Less is More? The Role of Demographic Author Information in Emotion Classification of Ambiguous Text

Sabine Weber, Lynn Greschner and **Roman Klinger**
Fundamentals of Natural Language Processing, University of Bamberg, Germany
{firstname.lastname}@uni-bamberg.de

Abstract

Emotion annotation is a challenging task that often yields low inter-annotator agreement because texts are open to subjective interpretation. Beyond missing context and differences in world knowledge, extra-linguistic factors such as the author's identity influence how emotions are perceived. When the text alone does not provide sufficient information, additional details about the author may help resolve ambiguity. For instance, "This is sick." expresses a positive judgment when said by a young man at the skate park but a negative one when said by an old woman at the opera. We test the hypothesis that providing annotators with demographic information reduces disagreement in emotion annotation. We compare one group of annotators who sees each text alongside demographic information about its author, with a group who sees only the text itself. As a basis, we use the crowd-enVENT corpus, a disaggregated corpus of emotion annotations, directly sourced from people who lived through an event and then described it. This corpus also provides reader annotations, which we use to subselect instances with particularly low agreement. We find in our study with 500 annotators and 250 texts that displaying demographic information about the author of the text does not improve agreement between annotators, nor does it improve agreement with the gold label. These results show that more information for the annotators is not generally beneficial. The only exception in our study are cases where the emotion polarity (positive or negative) is unclear. We also find that annotators perform overall better at identifying the correct emotion label when it aligns with gender stereotypes. Zero-shot prompting experiments with large language models do resemble the human annotation experimental results. Our findings suggest that providing demographic information is not a straightforward remedy for ambiguity in emotion annotation and careful consideration is needed when incorporating such data.

Keywords: emotion, ambiguity, extralinguistic context

1. Introduction

Conveying an emotion in text is both a finely honed skill that professional writers aim at perfecting and a mundane task we perform every time we write a friend about an emotional event. The interest in the topic as a natural language processing (NLP) task has steadily increased within the last years, as Plaza-del-Arco et al. (2024a) show in their study on emotion classification and analysis.

In written communication, authors typically share only the information needed for their message to be understood (Grice's Maxim of Quantity see, e.g., Krause and Vossen (2024)). Readers, however, may know or infer details about the author's identity and use these details to interpret the emotional connotation of a text. In NLP, most work on emotion analysis has focused on linguistic features alone, without considering such meta-textual information (e.g., Étienne et al. (2024); Wemmer et al. (2024)). This raises the question of how author-related context affects emotion interpretation, especially in ambiguous cases where several emotion interpretations might coexist.

To answer this question we conduct a study with two groups of annotators, showing one of them emotionally ambiguous text along with demographic information about the author, while the other group only sees the text by itself (an illustration of the setup can be seen in Figure 1). Possibly, the additional information could help annotators to form a mental image of the author and come to a more consistent and truthful judgment of the expressed emotion.

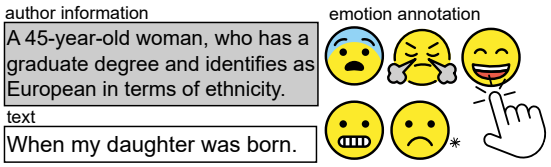


Figure 1: Experimental setup for the annotators who see both author information (gray) and the text. The comparison group sees only the text without additional information.

In our study with 500 annotators we find that showing demographic information does not improve the agreement between annotators or correct prediction of the emotion label. To better understand how and when demographic cues might influence judgments, we construct targeted subsets of the dataset based on two theoretically motivated dimensions. First, we identify instances involving gender-stereotypical emotion associations by categorizing examples according to whether the gold

emotion label alignes with gender stereotypes in emotion perception (e.g., fear being stereotypically associated with women and anger with men (Plant et al., 2000; Plaza-del-Arco et al., 2024b)). An example of this is annotators ascribing fear to the text "My teen age son wasn't in his bed after bedtime" when knowing that the author is a woman, but surprise when no gender information is provided. We observe that annotators perform slightly better when the gold label aligns with these stereotypes.

Second, we categorize texts based on polarity clarity, distinguishing between examples with clearly positive or negative sentiment and those with ambiguous sentiment. We observe a slight increase in F1 score when demographic data are displayed with the ambiguous texts.

To see if these human tendencies translate into modeling, we conduct experiments with four recent, open-weight Large Language Models. Our zero-shot LLM experiments show that the tested LLMs perform worse than humans at predicting the correct gold labels. The display of demographic data often leads to a decrease in prediction performance, the only exceptions being cases where the labels align with gender stereotypes or when the polarity of the emotion expressed in the text is unclear, showing similar trends to human annotators.

Our findings caution that the display of author demographic information is not a simple solution to annotation ambiguity: while potentially informative, it can also introduce biases and increase cognitive load, leading to overall lower annotation quality. Further research is needed to determine under which conditions demographic author information should be used, balancing positive and negative effects.

Our research questions are:

- **RQ1:** Does demographic information about the author of a text improve inter-annotator agreement? *(No)*
- **RQ2:** Does demographic information about the author of a text improve agreement with the original author (i.e., the gold label)? *(No)*
- **RQ3:** Do LLMs show the same tendencies as human annotators when faced with ambiguous text? *(Partially)*

We make all of our study data publicly available.[1]

2. Related Work

2.1. Perspectivism

Emotion classification from text is a well-researched task covering many languages and genres, and represented by shared tasks like Muhammad et al. (2025). While most studies on emotion classification focus on textual features alone, a growing line of work in NLP emphasizes interpersonal variation and subjectivity in annotation. This angle, sometimes referred to as a perspectivist approach (Frenda et al., 2025), treats disagreement between annotators not as noise but as informative signal about how individuals interpret emotion content. Several studies have examined emotion analysis from this viewpoint (Mieleszczenko-Kowszewicz et al., 2023; Miłkowski et al., 2022, 2021), including Kazienko et al. (2023), who train personalized models tailored to individual annotators.

