Marketing Science Institute Working Paper Series 2025

Report No. 25-124

Building Persuasive Stories with Emotion Sequences

Samsun Knight, Liu Liu, and Laura J. Kornish

Building Persuasive Stories with Emotion Sequences

Samsun Knight[*§] Liu Liu[†§] Laura J. Kornish[†]

Date: June 28, 2025

Abstract

What types of stories are most persuasive? In this paper, we introduce a new template for categorizing story types based on the specific emotional dynamics of text, or "emotion sequences"—for example, whether a story begins fearful and ends with sadness, or vice versa. We present this as a new way to capture distinct narrative progressions that is tractable even in short-form media, and then apply this method to analyze the persuasiveness of different story types in online fundraising. Using transformer-based emotion classification tools, we measure the two-part emotion sequences of 14,000 medical fundraising pitches from GoFundMe.org and show that, among other findings, medical fundraising pitches that begin with a sad tone and end on a caring tone are significantly more likely to succeed. We then develop a simple new approach for testing the generalizability of these observational findings by using crowd-sourced, LLM-assisted rewrites to introduce particular emotion sequences to a sample of 40 randomly-selected fundraisers. We show that human-only rewrites generally fail due to skill deficits (and LLM-only rewrites can introduce salient informational changes from originals), but demonstrate that crowd-sourced, LLM-assisted rewriting offers an effective method for testing the out-of-sample application of research results by everyday online users. With this, we establish that pitches rewritten to feature our focal emotion sequences see a significant boost in perceived persuasiveness, even for some sequences associated with lower success in observational data, while placebo rewrites produce null effects. Furthermore, we show that increased identification with the protagonist of the fundraiser is the primary mechanism driving the observed effects.

[*]University of Toronto; samsundknight@gmail.com

[†]University of Colorado, Boulder; liu.liu-1@colorado.edu, laura.kornish@colorado.edu

§ ≡ joint first author.

We thank Mahrukh Zahoor and Prasenjeet Gadhe for excellent research assistance. We are also grateful for helpful suggestions from John Lynch, Nicholas Reinholtz, and seminar participants at the Artificial Intelligence in Management Conference 2025, the INFORMS Marketing Science Conference 2025, the North American Chapter of the Association for Computational Linguistics Workshop on Narrative Understanding 2025, Yale University, and WU Vienna University of Economics and Business. All errors are our own.

1 Introduction

Storytelling is integral to marketing and persuasion. In advertising, stories can evoke emotions that stick with audiences long after a campaign ends, driving not only immediate sales but also long-term brand loyalty; while in fundraising, stories foster the personal connections that are essential for persuading readers to contribute (GoFundMe, 2023). Whether the goal is to sell a product or to build a customer relationship, storytelling remains one of the most potent tools for convincing listeners.

But not all stories hit the mark. Some narratives fail to resonate, while others may even alienate their intended audience. How can we determine which stories are most likely to connect with listeners and readers, versus those that will fall flat? And moreover, how can we establish that the story forms themselves are the causal driving factor, rather than other confounding factors that are merely correlated with certain narrative types?

Existing research offers some answers but leaves key gaps unexplored. To date, much of the work on narratives in marketing has examined how overall valence, average emotion levels, or other summary features of text drive audience responses, focusing primarily on static attributes.[1] A handful of more recent studies have shifted attention toward studying text dynamics and narrativity, investigating how fluctuations in aggregate positivity or overall emotionality, or average rates of movement in semantic embedding space, shape perceptions;[2] but these works generally focus on composite measures of emotionality or aggregated index

[1] For example, Brennan and Binney (2010); Cavanaugh, Bettman, and Luce (2015); Xu (2017); Paxton, Velasco, and Ressler (2020) and Berger, Moe, and Schweidel (2023), among others.

[2] For example, Van Laer et al. (2019); Berger, Kim, and Meyer (2021); Toubia, Berger, and Eliashberg (2021).

measures, overlooking the richer and more intricate ways that specific emotional states combine in sequence to characterize a narrative. Moreover, many of these dynamic measures require at least moderate volumes of text or video content in order to be computed, and may therefore be intractable in the very short-length or short-duration items that are found in many marketing applications, such as advertisements or short online appeals.

In this paper, we build on these foundations by introducing a novel template for characterizing narratives based on their particular emotional progressions. Instead of focusing solely on average emotion levels or aggregated measures, we analyze specific emotion sequences—distinct progressions of particular emotional states over the course of a text—introducing a new, structured way to measure story types that remains tractable even in short-form media.[3] Analyzing over 14,000 online medical fundraising pitches from GoFundMe.com, we establish that certain emotion sequences, such as pitches that start with sadness and end with caring, are strongly associated with higher success rates in our observational data. We then validate these findings out-of-sample through crowd-sourced, large language model-assisted (LLM-assisted) rewrites by everyday online participants, demonstrating that embedding these focal emotion sequences into fundraising pitches significantly improves their persuasiveness. Finally, in line with earlier work on the psychological mechanisms of narrative, we establish empirically that the observed positive effects of particular emotion sequences are predominantly driven by heightened identification with the protagonist of the fundraiser (Green and Brock, 2000; Cohen, 2001).

First, we leverage recent advances in computational language models to build rigorous,

[3]We focus on two-part emotion sequences in the present work for tractability, but we hope that future work may extend this framework to higher-order sequences as well, when examining contexts where content length is sufficient to reliably measure higher-order sequences.

context-aware classifications of emotion across the course of a text, and apply this to a large-scale corpus of online medical fundraising pitches. We use a specialized RoBERTa-based emotion classifier to label the presence of specific emotions in the first half and second half of online fundraising pitches for a large dataset of over 14,000 medical pitches from GoFundMe.com, focusing on three basic emotion categories that have been established as first-order important to online fundraising in prior literature, and that are also identified as core emotions in this context when using factor analysis: expressions of sadness[4], expressions of fear[5], and expressions of caring[6], as well as a fourth label for emotional neutrality to capture effects of overall emotionality. That said, while we consider this parsimonious set of both literature- and factor-analysis-derived emotions for our primary application, we emphasize that our method is readily usable with other emotion sets, and present robustness analyses examining alternative sets of emotions in Appendix Sections A.2 and A.3.

Next, to establish a baseline for comparing the relative contribution of our emotion sequences approach to prior methods, we analyze the effect of overall average emotion levels on fundraiser success. With this, we find that expressions of caring are positively associated with fundraiser success and that expressions of fear are negatively associated with fundraiser success, as measured by the log proportion of the fundraising goal raised. However, analyzing the effect of emotion labels in each separate half of the pitches, we show that this gestalt analysis overlooks important dynamic nuances, with expressions of sadness in the first half of the pitch and expressions of caring in the second half of the fundraising pitch shown to be especially important to fundraiser success. This suggests that individuals trying to act

[4]Zhao, Zhou, and Zhao (2022); Lu, Xu, and Fan (2024)
[5]Brennan and Binney (2010); Lu, Xu, and Fan (2024)
[6]Batson et al. (1981); Cavanaugh, Bettman, and Luce (2015); Telle and Pfister (2016); Chan and Septianto (2022)

on aggregate-feature analysis, if they lack any further insight into the dynamics underlying the overall patterns, may fail to achieve out-of-sample improvements in text persuasiveness if they e.g. change the tone in the wrong part. But if these positional emotion effects are codependent—that is, if the most appropriate ending emotion depends on the emotion at the beginning—this interdependence will still be overlooked in this simple dynamic analysis.

We therefore then break this down further into binary indicators of $4 \times 4 = 16$ specific first-half/second-half emotion sequences in order to inspect the role of specific emotional progressions in driving these overall patterns. We establish that the observed overall effects are indeed driven by just a small handful of specific progressions that are powerfully associated with differential rates of fundraising success in our sample: we find that pitches featuring $sadness \rightarrow caring$ progressions (that is, expressions of sadness in the first half and expressions of caring in the second half) are significantly associated with greater observed fundraiser success, while $sadness \rightarrow fear$ and $sadness \rightarrow sadness$ progressions feature null relationships with fundraising outcomes, highlighting the crucial importance of considering the particular emotional progression when analyzing storytelling patterns. To a lesser extent, we also find that $caring \rightarrow caring$ and $sadness \rightarrow neutral$ progressions are associated with greater fundraising success, and that $neutral \rightarrow fear$ progressions are associated with worse fundraising outcomes.

However, while these results recover strong associations between specific emotional progressions and fundraising success, we note that unobserved confounders may still meaningfully affect these field data estimates. For example, individuals may tend to write more fearful endings to their fundraising pitches when they reasonably fear that they are unlikely to achieve their goals, therefore leading to a spurious correlation between later expressions of

fear and likelihood of success. In order to demonstrate that our findings are not confounded and may indeed generalize to out-of-sample application, we require evidence that our measured emotions progressions can truly cause changes in how fundraising pitches are read and evaluated.

To address this, we develop and apply a simple new approach for out-of-sample testing of linguistics research findings based on crowd-sourced, LLM-assisted rewrites of randomly-selected sample texts. Specifically, we solicit human rewrites, LLM rewrites, and human-in-the-loop LLM-assisted rewrites from online participants, asking them to rewrite fundraisers to feature different emotion sequences while keeping content otherwise the same. We present this as a novel, intuitive way to effectively simulate and test the out-of-sample application of research results, directly mitigating concerns that researcher-written stimuli may be non-representative (Lynch Jr, 1982). With this, we show that randomly-selected fundraising pitches rewritten to feature expressions of sadness at the start and expressions of caring at the end do indeed see significant improvements in persuasiveness, while fundraising pitches rewritten with "placebo" rewriting prompts (without guidance to include any emotion sequence) cause null differences.[7] At the same time, we also show that pitches rewritten to begin emotionally neutral and end with fear also see significant positive improvements compared to the original, in contrast to the significant negative effects we estimate for this emotion sequence in our observational analysis; and pitches rewritten to begin with sadness and end emotionally neutral have null differences with originals, in contrast to significant positive

[7]We also show that human rewrites would lead to vastly different conclusions due to skill deficits among human rewriters, while LLM-only rewrite effects appear to be partly driven by changes in the informational content of rewrites, emphasizing the essential advance from using human-in-the-loop, LLM-assisted rewrites instead. N.B. that by "human rewriters", we refer to crowd-sourced, everyday individuals whom we are able to recruit using standard online platforms, rather than expert creative professionals.

effects in observational analyses. This demonstrates that confounds do drive some (but not all) of the observational associations between certain emotion progressions and fundraising outcomes, highlighting the importance of retesting findings with our out-of-sample rewriting approach. Given the especially consistent evidence in favor of *sadness* → *caring*, we present the positive impact of this particular emotion sequence in online fundraising as this paper's most robust empirical finding, with both clear ecological validity and strong experimental support.

Finally, we investigate the mechanism for our observed effects by examining two conceptual constructs drawn directly from prior research on the psychology of narrative: the mechanism of "narrative transportation", defined as a psychological state of heightened attention and engagement; and the mechanism of "identification", defined as empathic identification with the protagonist of the fundraiser. We adapt instruments for measuring these mechanisms directly from earlier research (Green and Brock, 2000; Cohen, 2001) and run a small follow-up study on our LLM-assisted rewrites to examine the extent to which either (or both) of these psychological mechanisms moderate the observed effects of the given rewrites. We find that, for the emotion sequences that we estimate to produce positive persuasiveness effects, the total effect is almost completely mediated by higher identification with the protagonist. We present this as an especially intuitive mechanism in the context of online fundraisers: that fundraisers rewritten to feature e.g. *sadness* → *caring* emotion sequences are more effective because they lead to heightened empathic identification with the subjects of the fundraiser among readers, which in turn leads to higher persuasiveness.

Taken together, our results offer a straightforward new framework for analyzing story types in short-form media, with numerous potential applications in advertising analysis and

other marketing contexts; as well as an intuitive new approach for validating findings for linguistics research using crowd-sourced, LLM-assisted rewrites. In the context of online fundraising, we demonstrate that earlier findings of overall emotion effects appear to be predominantly driven by a relatively small set of particular underlying emotion sequences, with *sadness* $\rightarrow$ *caring* (among other sequences) shown to be especially helpful in online medical philanthropy, both in large-scale field data analysis and when randomly-selected online participants rewrite randomly-selected fundraisers to feature this emotion sequence. We present both our conceptual structure of emotion sequences and our methodology of crowd-sourced rewriting as readily portable to other arenas, where managers may apply these tools to guide narrative analysis and subsequent (re)design across an array of marketing domains. Building on prior research on the importance of creative templates in advertising and product development (Goldenberg, Mazursky, and Solomon, 1999a,b), we hope that this proposed structure of "emotion sequences" may serve to establish new actionable templates for persuasive storytelling in narrative media design.

The rest of this paper is structured as follows. Section 2 reviews related literature. Section 3 presents details on our data sources. Section 4 presents the empirical results from our field dataset of over 14,000 GoFundMe online fundraisers. Section 5 presents the experimental results from our crowd-sourced rewriting approach. Section 6 presents the mechanism analysis. Section 7 concludes.

2 Relation to Prior Literature

This study builds on extensive research into the drivers of philanthropic giving and the role of linguistic features in marketing and fundraising. For instance, Min and Levina (2018) link the "identifiable victim effect" to prosocial microlending, while Guan et al. (2020) find that static measures of character and time-related words words can influence donor behavior based on a dictionary-based approach. Similarly, Li, Yang, and Sun (2022) explore how different average levels of emotional, logical, and personality-driven language affects crowdfunding outcomes, and Paxton, Velasco, and Ressler (2020) examine the impact of overall emotional tones on both donations and volunteering. Focusing specifically on medical fundraising, Zhang et al. (2024) underscore the role of linguistic style (as measured by the average abstractness or concreteness of language) in shaping donor behavior. More broadly, in marketing, Banerjee and Urminsky (2024) perform a systematic analysis of a large set of static linguistic features' effects on headline engagement, including averages of emotion-related linguistic features, in the very-short-form context of Upworthy.com headlines. Nguyen, Johnson, and Tsiros (2023) develop a powerful machine-learning based predictor of email open rates based on detected (static) emotion levels in email subject lines. Berger, Moe, and Schweidel (2023) show that language that is easy to process and features high average levels of uncertain or high-arousal emotion can increase attention and engagement. Netzer, Lemaire, and Herzenstein (2019) demonstrate that specific linguistic feature averages in loan narratives, such as the number of or presence of mentions of family, can reveal psychological states that predict financial outcomes. Extending these findings, our study introduces the concept of dynamic emotion

sequences as a new framework to characterize story types, allowing researchers to investigate how specific emotion trajectories influence engagement and fundraising success.

This research also contributes to the experimental literature on emotional language and persuasion. For example, Liang, Chen, and Lei (2017) find that high average levels of contrasting emotions increase donations, while Goenka and van Osselaer (2019) show that emotionally congruent appeals (aligned with a charity's moral objectives) enhance donor response. Tellis et al. (2019) show that emotionally engaging content, rather than information-focused or brand-prominent content, is the key driver of YouTube ad virality. Another related work is Fong, Kumar, and Sudhir (2024), who introduce a novel emotion-classification model based on music theory and demonstrate that this tool can be leveraged to improve congruence between video content and ad content, and thereby improve ad engagement in video. Our work introduces a new framework for analyzing story types based on the specific emotional progressions in text, and offers a simple new approach to combine large-scale field data with experimental evidence.

Furthermore, this study advances the cross-disciplinary literature on narrative structures, or other dynamic features of text, as key drivers of communication success. Research in psychology, such as Loewenstein and Prelec (1993) and Kahneman and Tversky (1979), has shown that improving sequences resonate with audiences through mechanisms like loss aversion and adaptation. Studies in narratology have documented that a small set of specific valence dynamics can characterize a large proportion of texts across contexts (Vonnegut, 1995; Reagan et al., 2016). In marketing, previous studies have shown how index measures of semantic or sentiment dynamics can predict success in fundraisers, reviews and movies, as well as other domains (Van Laer et al., 2019; Toubia, Berger, and Eliashberg, 2021; Berger,

Kim, and Meyer, 2021; Knight, Rocklage, and Bart, 2024). This paper builds most directly on Packard, Li, and Berger (2024), who demonstrate that there are key dynamic differences in the helpfulness of cognitive versus affective language on customer service calls depending on when such language occurs over the course of the call, as well as Knight, Rocklage, and Bart (2025), who examine average dynamic emotional variance in movies, podcasts, and songs. The present study introduces the framework of specific emotion sequences, such as $sadness \rightarrow fear$, to categorize story types in short-form media, and then demonstrates with multiple methods how these story types relate to persuasiveness. We also note that, while the large majority of the above approaches to measure narrative features are not tractable without large amounts of text content, our formula may be straightforwardly computed for short-form media such as advertising or short online appeals.

