An Individualized News Affective Response Dataset

Anonymous ACL submission

Abstract

The rise of sensationalism in news reporting, driven by market saturation and online competition, has compromised news quality and trust. At the core of sensationalism is the evocation of affective responses in the readers. Current NLP approaches to emotion detection often overlook the subjective differences in groups and individuals, relying on aggregation techniques that can obscure nuanced reactions. We introduce a novel large-scale dataset capturing subjective affective responses to news headlines. The dataset includes Facebook post screenshots from popular UK media outlets and uses a comprehensive annotation scheme. Annotators report their affective responses, provide discrete emotion labels, assess relevance to current events, and indicate sharing likelihood. Additionally, we collect demographic, personality, and media consumption data. This ongoing dataset aims to enable more accurate models of affective response by considering individual and contextual factors. This work is ongoing and we highly appreciate any feedback.

1 Introduction

The saturation of the traditional media market and increased competition in the online space have led to a rise in sensationalism in news reporting, appealing to readers' emotions to maximize click rate and sharing online (Kleemans and Hendriks Vettehen, 2009). This leads to a deterioration of news quality (Wang, 2012), a distorted perception of the state of the world among the public (Boyer, 2023), and declining trust in the news industry (Kleemans et al., 2017).

While often framed as an objective characteristic of news content and form (Kleemans and Hendriks Vettehen, 2009; Arbaoui et al., 2020), sensationalism is fundamentally about eliciting an **affective response** from the audience. This inherent subjectivity, akin to other psychological concepts, is influenced by a complex interplay of individual and group-level factors. Research on differential media effects demonstrates how diverse audiences, shaped by factors such as demographics, personality traits, and cultural backgrounds, respond to media content in distinct ways (Oliver, 2002; Valkenburg and Peter, 2013; Soroka et al., 2019). This variability in affective responses is further supported by emotion research highlighting the significant influence of individual characteristics like age, gender, and personality, alongside group-level variables like culture, on everyday emotional experiences (Kring and Gordon, 1998; Costa and McCrae, 2008; Charles and Carstensen, 2010; Mesquita and Frijda, 1992). Therefore, assessing sensationalism solely based on content analysis on the emotion used in the news, without accounting for the audience's subjective experience and individual differences, risks a simplistic and potentially inaccurate understanding of the phenomenon.

Numerous NLP studies aim to measure emotion in text, yet many fail to explicitly consider the perspective of the analysis (e.g., writer vs. reader) and rely on aggregation techniques like majority voting or averaging for annotation labels. However, research on subjectivity in NLP annotations, emphasizes the inherent subjectivity of these constructs (Ovesdotter Alm, 2011; Plank, 2022; Cabitza et al., 2023). Aggregating subjective responses without acknowledging individual variability and potential biases in perception risks obscuring nuanced emotional reactions and generating potentially misleading conclusions.

To address these limitations, we introduce a novel large-scale dataset focused on capturing the inherent subjectivity of affective responses to news content. Our dataset consists of screenshots from publicly available Facebook posts by the most popular UK media outlets (see Appendix for a full list). We employ a multi-faceted annotation scheme, requiring annotators to: (1) report their affective response using the valence-dominance-arousal frame-

Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), pages 555–563
August 11-16, 2024

work (Mehrabian and Russell, 1974), (2) provide discrete emotion annotations based on Plutchik's eight basic emotions (Plutchik, 1980), (3) assess the relevance of the post to current events, and (4) indicate their likelihood of sharing the post. Furthermore, we collect a comprehensive set of covariates for each annotator, encompassing demographic information, personality traits, and media consumption habits. This rich dataset will enable the development of more nuanced and accurate models of affective response to news, taking into account both individual differences and contextual factors. This dataset collection effort is still ungoing.