2.2. Author Information

The previously mentioned papers focus solely on properties of the reader of a text, while text interpretation occurs within a situational context that often includes information about the author. This influences the judgment of the presented text: Combs et al. (2023) show that annotators perceive arguments to be less convincing when they believe that they were written by a woman, while Sap et al. (2019) show that tweets using African American Vernacular English are perceived as less offensive if annotators are aware that the writer is Black. We follow these findings in our work by hypothesizing that demographic author information can influence the way annotators annotate emotions.

As to our knowledge, Troiano et al. (2023) present the only unaggregated corpus of text annotated for emotions that contains labels for the emotion felt by the author (gold label) and the emotion that the readers inferred (predicted label), as well as extensive demographic information of the author. It is therefore uniquely suited for answering our research questions. While Troiano et al. (2023) report overall high agreement between annotators, the corpus contains many examples of ambiguous texts where readers fail to come to a shared assessment of the text or to correctly guess the gold label. These texts are particularly relevant to our research, because this disagreement may indicate insufficient information for disambiguation. Leveraging the demographic data in the corpus, we can provide the readers of a text with a meta-textual signal to aid disambiguation.

2.3. Sociodemographic Prompting

The difference between LLM outputs depending on provided demographic information has been the focus of recent work on sociodemographic prompting (or persona-based prompting), also touching on subjective tasks like emotion classification. Plaza-del-Arco et al. (2024b) show that LLMs reproduce gender stereotypes about emotions. Sun et al. (2025) find that LLMs align most often with white annotators even if prompted otherwise and Lutz

[1] https://www.uni-bamberg.de/en/nlproc/resources/

et al. (2025) show that LLMs struggle to simulate members of marginalized groups.

Our LLM experiments differ from these studies in design and goal. Rather than prompting models to adopt the perspective of an annotator from a specific demographic group, we provide demographic information about the author of a text and ask the model to infer the emotion that author would have felt. This task focuses on interpreting emotion from text with contextual cues, not on simulating a social identity. Accordingly, our analysis aims to uncover general similarities between human and model behavior, rather than differences across annotator demographics.

3. Methods

We are interested in the annotators' judgment of the emotions expressed in text. In the following sections we outline the selection of data and labels and the full experimental setup. Lastly, we provide details on the LLM experiments.

3.1. Text Selection

As basis for our annotation setup we use the English crowd-enVENT corpus by Troiano et al. (2023). It was generated via a two step process: First, people were asked to describe an event in their lives and to label it using emotion and appraisal labels. In a second round, different people were presented with these texts and asked to assign emotion and appraisal labels. Every text was thus labeled by its author and 5 readers. This allows us to calculate both agreement with the gold label (assigned by the author) as well as agreement in between readers.

We consider low agreement between readers as an indicator that a text is ambiguous. Our study corpus consists of 250 texts selected from crowd-enVENT with the lowest agreement based on Krippendorf's alpha for the appraisal annotations and Fleiss Kappa for the emotion annotations, filtering first for low agreement on emotion and then on appraisal labels.

To analyze how the display of demographic information influences agreement with the gold label, our 250 selected texts are stratified by the difficulty of inferring the gold label: for 125 of them, the majority of annotators in crowd-enVENT correctly identified the gold emotion label, while for the other 125, the majority chose a different label.

3.2. Label Selection

We use the original emotion label set used by Troiano et al. (2023) but omit the no-emotion category. As Troiano et al. (2023) disclose, this label has been used by authors to label events that were emotionally charged, but where they did surprisingly not experience an emotion.

While the emotion label set is a combination of basic emotions and self-directed states, appraisal labels present a different angle under which emotion can be examined. Appraisals allow for an evaluation of the emotional significance of an event, e.g. whether it occurred suddenly, whether it was within the control of the experiencer or whether it aligned with the experiencer's values. We follow Troiano et al. (2023) to enable comparability with the gold appraisal labels assigned by the authors of the text. This leads to a set of 21 appraisal labels. A list of all emotion and appraisal labels and explanations is shown in Appendix A.

3.3. Annotator Selection

We recruit annotators using the platform Prolific[2]. To enable a wide spread of demographic features of the annotators, we recruit English-speaking participants worldwide. To assure data quality we select from them annotators with high language proficiency and a platform approval rating of 98–100%.

Every annotator annotates 5 texts, either with or without demographic data about the text's author displayed alongside the text, which results in a total of 1250 annotation per group and 2500 annotations in total. Each text receives a total of 10 annotations. The annotations are collected on three consecutive days. The total cost amounts to £1286. While Troiano et al. (2023) report an expected completion time of 8 minutes for 5 texts, our study takes 15 minutes (median) for the same amount of texts, presumably because their ambiguous nature makes them harder to annotate.

3.4. Experimental Setup Human Annotation

We vary one experimental condition: Half of the annotators see the texts with demographic information about the author, and half of them see the texts without demographic information. The demographic information consists of the self-reported age, gender, education level and ethnicity of the author of the text (see an example in Figure 1). Each annotator can only belong to one study group. By doing so we prevent the same annotator from seeing the same text multiple times. Apart from this condition, the study is the same for all annotators. We also collect demographic and personality information about the annotator. The full questionnaire is shown in Appendix D.