Related to this, this work also builds on the literature on creativity and innovation templates in advertising design and product design. Goldenberg, Mazursky, and Solomon (1999b) and Goldenberg, Mazursky, and Solomon (1999a) identified a set of well-defined, objectively verifiable, and quantifiable schemes, which they call templates, from historical analysis of new product innovations and quality creative ads, and demonstrate that these templates not only provide a foundation for understanding innovation and creativity, but also offer a systematic way to engineer and achieve creativity expertise. Similarly, emotion sequences may also be considered as structured, quantifiable templates for systematically designing narrative in short-form media.

Finally, this research contributes to the growing subliterature on the use of deep learning and large language models (LLMs) in social science and marketing research (Ziems et al., 2024; Goli and Singh, 2023; Li et al., 2024). The most closely related is Hong and

Hoban (2022), who develop a deep-learning based writing aid to help marketers craft more compelling appeals by scoring sentence quality and identifying weak sentences as targets for revision or removal, and furthermore decompose their model to infer the overall topics and word-level attributes that contribute to higher sentence scores, such as specificity, readability and concrete language. In the present work, we employ GPT-4 with chain-of-thought prompting to systematically rewrite fundraising pitches and examine dynamic narrative features of text, introducing specific emotion sequences while preserving narrative content. This methodology allows us to isolate the causal impact of specific emotional trajectories on narrative effectiveness. We furthermore show that human rewrites often fail due to skill defiicits, while LLM-only rewriting tends to introduce informational discrepencies with the original which then drives, in part, the results we observe for only-LLM rewrites; but human-in-the-loop, LLM-assisted rewrites capably rewrite fundraising pitches and are not driven by informational discrepancies. More broadly, this crowd-source, LLM-assisted rewriting method that we develop also relates to earlier research on addressing external validity concerns in marketing research (Lynch Jr, 1982; Lynch, 1999; Kappes, Gladstone, and Hershfield, 2021). By integrating LLM-assisted interventions with robust field and experimental data, this study demonstrates a scalable and rigorous approach for testing research findings on linguistics and communication across diverse contexts.

3 Data

Our field data come from direct webscraping of GoFundMe.org, from which we collect over 14,000 medical fundraisers. We gather these fundraisers through simple single-name iterative

searches over the top 100 male names and top 100 female names according to the Social Security Administration (2024), scraping all fundraisers that are returned on each search. From this, we gather a large dataset of fundraisers, which we subset to consider only medical fundraisers, the most popular category. To ensure reliable emotion sequence encoding, we focus on fundraisers between 100 and 250 words—sufficient for analysis while within typical context-window limits.

We score emotions using a new RoBERTa-based emotion classifier (Lowe, 2024; Demszky et al., 2020). This classifier is built using a base layer of RoBERTa, a recent large language model, and fine-tuned against the GoEmotion dataset, which is a large training dataset of labeled emotions in online Reddit posts. Given the curse of dimensionality that enters when we start to consider Cartesian pairs of emotion sets, we consider only a parsimonious set of emotions for our analysis, selected based on prior literature studying fundraising and emotion, which also dovetails with the emotion categories that we recover with factor analysis. Specifically, we track expressions of sadness, which have been shown to drive improved response to online fundraising appeals (Zhao, Zhou, and Zhao, 2022; Lu, Xu, and Fan, 2024), expressions of fear, which ave been shown to drive worse response to online fundraising appeals (Brennan and Binney, 2010; Lu, Xu, and Fan, 2024), and expressions of caring or empathy, which have long been theorized and studied as a core ingredient of online fundraising appeals (Batson et al., 1981; Cavanaugh, Bettman, and Luce, 2015; Telle and Pfister, 2016; Chan and Septianto, 2022). We also include a fourth label for emotional neutrality to measure effects of overall emotionality.[8] For further details on our emotion

[8]"Neutral" is a label also returned by our RoBERTa-based emotion classifier, analogous to the labels for the specific emotions of fear, caring, and sadness.

Table 1: Summary Statistics

	Mean (Std Dev)
Fundraising Goal Reached	0.103 (0.304)
Log(Prop. Goal Reached)	-1.577 (1.608)
Wordcount	169.431 (42.508)
Log(Fundraising Goal)	9.089 (1.205)
Fundraiser Age (Days)	164.608 (442.569)
Valence	0.647 (0.574)
First Half: Fear	0.010 (0.100)
First Half: Caring	0.128 (0.334)
First Half: Sadness	0.549 (0.498)
First Half: Neutral	0.477 (0.499)
Second Half: Fear	0.002 (0.050)
Second Half: Caring	0.414 (0.493)
Second Half: Sadness	0.086 (0.281)
Second Half: Neutral	0.115 (0.319)
Entire Text: Fear	0.004 (0.067)
Entire Text: Caring	0.406 (0.491)
Entire Text: Sadness	0.254 (0.435)
Entire Text: Neutral	0.116 (0.320)
N	14114

Notes: Summary statistics for sample of scraped GoFundMe.com fundraisers. Emotion scores from RoBERTa-based classifier (Lowe, 2024). Standard deviations in parentheses.

classification methodology, including validation of model-based emotion measures against human perceptions of emotion, see Appendix B.[9] As noted above, this set of four emotion categories also corresponds to core emotions that we identify empirically in our context using factor analysis; see Appendix A.3 for further details.

Fundraising goals and amounts raised are collected directly from the webscraped fundrais-

[9]Overall, we find that our measure correlates positively, but noisily, with human ratings of emotion in randomly-selected fundraising pitches, but find a highly similar level of noise in human-human rating comparisons, suggesting that ratings of emotion are highly subjective in this context and that our model is comparably performant to human raters. We expect that such noise in measurement would manifest in our analyses as attenuation bias, suggesting that our estimates may best be interpreted as conservative lower-bounds on true effects.

ers, as is the fundraiser age, computed from the reported date of fundraiser posting.[10] We exclude fundraisers that are detected to have been written in a language other than English. We compute valence based on the VADER lexicon (Hutto and Gilbert, 2014). Summary statistics for this sample of fundraisers are presented in Table 1.

4 Field Data Results

4.1 Overall Emotion Effects

To start, we examine the relationship between fundraiser success and baseline overall emotion labels, as measured using the full text of the fundraising pitch. This analysis is similar to that performed in prior work that examined the effect of overall emotion labels on fundraiser outcomes; we begin with this analysis to establish a clear baseline of comparison within our sample for our later, more fine-grained analyses. We use simple OLS regression and include controls for log fundraising goal, days that the fundraiser has been posted, and a quadratic for fundraising pitch wordcount. We estimate effects both in a logit regression against a binary indicator of fundraiser success ($Goal \leq TotalRaised$), shown in the first column, and an OLS regression against continuous indicator of the log proportion of goal reached ($log(\frac{TotalRaised}{Goal})$), shown in the second column. In Appendix Section A.1, we present robustness analyses for regressions against continuous scores of emotion intensity; results are very similar.

Results are presented in Table 2. Overall, we find strong evidence of effects from both

[10] We remark that GoFundMe fundraisers typically do not have hard cutoff dates, as Kickstarter campaigns do, and so regularly stay posted indefinitely.

Table 2: Overall Emotions and Fundraiser Success

	Success (Logit)	Log Proportion Fundraised (OLS)
Valence	0.163** (0.058)	0.121*** (0.024)
Fear	-0.397 (0.477)	-0.511** (0.190)
Caring	0.225*** (0.060)	0.282*** (0.027)
Sadness	-0.008 (0.068)	-0.008 (0.030)
Neutral	0.026 (0.096)	-0.064 (0.041)
Controls	X	X
Observations	14114	14114
R^2/Pseudo R^2	0.061	0.132

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

valence and particular emotions on fundraiser success, broadly in line with earlier work (Lu, Xu, and Fan, 2024). We find an especially strong positive association between whether the fundraising pitch exhibits "caring", or empathic expressions, and fundraiser success, as well as a significant negative association between fear and the proportion of fundraising goal achieved. We remark that these associations remain significant even with controls for the degree of overall positivity/negativity of the text, as well as other features of the fundraisers.

4.2 Dynamic (Half-by-Half) Emotion Effects

With these results in hand as a baseline, we then turn to investigating the importance of emotion dynamics in driving these effects. To perform this analysis, we split our text into N sentences and evaluate the emotion (and valence) in the first $\underline{N/2}$ sentences and then evaluate the emotion (and valence) in the remaining half. We then regress our measures of dynamic emotion and valence against overall outcome measures for each fundraiser.

Results are presented in Table 3. First, we find that for overall valence, dynamics don't

Table 3: Dynamic Emotion Measures and Fundraiser Success

	Success (Logit)	Log Proportion Fundraised (OLS)
1^{st} Half Valence	0.184*** (0.046)	0.122*** (0.020)
2^{nd} Half Valence	0.161* (0.078)	0.129*** (0.032)
1^{st} Half Neutral	-0.153* (0.062)	-0.057* (0.027)
2^{nd} Half Neutral	0.018 (0.101)	-0.065 (0.043)
1^{st} Half Fear	-0.127 (0.284)	-0.152 (0.128)
2^{nd} Half Fear	-0.816 (0.740)	-0.258 (0.254)
1^{st} Half Caring	-0.062 (0.089)	0.077† (0.040)
2^{nd} Half Caring	0.244*** (0.059)	0.257*** (0.026)
1^{st} Half Sadness	-0.016 (0.063)	0.048† (0.028)
2^{nd} Half Sadness	-0.082 (0.110)	-0.076 (0.047)
Controls	X	X
Observations	14114	14114
R^2/Pseudo R^2	0.064	0.134

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. † = 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

significantly matter—the difference between the coefficients for valence in the first or second half are statistically indistinguishable across outcome measures. However, for emotions, we find that the overall effects we observed in the prior analysis are highly positional: the general positive effect of caring appears to be driven by effects in the second half of the fundraising pitch. This suggests that emotion dynamics are a central feature of how these expressions of caring and empathy are associated with higher rates of fundraising success: there is a highly significant relationship between fundraising success and more caring in the second half of the pitch, but null or only 10% positive relationships between fundraiser success and expressions of caring in the first half of the pitch. For fear, we continue to recover negative

point estimates between expressions of fear and success outcomes but the relationship is no longer statistically significant, perhaps because of a moderate loss of power when examining two halves of effects instead of a single pooled effect. Finally, we see a marginally significant negative relationship between emotional neutrality and success outcomes in the first half, suggesting that any emotion in the first half may be moderately helpful, but only when expressed at the start of the text.

4.3 Emotion Sequence Effects

The above results clearly support our central claim that emotion dynamics are a significant component of the persuasive appeal of online fundraising pitches. But the framework remains imprecise, as there could be a multiplicity of specific emotion progressions that drive the same high-level findings, and this setup will overlook the possibility that positional emotion effects (such as caring in the second half) may be dependent on the specific preceding or succeeding emotion. Therefore, to inspect this more closely, we further divide our observed emotion measures even more finely into indicators for the $4 \times 4 = 16$ possible sequences of first-half and second-half emotions and test which precise progressions are the most effective.

Results are presented in Table 4. Here, we show that a small handful of specific emotion sequences appear to entirely drive the effects that we observe in the previous, coarser regressions. In particular, $sadness \rightarrow caring$ is highly significantly associated with better success outcomes across studied metrics. In the regression against log proportion fundraised, where the outcome variable preserves far more variation from which to identify effects, we also recover a significant positive effect of $sadness \rightarrow neutral$ and $caring \rightarrow caring$, and a

Table 4: Specific Emotion Sequences and Fundraiser Success

	Success (Logit)	Log Proportion Fundraised (OLS)
1^{st} Half Valence	0.217*** (0.043)	0.144*** (0.019)
2^{nd} Half Valence	0.169* (0.076)	0.137*** (0.031)
1^{st} Half Fear – 2^{nd} Half Fear	. (.)	. (.)
1^{st} Half Fear – 2^{nd} Half Caring	0.351 (0.398)	0.278 (0.197)
1^{st} Half Fear – 2^{nd} Half Sadness	0.260 (1.087)	-0.115 (0.445)
1^{st} Half Fear – 2^{nd} Half Neutral	. (.)	. (.)
1^{st} Half Caring – 2^{nd} Half Fear	. (.)	. (.)
1^{st} Half Caring – 2^{nd} Half Caring	0.136 (0.117)	0.130* (0.055)
1^{st} Half Caring – 2^{nd} Half Sadness	-0.163 (0.370)	-0.046 (0.160)
1^{st} Half Caring – 2^{nd} Half Neutral	0.306 (0.243)	0.034 (0.118)
1^{st} Half Sadness – 2^{nd} Half Fear	-0.154 (0.806)	0.266 (0.345)
1^{st} Half Sadness – 2^{nd} Half Caring	0.229*** (0.072)	0.208*** (0.032)
1^{st} Half Sadness – 2^{nd} Half Sadness	0.002 (0.152)	-0.095 (0.065)
1^{st} Half Sadness – 2^{nd} Half Neutral	-0.056 (0.168)	0.142* (0.071)
1^{st} Half Neutral – 2^{nd} Half Fear	-0.233 (1.124)	-1.043* (0.477)
1^{st} Half Neutral – 2^{nd} Half Caring	0.013 (0.079)	0.065† (0.035)
1^{st} Half Neutral – 2^{nd} Half Sadness	-0.152 (0.184)	-0.051 (0.074)
1^{st} Half Neutral – 2^{nd} Half Neutral	-0.038 (0.130)	-0.202*** (0.054)
Controls	X	X
Observations	14101	14114
R^2/Pseudo R^2	0.063	0.133

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. † = 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. ". (.)" indicates variables dropped due to multicollinearity.

significantly negative effect of $neutral \rightarrow fear$. We also find a significant negative effect of completely emotion-free progressions of $neutral \rightarrow neutral$.[11]

Taken together, these results suggest that much of the effects in our higher-level analyses

[11]To test for overfitting, we find that 5-fold cross-validated R^2 is approximately the same as the full sample R^2, of 0.128 and 0.133, respectively. This result means that our model's performance does not deteriorate out-of-sample, reducing any concern about overfitting.

are in fact being driven by just a few key emotion sequences. In particular, we find that *sadness* $\rightarrow$ *caring* is the most impactful story type both in terms of the probability of reaching one's fundraising goal and the log proportion of fundraising goal reached. This finding both helps us to understand the true underlying mechanism of the effects, and furthermore offers much more concrete guidance for those seeking to leverage our observed findings to improve the persuasive quality of their writing in the real world: when working to make a convincing medical fundraising pitch, write (or rewrite) your story in a *sadness* $\rightarrow$ *caring* arc.

However, one may remain concerned that the above findings, while ecologically valid and tested in large-scale field data, may be contaminated by confounding bias from unobserved omitted variables. For example, people writing fundraising pitches who have good reason to be fearful, perhaps because they already understand that their fundraiser is not likely to reach its goals, may also write pitches that tend to end on fear. This could lead to e.g. the observed significant negative association between success and $neutral \rightarrow fear$ that we observe above, even absent any causal negative relationship between this particular emotion sequence and actual persuasive appeal.

5 Crowd-Sourced Rewrites

Addressing the above concerns with experimental validation is especially difficult in the context of the complex computational linguistics analysis of this paper. While simpler features may be relatively easy to manipulate, emotion sequences are a high-level phenomenon of text that randomly-selected individuals from online experimental platforms may have trou-

ble rapidly mastering. This is a common problem with advanced computational linguistics; as recent work has begun to examine more abstract and nuanced phenomena of language and narrative, the difficulty of experimentally manipulating such features quickly makes such testing impractical. At best, researchers may sometimes be able to carefully manipulate such complex features themselves for a very small number of hand-picked stimuli and then test these altered stimuli for changes in outcomes. But this second-best method often raises concerns that selected stimuli may not be generally representative, or that manipulations of stimuli may not be generally representative; and as such, that the exercise does not actually validate the out-of-sample application of research findings by laypersons.

We therefore develop and test a new approach in this paper, leveraging recent new advances in large language models, to overcome these challenges. We propose this as a simple, rigorous method for directly testing the generalizability of linguistic research findings by enacting their out-of-sample application on real-world, randomly-selected stimuli, with real-world, randomly-selected rewriters. We describe this straightforward method as crowd-sourced rewriting.