2 Related Works

2.1 News and Emotion

While news content often leans negative, eliciting negative emotions and heightened arousal in readers (Soroka et al., 2019), individual responses can vary significantly based on demographics, personality, and other background factors (Oliver, 2002; Valkenburg and Peter, 2013; Soroka et al., 2019). This is crucial as emotional reactions to news can profoundly influence perception, cognition, and behavior. Affect, for instance, provides evaluative feedback on one's thoughts and inclinations, shaping reasoning and decision-making (Storbeck and Clore, 2008)]. Existing research on news perception predominantly focuses on the emotional tone of the news itself, rather than the emotions evoked in individual readers (de Hoog and Verboon, 2020). To address this gap, this work shifts perspective and introduces a large-scale dataset designed to analyze how diverse individuals emotionally respond to different news headlines.

2.2 Emotion Detection in NLP

Emotion detection has been a core task in NLP for nearly two decades (Strapparava and Mihalcea, 2007). Recent years have seen a large number of valuable resources on the task (see Demszky et al. (2020); Oberländer et al. (2020) for a overview). These efforts have significantly advanced the field, leading to more accurate and robust emotion detection systems.

However, most existing datasets rely on aggregated "gold labels", overlooking the inherent subjectivity and variation in human emotional perception (Ovesdotter Alm, 2011; Plank, 2022; Cabitza et al., 2023). Ample research demonstrates the impact of both individual characteristics (e.g., age, gender, personality) and group-level factors (e.g., culture) on how we perceive and interpret emotions (Kring and Gordon, 1998; Costa and McCrae, 2008; Charles and Carstensen, 2010; Mesquita and Frijda, 1992), most existing datasets rely on aggregated "gold labels." This approach, while simplifying annotation, overlooks the genuine variation and subjectivity inherent in human emotional responses (Ovesdotter Alm, 2011; Plank, 2022; Cabitza et al., 2023). Consequently, models trained on such data may struggle to capture the nuanced ways in which emotions are expressed and understood.

Limited attempts have been made to incorporate annotator information. For instance, Diaz et al. (2018) provides demographic data alongside sentiment annotations. However, this dataset only contains sentiment annotation, is restricted to a specific online community, and is thus unsuitable for our purpose.

3 Dataset Collection Protocol

Recognizing the limitations of existing emotion detection datasets, we develop a novel data collection protocol aimed at capturing individualized affective response to news headlines.

We first collect a selection Facebook news posts from a list of major UK news outlets from April 1 to April 20, 2024, using CrowdTangle. While acknowledging that social media content may not fully represent the entirety of a news outlet's output, we posit that the posts chosen for these platforms reflect the outlets' editorial decisions and public image. Typically, these posts consist of an image, a short description, and the headline, with the image linking to the full news article. An example can be seen in Figure 3. To ensure ecological validity and minimize bias, we took screenshots of the news posts, capturing the reaction counts while any comment information. These screenshots were then presented to the annotators.

We recruit our annotators from Prolific. We have around 5 annotators for each headlines. We make sure of features such as stratified sampling to ensure a balanced set of annotators in terms of gender, age and political learning. In total, each annotators annotator around 50 headlines and the two stage combined take around 45 minutes. We therefore pay the annotators £8.58, in accordance with the National Living Wage.

Our annotation process involved two stages:

Stage 1: Covariate Collection

In this initial stage (implemented in Qualtrics), we gather essential background information (which we will refer to as persona variables henceforth) about annotators. This includes:

- Demographics (age, gender, education, income level etc.)
- Ideology
- Questions about news consumption habits (e.g. How often do you fact-check news stories you come across; Which of the following platforms do you use for news nowadays)
- Trust in major UK news outlets: To gauge how trust in news sources (and hence as a proxy of consumption) might affect perception
- A short version of the Cognitive Reflection Test (Frederick, 2005): to measure the tendency to engage in reflective thinking versus intuitive thinking
- The Ten-Item Personality Measure (Gosling et al., 2003): To capture basic personality traits that may influence annotation behavior
- Selected questions from the Perth Emotional Reactivity Scale (Preece et al., 2018): To assess emotional reactivity which could affect judgment.
- Selected questions from the Positive and Negative Affect Schedule (Crawford and Henry, 2004): To evaluate the annotators' current affective state and its potential influence on their annotations.

We also present the annotation guideline[1], which are adapted from the seminal work of Bradley and Lang (2007), to the annotators at this stage but they always have access to it in the second stage as well.

