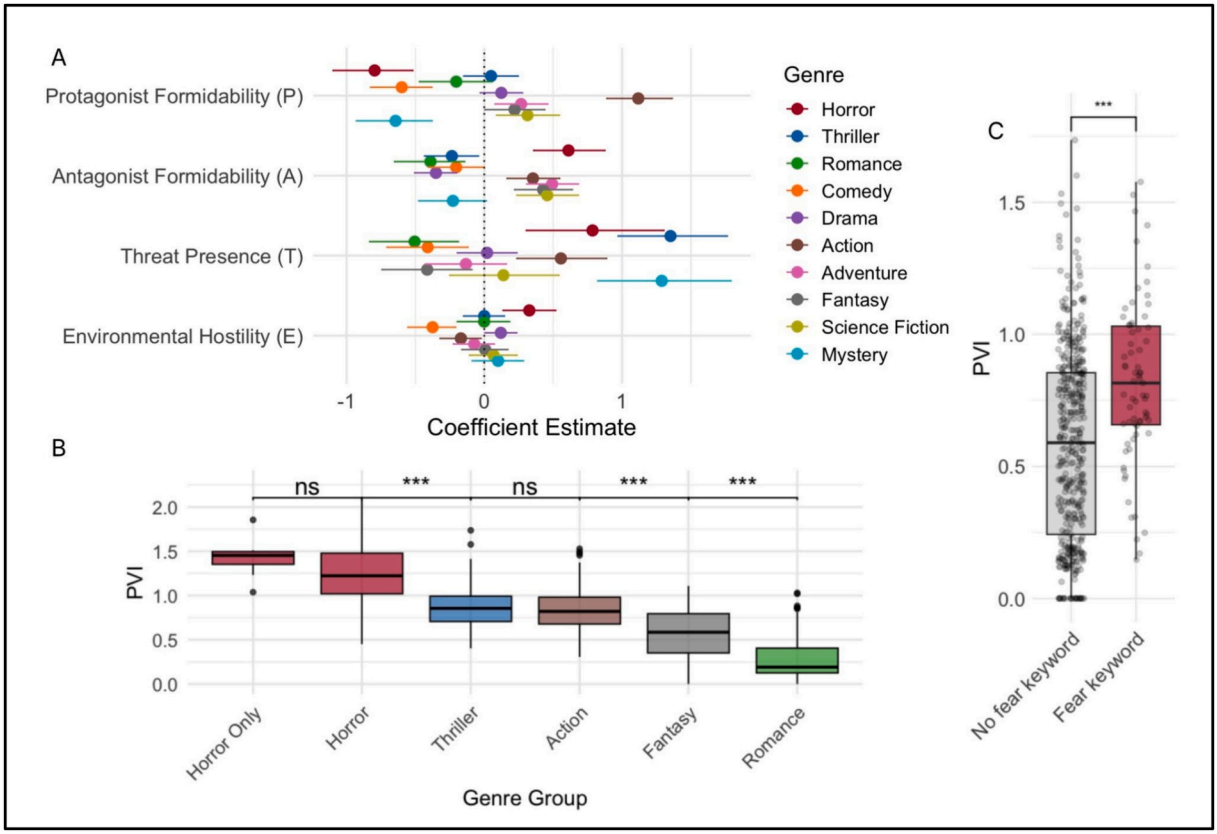

Fig. 1. A. Logistic regression coefficients predicting movie genre from PVI components. Horror is uniquely characterized by weaker protagonists, stronger antagonists, higher threat presence, and greater environmental hostility. **B.** Distribution of the PVI across genre groups. Horror movies show significantly higher PVI than all other genres. Horror Only refers to films classified exclusively as horror, while Horror refers to films tagged as horror alongside one or more additional genres. Significance bars are shown only for adjacent categories for visual clarity; see Supplementary Materials for the full statistical output. Both Horror Only and Horror films exhibit significantly higher PVI values than all other genres. **C.** Boxplot comparing log-transformed PVI scores across non-horror movies with and without fear-related keywords (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; ns = not significant).

aggregation, scaling, and transformation choices.