5.1 Methodology: Human, LLM, and Human-in-the-Loop

We explore three parallel methodologies for our crowd-sourced rewrites for each of the focal progressions that our field-data results highlight as predictive of fundraiser success—that is, $sadness \rightarrow caring$, $sadness \rightarrow neutral$, $caring \rightarrow caring$, and $neutral \rightarrow fear$.[12] For all of them, we follow the same overall structure: for each arc, we randomly select 10 fundraising

[12]We do not manipulation $neutral \rightarrow neutral$ directly, which is also found in field analysis to have a significantly negative effect, but remark that this is implicitly tested when participants rewrite $neutral \rightarrow neutral$ baseline fundraisers into one of the above emotional progression.

pitches without that emotion sequence from our large field-data sample, and then assign them to online participants to be rewritten.[13]

First, as a baseline, we ask human participants to rewrite each of the $10 \times 4 = 40$ randomly-selected fundraising pitches in our sample, instructing them specifically to rewrite them to newly feature the focal emotion sequence, but with the same overall content and sets of facts, and keeping the overall length to 90% to 110% of the original fundraising pitch. See Appendix Section C.1 for more details on the specific experimental design. We refer to these rewrites as "human", to denote that the rewrites come from an entirely human process.[14] Due to uncertain completion rates, we randomly assign each baseline pitch to multiple online human participants, collecting approximately 2 pitches per baseline pitch that a research assistant then examined for quality-assurance, discarding those that were very short, identical to the original, or otherwise erroneous, and selecting the best rewrite per original pitch.[15] These rewritten fundraising pitches then serve as our "human" crowd-sourced rewrites.

Next, we ask ChatGPT to rewrite each of the same 40 randomly-selected fundraising pitches, in the same spirit as above: we ask the model to rewrite the given fundraising pitch to feature the respective focal emotion sequence, preserving the general wordcount and the same set of factual details. We use chain-of-thought prompting, asking ChatGPT

[13]Specifically, for any given $emotion1 \rightarrow emotion2$ arc, we choose fundraising pitches that are neutral in the first half if $emotion1$ is not neutral, or that are not neutral in the first half if $emotion1$ is neutral; and that are neutral in the second half if $emotion2$ is not neutral, or that are not neutral in the second half if $emotion2$ is neutral. For example, for $sadness \rightarrow caring$, we randomly choose pitches that are neutral in both halves; while for $neutral \rightarrow fear$, we randomly choose pitches that are not neutral in the first half, but are neutral in the second half.

[14]We remark in the study prompt that use of large language models is strictly prohibited. Spot checks of responses (as well as subsequent comparison to actual LLM rewrites) strongly suggest that participants did not use LLMs in their rewriting for this condition.

[15]Our test is therefore not based on purely randomly-sampled rewrites, and should be interpreted as effects of a random sampling of rewrites from the upper half of the rewriting-ability distribution.

first to describe the fundraising pitch and consider the emotions in the pitch; to rewrite the pitch once, and then evaluate its own rewrite; and then to rewrite the pitch a second time. Full details on the prompting procedure are provided in Appendix Section C.2. With this procedure, we produce one GPT-generated rewrite for each of the 40 randomly-selected baseline fundraising pitches in our sample, which serve as our "LLM" rewrites.

Finally, we add a third treatment that combines both of the above. Specifically, we ask randomly-selected human participants to rewrite the same set of baseline fundraising pitches, with the exact same overall structure as in our "human" condition—but this time, at the step where we ask for rewrites, we show participants the GPT-generated rewrites and offer these to be used as a baseline. We describe these rewrites as GPT-generated and suggest to participants that they may use the text as starting points for their own rewrites. Further details on this condition are available in Appendix Section C.3. As with our first "human" condition, we randomly assign these baseline pitches to multiple online survey participants for rewriting, and a research assistant then discarded incomplete or poor-quality rewrites and selected the best among the remaining, choosing from a set of 2 rewrites on average per baseline pitch.[16] These rewrites serve as our "Human-in-the-Loop" crowd-sourced rewrites.

5.2 Outcome Measure Validation

Prior to testing our crowd-sourced rewrites, we first establish ground truth relationships between the survey measures that we will inspect in this next step and the actual fundraising outcomes that we investigated above. To do so, we randomly select 50 fundraisers from

[16] As above, this test should therefore be interpreted as effects of a random sampling of rewrites from the upper half of the rewriting-with-GPT-ability distribution.

our entire sample of 14,000+ and gather outcome data on participant evaluations of the GoFundMe.com pitches that we inspect.[17] We ask a parsimonious set of questions seeking to assess the overall persuasiveness of the fundraising pitch, especially as may relate to its emotion sequence, listed below:

- Taking 1 as not at all and 10 as very much, to what extent do you think the pitch is well-written?
- Taking 1 as not at all and 10 as very much, to what extent are you convinced by the pitch?
- Taking 1 as not at all and 10 as very much, to what extent are you moved by the pitch?
- Taking 1 as not at all and 10 as very much, to what extent does the pitch seem authentic?
- Taking 1 as not at all and 10 as very much, how likely would you be to donate to this fundraiser if you saw this online?

This study was pre-registered as AsPredicted #188910, available at `https://aspredicted.org/92xh-v6tc.pdf`. We ran the study with 100 participants, each asked to assess 3 fundraising pitches; our final sample consisted of individuals with $M_{age} = 38.1$, 49.3% female.[18]

We then analyze the relationships between these respondent answers and observed fundraising outcomes for these 50 randomly-selected fundraisers to validate that answers to our survey questions indeed reflect fundraiser persuasiveness. Results are presented in Table 5. We find that participant assessments of how moved they were by the fundraising pitch or how likely they are to donate significantly correlates with actual fundraising success. In subsequent sections, these therefore serve as our focal persuasiveness metrics for evaluating fundraising

[17] Note that these 50 randomly-selected fundraisers are from an independent draw as the 40 fundraisers that are the bases of our rewrites. We draw these 50 fundraisers from the full sample of online medical fundraisers.

[18] Some participants only completed fewer than 3 responses, leading to a slightly smaller overall sample size than 300.

Table 5: Instrument Validation Results

	Log Proportion Fundraised				
Moved	0.084** (0.029)	-	-	-	-
Well-written	-	0.022 (0.035)	-	-	-
Convinced	-	-	0.054 (0.030)†	-	-
Authentic	-	-	-	0.033 (0.035)	-
Likely to Donate	-	-	-	-	0.071* (0.032)
Controls	X	X	X	X	X
Observations	287	287	287	287	287
R^2	0.453	0.443	0.446	0.443	0.450

	Fundraising Target Reached				
Moved	0.016* (0.007)	-	-	-	-
Well-written	-	0.017† (0.009)	-	-	-
Convinced	-	-	0.014 (0.009)	-	-
Authentic	-	-	-	0.009 (0.009)	-
Likely to Donate	-	-	-	-	0.017* (0.008)
Controls	X	X	X	X	X
Observations	287	287	287	287	287
R^2	0.312	0.311	0.309	0.304	0.312

Notes: Standard errors in parentheses, clustered by participant. Controls include days since created, log goal, and a quadratic for wordcount. † = 0.1 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

pitch rewrites versus originals. (Results are highly similar when examining the full set of persuasiveness metrics, presented in Appendix E.)

5.3 Emotion Comparisons across Rewriting Modes

Next, we estimate the success of the rewrites in the primary terms of changing the emotional content of the fundraising pitches in the expected direction. We perform this validation at the start of our final persuasiveness evaluation survey. For this survey, we recruited participants on Prolific (M_{age} = 38.49, 49.2% female); the pre-registration plan is available

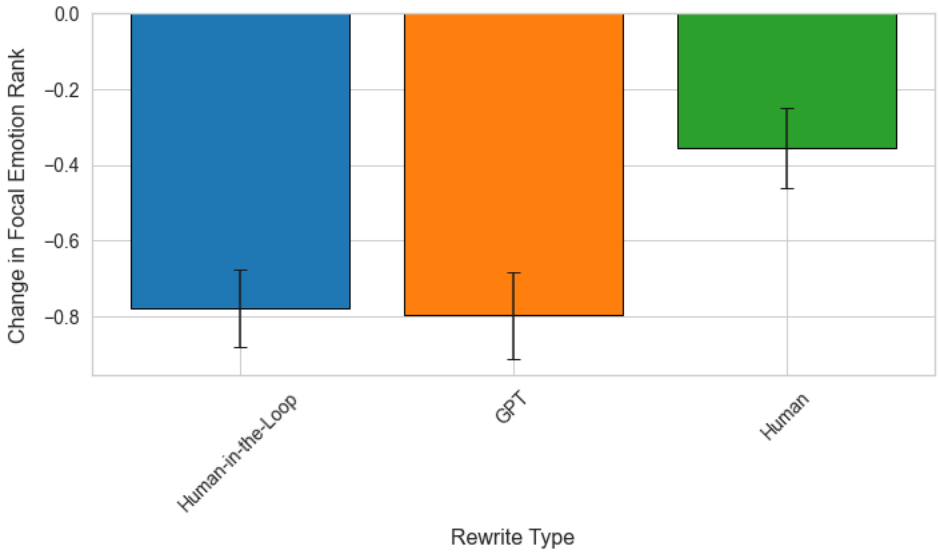
at `https://aspredicted.org/ycn9-kfzk.pdf`. We keep only observations with complete answers to all persuasiveness and emotion ranking questions and for participants that pass our attention check for a total N of 3579.[19]

We present participants with 3 sets of 4 fundraisers—both the original and the corresponding human, LLM, and human-in-the-loop rewrites (without remarking which are which)—and ask them to rank the sets of 4 pitches in terms of their respective emotions.[20] We rely on this ranking-based approach because our earlier validation of emotion ratings, presented in Appendix Section B.2, shows a high degree of noise across human emotion ratings, with considerable subjectivity in how humans rate the intensity of different emotions on 1-10 scales; by asking participants to rank emotions of the original and each rewrite, we are able to rigorously establish whether the rewrites were baseline successful in shifting the story's emotion while minimizing the measure's exposure to inconsistency of emotion-level perceptions across subjects. Results are presented in Table 6 and in Figure 1, where we perform regressions of indicators for the rank of specific types of rewrites, measured against the excluded category of pitch originals. We find that all rewrites are able to significantly shift emotion, but that LLM-assisted (or solely LLM) rewrites cause significantly more salient shifts, with effects of almost a half rank lower (i.e., higher in the focal emotion) on average than human rewrites. Nonetheless, both LLM-assisted and solely-human rewrites are both significantly associated with higher levels of emotion (lower rank) as compared to original

[19] We only keep observations with complete answers to keep the sample consistent across analyses, but as incomplete answer rates were high ($\geq 25\%$), we present regressions including respondent data from partially-complete survey responses in Appendix Section E.2. Results are qualitatively very similar, although slightly more precise given larger sample sizes.

[20] For each focal emotion sequence, we ask participants just about the respective emotion(s) that rewriters were instructed to put into the text. For the two-emotion shift of *sadness* → *caring*, we average the rank of sadness and caring to keep equal weight for each fundraiser and fundraiser type. For more details, see Appendix Section E.1.

Figure 1: Effect of Rewrites on Focal Emotion Rank



Notes: The effect of rewriting, across different modes, on average rank of focal emotion. Standard errors clustered by participant. 95% confidence intervals and coefficient estimates for the relative rank change from rewrite modes (compared to original). Results based on OLS.

Table 6: Rewrite Type and Focal Emotion Rank

	Focal Emotion Rank	
	(OLS)	(Ordinal Probit)
Human-in-the-Loop Rewrite	-0.779*** (0.052)	-0.803*** (0.057)
GPT Rewrite	-0.797*** (0.058)	-0.821*** (0.064)
Human Rewrite	-0.355*** (0.054)	-0.382*** (0.056)
Observations	3579	3579
R^2 / Log-Likelihood	0.099	-4774.7

Notes: Standard errors in parentheses, clustered by participant. Sample includes partial responses. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

fundraising pitches.

5.4 Persuasiveness Comparisons Across Sequences and Rewriting Modes

With these checks and first-order effect estimates in hand, we turn to evaluating the comparative persuasiveness of rewritten fundraisers relative to the original fundraising pitch. After asking online Prolific participants to evaluate the ranking of emotions for each rewrite, we then ask our pre-validated questions of fundraiser persuasiveness.[21] Results for persuasiveness in terms of our most strongly-validated question, of how "moved" participants were by the fundraising pitch, are presented in Table 7 and Figure 2.[22] We find strong evidence of improved persuasiveness across three of our tested emotion sequences, but only for LLM and human-in-the-loop rewrites: for human rewrites, effects are universally either null or negative. *sadness* $\rightarrow$ *caring* progressions are particularly low-quality when done with a solo human rewriter, but particular effective for other rewriting modes, emphasizing the importance of the skill augmentation from our LLM-assisted approach—without using LLM assistance, our experimental validation would have led to drastically different conclusions about the practical value of different emotion sequences.

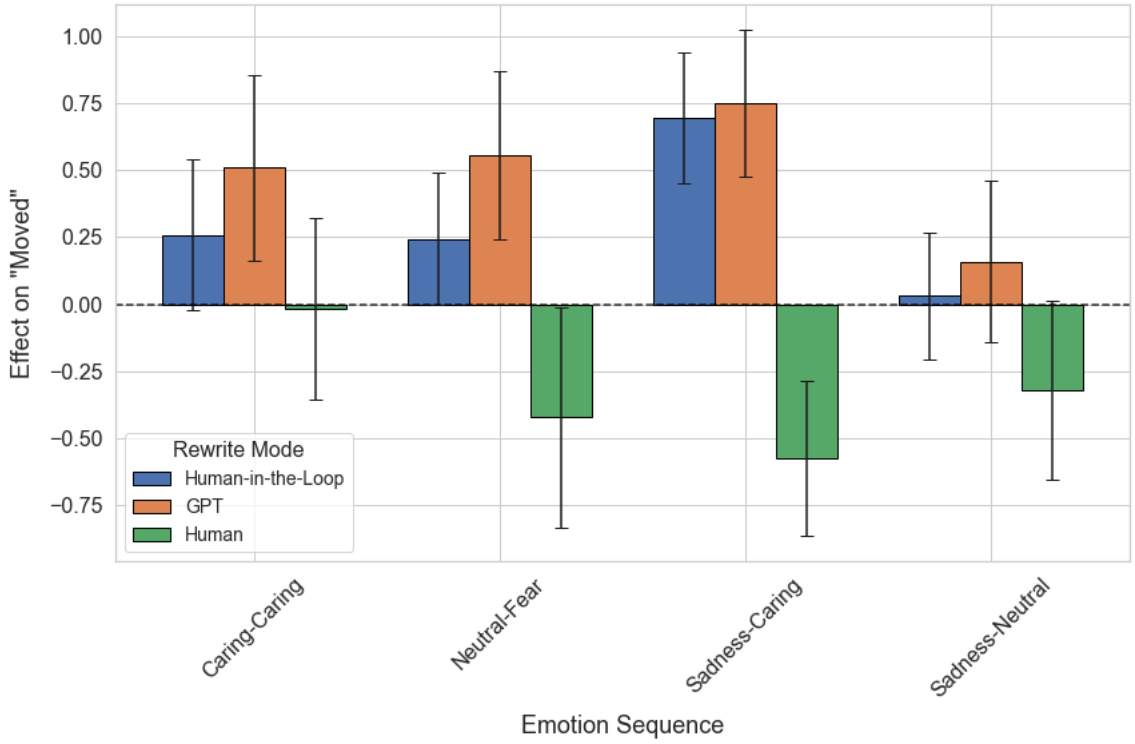
With LLM- or LLM-assisted rewrites, we find strong and large effects of rewriting fundraisers to feature the emotion sequences of *sadness* $\rightarrow$ *caring*, *caring* $\rightarrow$ *caring* and *neutral* $\rightarrow$ $fear$, with especially large effects for *sadness* $\rightarrow$ *caring*.[23] For *sadness* $\rightarrow$ *caring*

[21] This study was the second part of the pre-registration under AsPredicted #201218, available at `https://aspredicted.org/ycn9-kfzk.pdf`.