Stage 2: Headline Annotation

We then present the screenshots to the annotators with a website built on top of the the Potato annotation tool (Pei et al., 2022). For each screenshot, we ask the annotators to rate the valence, arousal and dominance they feel after reading the headline using the validated Self-Assessment Manikin (Bradley and Lang, 1994). We also ask the annotators to rate the discrete emotion categories based on Plutchik's eight basic emotions (Plutchik, 1980). This is because existing work have been using both and we would like to have a dataset that is comparable to either. We also ask the annotators the following three additional questions:

1. When considering your emotional reaction to this Facebook post, which element do you feel has the most influence?
2. Considering your personal experiences, interests, and the context of your life, how relevant do you find the following headline? Please select the option that best reflects your opinion.
3. Imagine you are seeing this headline for the first time on social media. How likely are you to share this news with others (e.g., through social media, messaging apps, or in person)? Please select the option that best reflects your opinion.

4 Preliminary Results

We have annotated 1,102 instances using a total of 113 annotators, averaging 5.27 annotations per sample.

Distribution of Annotators We show the distribution of our annotators among key persona variables in Table 1. Our data has a broad coverage in terms of the key persona variables listed.

Distribution of Annotations We present the distribution of the annotation variables we collect for each headline in Figure 1. In Figure 1a, 1b,and 1c, we observe that the neutral value of 4 is the most common for valence, arousal and dominance. As anticipated, the valence scores tend to skew negatively, arousal scores are predominantly high, and dominance scores skew slightly low.

For discrete emotions (Figure 1d), "neutral" is the most commonly selected emotion, followed by "sad". Interestingly, the next most frequent emotion is "happy," which is likely due to the limitation of having only one category for positive emotions.

Regarding relevance (Figure 1e), almost half of the annotations (44%) indicate "Not at all" relevant, with only 3.8% marked as "extremely relevant." For sharing inclination (Figure 1f), the distribution is even more skewed, with 54.5% of the annotations indicating "very unlikely" to share.

The majority of annotations (52.3%, Figure 1g) reveal that both the text and image significantly

[1] https://docs.google.com/document/d/1RPkjaPSksRbCy3y5d4WltidcUGhlH_np-aAuY2eH33c/

Variable	Category	Count	Percentage (%)	Mean (V)	Std (V)	Mean (A)	Std (A)
Gender	Man (including Trans Male/Trans Man)	59	53.15	3.62	1.50	**4.07**	1.42
	Woman (including Trans Female/Trans Woman)	52	46.85	3.38	1.61	**4.32**	1.39
Age Group	≤ 49	73	65.80	3.51	1.54	4.19	1.41
	> 50	38	34.20	3.50	1.60	4.18	1.42
Education Level	Below Bachelor's Degree	39	35.10	3.53	1.65	4.23	1.48
	Bachelor's Degree and Above	72	64.90	3.49	1.51	4.17	1.38
Personal Income Level	<£50,000	98	87.40	3.49	1.56	**4.22**	1.39
	≥£50,000	14	12.60	3.60	1.54	**3.99**	1.53
Political Leaning	Left	30	27.00	**3.32**	1.61	4.19	1.54
	Center	48	43.20	**3.53**	1.54	4.21	1.34
	Right	33	29.70	**3.61**	1.53	4.15	1.39
Neuroticism	Low	24	21.60	**3.65**	1.56	**3.98**	1.47
	Middle	74	66.70	**3.48**	1.54	**4.18**	1.40
	High	13	11.70	**3.39**	1.64	**4.60**	1.28
Current Affective State (PANAS)	Low	20	18.00	3.44	1.48	**4.44**	1.15
	Middle	73	65.80	3.51	1.55	**4.18**	1.43
	High	18	16.20	3.54	1.68	**3.93**	1.56
CRT	Low	49	44.10	3.57	1.47	4.13	1.41
	High	62	55.90	3.45	1.62	4.23	1.41

Table 1: Distribution of Annotators among Key Persona Variables

influence emotional reactions to news headlines. In contrast, approximately a third (36.7%) highlight the text alone as the primary factor. This indicates the importance of considering both the image and the text when modeling affective responses to news headlines on social media, rather than focusing solely on one or the other.