[2] https://www.prolific.com/

	IAA		GA	
	Emo. κ	Appr. α	Emo. F1	Appr. RMSE
T	−0.20	0.05	0.46	1.72
T+Demo.	−0.19	0.04	0.39	1.74
Δ	0.01	−0.01	−0.07	0.02

Table 1: Inter-annotator agreement (IAA) and agreement with the gold label (GA) across groups . T stands for the group who saw text only, T+Demo for the group who saw both text and author information. Δ stands for the difference in performance. While there are small differences, none are statistically significant.

3.5. Experimental Setup LLM Annotation

To compare the behavior of annotators with the predictions of LLMs, we select four recent LLMs (Gemma (Mesnard et al., 2024), Mistral (Jiang et al., 2023), Mixtral (Jiang et al., 2024), Llama3.2. (Grattafiori et al., 2024)) for our experiments (hyperparameter settings see Appendix C). We present the models with a simplified version of the setup presented to the human annotators, focusing only on annotating emotion labels. The full prompt can be seen in Appendix B, Figure 2.

LLMs are known to show inconsistent behavior, outputting different labels when prompted with the same text (Bartsch et al., 2023). This might be even more common when prompted with ambiguous text. We therefore prompt the LLM 10 times to verify the robustness of the assigned label. We parse the output for the emotion labels and perform majority voting, selecting the emotion label named most often by the LLM as the annotated label. When the most common prediction does not contain any of the emotion labels, we count the annotated label as “none”.

4. Results

In the following section we discuss the answers to RQ1 and RQ2, with the results of the human annotation study shown in Table 1. The results indicate that adding demographic data does not lead to better agreement between annotators or with the gold label. We present a closer analysis first of the whole data set cumulatively and then for specific subsets of the data set. We then present the results of the LLM experiments.

4.1. Cumulative Analysis of Human Annotation

Quantitative Analysis. We evaluate the agreement between the 5 annotators of a specific text using Fleiss’ Kappa for the categorical emotion labels and Krippendorff’s Alpha for the numerical appraisal values. All values are relatively low, reflecting the ambiguous nature of the text. While there are slight differences in values between study groups, the difference is not statistically significant (t-Test, p=0.87 and p=0.19). The same holds for the agreement with the gold labels, which we evaluate by pooling the votes from the 5 annotators of a text, applying majority voting and calculating the macro F1 score for emotion labels and root mean squared error (RMSE) for the appraisal labels.

Qualitative Analysis. Although showing demographic information does not improve overall performance, annotators exposed to author data label texts differently than those without such information. Table 2 shows instances with the highest differences in prediction accuracy between the study groups. These differences could be explained in part by prevailing gender stereotypes (Plaza-del-Arco et al., 2024b): When annotators know the author to be a woman ‘fear’ is chosen over ‘disgust’, ‘sadness’ or ‘surprise’ and ‘surprise’ is chosen over ‘pride’ (ID 2, 9, 15 and 16 in Table 2). Sometimes these gender stereotypes align with the correct label, e.g., when a 32-year old man feels relief rather than joy when getting a permanent job (ID 17), or a 22-year old woman feels shame rather than guilt for offending someone (ID 14). But this is not reliably the case, as with a mother who feels surprise at finding her son out of bed, rather than the annotator-predicted fear (ID 2). We take these examples to inspire our approach for subset analysis.

4.2. Subset Analysis of Human Annotation

Building on our qualitative analysis, we examine the influence of gender stereotypes on annotators’ judgments. Also, to deepen the analysis of ambiguity present in the texts, we want to examine cases in which the polarity of the text (emotionally positive or negative) is unclear.

These instances can be seen as cases of especially high ambiguity, where additional author information might help disambiguation. To do so, we divide the dataset into subsets and conduct statistical analyses. The assignment of instances to these categories was performed by one of the paper’s authors.

Gender Stereotypes. In our assessment of gender stereotypes about emotion we build upon previous literature on the topic. With regards to negative emotions, women are stereotypically associated with fear and sadness, and men with anger; for

	ID	Emotion			accΔ	text	age	gen	ethn
		$-$ inf.	$+$ inf	gold					
Lower Acc	1	sur	joy	sur	-0.8	I felt ... when I found a turkey on sale for only 15 dollars	57	M	Eu
	2	sur	fear	sur	-0.8	My teen age son wasn't in his bed after bedtime	37	F	Eu
	3	ang	sur	ang	-0.6	I was ignored by my last manager when I informed her of my leaving the company. They did not acknowledge my email I sent.	36	F	Eu
	4	rlf	joy	rlf	-0.6	I got a job after months of searching.	23	M	Eu
	5	bor	ang	bor	-0.6	I was in a mental health review and was being talked at by a psychiatrist and CPN. [...]	30	F	Eu
	6	sur	joy	sur	-0.6	When I gave birth to my baby girl last December, and we thought she was another baby boy.	32	F	Eu
	7	ang	ang	ang	-0.6	Although I work full time I still have to do all the housework	45	F	Eu
	8	disg	sad	disg	-0.6	I trod on the body of a dead bird	37	M	Eu
	9	prd	sur	prd	-0.6	my niece came out to me and told me she had a girlfriend before she told anyone else.	37	F	Eu
	10	rlf	fear	rlf	-0.6	When the webinar was over as I was concerned it wouldn't run smoothly	29	M	Eu
Higher Acc	11	rlf	rlf	rlf	0.6	I thought my boss was upset with me about a mistake I made at work but it turned out she was really understanding once I spoke with her.	23	F	NA
	12	bor	bor	bor	0.6	when im at home alone because im always either watching tv or my laptop	28	F	NA
	13	rlf	rlf	rlf	0.6	I had my first vaccine, as I have a fear of needles but it was not as bad as I thought it would be.	33	M	Eu
	14	glt	shm	shm	0.6	offended someone	22	F	Aust
	15	disg	fear	fear	0.6	A stranger in the street started shouting at me for not allowing him to approach my dog	45	F	Eu
	16	sad	fear	fear	0.6	I felt ... when I had a car accident in Puerto Plata Dominican republic, [...]	42	F	Eu
	17	joy	rlf	rlf	0.6	I felt ... when I got a job at my current position permanently.	32	M	NA
	18	rlf	fear	fear	0.6	I was nearly hit by a car while commuting to work	21	M	SA
	19	joy	sur	sur	0.6	Finding out I was pregnant with twins	42	F	Eu
	20	joy	sur	sur	0.8	when i won a £500 john lewis voucher	28	F	Eu