[22] In addition to the focal covariates for each emotion arc and rewrite type, we also include fixed effects for each fundraiser (i.e., an indicator for being either an original or rewrite of fundraising pitch number X in our sample, for each given fundraiser X in our sample) in these specifications in order to control for any potential baseline differences in reader perceptions of each pitch.

[23] That said, we do find that a number of effect estimates for our solo-LLM rewrites on the "moved" outcome drop to insignificance after dropping any rewrites flagged by online participants in post-validation as having

Figure 2: Effect of Rewrites on Fundraiser Persuasiveness: "Moved"



Notes: The effect of rewriting, across different modes, on participant answers to how "moved" participants are by the fundraising pitch on a scale of 1-10. Standard errors clustered by participant. 95% confidence intervals and coefficient estimates for the relative rank change from rewrite modes (compared to original). Results based on OLS.

and $caring \rightarrow caring$ this is a validation of the field data findings, but for $neutral \rightarrow fear$, the experimental validation results refute the pattern we recovered in the field data, where we found a significantly negative estimated effect of this sequence on observed fundraiser success. The point estimates of the effect of $neutral \rightarrow fear$ rewrites are of similar magnitude than those of the $caring \rightarrow caring$ sequence, and are only marginally less significant—but certainly not significantly negative. This pattern of results suggests that our field data may feature a source of confounding variation that correlates with the $neutral \rightarrow fear$ arc, such as a correlation between expressing fear at the end of a fundraiser is and a person having

added or missing salient information, suggesting that the solo-LLM effects may be driven in part by effects of changed information content; whereas our human-in-the-loop LLM rewrites are qualitatively extremely similar. See Appendix C.4 for more detail.

Table 7: Effect of Emotion Arc Rewrites on Fundraiser Quality

	How moved were you by... ?
$Caring \rightarrow Caring$: Human-in-the-Loop	$0.257^{\dagger}$ (0.144)
$Caring \rightarrow Caring$: GPT	0.509^{**} (0.177)
$Caring \rightarrow Caring$: Human	-0.016 (0.173)
$Neutral \rightarrow Fear$: Human-in-the-Loop	$0.244^{\dagger}$ (0.125)
$Neutral \rightarrow Fear$: GPT	0.556^{***} (0.162)
$Neutral \rightarrow Fear$: Human	-0.422^{*} (0.210)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.697^{***} (0.124)
$Sadness \rightarrow Caring$: GPT	0.752^{***} (0.140)
$Sadness \rightarrow Caring$: Human	-0.577^{***} (0.148)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	0.030 (0.120)
$Sadness \rightarrow Neutral$: GPT	0.159 (0.153)
$Sadness \rightarrow Neutral$: Human	$-0.322^{\dagger}$ (0.171)
Fundraiser Number Fixed Effects	X
Observations	3579
R^2	0.065

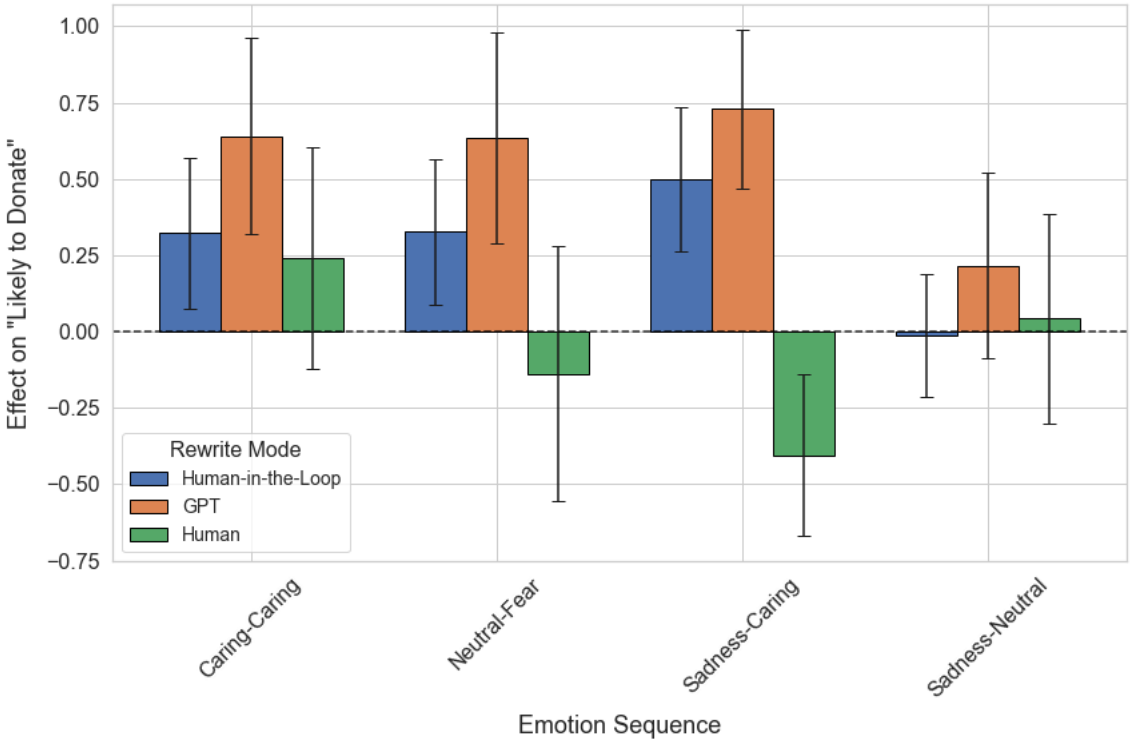
Notes: Standard errors in parentheses, clustered by participant. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

good reason to fear failure. Similarly, for *sadness* → *neutral* we find no significant positive effects from rewriting, contrary to the positive effect estimates we recovered from our observational field data, suggesting that this story type may conversely be associated with confounding factors that drive fundraising success.[24]

Taken together, these results provide strong support for the importance of implementing our simple crowd-sourced rewriting approach: without simulating the out-of-sample application of results, we would have mistakenly concluded that one of the helpful emotion sequences, $neutral \rightarrow fear$, was harmful to fundraiser performance. Moreover, if we'd only

[24]For example, if ending a fundraising pitch on a neutral tone is associated with having a higher confidence in one's probability of success due to outside knowledge of true higher success probability, this might confound observational effect estimates.

Figure 3: Effect of Rewrites on Fundraiser Persuasiveness: "Likely to Donate"



Notes: The effect of rewriting, across different modes, on participant answers to how "likely to donate" participants are by the fundraising pitch on a scale of 1-10. Standard errors clustered by participant. 95% confidence intervals and coefficient estimates for the relative rank change from rewrite modes (compared to original). Results based on OLS.

relied on solely-human rewrites, we could have mistakenly concluded that many potentially-helpful emotion sequences are not practically useful.

Results for our other significantly-validated persuasiveness metric, whether participants report being "likely to donate" to a fundraiser, are reported in Table 8 and Figure 3. Patterns are largely similar, although here we do find that human rewrites are insignificantly successful at improving persuasiveness for the *caring* → *caring* emotion sequence, albeit not in the other emotion sequence conditions. Otherwise, here we also find that LLM-assisted *sadness* → *caring*, *neutral* → *fear* and *caring* → *caring* rewrites all lead to higher fundraiser persuasiveness perceptions among survey participants, while *sadness* → *neutral*

Table 8: Effect of Emotion Arc Rewrites on Fundraiser Quality

	How likely would you be to donate... ?
Caring → *Caring* : Human-in-the-Loop	0.322*(0.126)
Caring → *Caring* : GPT	0.640***(0.164)
Caring → *Caring* : Human	0.241(0.186)
Neutral → *Fear* : Human-in-the-Loop	0.327**(0.121)
Neutral → *Fear* : GPT	0.634***(0.176)
Neutral → *Fear* : Human	-0.139(0.214)
Sadness → *Caring* : Human-in-the-Loop	0.500***(0.120)
Sadness → *Caring* : GPT	0.730***(0.132)
Sadness → *Caring* : Human	-0.405**(0.136)
Sadness → *Neutral* : Human-in-the-Loop	-0.011(0.103)
Sadness → *Neutral* : GPT	0.216(0.155)
Sadness → *Neutral* : Human	0.042(0.176)
Fundraiser Number Fixed Effects	X
Observations	3579
R^2	0.041

Notes: Standard errors in parentheses, clustered by participant. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

has null effects.

Results for the remaining outcome measures (those that did not significantly correlate with real-world outcomes in our metric validation step) are presented in Appendix Section E.3. We recover highly similar patterns of results across outcomes.

Finally, to ensure that results aren't driven by changes in the informational content of the pitches between originals and rewrites, we also perform extension analyses where we exclude rewrites that are flagged by subsequent validation surveys as featuring salient informational differences from the originals, in violation of our instructions. Full details on both this validation survey and restricted-sample analysis are presented in Appendix Section

C.4. Results for human-in-the-loop rewrites are qualitatively unchanged in this analysis, but positive effects of LLM-only rewrites are meaningfully attenuated and, particularly for the "moved" outcome, lose statistical significance, suggesting that solo-LLM rewrite effects are driven in part by informational differences in the rewrites. This suggests that human-in-the-loop rewriting, or LLM-assisted rewriting, is the most reliable rewriting mode for implementing our crowd-sourced rewriting approach.

5.5 "Placebo" Rewrites

While the above results show clearly that rewriting pitches to include our emotion sequences cause significantly positive improvements in persuasiveness, one may also be concerned about the persistently positive effects we see for human-in-the-loop and, to a lesser extent, LLM-only rewrites. In particular, one may worry that our rewriting exercise introduces a novel confound of the possible positive benefit of LLM-only or human-in-the-loop rewriting at baseline, even without any emotion sequence. While this would not entirely be consistent with the pattern that we find in our data—we still find differential responses across emotion sequences, with insignificant and much smaller effects for $sadness \rightarrow neutral$—one may find it surprising that even progressions that were estimated to have negative impacts in our field data are estimated here to have positive impacts on rewriting, and suggestive that there may be a baseline effect of LLM-assisted rewriting in general, outside of any particular effects of emotion sequences specifically. Similarly, given the frequently negative effects of human rewrites, one may worry that there is simply a generalized persuasiveness decline associated with solely-human rewrites.

Table 9: Effect of Placebo Rewrites on Fundraiser Quality

	How moved were you by... ?
Placebo: Human-in-the-Loop	-0.083 (0.121)
Placebo: ChatGPT	-0.191 (0.136)
Placebo: Human	-0.545*** (0.160)
Fundraiser Number Fixed Effects	X
Observations	1,212
R^2	0.050

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

We empirically test this possibility with a simple "placebo" test of our entire rewriting pipeline. For a set of 10 *neutral* → *neutral* progressions from our set of 40 randomly-selected original pitches, we ask ChatGPT to rewrite the fundraising pitches as in our earlier first step, with identical prompts except that all language relating to emotions is stripped out. We then replicate our recruited human participant and human-in-the-loop rewriting procedure after also stripping out language related to emotions or emotion sequences in survey questions and prompts. Finally, we evaluated the persuasiveness of the rewrites against the original in an identical survey to those we used to evaluate the baseline emotion sequence rewrites above, administered to 103 participants ($M_{age} = 37.26$, 49.5% female).[25]

Results are presented in Table 9. We find null effects of rewrites that do not include emotion sequences with either our human-in-the-loop rewrites or solely-LLM rewrites, but informatively, we do find that there is a baseline persuasiveness decline that's caused by human rewrites. This demonstrates that crowd-sourced human rewrites are unreliable as a

[25]This study was pre-registered as AsPredicted #201617, available at `https://aspredicted.org/2nw3-yt3k.pdf`.

testing mechanism due to skill deficits; but LLM and LLM-assisted rewrites show null effects in our placebo check, suggesting that the strong results we find from our emotion sequence rewrites are indeed driven by the given emotion sequence.

6 Mechanism: Identification or Transportation?

Lastly, we investigate the possible underlying psychological mechanisms that may drive this persuasiveness effect. We focus specifically on two candidate mechanisms emphasized by prior literature on the psychology of narrative: that stories may be more affecting through inducing higher "narrative transportation" (Green and Brock, 2000)—defined as a psychological state of heightened attention to, and emotional engagement with, the story at hand—or alternatively may be more affecting through inducing greater empathic identification with the protagonist (Cohen, 2001). We hypothesize that emotion sequence effects are mediated by how effective each given sequence is as a story, as measured in terms of these psychology-of-narrative dimensions. We investigate this hypothesis empirically in a small follow-up study.

For this study, we recruited 400 participants on Prolific, 396 of which passed our attention check (μ_{age} = 39.6, 50.0% female) and were included in the analysis sample. This study mirrored our primary study except in that we only showed participants original fundraising pitches and human-in-the-loop rewrites to conserve costs; each participant was shown a pair of original fundraising pitch and human-in-the-loop rewrite and then asked to score the pitches on both the prior persuasion outcome metrics[26] as well as a set of 10 questions

[26]That is, the same set of questions as detailed in section 5.2 above, e.g. "how moved were you...?" and "how likely would you be to donate...?"

adapted for our setting from the Green and Brock (2000) instrument for measuring narrative transportation[27] and 6 questions adapted for our setting from the Cohen (2001) instrument for measuring identification in narratives, with all questions presented in fully randomized order.[28] For the full list of questions included in these instruments and further detail on the study design, see Appendix Section D.1.

We present the mediation analysis from this study for the *sadness* → *caring* emotion sequence in Table 10. Mediation analyses for other emotion sequences are presented in Appendix Section D.2.[29] We performed this analysis using standard structural equation modeling (Rosseel, 2012), modeling the effect of *sadness* → *caring* on donation behavior ("likely to donate" ratings) as composed of a direct effect and parallel mediators of narrative transportation and identification. In the first panel of Table 10, we demonstrate that we find that the indirect effects significantly mediate the overall observed effect, and furthermore that this indirect effect is driven entirely by a highly significant effect on identification, compared to a null effect on narrative transportation. This shows that *sadness* → *caring* sequences are significantly more effective in online fundraising because they drive heightened empathic identification with the protagonists of the fundraiser, and not because such story types are necessarily more narratively engaging. Furthermore, we find that the direct effect, presented in the bottom panel of Table 10, is insignificant after accounting for these indirect effects, and with a point estimate less than half the size of the mediated effect through identification,

[27] For example, "While I was reading the fundraiser, I could easily picture the events in it taking place", or "[w]hile reading the fundraiser, I forgot myself and was fully absorbed", rated on a scale of 1-10.

[28] For example, "At key highlights in the fundraiser, I felt I knew exactly what [the protagonist] was going through," or "I think I have a good understanding of [the protagonist]", rated on a scale of 1-10, where [the protagonist] is replaced by the main person(s) whose plight is (are) featured in the fundraising pitch.

[29] Overall, we find directionally similar mediation effects for *neutral* → *fear*, *caring* → *caring* and *sadness* → *neutral* emotion sequences in this follow-up study, but with the 75% smaller sample size as compared to our full study, we do not recover effects with statistical significance for these sequences.

Table 10: Mediation Analysis

Indirect Effects	
Transportation	0.030 (0.018)
Identification	0.543 (0.128)***
Total Indirect	0.573 (0.139)***

Direct and Total Effects	
Direct Effect	0.275 (0.192)
Total Effect	0.848 (0.260)**

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Results shown for *sadness* → *caring* emotion sequence effects, other emotion sequence mediation analyses shown in Appendix D.2.

demonstrating that identification is the predominant mechanism of the observed total effect.

7 Conclusion

In sum, this study introduces a new template for categorizing narrative types that is tractable even in short-form media, based on the specific emotional progression of the given content, which we term the emotion sequence. We then empirically demonstrate the potential for this template in narrative analysis in the context of medical fundraising, analyzing over 14,000 GoFundMe campaigns as well as crowd-sourced rewrites of randomly-selected fundraisers, and establish that specific emotion progressions strongly influence fundraiser persuasiveness, highlighting the key importance of not just what emotions are present in a narrative but when and how they unfold.

First, using a RoBERTa-based emotion classifier, we measured the emotional dynamics within each half of each fundraising pitch, uncovering patterns that simpler analyses of overall emotional content would miss or only incompletely characterize. We find that *sadness* →

caring arc, in particular, is associated with significantly higher probability of fundraising success. Similarly, progressions like $caring \rightarrow caring$ and $sadness \rightarrow neutral$ also show positive associations with fundraising outcomes, while progressions such as $neutral \rightarrow fear$ tend to perform poorly.