Relationship Between Arousal and Valence "Figure 2 depicts the average valence and arousal scores per headline, revealing a V-shaped distribution. This pattern, characterized by high arousal at both low and high valence levels, aligns with previous findings [Lang1997, Kurdi2017]. However, our results differ from those of [Kurdi2017] in exhibiting a greater concentration of data points at higher arousal levels (above 6, particularly in the second quadrant, which corresponds to low valence and high arousal). This discrepancy may be attributed to the inherent negativity bias prevalent in news headlines, as compared to the more diverse range of scenes and objects typically included in image-based studies."

We calculate the average valence and arousal for each headline and present the results in Figure 2. The distribution follows a V-shaped pattern, where arousal levels are high at both low and high extremes of valence, consistent with prior research (Lang et al., 1997; Kurdi et al., 2017). Notably, our data diverges somewhat from the findings of Kurdi et al. (2017), displaying a higher concentration of points at elevated arousal levels (above 6) in both the first and second quadrants. This trend is particularly pronounced in the second quadrant, characterized by very low valence and very high arousal. We hypothesize that this discrepancy arises from the inherently negative nature of news headlines, in contrast to the more varied emotional content typically found in datasets comprising images of scenes and objects.

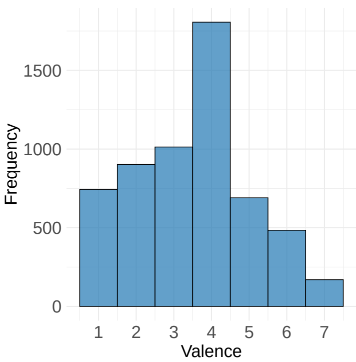
Group Level Differences We show the group-level mean and standard deviation of the valence and arousal annotation in Table 1.

Men exhibited a slightly higher mean valence (Mean (V) = 3.62) compared to women (Mean (V) = 3.38). Conversely, women showed a higher mean arousal (Mean (A) = 4.32) compared to men (Mean (A) = 4.07).

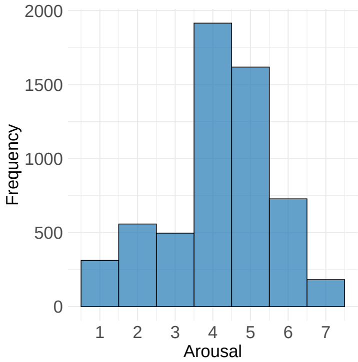
Left-leaning participants reported the lowest mean valence (Mean (V) = 3.32) and the highest variability in arousal (Std (A) = 1.54).

A particularly notable finding is within the neuroticism variable. Annotators with high neuroticism had a significantly higher mean arousal (Mean (A) = 4.60), consistent with well-documented associations between neuroticism and higher emotional reactivity (Costa and McCrae, 1980).

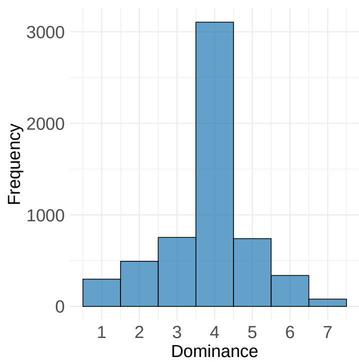
There is a large different in the group-level mean in annotators with different levels of current affective state (PANAS Positive - Panas Negative). The mean arousal score ranges from 4.44 to 4.18 to 3.93 from the lowest to highest level of current affective state. Also interestingly, annotators with the lowest current affective state report the lowest standard devitation in arousal level. This is despite the standard deviation of arousal level being largely the same in any other groupings.