Emotion abbreviations: ang = anger, bor = boredom, disg = disgust, fear = fear, glt = guilt, joy = joy, prd = pride, rlf = relief, sad = sadness, shm = shame, sur = surprise. **Region abbreviations:** Eu = Europe, NA = North America, SA = South America, Aust = Australia

Table 2: Examples of the sentences where the agreement with the gold label between the two experimental group differed most. The top half shows examples where annotators who saw author information guessed worse than annotators who saw only the text, the bottom half shows examples where annotators who saw author information guessed better. -inf shows the majority guessed label when no author information was displayed, $+$inf the majority label when author information was shown. accΔ shows the difference in accuracy score on the text between the two groups of annotators. ID is used for referencing the examples in Section 4.1. Rows where gender biases might have influenced annotators' judgments are highlighted in yellow.

positive emotions, women are stereotypically expected to experience joy and men pride (Plant et al., 2000; Plaza-del-Arco et al., 2024b).

To investigate the role of gender stereotypes we want to select examples where the gender of the authors is not given in the text itself, so that explicitly displayed author gender adds extra-textual information. We also want to select examples, where gender information might lead to stereotypically different emotion labels (e.g., fear vs. anger). To do so, we first filter out text instances where the gender of the speaker is disclosed in the text e.g., using

Model	Text	Text + Demo.	Δ
Gemma	0.25	0.25	0.00
Mistral	0.26	0.24	−0.02
Mixtral	0.25	0.25	0.00
Llama 3.2	0.27	0.26	−0.01
Human	**0.46**	**0.39**	−0.07

Table 3: Mean F1 scores for different LLMs with and without disclosed author data as well as human performance. Δ = (With – Without). Red indicated worse performance when author data is displayed. None of the performance differences are statistically significant. Human performance is stronger than zero-shot LLM performance.

self-descriptors as "mother" or "wife". We also filter out texts with gendered references to partners e.g., "my girlfriend" or "my husband", as in a heteronormative societal context this too gives a clue towards the gender of the speaker. This procedure removes 44 texts from our corpus of 250 texts.

We then split the corpus into cases where gender stereotypes might influence emotion assignments. Neutral cases include statements like "I went to Los Angeles." while stereotype-relevant cases include "I was shouted at by a stranger." While the first sentence depends on the author's liking of Los Angeles, the latter sentence could (in a stereotypical interpretation) elicit fear for women and anger for men. The result consists of 126 neutral texts, and 80 texts where stereotypes could play a role. Lastly, we determine if the gold labels provided by the original authors of the text align with the stereotypes or contradict them. This procedure creates a set of 56 pro-stereotypical and 24 anti-stereotypical texts-label pairs.

To analyze the influence of the display of demographic data in these subsets, we calculate the inter-annotator agreement as Fleiss' Kappa and the agreement with the gold label as F1 score. Inter-annotator agreement slightly decreases when demographic data is displayed, as do F1 scores for predicting the authors' emotion labels (by 0.03 points for pro-stereotypical and 0.02 points for anti-stereotypical cases, see last row in Table 4). However, annotators who see demographic information are more likely to correctly identify emotions that align with gender stereotypes, with F1 scores of 0.26 for pro-stereotypical and 0.21 for anti-stereotypical examples (see last row in Table 4).

The decrease of inter-annotator agreement and prediction performance in both the pro- and anti-stereotypical subset when demographic data is displayed could be explained by the fact that the displayed demographic information does not only contain information about gender, but also about age, education level and ethnicity of the author, providing more information irrelevant to the task at hand and adding to the annotators' cognitive load. This effect is not outweighed by the additional information that the author gender provides. Some annotators might also be aware of prevailing stereotypes and therefor actively avoid them as a base for their judgment.

Level of Ambiguity. To further examine the role of author information in ambiguous texts, we separate the our data set in to two subsets: One where the emotion polarity (positive or negative) is clear and one in which it is unclear. We want to determine whether the display demographic information has an influence in these specific cases (RQ1).

To answer this question we manually sort our data set in polarity-clear and polarity-ambiguous cases. A clear case is for example "I was shouted at by a stranger" (negative) and an ambiguous case is "I went to Los Angeles" (could be positive or negative depending on whether the author likes Los Angeles). We also remove cases where gender is disclosed to avoid possible conflicting influence of gender stereotypes. After this filter, we find 41 texts to be polarity-ambiguous and 165 texts polarity-clear.