However, given that these observational data may feature meaningful confounding variation from unobserved variables that are associated with both the presence of particular emotion sequences and fundraiser success, we next develop an intuitive new approach—crowd-sourced rewriting—to validate these findings. In short, we assign recruited online participants to rewrite randomly-selected online fundraising pitches to feature a given emotion sequence, allowing us to effectively test the out-of-sample application of our research results; in other words, we test whether online participants can apply our findings and indeed improve the perceived persuasiveness of a given fundraising pitch. We implement this with three modes to test the relative validity of either solely-human, solely-LLM, or LLM-assisted (human-in-the-loop) rewrites and find that, while human rewrites often fell short due to skill gaps and solo-LLM rewrites are found to be partly driven by informational differences between the original and the rewrite, LLM-assisted rewrites successfully embedded the focal progressions and demonstrated robust improvements in perceived persuasiveness—for certain sequences. In particular, we find that $sadness \rightarrow caring$ meaningfully improves persuasiveness when embedded in fundraising pitches, replicating the strong positive association we find in our field data; but we also find that e.g. $neutral \rightarrow fear$ does not replicate the negative association with fundraising success that we estimate in field data, suggesting that some (but not all) of our observational results were indeed affected by confounds and highlighting the importance of this crowd-sourced-rewriting validation. Together, this combination of large-scale

field data analysis and experimental validation together offers strong evidence in favor of the significant positive impact of *sadness* → *caring* and *caring* → *caring* emotion sequences, in particular, on persuasiveness.

Finally, we establish through mediation analysis that empathic identification with the protagonist of the fundraiser is the primary mechanism of the effect, with identification significantly mediating the total emotion sequence effects, building directly on earlier work on the psychological mechanisms of narrative (Cohen, 2001).

At the same time, we note that several limitations remain. This study focuses exclusively on medical fundraising pitches, raising the question of whether the observed effectiveness of the *sadness* → *caring* arc is specific to this context. It likely is. Medical fundraisers predominantly seek to evoke empathy and emotional connection, which may make *sadness* → *caring* particularly powerful here. Future research should explore which emotion sequences are most effective in other domains, such as political campaigns, product marketing, or public health messaging, where the most-appropriate emotion dynamics may differ due to differences in the particulars of the context and/or audience expectations. However, while the findings on specific emotion sequences may not generalize to other contexts, we hope that our human-in-the-loop, LLM-assisted methodology can easily be applied to study emotion sequences (and other linguistic features) across different settings.

Extending this work to other contexts and cultural settings could also yield broader insights into the particular mechanisms of storytelling success in other domains. Emotional dynamics may resonate for a variety of reasons in other contexts, which could include greater narrative transportation or combinations of transportation and identification from specific sequences, which may vary along in other contexts with the effectiveness of different emotion

sequences. Further work is needed to analyze why some progressions are most appropriate, and in which settings, than others. Future work may also seek to establish systematic relations between which particular sequences are most effective in which contexts, and which narrative-related psychological mechanism operates as the primary mechanism, depending. Finally, while we focus on two-part emotion sequences for tractability in our present context, richer insights may be gleaned in future work that analyzes three- or four-part, or even higher-order, sequences, particular in longer-form contexts where content data are rich enough to support these higher-dimensional analyses.

Taken together, this study shows that emotional dynamics are core to storytelling and offers a new template for studying story types across contexts. By combining large-scale observational analysis with experimental validation, we provide actionable insights for fundraisers and a roadmap for future studies on storytelling, persuasion, and the role of particular emotion progressions in communication.

References

Banerjee, Akshina and Oleg Urminsky (2024)„ "The Language That Drives Engagement: A Systematic Large-scale Analysis of Headline Experiments,", *Marketing Science*, https://doi.org/10.1287/mksc.2021.0018, forthcoming.

Batson, C. Daniel, Bruce D. Duncan, Paula Ackerman, Terese Buckley, and Kimberly Birch (1981)„ "Is Empathic Emotion a Source of Altruistic Motivation?,", *Journal of Personality and Social Psychology*, 40 (2), 290–302.

Berger, Jonah, Y. Dan Kim, and Robert Meyer (2021)„ "What makes content engaging? How emotional dynamics shape success,", *Journal of Consumer Research*, 48 (2), 235–250.

Berger, Jonah, Wendy W Moe, and David A Schweidel (2023)„ "What holds attention? Linguistic drivers of engagement,", *Journal of Marketing*, 87 (5), 793–809.

Brennan, Linda and Wayne Binney (2010)„ "Fear, guilt, and shame appeals in social marketing,", *Journal of Business Research*, 63 (2), 140–146, https://doi.org/10.1016/j.jbusres.2009.02.006.

Buchanan, Joy, Stephen Hill, and Olga Shapoval (2023)„ "ChatGPT Hallucinates Nonexistent Citations: Evidence from Economics,", *The American Economist*, 69 (1), TBD, https://doi.org/10.1177/05694345231218454, first published online November 23, 2023.

Cavanaugh, Lisa A., James R. Bettman, and Mary Frances Luce (2015)„ "Feeling Love and Doing More for Distant Others: Specific Positive Emotions Differentiate Prosocial Behavior Toward Distant Versus Close Others,", *Journal of Marketing Research*, 52 (5), 655–669.

Chan, Eugene Y. and Felix Septianto (2022)„ "Disgust predicts charitable giving: The role of empathy,", *Journal of Business Research*, 142, 946–956.

Cohen, Jonathan (2001)„ "Defining Identification: A Theoretical Look at the Identification of Audiences With Media Characters,", *Mass Communication and Society*, 4 (3), 245–264, published online: 17 Nov 2009.

Demszky, Dorottya, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi (2020)„ "GoEmotions: A Dataset of Fine-Grained Emotions,", *arXiv preprint arXiv:2005.00547*, https://doi.org/10.48550/arXiv.2005.00547, accepted to ACL 2020.

Ekman, Paul (1992)„ "An argument for basic emotions,", *Cognition Emotion*, 6 (3-4), 169–200.

Fong, Hortense, Vineet Kumar, and K. Sudhir (2024)„ "A Theory-Based Explainable Deep Learning Architecture for Music Emotion,", *Marketing Science*, 44 (1), 153–173, https://doi.org/10.1287/mksc.2022.0323.

Goenka, Shreyans and Stijn M J van Osselaer (2019)„ "Charities can increase the effectiveness of donation appeals by using a morally congruent positive emotion,", *Journal of Consumer Research*, 46 (4), 774–790, https://doi.org/10.1093/jcr/ucz012.

GoFundMe, "How to Tell Your Campaign Story: A GoFundMe Guide," (2023), https://www.gofundme.com/c/blog/campaign-story, accessed: 2024-12-26.

Goldenberg, Jacob, David Mazursky, and Sorin Solomon (1999a)„ "The Fundamental Templates of Quality Ads,", *Marketing Science*, 18 (3), 333–351.

Goldenberg, Jacob, David Mazursky, and Sorin Solomon (1999b)„ "Toward Identifying the Inventive Templates of New Products: A Channeled Ideation Approach,", *Journal of Marketing Research*, 36 (2), 200–210.

Goli, Ali and Amandeep Singh (2023)„ "Can LLMs Capture Human Preferences?,", *arXiv preprint arXiv:2305.02531*.

Green, Melanie C. and Timothy C. Brock (2000)„ "The role of transportation in the persuasiveness of public narratives,", *Journal of Personality and Social Psychology*, 79 (5), 701–721.

Guan, Zhengzhi, Fangfang Hou, Boying Li, Zhao Cai, and Alain Yee-Loong Chong, "Evaluating Charitable Crowdfunding Based on Storytelling Linguistic Cues: A Narrative Persuasion Perspective,", "SIGHCI 2020 Proceedings, Special Interest Group on Human-Computer Interaction," Association for Information Systems (AIS) Electronic Library (AISeL) (2020).

Hong, Jiyeon and Paul R. Hoban (2022)„ "Writing More Compelling Creative Appeals: A Deep Learning-Based Approach,", *Marketing Science*, 41 (5), 845–865, https://doi.org/10.1287/mksc.2022.1351.

Hutto, C. and Eric Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text,", "Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media," (2014), https://doi.org/10.1609/icwsm.v8i1.14550.

Kahneman, Daniel and Amos Tversky (1979)„ "Prospect Theory: An Analysis of Decision under Risk,", *Econometrica*, 47 (2), 263–292.

Kappes, Heather Barry, Joe J Gladstone, and Hal E Hershfield (2021)„ "Beliefs about whether spending implies wealth,", *Journal of Consumer Research*, 48 (1), 1–21.

Knight, Samsun, Matthew D. Rocklage, and Yakov Bart (2024)„ "Narrative reversals and story success,", *Science Advances*, 10 (34), eadl2013, https://doi.org/10.1126/sciadv.adl2013.

Knight, Samsun, Matthew D. Rocklage, and Yakov Bart (2025)„ "Emotion Dynamics and the Consumption Paradox: Emotional Variance Predicts High Ratings but Low Popularity,", *Working paper*.

Li, Peiyao, Noah Castelo, Zsolt Katona, and Miklos Sarvary (2024)„ "Frontiers: Determining the validity of large language models for automated perceptual analysis,", *Marketing Science*, 43 (2), 254–266.

Li, Wei, Dongshan Yang, and Yuxin Sun (2022)„ "Analysis of text factors impacting donation behavior in public welfare crowdfunding projects,", *Human Systems Management*, 42 (1), https://doi.org/10.3233/HSM-220024, first published online July 16, 2022, Free access.

Liang, Jianping, Zengxiang Chen, and Jing Lei (2017)„ "Inspirational Appeals Are More Effective: The Influence of Strength Emotions on Persuasion and Donation,", *Rutgers Business Review*, 2 (2), 212–216.

Loewenstein, George F. and Drazen Prelec (1993)„ "Preferences for sequences of outcomes,", *Psychological Review*, 100 (1), 91–108.

Lowe, Sam, "RoBERTa-Base-GoEmotions Model on Hugging Face,", https://huggingface.co/SamLowe/roberta-base-go_emotions (2024), accessed: December 15, 2024.

Lu, Baozhou, Tailai Xu, and Weiguo Fan (2024)„ "How do emotions affect giving? Examining the effects of textual and facial emotions in charitable crowdfunding,", *Financial Innovation*, 10, Article 108.

Lynch, John G (1999)„ "Theory and external validity,", *Journal of the Academy of Marketing Science*, 27, 367–376.

Lynch Jr, John G (1982)„ "On the external validity of experiments in consumer research,", *Journal of consumer Research*, 9 (3), 225–239.

Min, Semi and Natalia Levina, "Linguistic Features of Identifiable Victim Effect in Microlending,", "Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing," (2018).

Netzer, Oded, Annabelle Lemaire, and Michal Herzenstein (2019)„ "When words sweat: Identifying signals for loan default in the text of loan applications,", *Journal of Marketing Research*, 56 (6), 960–980.

Nguyen, Nguyen, Joseph Johnson, and Michael Tsiros (2023)„ "Unlimited Testing: Let's Test Your Emails with AI,", *Marketing Science*, 43 (2), 240–256, https://doi.org/10.1287/mksc.2021.0126.

Packard, Grant, Yang Li, and Jonah Berger (2024)„ "When Language Matters,", *Journal of Consumer Research*, 51 (3), 634–653, https://doi.org/10.1093/jcr/ucad080.

Paxton, Pamela, Kristopher Velasco, and Robert W. Ressler (2020)„ "Does Use of Emotion Increase Donations and Volunteers for Nonprofits?,", *American Sociological Review*, 85 (6), 1051–1073, https://doi.org/10.1177/0003122420960104.

Reagan, Andrew J., Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds (2016)„ "The emotional arcs of stories are dominated by six basic shapes,", *EPJ Data Science*, 5 (1), 31, https://doi.org/10.1140/epjds/s13688-016-0093-1.

Rosseel, Yves (2012)„ "lavaan: An R Package for Structural Equation Modeling,", *Journal of Statistical Software*, 48 (2), 1–36, http://www.jstatsoft.org/v48/i02/.

Social Security Administration, "Popular Baby Names,", https://www.ssa.gov/oact/babynames/ (2024), accessed: 2024-12-15.

Telle, Nils-Torge and Hans-Rüdiger Pfister (2016)„ "Positive Empathy and Prosocial Behavior: A Neglected Link,", *Emotion Review*, 8 (2), 154–163.

Tellis, Gerard J, Deborah J MacInnis, Seshadri Tirunillai, and Yanwei Zhang (2019)„ "What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence,", *Journal of marketing*, 83 (4), 1–20.

Toubia, Olivier, Jonah Berger, and Jehoshua Eliashberg (2021)„ "How Quantifying the Shape of Stories Predicts Their Success,", *Proceedings of the National Academy of Sciences*, 118 (26), e2011695118, https://doi.org/10.1073/pnas.2011695118.

Van Laer, Tom, Jennifer Edson Escalas, Stephan Ludwig, and Ellis A. van den Hende (2019)„ "What Happens in Vegas Stays on TripAdvisor? A Theory and Technique to Understand Narrativity in Consumer Reviews,", *Journal of Consumer Research*, 46 (2), 267–285.

Vonnegut, Kurt, "Shapes of Stories,", Video (1995), available at: https://youtube.com/watch?v=oP3c1h8v2ZQ.

Xu, Jie (2017)„ "Moral Emotions and Self-Construal in Charity Advertising: Communication on Focus,", *Journal of Promotion Management*, 23 (4), 557–574, https://doi.org/10.1080/10496491.2017.1297976.

Zhang, Xing, Xinyue Wang, Durong Wang, Quan Xiao, and Zhaohua Deng (2024)„ "How the Linguistic Style of Medical Crowdfunding Charitable Appeal Influences Individuals' Donations,", *Technological Forecasting and Social Change*, 203, 123394, https://doi.org/10.1016/j.techfore.2024.123394.

Zhao, Kexin, Lina Zhou, and Xia Zhao (2022)„ "Multi-modal emotion expression and online charity crowdfunding success,", *Decision Support Systems*, 163, 113842.

Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang (2024)„ "Can large language models transform computational social science?,", *Computational Linguistics*, 50 (1), 237–291.

Appendix A:

Robustness Analyses for Field Data Results

In this section, we present robustness analyses for our field data regressions of emotion against fundraiser success using different modeling choices and specifications. We also present extension analyses using different emotion sets to examine other potential patterns in our data, and discuss these results both in light of our own findings and in terms of potential directions for future work.

A.1 Continuous Score Regressions

Our RoBERTa-based classifier is trained on binary 0/1 labels of emotion, and in implementation, we follow standard practice and translate the continuous emotion scores returned by the classifier model into binary 0/1 predicted labels based on a threshold of 0.25 (Lowe, 2024). In this section, we present regressions of detected emotions against fundraiser success using continuous emotion scores from our RoBERTa-based classifier instead of binary labels. Across the board, results are highly similar when using the continuous scores directly instead, and effects are generally slightly larger and more significant, which intuitively matches the fact that continuous scores preserve more statistical signal by retaining the richer original variation. While we rely on binary labels because they are easier to interpret, especially in the context of the emotion sequences that are the eventual focus of our analysis, we interpret these results as demonstrating strong robustness to alternative model specifications.

Results for overall emotion scores against fundraiser success are presented in Table A1.

Table A1: Overall Emotions and Fundraiser Success:
Continuous Emotion Scores

	Success (Logit)	Log Proportion Fundraised (OLS)
Valence	0.141* (0.060)	0.104*** (0.025)
Fear	-1.358 (1.085)	-0.854* (0.414)
Caring	0.519*** (0.134)	0.755*** (0.061)
Sadness	0.022 (0.155)	0.073 (0.067)
Neutral	-0.114 (0.179)	-0.142† (0.073)
Controls	X	X
Observations	14114	14114
R^2/Pseudo R^2	0.061	0.134

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. $\dagger = 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.$

Results are stronger for the positive effect of caring but are slightly weaker for the negative effect of fear (effects are just beyond marginally significant, with $p = 0.105$), and we now see a marginally significant negative effect of neutrality on the log proportion fundraised outcome.

Results for half-by-half emotion scores against fundraiser success are presented in Table A2. Results are stronger for the positive effect of caring in both halves and for the negative effect of neutrality in the first half, although effects for sadness are slightly different; here we see a significant negative effect of sadness in the second half and null effects of sadness in the first half, compared to the marginally significant positive effect of sadness in the first half we see with the label regression.[30]

Results for half-by-half emotion interaction terms against fundraiser success are presented in Table A2. Here, since we no longer have binary labels for each half, instead of sequences,

[30]We also see slightly smaller effects for second half valence in particular, which measure is unchanged, suggesting that the continuous emotion measures account for some of the effect that we recover for valence in our primary regressions.