4. Study 2: Predicting fear

4.1. Introduction

Fear can be defined as a conscious awareness that you are in danger (LeDoux, 2014). Although they are neurologically separable, feeling afraid tends to co-occur with threat detection because subcortical circuits involved in threat management are a key input to the cortical circuit that mediates the fear experience. The outputs of this system are autonomic responses such as increased heart rate, pupil dilation, and perspiration. Heart rate in particular is a reliable and easily measurable output of threat management activation and cognitive feelings of fear. Therefore, we should expect that horror films with higher PVI (i.e., more threatening antagonists or situations) will elicit more fear and higher heart rate in viewers.

4.2. Data and method

To test whether the PVI captures fear-inducing narrative content beyond the boundaries of genre classification, we turned to IMDb's keyword system—a user-generated tagging feature that allows contributors to associate descriptive terms with films, including emotions, themes, or plot elements. This system enabled us to identify non-horror movies that still involve fearful content, as perceived by viewers. Focusing on the subset of films in our dataset that were not classified as horror ($n = 512$), we examined whether movies tagged with the "fear" keyword had higher PVI scores than those without such tags.

To test that higher PVI elicit higher heart rate in viewers, we used data from the Science of Scare Project (MoneySuperMarket, 2025). The aim of this project is to identify the scariest horror movies based on the average heart rate of its viewers. For each movie, the dataset reports the mean heart rate (BPM) and mean heart-rate increase relative to a resting baseline, averaged across approximately 250 adult participants. During controlled screenings, participants wore heart-rate monitors throughout each film, and their BPM was continuously recorded. Individual-level data are not publicly available, so each film is treated as a single aggregate observation in our analyses. While this limits the possibility of estimating within-film variance or participant-level effects, the consistency of the methodology across all films minimizes noise and allows valid cross-film comparisons.

Because heart rate (beats per minute, or BPM) is a reliable measure of fear, films where viewers had higher average BPM are considered to be scarier. Films are chosen each year based on lists by critics, fans, and expert recommendations, and include films from many different decades. The Science of Scare project includes BPM data for studies that took place each year from 2020 to 24. We excluded data from the 2024 study since it included films released in 2024 and GPT's training data only includes knowledge up to mid-2024. We ended up with a final sample of 56 horror movies for which we had both average BPM scores and the PVI scores.

4.3. Analyses and results

4.3.1. PVI scores as predictors of the presence of fear in non-horror movies

A Welch two-sample *t*-test revealed a significant difference in log-transformed PVI scores between non-horror movies with fear-related keywords (mean = 0.82) and those without (mean = 0.58; t(103.1) = −5.73, $p < 0.001$, 95 % CI [−0.32, −0.15]; see Fig. 1.**C**). These findings indicate that the PVI is not only diagnostic of genre membership but also tracks the presence of fear-evoking content at a finer narrative level.

Across the 36 alternative PVI computations (see Section 2.3.), 32 produced significant effects (**Table S1**); the only non-significant cases were the unaugmented A/P ratio without T or E (mean aggregation with none/log-shift/z-score, and RMS with log-shift).

4.3.2. PVI scores as predictors of heart rate in viewers

To assess whether the PVI score predicts the intensity of fear elicited by horror films, we regressed average heart rate increase from baseline (a proxy for fear) on PVI across the 56 movies. PVI significantly predicted the average increase in heart rate during horror films (ß = 0.50, $p = 0.019$). This association remained significant when controlling for the release year of the movie (ß = 0.47, $p = 0.025$), while the effect of year itself was not significant in the model (ß = 0.06, $p = 0.13$). These results suggest that our narrative-based PVI score that captures the psychological structure of threat is a reliable predictor of physiological fear responses across horror films (Fig. 2.**A**). This result was again robust to changes in the formula used to compute PVI. When we removed the situational variables and used only the ratio of antagonist to protagonist formidability, this simplified measure still significantly predicted the average heart rate (BPM) elicited by each film (p = 0.019; see Supplementary Materials).