We evaluate inter-annotator agreement by calculating Fleiss' Kappa. We find that in the polarity-ambiguous cases the display of annotator demographic information leads to slightly lower inter-annotator agreement. But when calculating F1 score for the agreement with the gold emotion labels we find that there is a small improvement in F1 score when demographic data is displayed, lifting it from 0.2 with no demographic information to 0.25 with demographic information (see last row in Table 4). This can be seen as a point in evidence that when several contradicting emotion assignments are possible, the demographic data provides a weak signal for disambiguation and possibly outweighs the additional cognitive load.

4.3. LLM Experiments

We want to determine if LLMs replicate the described human judgments in a zero-shot prompting setup (RQ3). To do so we provide the LLMs with the same information as was shown to the human annotators. We prompt the LLMs either with text only or with text and additional demographic author information. We then calculate the F1 score comparing the LLM-assigned emotion label with the gold label given by the original author of the text.

Overall Performance. Table 3 shows that human annotators outperform all LLMs in predicting the gold labels. All models perform above random

Model	Pro Stereotypical			Anti Stereotypical			Unclear Polarity			Clear Polarity		
	T	T+D	Δ	T	T+D	Δ	T	T+D	Δ	T	T+D	Δ
Gemma	0.17	0.22	+0.06	0.18	0.23	+0.05	0.20	0.22	+0.02	0.24	0.27	+0.03
Mistral	0.24	0.28	+0.04	0.18	0.18	0.00	0.20	0.20	0.00	0.28	0.29	+0.01
Mixtral	0.20	0.22	+0.02	0.23	0.18	−0.05	0.17	0.22	+0.05	0.27	0.25	−0.02
LLaMA 3.2	0.30	0.31	+0.02	0.14	0.14	0.00	0.24	0.34	+0.10	0.29	0.26	−0.03
Human	0.29	0.26	−0.03	0.23	0.21	−0.02	0.20	0.25	+0.05	0.39	0.35	−0.04

Table 4: Mean F1 scores for different LLMs across conditions varying in gender stereotypicality (pro-stereotypical vs. anti-stereotypical) and emotional ambiguity (polarity clear vs. ambiguous). Results are reported with (T+D) and without (T) disclosed demographic data of the author. Δ stands for the difference in performance. Green indicates improvement with author data, red indicates a decrease.

(F1=0.08). Adding demographic data does not improve LLM performance and, in some cases, even reduces it.

Subset Performance. When evaluating on the subsets described in the previous section (one split by gender stereotypes and the other split by clear versus unclear emotional polarity), we can see varying differences to human performance (see Table 4). Especially in the examples where gold labels align with gender stereotypes of emotion, all LLMs show improvement when gender information is displayed, which is different to human annotation. With anti-stereotypical examples and examples where the emotion polarity is clear adding demographic data leads to inconsistent results, slightly improving performance in some LLMs and showing no effect or negative effects in others. In contrast, in human annotation the display of demographic data in these cases always leads to a decrease in performance. Only in the examples with unclear polarity does displaying demographic data have a positive effect both with human annotators and LLMs.

5. Discussion

These results show that demographic information about the author does not improve agreement among annotators and gold label prediction. The presented texts are emotionally ambiguous, but the presented demographic details about the author do not directly contribute to a clarification. This leaves the annotator wondering whether there is a connection they are not finding, leading to higher cognitive load and lower prediction performance. Our findings stand in contrast to previous work where annotator demographics had a more obvious bearing on the task at hand, e.g., in Sap et al. (2019) who study offensiveness detection. The usage of certain terms is offensive when the speaker is white but not offensive when the speaker is Black, so that the judgment of offensiveness could be clearly informed by a specific demographic detail.

Confusion and frustration about the demographic information was also expressed explicitly by the annotators themselves. In the optional comment field at the end of the study several annotators commented upon the difficulty of the task, wishing for a 'don't know' option among the emotion and appraisal labels or pointing out axes of ambiguity not addressed by demographic information. One annotator wrote: '[...] Perhaps the shorter [texts] could be made longer to give more context [...]', showing that the provided demographic context was not sufficient.

Our qualitative analysis of the results shows that demographic information makes a difference in the annotators decision, but this difference does not result in more correct or consistent judgment, possibly because it relies on stereotypes about the displayed demographic groups. While annotators are generally better at inferring emotions that align with gender stereotypes, the display of demographic information hurts performance in all cases, except for very slight improvements in cases where emotion polarity is unclear.

Our study of LLMs shows that for LLMs, too, the display of demographic information does lead to worse or unchanged performance. The display of demographic information in cases where gold labels aligned with gender stereotypes helps, which confirms the findings of Plaza-del-Arco et al. (2024b), but in other cases the demographic data possibly acts as a distractor.

Our findings point to several directions for future research. First, demographic attributes provided in our study may have been too coarse-grained to meaningfully support emotion disambiguation. Sorensen et al. (2025) show that value profiles help to explain differences when annotators with the same demographics disagree on labels – an approach which could be extended to author information. Second, the display of author information as metadata alongside the text is not very natural and annotators may benefit more from contextualized social information, e.g., via profile pictures

as used in Combs et al. (2023). Third, our results suggest the need to broaden the scope beyond demographic cues to examine other sources of systematic bias in emotion perception, such as stereotypes related linguistic style (e.g., hedging versus assertive expression). Comparative studies along these dimensions could help clarify when supplementary information genuinely aids interpretation and when it instead reinforces stereotypical judgments. Future work could also investigate interactive approaches, for example by directly eliciting annotators' perceptions of what additional contextual information they would need to resolve ambiguity.