Table A2: Dynamic Emotion Measures and Fundraiser Success: Continuous Emotion Scores

	Success (Logit)	Log Proportion Fundraised (OLS)
1^{st} Half Valence	0.170*** (0.050)	0.105*** (0.022)
2^{nd} Half Valence	0.111 (0.082)	0.073* (0.034)
1^{st} Half Neutral	-0.467*** (0.137)	-0.176** (0.060)
2^{nd} Half Neutral	0.001 (0.168)	-0.112 (0.070)
1^{st} Half Fear	0.117 (0.511)	-0.052 (0.242)
2^{nd} Half Fear	-4.098† (2.125)	-1.364** (0.471)
1^{st} Half Caring	-0.130 (0.196)	0.277** (0.088)
2^{nd} Half Caring	0.500*** (0.115)	0.605*** (0.052)
1^{st} Half Sadness	-0.063 (0.125)	0.047 (0.056)
2^{nd} Half Sadness	-0.316 (0.233)	-0.316** (0.096)
Controls	X	X
Observations	14114	14114
R^2/Pseudo R^2	0.066	0.139

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. † = 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

we estimate baseline and interaction terms for specific first-half emotion scores and specific second-half emotion scores for each respective focal emotion in either half; the table present results for the joint effect of the first half emotion, the second half emotion, and the first half emotion X second half emotion interaction.[31] Results are broadly similar for the positive effect of *sadness* → *caring*, *caring* → *caring*, and *sadness* → *neutral*, but are just below 5% significance for *neutral* → *fear*, possibly due to heightened multicollinearity with this approach. We also find a handful of significant effects that were not present previously. We

[31]Note that coefficient estimate magnitudes sometimes vary widely in this specification as average covariate magnitudes vary considerably, whereas in our primary specification covariate magnitudes are all either 0 or 1.

rely on our binarized "sequence" specification for our main results for ease of interpretation and because using binary labels from machine learning classifiers is generally more standard, but we present these results as evidence that these patterns are generally highly robust to alternative specifications, and in particular the positive effects of *sadness* → *caring*, *caring* → *caring*, and *sadness* → *neutral*.

A.2 Alternative Emotion Set: Ekman (1992) Emotions

We also present here the results from an alternative set of focal emotions based on the Ekman 6 emotions of Ekman (1992). As described in the main text, we choose our focal emotions as a parsimonious set of three emotion categories that prior literature has highlighted as especially salient in the online fundraising context, as well as a category for emotional neutrality, which also corresponds to 4 of the 5 primary emotion factors that we recover from factor analysis and test above.[32] While our RoBERTa-based emotion classification tool examines 28 emotion categories (including neutrality), we seek to avoid multicollinearity—a particularly salient concern given our focus on emotion sequences, which increases the dimensionality of our covariates at a quadratic rate—by focusing on a smaller set of plausibly-orthogonal emotion categories that we expect may nonetheless capture important variation in story types in online fundraisers. We also sought to focus on emotion categories that online participants would be able to intuitively rewrite, avoiding non-intuitive primary emotions such as "joy" due to concerns of confusing participants with requests to make online medical fundraisers "more joyful." Given this, we position the present work as a demonstrative study of some of the

[32] Batson et al. (1981); Brennan and Binney (2010); Cavanaugh, Bettman, and Luce (2015); Telle and Pfister (2016); Chan and Septianto (2022); Zhao, Zhou, and Zhao (2022); Lu, Xu, and Fan (2024)

Table A3: Specific Emotion Interactions and Fundraiser Success: Continuous Emotion Scores

	Success (Logit)	Log Proportion Fundraised (OLS)
1^{st} Half Valence	0.184^{***}	0.124^{***}
2^{nd} Half Valence	$0.134^{\dagger}$	0.079^{*}
Joint Effects:		
1^{st} Half Fear – 2^{nd} Half Fear	-24.63	-8.56^{**}
1^{st} Half Fear – 2^{nd} Half Caring	$3.10^{\dagger}$	3.08^{***}
1^{st} Half Fear – 2^{nd} Half Sadness	-2.17	0.12
1^{st} Half Fear – 2^{nd} Half Neutral	-0.92	-0.50
1^{st} Half Caring – 2^{nd} Half Fear	-6.68	2.21
1^{st} Half Caring – 2^{nd} Half Caring	0.22	0.52^{*}
1^{st} Half Caring – 2^{nd} Half Sadness	-0.13	0.19
1^{st} Half Caring – 2^{nd} Half Neutral	0.54	-0.15
1^{st} Half Sadness – 2^{nd} Half Fear	1.24	-3.08^{**}
1^{st} Half Sadness – 2^{nd} Half Caring	0.69^{**}	0.65^{***}
1^{st} Half Sadness – 2^{nd} Half Sadness	-0.48	-0.72^{***}
1^{st} Half Sadness – 2^{nd} Half Neutral	-0.44	0.66^{**}
1^{st} Half Neutral – 2^{nd} Half Fear	-12.75	$-3.19^{\dagger}$
1^{st} Half Neutral – 2^{nd} Half Caring	-0.35	0.28^{*}
1^{st} Half Neutral – 2^{nd} Half Sadness	0.64	-0.60^{*}
1^{st} Half Neutral – 2^{nd} Half Neutral	$-0.50^{\dagger}$	-0.49^{***}
Controls	X	X
Observations	14,114	14,114
Pseudo R^2/R^2	0.0659	0.1405

Notes: Table reports joint effect estimates of emotion sequence interactions. Significance stars based on joint test of the sum of the first-half, second-half, and interaction term. $^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Controls include days posted, log fundraising goal, and a quadratic for wordcount.

emotion sequences that may drive differential success rates in online fundraisers, and make no claims that these are the only, or even the best, emotion sequences for this application; we consider a more exhaustive study of all possible emotion sequences as a promising direction for future work.

However, to help provide guidance for this future work and to examine potential other effects that this study does not focus on, we also present analyses using the Ekman (1992) set of six primary emotions instead of our three focal emotions + neutrality. This set of emotions satisfies the first of our desired conditions, of plausibly-orthogonal emotion categories to avoid overfitting, but retains emotions that would be unintuitive for participants to rewrite in our context, namely joy but also to a lesser extent anger and disgust. It also leaves out some emotions that prior literature highlights as potentially first-order important, namely expressions of empathy / caring. Nonetheless, as these are the canonical core emotion set in the emotion literature, we present analyses using these as our focal categories below, both for our overall regression regression analysis and half-by-half regression analysis.[33]

Results for the effect of overall emotion levels on fundraiser success for the Ekman (1992) emotions are presented in Table A4.[34] Here we see one of the effects that we find in our prior analysis, namely a negative association between overall levels of fear in a fundraising pitch and fundraiser success, but also find that we do not have sufficient variation in a handful of the Ekman (1992) categories to estimate effects in our logistic regression. We find a significant negative effect of disgust on fundraiser success, but as is made clear by the fact that this covariate is dropped in the logistic regression, this is based on a very small

[33]Since there are 36 interaction terms for the 6x6 matrix of potential sequences, we omit that analysis here due to space constraints.

[34]A very small number of observations are dropped for being completely collinear with particular Ekman (1992) emotion covariates in the logistic regression.

Table A4: Overall Emotions and Fundraiser Success:
Ekman (1992) Emotions

	Success (Logit)	Log Proportion Fundraised (OLS)
Valence	0.195*** (0.056)	0.175*** (0.023)
Sadness	-0.021 (0.068)	-0.022 (0.030)
Joy	0.086 (0.407)	0.045 (0.189)
Disgust	. (.)	-5.728*** (1.505)
Fear	-0.469 (0.477)	-0.581** (0.190)
Anger	. (.)	1.240 (1.505)
Surprise	-0.600 (1.076)	0.137 (0.389)
Controls	X	X
Observations	14112	14114
R^2/Pseudo R^2	0.059	0.125

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. ". (.)" indicates variables dropped due to multicollinearity.

number of online medical fundraisers that contain disgust expressions, which recommends caution in overinterpreting this observational finding. Nonetheless, this result is suggestive that expressions of disgust may also be effective in online fundraising pitches.

Results for the effect of half-by-half emotion levels are presented in Table A5. The takeaways are broadly similar as in the prior analysis; we find a high number of covariates that are dropped due to multicollinearity and insufficient variation, and find that the previously estimated significant negative effect of disgust is driven in particular by expressions of disgust in the first half of the fundraiser. We also find robustness to our prior estimated positive effect of sadness in the first half of the fundraiser, with a highly significant positive effect of sadness at the beginning of the fundraising pitch.

Taken together, this analysis is broadly supportive of our main analyses in the dimensions that they overlap, and also suggests that our focal emotions are indeed—as expected based

Table A5: Dynamic Emotion Measures and Fundraiser Success: Ekman (1992) Emotions

	Success (Logit)	Log Proportion Fundraised (OLS)
1^{st} Half Valence	0.211*** (0.045)	0.146*** (0.020)
2^{nd} Half Valence	0.195** (0.073)	0.187*** (0.030)
1^{st} Half Sadness	0.031 (0.061)	0.080** (0.027)
2^{nd} Half Sadness	-0.071 (0.110)	-0.071 (0.047)
1^{st} Half Joy	-0.058 (0.244)	0.077 (0.108)
2^{nd} Half Joy	-0.032 (0.439)	-0.028 (0.201)
1^{st} Half Disgust	. (.)	-2.571** (0.868)
2^{nd} Half Disgust	. (.)	. (.)
1^{st} Half Fear	-0.051 (0.283)	-0.126 (0.128)
2^{nd} Half Fear	-0.842 (0.739)	-0.280 (0.276)
1^{st} Half Anger	. (.)	-0.313 (1.064)
2^{nd} Half Anger	. (.)	1.391 (1.504)
1^{st} Half Surprise	0.397 (0.350)	0.149 (0.174)
2^{nd} Half Surprise	. (.)	0.066 (0.402)
Controls	X	X
Observations	14094	14114
R^2/Pseudo R^2	0.062	0.128

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. † = 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. ". (.)" indicates variables dropped due to multicollinearity.

on prior literature—plausibly the most salient emotions for determining the success or failure of online fundraising pitches. They also suggest that a potentially promising direction for future research could be examining the role of disgust in persuasion and online fundraising.

A.3 Alternative Emotion Set: 5-Emotion Set Motivated by Factor Analysis

We also present here the results from an alternative set of focal emotions derived using factor analysis. We also present this analysis as providing empirical support for the emotion categories that we focus on in our main text, as all 4 of our focal emotion categories correspond to latent factors that we recover in our 5-factor model.

A.3.1 Factor Analysis Results

We use standard factor analysis on the full set of all 28 emotion categories measured by our RoBERTa-based classifier (Lowe, 2024) to recover a limited number of latent factors that can explain the observed variation in emotions in our sample of GoFundMe fundraising pitches.[35] We focus on a 5-factor model for this exercise, but the implied emotion set is highly similar (albeit with potentially fewer or higher numbers of emotions) when focusing on different factor degrees.[36] The latent vectors of factor loadings that we estimate, using maximum likelihood estimation, are presented in Table A6. Here, we find that sets of similar emotions do tend to broadly correlate with one another, and interpret these results as

[35] We perform this factor analysis on the continuous emotion scores.

[36] We choose 5 heuristically, as we find that 6 and higher numbers of factors leads to multiple highly similar factors (such as separate factors with high loadings on sadness and with high loadings on disappointment) which runs counter to our purpose of recovering distinct focal emotion categories. Conversely, smaller numbers of factors lead to sets of vectors that are similar to subsets of the 5-factor analysis vectors.

supporting evidence for our focus on our four-emotion set: the first factor clearly corresponds to sadness, the third to caring and a broad set of positive emotions related to caring, the fourth (inversely) to overall emotional neutrality, and the fifth to fear. As such, we present this factor analysis as supporting empirical evidence that these emotion dimensions are indeed primary in our particular setting, and parsimoniously capture a majority of the important dimensions of emotion in these data.

That said, this analysis also highlights that a significant portion of the variation in our emotion metrics is explained by another factor corresponding to anger and anger-related emotions, represented in our table by Factor 2, for which we do not include any corresponding measure in our main analyses. As such, we present robustness tables here where we add anger as a fifth emotion category and re-run our baseline analyses with this category included. (We note as well that, as later presented in Appendix B, anger is not as reliably measured by our RoBERTa-based classifier, and so the below analysis comes with the caveat that effect estimates for anger may be especially attenuated by measurement error.)

Results are presented in Tables A7 and A8. Overall, the effects of anger are very noisily estimated and do not meaningfully alter the estimated effects from other emotion or emotion sequence categories; as anger is fairly infrequently observed in our context, while our factor analysis highlights that the anger-related factor can explain a degree of variation in the continuous emotion scores, the emotion rarely exceeds the 0.25 threshold for any given fundraiser in our data and does not appear to retain significant explanatory power for the outcomes that we inspect. Other results remain similar when adding anger as a fifth covariate category.

Table A6: Factor Loadings for 5 Factors

height	(1)	(2)	(3)	(4)	(5)
Admiration	-0.026	-0.027	0.595	-0.036	-0.112
Amusement	-0.013	0.029	0.006	-0.010	0.001
Anger	0.092	0.686	0.031	0.108	-0.051
Annoyance	0.069	0.978	-0.161	-0.080	-0.045
Approval	-0.111	-0.027	0.631	-0.290	-0.028
Caring	-0.035	-0.087	0.455	-0.062	-0.019
Confusion	-0.036	0.024	-0.040	0.004	0.282
Curiosity	-0.013	-0.015	-0.031	-0.030	0.099
Desire	-0.076	0.034	0.065	-0.174	-0.037
Disappointment	0.451	0.401	-0.152	-0.179	0.162
Disapproval	0.054	0.438	0.040	-0.156	-0.005
Disgust	0.179	0.340	-0.056	0.054	0.025
Embarrassment	0.004	0.235	-0.021	0.017	0.038
Excitement	-0.065	-0.026	0.092	0.003	0.050
Fear	0.009	0.062	0.045	0.065	0.701
Gratitude	-0.405	-0.158	-0.318	0.820	-0.180
Grief	0.829	-0.026	0.110	0.305	0.016
Joy	-0.011	0.034	0.331	0.076	0.021
Love	0.047	0.015	0.288	-0.077	0.007
Nervousness	0.160	0.124	0.029	0.043	0.975
Optimism	-0.219	-0.071	0.187	-0.043	-0.027
Pride	-0.067	0.038	0.624	0.124	-0.059
Realization	0.146	0.093	-0.003	-0.119	0.169
Relief	-0.008	-0.009	0.405	0.314	0.046
Remorse	0.202	0.043	-0.046	0.344	-0.028
Sadness	0.982	0.076	-0.120	0.056	0.081
Surprise	0.018	-0.008	0.010	0.004	0.054
Neutral	-0.049	0.054	-0.253	-0.700	0.010

Notes: Factor loadings from 5-factor analysis. Highlighted cells in yellow indicate high (> 0.4) positive loadings, while red highlights indicate high (< -0.4) negative loadings.

Table A7: Overall Emotions and Fundraiser Success:
5-Emotion Set Motivated by Factor Analysis

	Success (Logit)	Log Proportion Fundraised (OLS)
Valence	0.163** (0.058)	0.122*** (0.024)
Fear	-0.397 (0.477)	-0.511** (0.190)
Caring	0.225*** (0.060)	0.283*** (0.027)
Sadness	-0.009 (0.068)	-0.007 (0.030)
Neutral	0.026 (0.096)	-0.063 (0.041)
Anger	. (.)	1.262 (1.500)
Controls	X	X
Observations	14,113	14,114
R^2/Pseudo R^2	0.061	0.132

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. ". (.)" indicates variables omitted due to collinearity.