All 36/36 PVI formulations significantly predicted both outcomes (average heart rate and the difference between each film's average BPM and the participants' resting BPM) whether or not the release year was included as a covariate (i.e., 144 unique models). This complete convergence across aggregation, scaling, and transformation choices confirms that the association between protagonist vulnerability and fear responses is robust.

5. Study 3: Predicting audience profile

5.1. Data

To investigate whether our PVI scores are associated with audience characteristics, we combined the sample of 691 films annotated with our method in Study 1 with the dataset compiled by Nave et al. (2020), which links over 3.5 million Facebook users' movie preferences to their Big Five personality traits, gender, and age.

5.2. Analyses and results

To assess whether individual differences in personality and socio-demographics predict preferences for frightening stories, we examined bivariate correlations between log-transformed PVI scores and each predictor. Openness to experience was positively associated with a preference for frightening stories ($r = 0.13$, $p = 0.0019$), suggesting that individuals high in curiosity and novelty-seeking are more likely to enjoy high-threat narratives. In contrast, agreeableness ($r = -0.42$, $p < 0.001$), gender ($r = -0.31$, $p < 0.001$; films with more male-skewed audiences tend to have lower PVI), conscientiousness ($r = -0.16$, $p < 0.001$), and extraversion ($r = -0.37$, $p < 0.001$) were all negatively correlated with frightening story preference. Neuroticism showed a marginally negative association ($r = -0.08$, $p = 0.053$), and age had no significant effect ($r = -0.02$, $p = 0.64$; although note that the age range in this dataset is restricted to 18–35).

We then entered all predictors into a single model. Crucially, we included an interaction between neuroticism and openness to test whether the effect of threat sensitivity (neuroticism) on movie preference (indexed by PVI) depends on the individual's tendency to be curious (openness). We included this interaction to test a key theoretical prediction: that neither trait alone is sufficient to explain preference for fear-inducing narratives. Individuals high in neuroticism but low in openness may find such content overwhelming or distressing, leading them to avoid it. Conversely, individuals high in openness but low in neuroticism may lack the heightened vigilance or arousal that makes threatening stimuli psychologically engaging, rendering horror