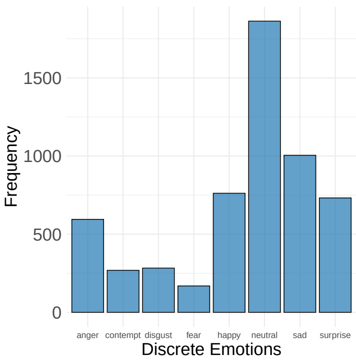
(a) Valence Distribution



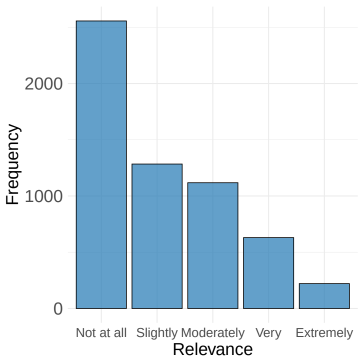
(b) Arousal Distribution



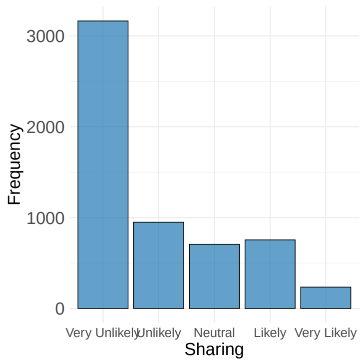
(c) Dominance Distribution



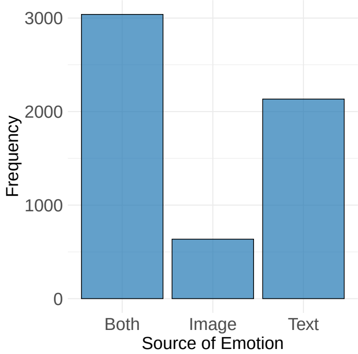
(d) Discrete Distribution



(e) Relevance Distribution



(f) Sharing Distribution



(g) Source Distribution

Figure 1: Distribution of Annotations

We further conduct a fixed effect linear regression analysis[2], including all the persona variables mentioned in Table 1. The effect of gender, political leaning, neuroticism and current affective state are significant ($p<0.05$).

Conclusion and Future Work In this paper, we describe an ongoing project to collect a large-scale individualized affective news response dataset, enriched with various persona variables about individual annotators. We envision this dataset to be useful for multiple purposes, for both psychology and natural language processing. For example, it could be helpful for understanding the group-level and individual-level covariates that would be important to explain the varied affective response to news headlines and the underling mechanism that leads to such differences. It could be valuable for NLP researchers focused on developing culturally-aware, pluralistic systems that account for global diversity in human responses. The dataset also has the potential to facilitate the creation of algorithms designed to accommodate individual differences, paving the way for personalized language models that could greatly enhance applications like personal assistants. As this project is still in progress, we highly welcome any feedback.

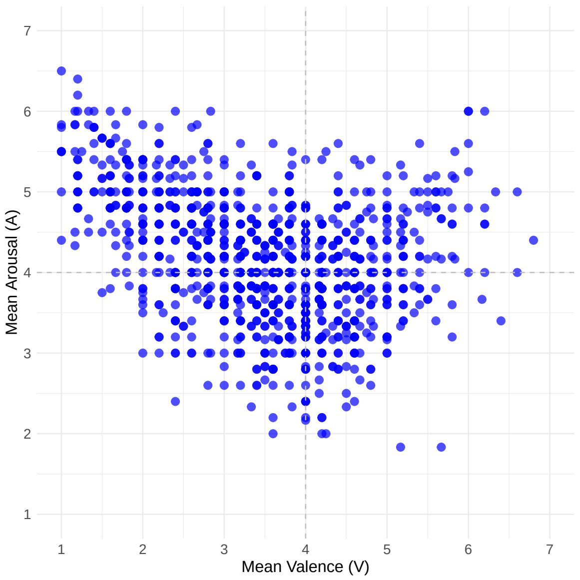


Figure 2: Mean Arousal vs. Mean Valence per Headline. Darker color reflects overlapping points

[2]In R notation, `annotation ~ persona variables`

References

Bouchra Arbaoui, Knut De Swert, and Wouter van der Brug. 2020. Sensationalism in news coverage: A comparative study in 14 television systems. *Communication Research*, 47(2):299–320.