6. Conclusions

Our study tests whether knowing an author's demographic data helps annotators disambiguate ambiguous text in the task of emotion classification (assigning one emotion label reflecting the emotion expressed in a text). Additional information could aid in forming a mental image of the author, improving both inter-annotator agreement and alignment with the gold label. However, our findings show no such improvements.

Our qualitative analysis reveals that author information influences annotation decisions, occasionally but not consistently improving gold label accuracy. This could be due to the stereotypes about specific social groups that annotators apply to the text. Overall the display of demographic data leads to worse performance, possibly due to the additional cognitive load. These findings are mirrored, partially, by our LLM experiments.

Our main findings are: (1) demographic information is not a simple solution to annotation ambiguity and can, in some cases, introduce bias reflecting social stereotypes and reduce performance; and (2) it is crucial to distinguish between situations where demographic information genuinely aids disambiguation and those where other forms of contextual information are necessary. This opens an avenue for future work: given an ambiguous text and a specific task, how can we determine what additional information is most beneficial for completing it?

7. Limitations

When designing our experimental setup we needed to make the decision on how to display the demographic information to measure its effect on the annotators. An alternative setup to the reported one shows an annotator the text without author information first, collects emotion and appraisal labels, and then shows the annotator the text again with author information to assess which labels (if any) would change. We reject this setup because the double exposure to the same text could introduce unwanted effects that would be difficult to disentangle.

A second limitation is the relatively small sample size, which limits the options for statistical analysis of different groups of annotators (e.g., comparing performance by gender or age), especially with regards to the various subsets. Additionally, the subsets are not balanced with some classes being larger than others which can also lead to a distortion of results. While the texts were selected from a larger corpus using criteria we outlined in Section 3, there is still the chance that concentrating on these examples might have skewed our data in ways we did not anticipate. Replicating the experiment with more and different texts would be an avenue for future work.

Conducting this study in English allowed us to use an existing text corpus and to recruit participants over a world-wide operating annotation service (Prolific). But this also limits the validity of our results, because inter-cultural differences could not be captured in as much depth as a multi-lingual study.

8. Ethical Considerations

Our study uses data from the crowd-enVENT corpus (Troiano et al., 2023), for which the data acquisition and annotation process passed an ethics board review process. We only use data from this corpus and replicate the annotation procedure by Troiano et al. (2023), with minor changes to the acquired variables. At the beginning of our study, we display a content warning for the topics of illness, infidelity, divorce, injury, pregnancy and homelessness. Therefore, our annotators are not at greater risk of being exposed to sensitive topics compared to their day-to-day lives or while consuming online content.

Further, all participants in our study are informed about their annotations being used in scientific publications, and we explicitly obtain their consent for this. We publish an anonymized version of our collected data (removing the participants IDs). The findings of our study highlight the need for a nuanced and sensitive handling of demographic data, cautioning against the unreflected collection and display of such data.

Our study contains attention checks, and annotators are not paid if they fail more than two of them, which is in line with Prolific's payment policy. We explicitly state this at the beginning of our study. We are aware that labeling text with emotions can be emotional for annotators; however, the task is to assign the emotion the author felt, which decreases the cognitive effort of processing their own emotions and allows annotators to distance themselves

from the described events.

While we do not design any new models for emotion classification or analysis, we are aware that the results and research presented in our study can be used to build such systems, for instance, automated systems to detect authors' emotions in event descriptions. Prior research shows that automatic emotion detection systems are biased for multiple reasons (Kiritchenko and Mohammad, 2018), highlighting the need for careful and thoughtful experiments in the scope of emotion research.

9. Acknowledgment

This work has been supported by the German Research Foundation (DFG) in the project KL2869/1–2 (CEAT, project number 380093645) and KL2869/12–1 (EMCONA, project number 516512112).

10. Bibliographical References

Henning Bartsch, Ole Jorgensen, Domenic Rosati, Jason Hoelscher-Obermaier, and Jacob Pfau. 2023. Self-consistency of large language models under ambiguity. In *Proceedings of the 6th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pages 89–105, Singapore. Association for Computational Linguistics.

Aidan Combs, Graham Tierney, Fatima Alqabandi, Devin Cornell, Gabriel Varela, Andrés Castro Araújo, Lisa P Argyle, Christopher A Bail, and Alexander Volfovsky. 2023. Perceived gender and political persuasion: a social media field experiment during the 2020 us democratic presidential primary election. *Scientific Reports*, 13(1):14051.

Aline Étienne, Delphine Battistelli, and Gwénolé Lecorvé. 2024. Emotion identification for French in written texts: Considering modes of emotion expression as a step towards text complexity analysis. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 168–185, Bangkok, Thailand. Association for Computational Linguistics.

Simona Frenda, Gavin Abercrombie, Valerio Basile, Alessandro Pedrani, Raffaella Panizzon, Alessandra Teresa Cignarella, Cristina Marco, and Davide Bernardi. 2025. Perspectivist approaches to natural language processing: a survey. *Language Resources and Evaluation*, 59:1719–1746.

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

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, and (many others). 2024. The Llama 3 herd of models. *arXiv preprint*, arXiv:2407.21783.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *arXiv preprint*, arXiv:2310.06825.

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts. ArXiv:2401.04088 [cs].

Przemysław Kazienko, Julita Bielaniewicz, Marcin Gruza, Kamil Kanclerz, Konrad Karanowski, Piotr Miłkowski, and Jan Kocoń. 2023. Human-centered neural reasoning for subjective content processing: Hate speech, emotions, and humor. *Information Fusion*, 94:43–65.