Table A8: Dynamic Emotion Measures and Fundraiser Success:
5-Emotion Set Motivated by Factor Analysis

	Success (Logit)	Log Proportion Fundraised (OLS)
1^{st} Half Valence	0.184*** (0.046)	0.122*** (0.020)
2^{nd} Half Valence	0.158* (0.078)	0.129*** (0.032)
1^{st} Half Neutral	-0.154* (0.062)	-0.057* (0.027)
2^{nd} Half Neutral	0.016 (0.101)	-0.065 (0.043)
1^{st} Half Fear	-0.128 (0.285)	-0.152 (0.128)
2^{nd} Half Fear	-0.817 (0.740)	-0.257 (0.254)
1^{st} Half Caring	-0.063 (0.089)	0.077† (0.040)
2^{nd} Half Caring	0.244*** (0.059)	0.257*** (0.026)
1^{st} Half Sadness	-0.017 (0.063)	0.048† (0.028)
2^{nd} Half Sadness	-0.082 (0.110)	-0.075 (0.047)
1^{st} Half Anger	. (.)	-0.372 (1.060)
2^{nd} Half Anger	. (.)	1.333 (1.498)
Controls	X	X
Observations	14,111	14,114
R^2/Pseudo R^2	0.064	0.134

Notes: Standard errors in parentheses. Controls include days posted, log fundraising goal, and a quadratic for wordcount. † = 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. ". (.)" indicates variables omitted due to collinearity.

Appendix B:

Validation of RoBERTa-Based Emotion Scorer

B.1 Validation Survey

In this section, we present validation results for our RoBERTa-based emotion scorer, which we use throughout the above analysis. For this exercise, we use human evaluations of emotion presence and intensity for 50 randomly-selected fundraisers from our sample, based on additional, emotion-focused questions in the same survey that we used to validate our persuasiveness metrics, presented in Section 4.2 in the main draft.

To gather benchmark emotions scores from human participants, we asked participants to rate how emotional the pitch was, and then to rate the intensity of each of our focal emotion categories of sadness, caring, neutral, and fear, as well as anger, another emotion highlighted in our factor analysis of the RoBERTa emotions. All of these ratings were on a scale of 1 (not at all) to 10 (very).

We randomly divided the 50 pitches into 5 blocks of 10. Each participant was randomly assigned to a block. This blocked design allows us to compute a measure of inter-rater reliability based on raters that all evaluated the same sets of pitches. 99 Prolific respondents completed the survey ($M_{age} = 36.7$, 67% female) and completed our attention check, resulting in approximately 20 raters for each 10-pitch block.

Table B1: Human-Scorer Correlation in Emotion Ratings
for 50 Randomly-Selected Fundraisers

	Pearson's r
Neutral	0.156
Fear	0.317
Sadness	0.158
Caring	0.507
Anger	0.081
Inter-Subject Correlation (Human-Human)	0.212

Notes: $N = 50$ fundraisers and $N_{raters} = 990$ human ratings; correlations computed from RoBERTa-based scores against the average of human ratings for each of these 50 fundraising pitches. Pearson's r values reflect the strength of correlation between human ratings and RoBERTa-based classifier scores. Inter-subject correlation calculated as the average inter-subject correlation at the level of each emotion, and then averaged across each emotion.

B.2 Human-Scorer Correlation & Human-Human Correlation

With this, we validate our RoBERTa-based scorer by estimating the correlation of the average human ratings of the 50 pitches against the RoBERTa scores. Table B1 presents the results.

Overall, we find positive, albeit somewhat low, correlations between average human ratings and the scores from our RoBERTa-based classifier. However, we also find that inter-subject correlations between humans is similarly low, if not lower for certain emotion categories. This suggests that our RoBERTa-based classifier is comparably performant as human encoders, achieving a similar level of consistency that one might expect from a human hand-coder.

Nonetheless, we note that this implies that our paper's results are likely to suffer from a degree of attenuation bias, as classical measurement error is likely to arise from this noisy emotion measurement. Moreover, we remark that this attenuation bias may differ across emotion categories—as shown above, anger is the least consistently classified by our RoBERTA-

based scorer, suggesting that attenuation bias may be especially prevalent for those effect estimates. Conversely, "caring" emotion is the most consistently classified, suggesting that our relatively strong effect estimates for "caring" in our field data analysis may arise in part because of less measurement error in that emotion's measurement.

Finally, these relatively low inter-subject correlations also motivates our use of relative rankings to validate the first-order effect of our rewrites, rather than direct scores; as even human ratings of emotion are shown to be fairly subjective in this context, we designed our primary survey to test for comparative changes in emotion by ranking rewrites and originals against one another to mitigate the potential influence of noise across subjects.

Appendix C:

Further Details on Rewriting Methodology

In this section, we provide further detail on our methodology for soliciting rewrites from online Prolific participants ("human" rewrites), ChatGPT ("LLM" rewrites) and online Prolific participants using ChatGPT ("human-in-the-loop" rewrites).

C.1 Details on Human Rewrites

To solicit human rewrites, we recruited online participants from Prolific and gave them the following prompt:

> For this survey, you are asked to rewrite the dominant emotional tone of an online fundraising pitch. This rewriting should preserve the content of the writing, but change the word choices, phrasings, and emphases in order to strongly convey and express different emotions in the fundraising pitch. The objective is to rewrite the fundraising pitch as if you're the original author, writing about the same original issue, but in a very different mood and with very different tones, making different stylistic choices for what type of emotion to convey and express.
>
> For example, for the following fundraising pitch...
>
> [Here we placed an example fundraising pitch and an example rewrite towards "anger", a non-focal emotion in our study, to demonstrate what we intend participants to accomplish.]
>
> Using the same content and sentences as a basis, we've rewritten the pitch to be predominantly angry in tone, using different wordings, formatting, and punctuation, as well as different metaphors in certain places ([example phrase 1] –> [example phrase 1 rewritten]). We also carefully changed a small handful of the more pro-forma phrases that highlight other emotions ([example phrase 2]) into angrier pro-forma phrases ([example phrase 2 rewritten]) that perform the same 'function' of a final exhortation at the end of the pitch. As stated above, the objective here is to write the fundraising pitch as if you're the same person as the original author, writing about the same issue, but in a very different style.

Above, we offer an example baseline pitch and example rewrite to ensure clarity in our intentions, but focus on "anger", an emotion that is never a focal emotion of our rewrites, to ensure that we're not inadvertently priming participants to rewrite a particular emotion in the way of our example.

We then show the participant the randomly-selected fundraising pitch that they have been assigned to rewrite with the following prompt:

> Rewrite the provided pitch to convey different dominant emotional tones in what you consider to be appropriate but approximate halves of the fundraising pitch, with roughly the same number of sentences in each half.
>
> Strongly convey the emotion of [focal emotion in first half] in the first half (try to use emotional language of [focal emotion in first half], specifically– not language of [other focal emotion 1] or [other focal emotion 2]).
>
> Strongly convey the emotion of [focal emotion in second half] in the second half (try to use emotional language of [focal emotion in second half], specifically– not language of [other focal emotion 1] or [other focal emotion 2]).
>
> This rewriting should preserve the content of the writing, but change the word choices, wordings, phrasings and emphases in order to strongly convey [focal emotion in first half] and [focal emotion in second half] in the appropriate sections of the fundraising pitch. The objective is to rewrite the fundraising pitch as if you're the same person as the original author, writing about the same original issue, but in a very different mood and with very different tones, making different stylistic choices for what type of emotion to express and convey. Try to keep the length approximately the same, within 90% to 110% of the original length.
>
> The baseline fundraising pitch is copied below:
>
> [Randomly-selected fundraising pitch copied here.]

Above, [focal emotion in first half] is replaced by the respective focal emotion of that half, i.e. "sadness" for $sadness \rightarrow caring$; [focal emotion in second half] is replaced by the respective focal emotion of that half, i.e. "caring" for $sadness \rightarrow caring$; [other focal emotion 1] and [other focal emotion 2] are replaced by the other focal emotions that we do not wish

the rewrite to increase (i.e. "fear" or "caring", if the focal emotion in the first half is "sadness", and so forth).

We randomly assign fundraising pitches to rewriters for our $10 \times 4 = 40$ randomly-selected pitches that do not contain the respective focal emotion sequence in the original pitch, gathering an average of approximately 2 rewrites for each original pitch. Our research assistant then examined rewrites for "good faith" completion, discarding rewrites that were either very short or featured very high rates of typos or difficult-to-read sentences, and selected the best rewrite available out of those collected for each original pitch. As noted in the main text, this final selection step means that our results for human rewrites should not be interpreted as the effects from rewrites of randomly-selected human rewriters, but rather as the effects of rewrites of human rewriters from approximately the top 50% of the rewriting skill distribution among Prolific participants who enrolled in our study.

Finally, we also note that participants, in their consent form, are asked to agree to the following statement:

> Use of Outside Writing Tools: By agreeing to this survey, you also agree to fill out this survey without the assistance of outside writing tools or generative AI. You commit to producing writing responses in text boxes that are your own produced writing.

As such, we believe that our human rewrites do not include LLM usage; the differences between these rewrites and our later LLM and LLM-assisted rewrites further corroborates that participants likely did not use LLMs to perform these rewrites.

C.2 Details on LLM (ChatGPT) Rewrites

For our LLM rewrites, we use the same baseline fundraising pitch with a chain-of-thought prompt to recover single LLM rewrites for each of our $10 \times 4 = 40$ randomly-selected original pitches. We relied on the "gpt-4o" model.

Our exact prompt, embedded in the Python function that we used for implementation, is displayed in Figure C1.