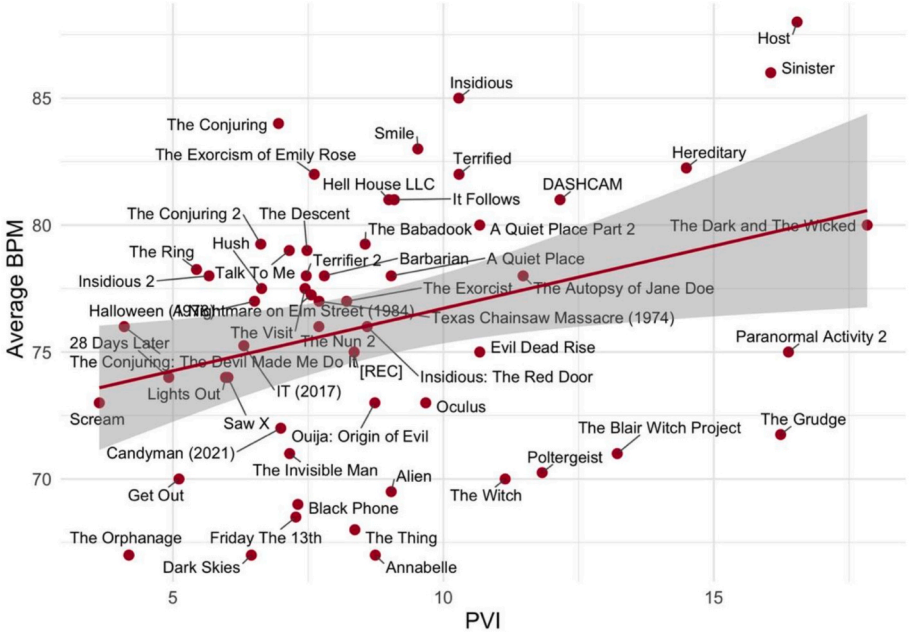

Fig. 2. Relationship between Protagonist Vulnerability Index (PVI) and average heart rate (BPM) across horror films from the Science of Scare dataset. Each point represents one movie. Higher PVI scores are associated with higher heart rate responses, suggesting that fear intensity increases when vulnerable protagonists face powerful antagonists in threatening environments. **B.** Standardized coefficient estimates predicting PVI (log) from personality traits, gender, and age. Error bars represent 95 % confidence intervals.

narratives emotionally flat or uninteresting. We expect the combination of both traits to lead to morbid curiosity.

With all variables in a single model, the analysis revealed significant main effects of neuroticism (ß = −1.05, p < 0.001), agreeableness (ß = −1.23, p < 0.001), conscientiousness (ß = 0.79, p < 0.001), extraversion (ß = −0.77, p < 0.001), gender (ß = −0.75, p < 0.001), and age (ß = −0.03, $p = 0.003$). Openness had a marginally significant negative effect (ß = −0.19, $p = 0.095$). Most importantly, the neuroticism × openness interaction was significant (ß = 1.12, $p = 0.039$), indicating that the relationship between neuroticism and PVI depends on openness (Fig. 3. **A.**).

The significant positive interaction between neuroticism and openness (ß = 1.39, $p = 0.013$) indicates that the effect of neuroticism on preference for frightening movies depends on openness. Specifically, individuals high in neuroticism tend to avoid high-threat movies. However, this aversion is attenuated in individuals who are also high in openness. In other words, curiosity appears to buffer the effect of neuroticism: curious individuals who are also threat-sensitive may still engage with frightening narratives, possibly to satisfy a form of morbid curiosity.

6. Discussion

Across three studies, we investigated the structural and psychological nature of fear in narrative fiction through a novel computational measure, the Protagonist Vulnerability Index (PVI). PVI quantifies the vulnerability of a protagonist as a function of their formidability compared to the antagonist and the risk of attack by the antagonist. Study 1 showed that horror films are structurally distinct from other genres, featuring weaker protagonists, stronger antagonists, persistent threats, and more hostile environments—elements that elevate PVI and reliably differentiate horror from other narrative categories. Study 2 demonstrated that PVI scores predict physiological markers of fear responses in viewers, with higher PVI values associated with increased heart rates during horror film viewing. PVI also predicted whether non-horror films were tagged with the fear keyword, suggesting that it captures fear-relevant structure beyond genre labels. Study 3 revealed that preferences for high-PVI films vary systematically across personality profiles: individuals high in openness and low in agreeableness, conscientiousness, and extraversion were more likely to enjoy higher-PVI movies. Most importantly, a significant interaction between neuroticism and openness indicated that curiosity can modulate avoidance tendencies, enabling anxious individuals to engage with threatening content. This corresponds with the concept of morbid curiosity and previous work showing that some anxious people enjoy horror entertainment (Scrivner, 2021a, 2021b).

These findings provide converging evidence that the fear elicited by horror stories arises from an alignment between evolved threat-management mechanisms and the narrative design of fictional fear. Horror stories reliably fulfill the input conditions that trigger these mechanisms: they present vulnerable protagonists confronted with persistent, highly formidable threats. These exaggerated features mirror, and often amplify, cues of danger associated with predators and hostile conspecifics. In this sense, horror fiction operates as a kind of supernormal stimulus, an intensified simulation that exploits the relevant cues our brains are tuned to detect (Clasen, 2012; Dubourg & Baumard, 2022). By activating our threat-management systems in a cost-free and controlled environment, fictional fear may serve as a cognitive training ground for detecting, interpreting, and emotionally responding to dangerous situations (Clasen, 2017; Chu & Schulz, 2020; Steen & Owens, 2001).