Svetlana Kiritchenko and Saif Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 43–53, New Orleans, Louisiana. Association for Computational Linguistics.

Lea Krause and Piek T.J.M. Vossen. 2024. The Gricean maxims in NLP - a survey. In *Proceedings of the 17th International Natural Language Generation Conference*, pages 470–485, Tokyo, Japan. Association for Computational Linguistics.

Marlene Lutz, Indira Sen, Georg Ahnert, Elisa Rogers, and Markus Strohmaier. 2025. The

prompt makes the person(a): A systematic evaluation of sociodemographic persona prompting for large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 23212–23237, Suzhou, China. Association for Computational Linguistics.

Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, Pouya Tafti, Léonard Hussenot, Pier Giuseppe Sessa, Aakanksha Chowdhery, Adam Roberts, Aditya Barua, Alex Botev, Alex Castro-Ros, Ambrose Slone, Amélie Héliou, Andrea Tacchetti, Anna Bulanova, Antonia Paterson, Beth Tsai, Bobak Shahriari, Charline Le Lan, Christopher A. Choquette-Choo, Clément Crepy, Daniel Cer, Daphne Ippolito, David Reid, Elena Buchatskaya, Eric Ni, Eric Noland, Geng Yan, George Tucker, George-Christian Muraru, Grigory Rozhdestvenskiy, Henryk Michalewski, Ian Tenney, Ivan Grishchenko, Jacob Austin, James Keeling, Jane Labanowski, Jean-Baptiste Lespiau, Jeff Stanway, Jenny Brennan, Jeremy Chen, Johan Ferret, Justin Chiu, Justin Mao-Jones, Katherine Lee, Kathy Yu, Katie Millican, Lars Lowe Sjoesund, Lisa Lee, Lucas Dixon, Machel Reid, Maciej Mikuła, Mateo Wirth, Michael Sharman, Nikolai Chinaev, Nithum Thain, Olivier Bachem, Oscar Chang, Oscar Wahltinez, Paige Bailey, Paul Michel, Petko Yotov, Rahma Chaabouni, Ramona Comanescu, Reena Jana, Rohan Anil, Ross McIlroy, Ruibo Liu, Ryan Mullins, Samuel L. Smith, Sebastian Borgeaud, Sertan Girgin, Sholto Douglas, Shree Pandya, Siamak Shakeri, Soham De, Ted Klimenko, Tom Hennigan, Vlad Feinberg, Wojciech Stokowiec, Yu-hui Chen, Zafarali Ahmed, Zhitao Gong, Tris Warkentin, Ludovic Peran, Minh Giang, Clément Farabet, Oriol Vinyals, Jeff Dean, Koray Kavukcuoglu, Demis Hassabis, Zoubin Ghahramani, Douglas Eck, Joelle Barral, Fernando Pereira, Eli Collins, Armand Joulin, Noah Fiedel, Evan Senter, Alek Andreev, Kathleen Kenealy, et al. 2024. Gemma: Open models based on gemini research and technology. *arXiv preprint*, arXiv:2403.08295.

Wiktoria Mieleszczenko-Kowszewicz, Kamil Kanclerz, Julita Bielaniewicz, Marcin Oleksy, Marcin Gruza, Stanisław Woźniak, Ewa Dzięcioł, Przemysław Kazienko, and Jan Kocoń. 2023. Capturing human perspectives in NLP: Questionnaires, annotations, and biases. In *European Conference on Artificial Intelligence ECAI 2023*.

Piotr Miłkowski, Marcin Gruza, Kamil Kanclerz, Przemyslaw Kazienko, Damian Grimling, and Jan Kocoń. 2021. Personal bias in prediction of emotions elicited by textual opinions. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: student research workshop*, pages 248–259.

Piotr Miłkowski, Stanisław Saganowski, Marcin Gruza, Przemysław Kazienko, Maciej Piasecki, and Jan Kocoń. 2022. Multitask personalized recognition of emotions evoked by textual content. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pages 347–352. IEEE.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Seid Muhie Yimam, Jan Philip Wahle, Terry Lima Ruas, Meriem Beloucif, Christine De Kock, Tadesse Destaw Belay, Ibrahim Said Ahmad, Nirmal Surange, Daniela Teodorescu, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino Dario Mario Ali, Vladimir Araujo, Abinew Ali Ayele, Oana Ignat, Alexander Panchenko, Yi Zhou, and Saif Mohammad. 2025. SemEval-2025 task 11: Bridging the gap in text-based emotion detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, pages 2558–2569, Vienna, Austria. Association for Computational Linguistics.

E. Ashby Plant, Janet Shibley Hyde, Dacher Keltner, and Patricia G Devine. 2000. The gender stereotyping of emotions. *Psychology of women quarterly*, 24(1):81–92.

Flor Miriam Plaza-del-Arco, Alba Cercas Curry, Amanda Cercas Curry, and Dirk Hovy. 2024a. Emotion analysis in NLP: Trends, gaps and roadmap for future directions. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5696–5710, Torino, Italia. ELRA and ICCL.

Flor Miriam Plaza-del-Arco, Amanda Cercas Curry, Alba Curry, Gavin Abercrombie, and Dirk Hovy. 2024b. Angry men, sad women: Large language models reflect gendered stereotypes in emotion attribution. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7682–7696, Bangkok, Thailand. Association for Computational Linguistics.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association*

for Computational Linguistics, pages 1668–1678, Florence, Italy. Association for Computational Linguistics.