Ming M. Boyer. 2023. Aroused argumentation: How the news exacerbates motivated reasoning. *The International Journal of Press/Politics*, 28(1):92–115.

Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

Margaret M. Bradley and Peter J. Lang. 2007. Affective norms for english text (anet): Affective ratings of text and instruction manual. Technical report D-1, University of Florida, Gainesville, FL.

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a perspectivist turn in ground truthing for predictive computing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6):6860–6868.

Susan T Charles and Laura L Carstensen. 2010. Social and emotional aging. *Annual review of psychology*, 61:383–409.

Paul T Costa and Robert R McCrae. 1980. Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *Journal of personality and social psychology*, 38(4):668.

Paul T Costa and Robert R McCrae. 2008. The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, 2(2):179–198.

John R Crawford and Julie D Henry. 2004. The positive and negative affect schedule (panas): Construct validity, measurement properties and normative data in a large non-clinical sample. *British journal of clinical psychology*, 43(3):245–265.

Natascha de Hoog and Peter Verboon. 2020. Is the news making us unhappy? the influence of daily news exposure on emotional states. *British Journal of Psychology*, 111(2):157–173.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Mark Diaz, Isaac Johnson, Amanda Lazar, Anne Marie Piper, and Darren Gergle. 2018. Addressing age-related bias in sentiment analysis. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–14, New York, NY, USA. Association for Computing Machinery.

Shane Frederick. 2005. Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42.

Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

Mariska Kleemans, Paul G. J. Hendriks Vettehen, Johannes W. J. Beentjes, and Rob Eisinga. 2017. The influence of sensationalist features in television news stories on perceived news quality and perceived sensationalism of viewers in different age groups. *Studies in Communication Sciences*, 17(2):183–194.

Mariska Kleemans and PGJ Hendriks Vettehen. 2009. Sensationalism in television news: A review.

Ann M Kring and Albert H Gordon. 1998. Sex differences in emotion: expression, experience, and physiology. *Journal of personality and social psychology*, 74(3):686.

Benedek Kurdi, Shayn Lozano, and Mahzarin R. Banaji. 2017. Introducing the open affective standardized image set (oasis). *Behavior Research Methods*, 49(2):457–470.

Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. 1997. International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1(39-58):3.

Albert Mehrabian and James A Russell. 1974. *An approach to environmental psychology.* the MIT Press.

Batja Mesquita and Nico H Frijda. 1992. Cultural variations in emotions: a review. *Psychological bulletin*, 112(2):179.

Laura Ana Maria Oberländer, Evgeny Kim, and Roman Klinger. 2020. Goodnewseveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1554–1566.

Mary Beth Oliver. 2002. Individual differences in media effects. In *Media effects*, pages 517–534. Routledge.

Cecilia Ovesdotter Alm. 2011. Subjective natural language problems: Motivations, applications, characterizations, and implications. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 107–112, Portland, Oregon, USA. Association for Computational Linguistics.

Jiaxin Pei, Aparna Ananthasubramaniam, Xingyao Wang, Naitian Zhou, Apostolos Dedeloudis, Jackson Sargent, and David Jurgens. 2022. POTATO: The portable text annotation tool. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 327–337, Abu Dhabi, UAE. Association for Computational Linguistics.

Barbara Plank. 2022. The “problem” of human label variation: On ground truth in data, modeling and evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10671–10682, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.

David Preece, Rodrigo Becerra, and Guillermo Campitelli. 2018. Assessing emotional reactivity: Psychometric properties of the perth emotional reactivity scale and the development of a short form. *Journal of Personality Assessment*.

Stuart Soroka, Patrick Fournier, and Lilach Nir. 2019. Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, 116(38):18888–18892.

Justin Storbeck and Gerald L Clore. 2008. Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and personality psychology compass*, 2(5):1824–1843.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007)*, pages 70–74.

Patti M Valkenburg and Jochen Peter. 2013. The differential susceptibility to media effects model. *Journal of communication*, 63(2):221–243.

Tai-Li Wang. 2012. Presentation and impact of market-driven journalism on sensationalism in global tv news. *International Communication Gazette*, 74(8):711–727.