Some paranormal antagonists such as ghosts may at first blush appear to be counterexamples to our formidability-based argument of horror. However, paranormal antagonists still conform to this structure when formidability is understood functionally. In these cases, ghosts or unseen forces evoke vulnerability-induced fear because they (1) persistently antagonize or threaten the protagonist, (2) maintain high threat presence and create hazardous or unpredictable environments, and (3) resist control or counteraction, thereby exceeding the protagonist's capacity to defend or escape. The perception of their formidability derives not from physical size or strength but from persistence, motivation, and the (supernatural) capacity to inflict harm despite the protagonist's efforts to intervene.

Beyond explaining variation in fear responses across genres and individuals, these studies also contribute to understanding the computational architecture of fear itself. In line with the internal regulatory framework proposed by Tooby et al. (2014), fear can be understood as being regulated by internal variables that track the relative danger posed by others (see also Dubourg, Chambon, & Baumard, 2025). Our findings suggest that the formidability ratio between oneself (the

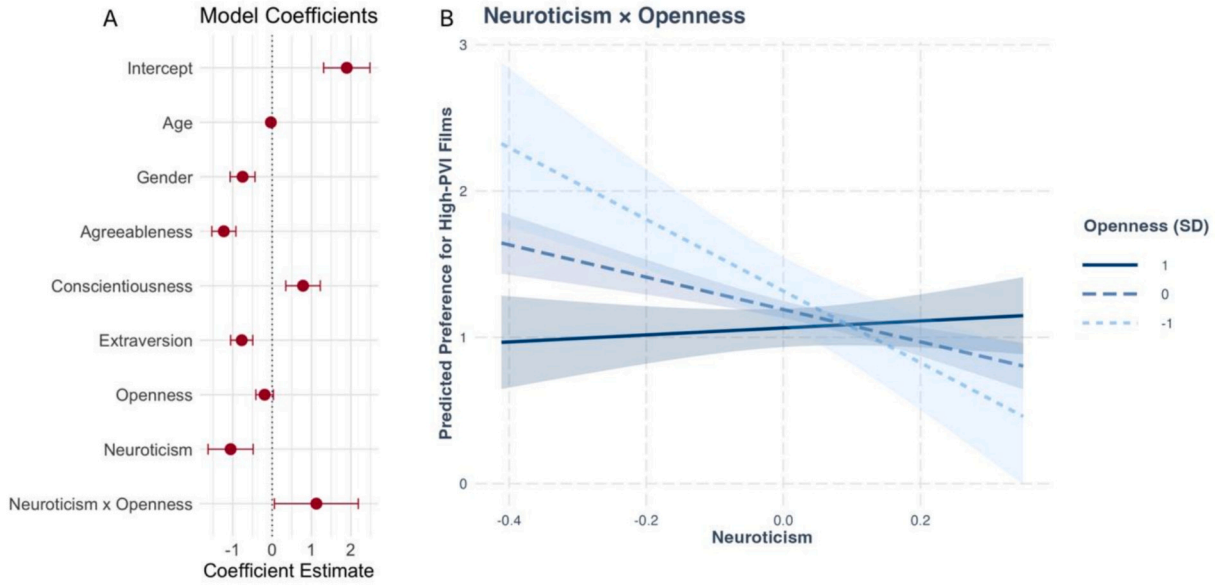

Fig. 3. A. Standardized coefficient estimates predicting PVI (log) from personality traits, gender, and age. Error bars represent 95 % confidence intervals. **B.** Neuroticism predicts lower preference for high-PVI films only among individuals low in Openness; at high Openness, the association reverses.

protagonist) and another individual (an antagonist) functions as one such internal regulatory variable: it summarizes critical input features, like the ability to inflict harm versus defend oneself, and provides structured input to evolved threat-avoidance mechanisms. When this variable exceeds a certain threshold, it triggers downstream fear responses such as increased vigilance and physiological arousal. By identifying which narrative features predict both subjective fear and physiological markers such as heart rate, our work begins to specify the precise input conditions under which real-life threats become compelling and frightening. In this view, horror stories "work" because they exploit parameters of fear response activation that have been carved throughout evolutionary time.