Figure C1: Python function and prompt for rewriting fundraising pitches to different emotions

```
def change_both_emotion(text, emotion1, emotion2):
    """Changes the emotional tone of a fundraising pitch into two parts.

    Args:
        text (str): The original text of the fundraising pitch.
        emotion1 (str): The emotion to convey in the first half.
        emotion2 (str): The emotion to convey in the second half.

    Returns:
        str: The rewritten fundraising pitch with the specified emotional tones.
    """
    completion = client.chat.completions.create(
        model="gpt-4o",
        messages=[
            {
                "role": "system",
                "content": "You are an expert fundraising writer with a deep
                    understanding of emotional psychology, skilled at crafting
                    compelling and authentic pitches that elicit strong emotional
                    responses."
            },
            {
                "role": "user",
                "content": f"""I need you to rewrite the fundraising pitch
                    provided between the triple backticks, following the
                    instructions below.

                    Instructions:

                    - Rewrite the pitch, dividing it into two approximately equal
                      halves (either by sentence count or word count).
                    - In the first half, **strongly convey the emotion of {
                      emotion1}**.
                    - In the second half, **strongly convey the emotion of {
                      emotion2}**.
                    - Preserve the core message and content of the original pitch,
                       but adjust word choices, phrasing, and emphasis to reflect
                       the specified emotions.
                    - Use vivid and emotive language to enhance the emotional
                      impact.
                    - Ensure the tone and style remain authentic, as if written by
                       the original author.
                    - Keep the overall length within 90% to 110% of the original.
                    - **Do not include any explanations or additional commentary;
                      provide only the rewritten pitch.**
                    - Your goal is to create a compelling and authentic
                      fundraising pitch that elicits strong emotional responses
                      appropriate to the specified emotions, encouraging readers
                      to donate.

                    Input Pitch:
                    ```{text}```
                """
            }
        ],
        temperature=0.4,
        seed=42,
    )
    return completion.choices[0].message.content
```
```

We then use this single rewrite as our LLM rewrite for each respective fundraising pitch.

C.3 Details on Human-in-the-Loop Rewrites

Our human-in-the-loop rewrites followed the same procedure as our human rewrites, detailed in Appendix Section C.1 above, except for a single deviation. For the baseline layout of the prompt and methodology, see that section above.

To that prior procedure we simply add, after presenting the baseline original fundraising pitch, the following text:

> A starter rewrite from ChatGPT is copied below that you may use as a starting-place, template or guide, if helpful:
>
> [GPT rewrite of respective fundraising pitch copied here.]

Note that this comes after this message to participants that they were required to agree to in their consent form:

> Use of Outside Writing Tools: By agreeing to this survey, you also agree to fill out this survey without the assistance of outside writing tools or generative AI. You commit to producing writing responses in text boxes that are your own produced writing.

so we assert that participants only used LLMs to the extent of referencing our LLM rewrite in the above block of text.

After collecting these rewrites, our research assistant then selected the best human-in-the-loop rewrite, if multiple valid rewrites were available, for use in the final persuasiveness evaluation. As above, this is identical to the procedure for the human rewrites detailed in Appendix Section C.1.

C.4 Rewrite Validation

We remark here that we are using ChatGPT in our rewrites, and one of the most salient known issues with large language models is the tendency to "hallucinate" details (Buchanan, Hill, and Shapoval, 2023). This issue tends to be most salient when creating original text,

and as such we expect that this issue may not be nearly as problematic for rewrites, when the language model is explicitly given baseline text to adjust and edit; however, to ensure that this is not an issue, we also perform an ancillary post-validation exercise to inspect whether the salient information in the fundraising pitch changes across different rewriting modes, and whether this impacts our final estimates. This validation study was pre-registered as AsPredicted #188563, available at `https://aspredicted.org/txvz-y3np.pdf`. For every rewrite, we present the original fundraising pitch to online participants and then present the rewritten fundraising pitch, described as "a new version, rewritten by a human" (for all pitches, including those not written by humans), with pitch rated by at least 3 online participants. We then asked participants to rate the validity of the fundraising pitch in terms of the following questions:

- Did the rewrite add any salient information or details that were not present in the original?
- Is the rewrite missing any salient information or details from the original?

With these questions, we seek to systematically evaluate any changes to the factual presentation in the rewrites, either in the form of missing information or added information, especially in LLM or LLM-assisted rewrites. We then test directly whether rewrites with missing or added salient info may have impacted our results by dropping from our sample any rewrites that are flagged by a majority of online participants to have either added or missing salient information, and rerunning our analyses with these potentially-problematic rewrites removed.

Results are presented in Tables C1, C2 and C3. We find nearly identical first-order effects of rewriting on emotion ranking, but a number of the significant effects for solely-LLM rewrites on how "moved" participants are drops to statistical insignificance when removing any rewrites flagged as missing or adding salient information, although effects are largely robust in regressions against the "likely to donate" outcome with the exception of the LLM rewrites of the *caringcaring* arc. Results for human-in-the-loop rewrites are in some cases (*caring* $\rightarrow$ *caring*) even larger and more significant when dropping rewrites with salient

Table C1: Rewrite Type and Focal Emotion Rank:
Excluding Rewrites with Salient Information Changes

	Focal Emotion Rank	
	(OLS)	(Ordinal Probit)
Human-in-the-Loop Rewrite	-0.772*** (0.063)	-0.808*** (0.068)
GPT Rewrite	-0.767*** (0.065)	-0.782*** (0.069)
Human Rewrite	-0.444*** (0.062)	-0.437*** (0.065)
Observations	2397	2397
R^2 / Log-Likelihood	0.099	-3193.2

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

information either added or missing, and in other cases are highly similar (significant and positive for $sadness \rightarrow caring$, null for $sadness \rightarrow neutral$). We see a modest change in the measured effect of the $neutral \rightarrow fear$ emotion sequence for human-in-the-loop rewrites on the "moved" outcome, but continue to see significant positive effects on the "likely to donate" outcome.

Overall, we suggest that this evidence strongly supports the validity of our estimated effects for human-in-the-loop rewrites in particular. We do find evidence that some our LLM-rewrite effects on how "moved" participants are may be partly driven by salient information that has been added into the rewrite that was not present in the original pitch, but we recover highly similar effects on how "likely to donate" participants are, again particularly for our human-in-the-loop rewrites. This suggests to us that our human-in-the-loop approach is the most reliable for researchers when implementing our technique.

Table C2: Effect of Emotion Arc Rewrites on Fundraiser Quality:
Excluding Rewrites with Salient Information Changes

	How moved were you by... ?
$Caring \rightarrow Caring$: Human-in-the-Loop	0.743^{***} (0.189)
$Caring \rightarrow Caring$: GPT	$0.411^{\dagger}$ (0.240)
$Caring \rightarrow Caring$: Human	0.172 (0.186)
$Neutral \rightarrow Fear$: Human-in-the-Loop	0.118 (0.221)
$Neutral \rightarrow Fear$: GPT	$0.361^{\dagger}$ (0.207)
$Neutral \rightarrow Fear$: Human	-0.678^{*} (0.278)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.423^{**} (0.163)
$Sadness \rightarrow Caring$: GPT	0.344^{*} (0.166)
$Sadness \rightarrow Caring$: Human	-0.101 (0.223)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	-0.165 (0.151)
$Sadness \rightarrow Neutral$: GPT	0.251 (0.179)
$Sadness \rightarrow Neutral$: Human	-0.282 (0.175)
Fundraiser Number Fixed Effects	X
Observations	2397
R^2	0.053

Notes: Standard errors in parentheses, clustered by participant. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

Table C3: Effect of Emotion Arc Rewrites on Fundraiser Quality
Excluding Rewrites with Salient Information Changes

	How likely would you be to donate... ?
$Caring \rightarrow Caring$: Human-in-the-Loop	0.741^{***} (0.167)
$Caring \rightarrow Caring$: GPT	0.461^{*} (0.206)
$Caring \rightarrow Caring$: Human	0.401^{*} (0.203)
$Neutral \rightarrow Fear$: Human-in-the-Loop	$0.295^{\dagger}$ (0.176)
$Neutral \rightarrow Fear$: GPT	0.554^{*} (0.234)
$Neutral \rightarrow Fear$: Human	-0.289 (0.298)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.392^{*} (0.157)
$Sadness \rightarrow Caring$: GPT	0.369^{*} (0.145)
$Sadness \rightarrow Caring$: Human	-0.041 (0.194)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	-0.164 (0.123)
$Sadness \rightarrow Neutral$: GPT	$0.329^{\dagger}$ (0.181)
$Sadness \rightarrow Neutral$: Human	0.076 (0.182)
Fundraiser Number Fixed Effects	X
Observations	2397
R^2	0.031

Notes: Standard errors in parentheses, clustered by participant. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

Appendix D:

Further Details on Mechanism Study and Mediation

In this section, we provide additional details on our follow-up mechanism study. As noted in the main paper, this study mirrored our main study except in that we only tested originals and human-in-the-loop rewrites, and we supplemented our original persuasion-metric outcomes with a set of questions derived from prior literature on the psychological mechanisms of persuasive narratives, detailed below.

D.1 "Narrative Transportation" and "Identification" Instruments

To measure "narrative transportation", we adapted an instrument of ten questions derived directly from the foundational prior work on this psychological mechanism, Green and Brock (2000), for the context of online fundraising. The questions we asked survey participants were as follows:

1. While I was reading the fundraiser, I could easily picture the events in it taking place.
2. While I was reading the fundraiser, activity going on in the room around me was on my mind. (reverse-coded)
3. I could picture myself in the scene of the events described in the fundraiser.
4. I was mentally involved in the fundraiser while reading it.
5. After finishing the fundraiser, I found it easy to put it out of my mind. (reverse-coded)
6. I wanted to learn how the fundraiser ended.
7. The fundraiser affected me emotionally.
8. I found myself thinking of ways the fundraiser could have turned out differently.
9. I found my mind wandering while reading the fundraiser. (reverse-coded)
10. The events in the fundraiser are relevant to my everyday life.

To measure "identification", we adapted an instrument of 6 questions derived directly from the foundational prior work on this psychological mechanism, Cohen (2001), for the context of online fundraising. The questions we asked survey participants were as follows:

1. I was able to understand the events in the fundraiser in a manner similar to the way [the protagonist] understood them.
2. I think I have a good understanding of [the protagonist].
3. While reading the fundraiser, I forgot myself and was fully absorbed.
4. While reading the fundraiser, I could feel the emotions [the protagonist] portrayed.
5. At key highlights in the fundraiser, I felt I knew exactly what [the protagonist] was going through.
6. While reading the fundraiser, I wanted [the protagonist] to achieve their goal.

where [the protagonist] is replaced by the subject (or "main character") of the given fundraising pitch.

D.2 Full Set of Mediation Analyses

In this section, we present mediation analyses for all four focal fundraising arcs, in addition to the mediation analysis for *sadness* $\rightarrow$ *caring* that we presented in Section 6. Results are presented in Tables D1, D2, D3, and D4. As noted in Section 6, results for *sadness* $\rightarrow$ *caring* are dominantly mediated by mechanism of identification with the protagonist of the fundraiser, with roughly 70% of the effect estimated to be driven by the indirect effect through identification. Similarly, for all other emotion arcs, point estimates suggest a highly similar story, with around 70%-75% of the effect being driven by identification with the protagonist(s) of the fundraiser, although results are insignificant. We remark that while we only find significant effects for our largest emotion sequence effect, *sadness* $\rightarrow$ *caring*, in this smaller follow-up study, our sample size is roughly 25% in this follow-up as compared to our main study; and although insignificant, the point estimates of the total effects that we recover for the other arcs are within the confidence intervals we estimate in the main study, and vice versa.

Table D1: Mediation Analysis: Sadness → Caring Effects

Indirect Effects	
Transportation	0.030 (0.018)
Identification	0.543 (0.128)***
Total Indirect	0.573 (0.139)***

Direct and Total Effects	
Direct Effect	0.275 (0.192)
Total Effect	0.848 (0.260)**

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Results shown for *sadness* → *caring* emotion sequence effects.

Table D2: Mediation Analysis: Neutral → Fear Effects

Indirect Effects	
Transportation	0.015 (0.015)
Identification	0.136 (0.168)
Total Indirect	0.151 (0.179)

Direct and Total Effects	
Direct Effect	0.084 (0.183)
Total Effect	0.235 (0.262)

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Results shown for *neutral* → *fear* emotion sequence effects.

Taken together, we interpret these results as converging evidence that identification with the protagonist(s) of the given fundraiser is the primary psychological mechanism of the observed total emotion sequence effects.

Table D3: Mediation Analysis: Caring → Caring Effects

Indirect Effects	
Transportation	-0.003 (0.011)
Identification	0.076 (0.141)
Total Indirect	0.073 (0.150)

Direct and Total Effects	
Direct Effect	0.065 (0.172)
Total Effect	0.137 (0.240)

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Results shown for *caring* → *caring* emotion sequence effects.

Table D4: Mediation Analysis: Sadness → Neutral Effects

Indirect Effects	
Transportation	0.018 (0.016)
Identification	0.123 (0.172)
Total Indirect	0.142 (0.184)

Direct and Total Effects	
Direct Effect	0.033 (0.154)
Total Effect	0.175 (0.269)

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Results shown for *sadness* → *neutral* emotion sequence effects.

Appendix E:

Further Details on Persuasiveness Evaluation Results

In this section, we provide further details and documentation on the final persuasiveness evaluations of our rewritten fundraising pitches. All of the below relates to the persuasiveness survey pre-registered at `https://aspredicted.org/ycn9-kfzk.pdf`.

E.3 Details on Emotion Ranking Question

To evaluate the comparative shifts in emotion, we rely on a ranking-based approach, where we present participants with both the original fundraising pitch and each of the 3 rewrites (human, LLM, human-in-the-loop) and ask them to rank them in terms of their comparative levels of the focal emotion. Specifically, we ask the following question:

> Rank the fundraising pitches below based on how strongly they convey [focal emotion], with 1 being the [most focal emotion] and 4 being the [least focal emotion].

For $caring \rightarrow caring$, $neutral \rightarrow fear$, and $sadness \rightarrow neutral$, we ask only about the single non-neutral emotion; for $sadness \rightarrow caring$, we ask two ranking questions, one for sadness and one for caring. For our regressions, we then average the ranking of sadness and caring for each given participant $\times$ fundraising pitch set, to maintain equal weighting where each participant $\times$ fundraising pitch set translates to a single observation.

We rely on this ranking-based approach because our earlier validation of emotion ratings, presented in Appendix Section B.2, shows a high degree of noise across human emotion ratings, with considerable subjectivity in how humans rate the intensity of different emotions on 1-10 scales; by asking participants to rank emotions of the original and each rewrite, we are able to rigorously establish whether the rewrites were baseline successful in shifting the story's emotion while minimizing the measure's exposure to inconsistency of emotion

Table E5: Rewrite Type and Focal Emotion Rank

	Focal Emotion Rank (OLS)	Focal Emotion Rank (Ordinal Probit)
Human-in-the-Loop Rewrite	-0.779*** (0.052)	-0.803*** (0.057)
GPT Rewrite	-0.778*** (0.053)	-0.783*** (0.058)
Human Rewrite	-0.370*** (0.052)	-0.384*** (0.054)
Observations	4148	4148
R^2 / Log-Likelihood	0.090	-5553.3

Notes: Standard errors in parentheses, clustered by participant. Sample includes partial responses. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance.

perceptions across subjects. We ask about emotions overall, rather than half-by-half, in order to maintain simplicity in the question and evaluation and avoid confusion among participants; since the rewrites were built to shift emotion in specific halves, we interpret these results as arising from shifts in respective halves.

Finally, we estimate the first-order effects of rewrites on the density of respective emotions in the pitches by regressing indicators for rewrite mode against the rank, compared to the excluded category of the original pitch. Results are presented in the main draft in Table 6 and Figure 1.

E.4 Rewrite Effect Analyses Including Partial Responses

In our main analyses, we present results from our sample only including observations for pitches that have complete responses to both our persuasiveness metrics and ranking. As we have somewhat high rates of partial responses, we here present results from regressions that retain partial responses in the sample, as long as they have the requisite data on the focal outcome of the given analysis.

Results on emotion rank effects are presented in Table E5; results are qualitatively identical. Results on persuasiveness are presented in Table E6 and Table E7. Sample sizes

Table E6: Effect of Emotion Arc Rewrites on Fundraiser Quality

	How moved were you by... ?
$Caring \rightarrow Caring$: Human-in-the-Loop	0.125 (0.133)
$Caring \rightarrow Caring$: GPT	0.416** (0.134)
$Caring \rightarrow Caring$: Human	0.136 (0.159)
$Neutral \rightarrow Fear$: Human-in-the-Loop	0.306** (0.112)
$Neutral \rightarrow Fear$: GPT	0.462*** (0.116)
$Neutral \rightarrow Fear$: Human	-0.184 (0.152)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.731*** (0.117)
$Sadness \rightarrow Caring$: GPT	0.718*** (0.129)
$Sadness \rightarrow Caring$: Human	-0.548*** (0.137)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	0.116 (0.109)
$Sadness \rightarrow Neutral$: GPT	0.171 (0.111)
$Sadness \rightarrow Neutral$: Human	-0.187 (0.137)
Fundraiser Number Fixed Effects	X
Observations	4965
R^2	0.057

Notes: Standard errors in parentheses, clustered by participant. Sample includes partial responses. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

are slightly larger and so effect estimates are slightly more precise and stronger, but otherwise effects are qualitatively very similar. We do find a significantly positive effect of LLM rewrites with *sadness* $\rightarrow$ *neutral* on the likelihood to donate, but null effects on how moved participants report they are from reading the fundraising pitch.

Table E7: Effect of Emotion Arc Rewrites on Fundraiser Quality

	How likely would you be to donate... ?
$Caring \rightarrow Caring$: Human-in-the-Loop	0.258^{*} (0.121)
$Caring \rightarrow Caring$: GPT	0.610^{***} (0.157)
$Caring \rightarrow Caring$: Human	0.392^{*} (0.173)
$Neutral \rightarrow Fear$: Human-in-the-Loop	0.416^{***} (0.111)
$Neutral \rightarrow Fear$: GPT	0.612^{***} (0.159)
$Neutral \rightarrow Fear$: Human	0.058 (0.168)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.519^{***} (0.112)
$Sadness \rightarrow Caring$: GPT	0.683^{***} (0.120)
$Sadness \rightarrow Caring$: Human	-0.423^{***} (0.124)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	0.087 (0.093)
$Sadness \rightarrow Neutral$: GPT	0.346^{*} (0.144)
$Sadness \rightarrow Neutral$: Human	0.031 (0.153)
Fundraiser Number Fixed Effects	X
Observations	4835
R^2	0.042

Notes: Standard errors in parentheses, clustered by participant. Sample includes partial responses. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

E.5 Effects of Rewrites on Other Persuasiveness Measures

Throughout our persuasiveness evaluation, we focus on the two online participant evaluation metrics that we successfully validate against real-world field data outcomes: how "moved" participants report being, and how "likely to donate" participants report they are to a given fundraising pitch, which we show in the main draft, in Table 5, is significantly correlated with actual fundraising success outcomes in a randomly-selected sample of 50 real GoFundMe.com medical fundraisers.

In this section, we present extension analyses of the effects of rewriting, across different modes and emotion sequences, on our remaining metrics of persuasiveness: how "well-written" the fundraising pitch is, how "authentic" the fundraising pitch is, and how "convinced" participants are by the fundraising pitch. For this analysis, as in Appendix Section E.2, we use all data with responses for these particular responses.

Results for "convinced" are presented in Table E8. Results for "authentic" are presented in Table E9. Results for "well-written" are presented in Table E10. Effects are consistent with our main analyses for "convinced" and "well-written", but are slightly weaker for the "authentic" outcome; for this last outcome, we find weaker positive effects from LLM rewrites, but some null effects from human-in-the-loop effects—with the important exception of *sadness* $\rightarrow$ *caring*, which remains highly significant. Overall, we find strong consistency in our pattern of results across both our primary, focal outcomes and these ancillary outcome metrics.

Table E8: Effect of Emotion Sequence Rewrites on Fundraiser Quality

	Alternative Outcome: "Convinced"
$Caring \rightarrow Caring$: Human-in-the-Loop	0.075 (0.128)
$Caring \rightarrow Caring$: GPT	0.376** (0.128)
$Caring \rightarrow Caring$: Human	-0.717*** (0.194)
$Neutral \rightarrow Fear$: Human-in-the-Loop	0.217† (0.121)
$Neutral \rightarrow Fear$: GPT	0.483*** (0.121)
$Neutral \rightarrow Fear$: Human	-0.603*** (0.159)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.571*** (0.118)
$Sadness \rightarrow Caring$: GPT	0.641*** (0.125)
$Sadness \rightarrow Caring$: Human	-0.679*** (0.139)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	-0.069 (0.105)
$Sadness \rightarrow Neutral$: GPT	0.064 (0.114)
$Sadness \rightarrow Neutral$: Human	-0.750*** (0.149)
Fundraiser Number Fixed Effects	X
Observations	4965
R^2	0.077

Notes: Standard errors in parentheses, clustered by participant. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

Table E9: Effect of Emotion Sequence Rewrites on Fundraiser Quality

	Alternative Outcome: "Authentic"
$Caring \rightarrow Caring$: Human-in-the-Loop	-0.168 (0.132)
$Caring \rightarrow Caring$: GPT	0.168 (0.123)
$Caring \rightarrow Caring$: Human	-0.174 (0.153)
$Neutral \rightarrow Fear$: Human-in-the-Loop	-0.018 (0.108)
$Neutral \rightarrow Fear$: GPT	0.281^{*} (0.110)
$Neutral \rightarrow Fear$: Human	-0.449^{**} (0.148)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.420^{***} (0.114)
$Sadness \rightarrow Caring$: GPT	0.413^{***} (0.124)
$Sadness \rightarrow Caring$: Human	-0.663^{***} (0.142)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	-0.214^{*} (0.105)
$Sadness \rightarrow Neutral$: GPT	-0.049 (0.111)
$Sadness \rightarrow Neutral$: Human	-0.350^{*} (0.138)
Fundraiser Number Fixed Effects	X
Observations	4965
R^2	0.049

Notes: Standard errors in parentheses, clustered by participant. * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.

Table E10: Effect of Emotion Sequence Rewrites on Fundraiser Quality

	Alternative Outcome: "Well-Written"
$Caring \rightarrow Caring$: Human-in-the-Loop	0.158 (0.136)
$Caring \rightarrow Caring$: GPT	0.667^{***} (0.134)
$Caring \rightarrow Caring$: Human	0.067 (0.177)
$Neutral \rightarrow Fear$: Human-in-the-Loop	0.636^{***} (0.134)
$Neutral \rightarrow Fear$: GPT	0.884^{***} (0.130)
$Neutral \rightarrow Fear$: Human	-0.129 (0.167)
$Sadness \rightarrow Caring$: Human-in-the-Loop	0.772^{***} (0.124)
$Sadness \rightarrow Caring$: GPT	0.923^{***} (0.128)
$Sadness \rightarrow Caring$: Human	-0.593^{***} (0.139)
$Sadness \rightarrow Neutral$: Human-in-the-Loop	$0.197^{\dagger}$ (0.117)
$Sadness \rightarrow Neutral$: GPT	0.474^{***} (0.122)
$Sadness \rightarrow Neutral$: Human	-0.174 (0.148)
Fundraiser Number Fixed Effects	X
Observations	4965
R^2	0.064

Notes: Standard errors in parentheses, clustered by participant. †= 0.10 significance, * = 0.05 significance, ** = 0.01 significance, *** = 0.001 significance. Fundraiser number fixed effects included but not shown.