A Appendix

The list of news outlets that we sample from include:

- Daily Mail
- The Telegraph
- The Mirror
- Metro
- The Sun
- Daily Star
- The Independent
- Daily Express

- The i Paper
- GB News
- LADbible
- The Economist
- The Times and The Sunday Times
- The Guardian
- ITV News
- BBC News
- Sky News
- Reuters UK
- LBC
- Financial Times
- Channel 4 News

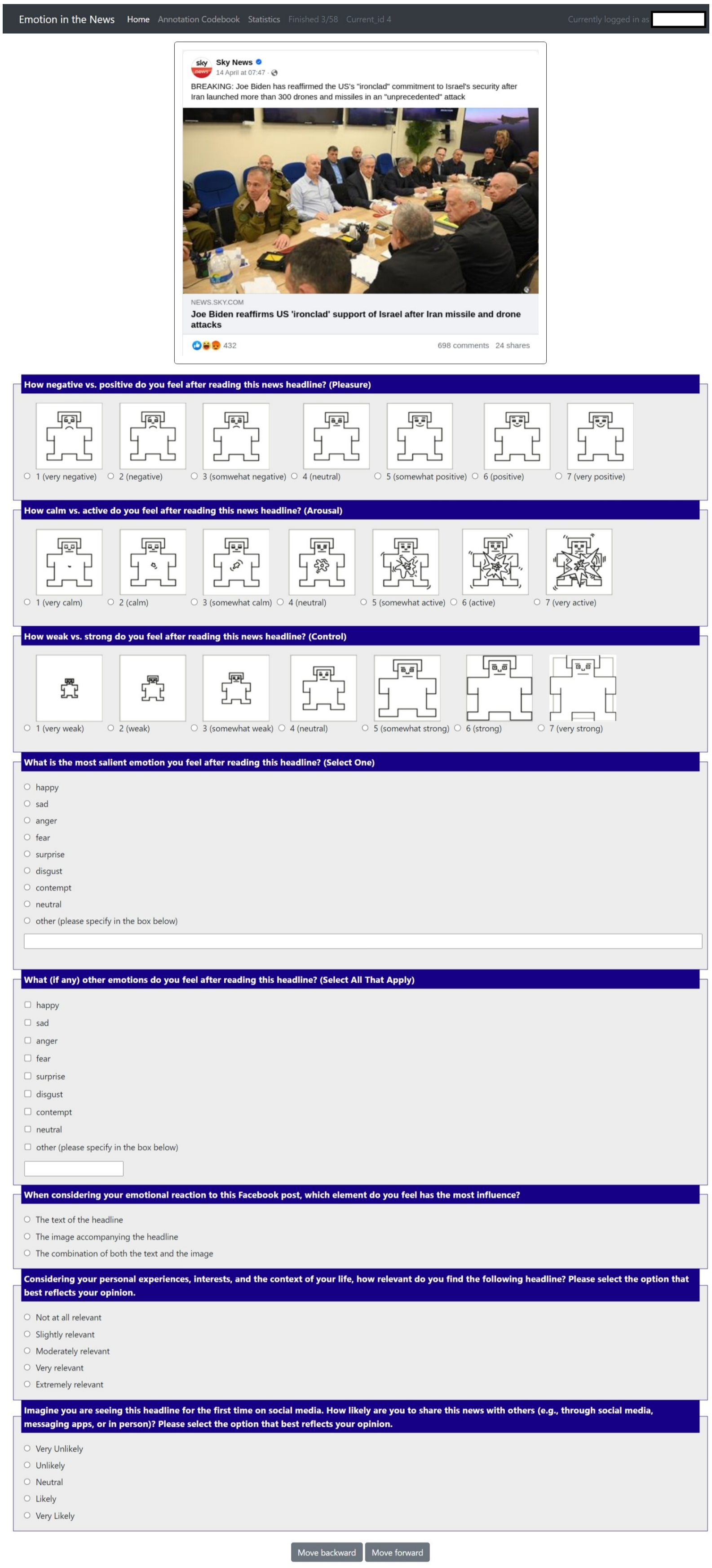


Figure 3: An example headline.