6.1. Limitations

Despite the strengths of our approach, some limitations should be noted. First, our annotation method relies on large language models, which, while powerful and increasingly validated, still introduce potential bias or hallucination in their outputs. However, to help mitigate this, we conducted multiple robustness checks and found high replicability across simplified formulations of the PVI. Second, our film corpus, while large and genre-diverse, may still reflect biases in IMDb tagging practices, especially given its reliance on English-language films and primarily Western user data.

Although the Science of Scare data in Study 2 provides converging physiological evidence, the sample size was limited and may not fully capture long-term or individual-level fear responses. Future work could expand this line of inquiry using larger physiological datasets or biometric monitoring during horror film viewing. Finally, while our studies offer a strong correlational framework linking narrative structure to fear and personality, experimental manipulation of formidability dynamics within controlled narrative stimuli would be a crucial next step in testing causal claims. Such work could clarify how specific narrative changes modulate emotional engagement and deepen our understanding of fear's computational triggers.

6.2. Conclusion

Fear is a phylogenetically ancient emotion that processes specific input cues. By mapping the computational logic of fear onto narrative structure, this work reveals how horror stories activate evolved threat-management systems through specific cues of vulnerability and formidability asymmetry. The Protagonist Vulnerability Index bridges evolutionary theory, cultural analysis, and computational modeling to offer a novel, generalizable tool for understanding fear in media. As narrative technologies continue to evolve, this framework offers a roadmap for predicting, and perhaps precisely engineering, what frightens us most.

CRediT authorship contribution statement

Edgar Dubourg: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Coltan Scrivner:** Writing – review & editing, Methodology, Conceptualization.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2025.106813.

References

Archer, J. (1988). *The behavioural biology of aggression*. Pr: Cambridge Univ.

Armfield, J. M. (2006). Cognitive vulnerability: A model of the etiology of fear. *Clin. Psychol. Rev., 26*(6), 746–768. https://doi.org/10.1016/j.cpr.2006.03.007

Barlow, G. W., Rogers, W., & Fraley, N. (1986). Do Midas cichlids win through prowess or daring? It depends. *Behav. Ecol. Sociobiol., 19*(1), 1–8. https://doi.org/10.1007/BF00303836

Bongini, P., Becattini, F., & Del Bimbo, A. (2023). Is GPT-3 all you need for visual question answering in cultural heritage? In L. Karlinsky, T. Michaeli, & K. Nishino (Eds.), *Vol. 13801. Computer vision – ECCV 2022 workshops* (pp. 268–281). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25056-9_18.

Carroll, N. (1990). *The philosophy of horror, or paradoxes of the heart*. Routledge.

Chu, J., & Schulz, L. E. (2020). Play, curiosity, and cognition. *Annual Review of Developmental Psychology, 2*(1), 317–343. https://doi.org/10.1146/annurev-devpsych-070120-014806

Clasen, M. (2012). Monsters evolve: A biocultural approach to horror stories. *Rev. Gen. Psychol., 16*(2), 222–229. https://doi.org/10.1037/a0027918

Clasen, M. (2017). *Why horror seduces*. Oxford University Press.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc. Natl. Acad. Sci., 114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Ding, B., Qin, C., Liu, L., Chia, Y. K., Joty, S., Li, B., & Bing, L. (2023). Is GPT-3 a good data annotator? (no. arXiv:2212.10450). *arXiv*. http://arxiv.org/abs/2212.10450.

Dubourg, E., & Baumard, N. (2022). Why and how did narrative fictions evolve? Fictions as entertainment technologies. *Front. Psychol., 13*, Article 786770. https://doi.org/10.3389/fpsyg.2022.786770

Dubourg, E., & Chambon, V. (2025). DEEP: A model of gaming preferences informed by the hierarchical nature of goal-oriented cognition. *Entertainment Computing, 53*, Article 100930. https://doi.org/10.1016/j.entcom.2025.100930

Dubourg, E., Chambon, V., & Baumard, N. (2025). Human motivation is organized hierarchically, from proximal (means) to ultimate (ends). *The Behavioral and Brain Sciences, 48*, Article e31. https://doi.org/10.1017/S0140525X24000542

Dubourg, E., Safan, R., Thouzeau, V., & Baumard, N. (2025). Charting the rise of imaginary worlds in history. Humanities and Social Sciences Communications).

Dubourg, E., Thouzeau, V., Beuchot, T., Bonard, C., Boyer, P., Clasen, M., … Baumard, N. (2024). The cognitive foundations of fictional stories. *OSF*. https://doi.org/10.31219/osf.io/me6bz

Dubourg, E., Thouzeau, V., Borredon, Q., & Baumard, N. (2025). *Quantifying and explaining the rise*. Evolutionary Human Sciences.

Dubourg, E., Valentin, T., & Baumard, N. (2024). A step-by-step method for cultural annotation by LLMs. *Frontiers in Artificial Intelligence, 7*. https://doi.org/10.3389/frai.2024.1365508

Durkee, P. K., Goetz, A. T., & Lukaszewski, A. W. (2018). Formidability assessment mechanisms: Examining their speed and automaticity. *Evol. Hum. Behav., 39*(2), 170–178. https://doi.org/10.1016/j.evolhumbehav.2017.12.006

Elwood, R. (2009). Difficulties remain in distinguishing between mutual and self-assessment in animal contests. *Anim. Behav.*. https://doi.org/10.1016/j.anbehav.2008.11.010

Emlen, D. J. (2008). The Evolution of Animal Weapons. *Annu. Rev. Ecol. Evol. Syst., 39*, 387–413. https://doi.org/10.1146/annurev.ecolsys.39.110707.173502

Fessler, D. M., & Holbrook, C. (2014). Marching into battle: Synchronized walking diminishes the conceptualized formidability of an antagonist in men. *Biol. Lett., 10*(8), 20140592. https://doi.org/10.1098/rsbl.2014.0592

Fessler, D. M. T., Holbrook, C., & Snyder, J. K. (2012). Weapons make the man (larger): Formidability is represented as size and strength in humans. *PLoS One, 7*(4), Article e32751. https://doi.org/10.1371/journal.pone.0032751

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proc. Natl. Acad. Sci., 120*(30), Article e2305016120. https://doi.org/10.1073/pnas.2305016120

Grodal, T. (2009). *Embodied visions: Evolution, emotion, culture, and film*. Oxford University Press.

Hofmann, H. A., & Schildberger, K. (2001). Assessment of strength and willingness to fight during aggressive encounters in crickets. *Anim. Behav., 62*(2), 337–348. https://doi.org/10.1006/anbe.2001.1746

Hurd, P. L. (2006). Resource holding potential, subjective resource value, and game theoretical models of aggressiveness signalling. *J. Theor. Biol., 241*(3), 639–648. https://doi.org/10.1016/j.jtbi.2006.01.001

Kuzman, T., Mozetič, I., & Ljubešić, N. (2023). ChatGPT: Beginning of an end of manual linguistic data annotation? Use case of automatic genre identification (no. arXiv: 2303.03953). *arXiv*. http://arxiv.org/abs/2303.03953.

Laundré, J. W., Hernandez, L., & Ripple, W. J. (2010). The landscape of fear: Ecological implications of being afraid. *The Open Ecology Journal, 3*, 1–7. https://doi.org/10.1890/13-1083.1

LeDoux, J. E. (2014). Coming to terms with fear. *PNAS, 111*(8), 2871–2878. https://doi.org/10.1073/pnas.1400335111

Lynch, T. (2018). Evolutionary formidability mechanisms as moderators of fear experience. In J. Breuer, D. Pietschmann, B. Liebhold, & B. P. Lange (Eds.), *Evolutionary psychology and digital games: Digital hunter-gatherers*. London: Routledge.

Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, time, and fear: Survival computations along defensive circuits. *Trends Cogn. Sci., 24*(3), 228–241. https://doi.org/10.1016/j.tics.2019.12.016

Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., … Frith, C. D. (2009). From threat to fear: The neural organization of defensive fear systems in humans. *J. Neurosci., 29*(39), 12236–12243. https://doi.org/10.1523/JNEUROSCI.2378-09.2009

MoneySuperMarket. (2025). The Science of Scare Project. https://www.moneysupermarket.com/science-of-scare/.
Nave, G., Rentfrow, J., & Bhatia, S. (2020). We are what we watch: Movie plots predict the personalities of those who "like" them [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/wsdu8
Oatley, K. (1999). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Rev. Gen. Psychol., 3*(2), 101–117. https://doi.org/10.1037/1089-2680.3.2.101
Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *J. Theor. Biol., 47*(1), 223–243. https://doi.org/10.1016/0022-5193(74)90111-8
Pei, X., Li, Y., & Xu, C. (2023). GPT self-supervision for a better data annotator (no. arXiv: 2306.04349). *arXiv*. http://arxiv.org/abs/2306.04349.
Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C., & Van Bavel, J. J. (2023). GPT is an effective tool for multilingual psychological text analysis [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/sekf5
Scrivner, C. (2021a). Scaring away anxiety: Therapeutic avenues for horror fiction to enhance treatment for anxiety symptoms. PsyArXiv Preprints https://osf.io/preprints/psyarxiv/7uh6f_v3.
Scrivner, C. (2021b). The psychology of morbid curiosity: Development and initial validation of the morbid curiosity scale. *Personal. Individ. Differ., 183*(111139), 52. https://doi.org/10.1016/j.paid.2021.111139
Scrivner, C. (2024). Bleeding-heart horror fans: Enjoyment of horror media is not related to lower empathy or compassion. *Journal of Media Psychology: Theories, Methods, and Applications.*. https://doi.org/10.1027/1864-1105/a000405
Scrivner, C., Holbrook, C., Fessler, D. M., & Maestripieri, D. (2020). Gruesomeness conveys formidability: Perpetrators of gratuitously grisly acts are conceptualized as larger, stronger, and more likely to win. *Aggress. Behav., 46*(5), 400–411. https://doi.org/10.1002/ab.21907
Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proc. Natl. Acad. Sci. USA, 106*(35), 15073–15078. https://doi.org/10.1073/pnas.0904312106
Singh, M. (2021). The sympathetic plot, its psychological origins, and implications for the evolution of fiction. *Emot. Rev., 13*(3), 183–198. https://doi.org/10.1177/17540739211022824
Smith, M. (1995). *Engaging characters: Fiction, emotion, and the cinema*. Oxford University Press.
Steen, F., & Owens, S. (2001). Evolution's pedagogy: An adaptationist model of pretense and entertainment. *J. Cogn. Cult., 1*(4), 289–321. https://doi.org/10.1163/156853701753678305
Tan, E. S. (1996). *Emotion and the structure of narrative film: Film as an emotion machine*. Lawrence Erlbaum.
Taylor, P. M., & Uchida, Y. (2022). Horror, fear, and moral disgust are differentially elicited by different types of harm. *Emotion, 22*(2), 346. https://doi.org/10.1037/emo0001061
Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2014). Internal regulatory variables and the Design of Human Motivation: A computational and evolutionary approach. In *Handbook of Approach and Avoidance Motivation*. Routledge. https://doi.org/10.4324/9780203888148.ch15.