Taylor Sorensen, Pushkar Mishra, Roma Patel, Michael Henry Tessler, Michiel A. Bakker, Georgina Evans, Iason Gabriel, Noah Goodman, and Verena Rieser. 2025. Value profiles for encoding human variation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 2047–2095, Suzhou, China. Association for Computational Linguistics.

Huaman Sun, Jiaxin Pei, Minje Choi, and David Jurgens. 2025. Sociodemographic prompting is not yet an effective approach for simulating subjective judgments with LLMs. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 845–854, Albuquerque, New Mexico. Association for Computational Linguistics.

Enrica Troiano, Laura Oberländer, and Roman Klinger. 2023. Dimensional modeling of emotions in text with appraisal theories: Corpus creation, annotation reliability, and prediction. *Computational Linguistics*, 49(1):1–72.

Eileen Wemmer, Sofie Labat, and Roman Klinger. 2024. EmoProgress: Cumulated emotion progression analysis in dreams and customer service dialogues. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5660–5677, Torino, Italia. ELRA and ICCL.

A. Emotion and Appraisal Labels

To ensure comparability, we use the same emotion and appraisal label set at Troiano et al. (2023). This leaves us with the emotion labels 'anger', 'boredom', 'disgust', 'fear', 'guilt', 'joy', 'pride', 'relief', 'sadness', 'shame', 'surprise' and 'trust'. The appraisal label set contains the labels 'suddenness', 'familiarity', 'predict event', 'pleasantness', 'unpleasantness', 'goal relevance', 'chance responsibility', 'self responsibility', 'other responsibility', 'predict consequences', 'goal support', 'urgency', 'self control', 'other control', 'chance control', 'accept consequences', 'standards', 'social norms', 'attention', 'not consider' and 'effort'. Rather than asking for emotion categories themselves, these labels ask the annotator to evaluate the event. Emotion states can then be deduced from the evaluations. For a more thorough view on emotion and appraisal labels in NLP, see Troiano et al. (2023).

B. LLM prompt

Figure 2 shows the prompt used for either of the two experimental setups. The author description is provided in one of them and left out in the other.

```
<|system|>
You are a human annotator. Identify the
    emotion expressed in the given text.
Use ONLY emotions from the label set:
Anger, Boredom, Disgust, Fear, Guilt,
    Joy, Pride, Relief, Sadness, Shame,
    Surprise, Trust.
Do not provide any explanation, only
    output the single emotion label.

<|user|>
This is a description of the author: A
    {age}-year-old {gender} person who
    has a {education} and identifies as
     {ethnicity} in terms of ethnicity.
Text: "{text}"
What do you think the writer of the text
     felt when experiencing this event?
Emotion:

<|assistant|>
```

Figure 2: Prompt with author metadata

C. Model Details

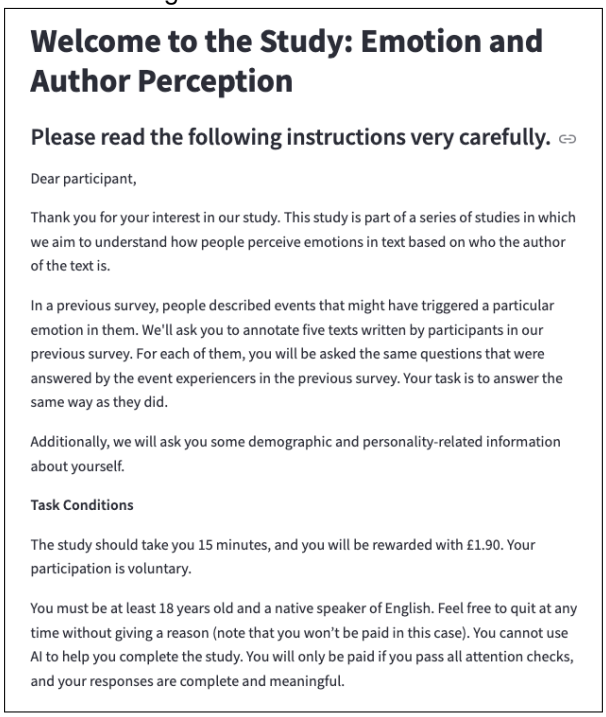
To examine the capabilities of light-weight LLMs we choose four open-weight models (Gemma 8.5B (Mesnard et al., 2024), Mistral 7.2B (Jiang et al., 2023), Mixtral 46.7B (Jiang et al., 2024), 3.2B (Grattafiori et al., 2024)). All models are hosted on our own server via Ollama[3]. We use Ollama default parameters for all models: a temperature of 0.7, token limit of 1024 and Top P sampling set to 0.95.

D. Study Design and Questionnaire

We implemented the study using the Streamlit package for python[4]. The study design is the same for both groups of annotators, except for the text presentation. The study is structured as follows: After an introductory page that contains a general description of the study topic and the participation conditions, we first ask the annotators to provide information about their current emotion and ask them to fill out a Big 5 personality test (Gosling et al., 2003). We do so because previous work (Troiano et al., 2023) has shown that these two factors can influence the way people annotate emotion in text.

[3] https://github.com/ollama/ollama
[4] https://streamlit.io/

We then show annotators texts that describe events that caused an emotion and ask the annotators which emotion the author of the text felt and how the author would have evaluated the event (appraisal). Annotators were asked to report their confidence. We found our selected scale (1 = highest, 5 = lowest) to be counterintuitive during data analysis, and therefore the annotators' responses to be unreliable and we excluded them from the analysis. Text presentation repeats five times, showing a new text each time (and for one of the two study groups, the respective author information). After this loop the annotators are asked to provide their own demographic information and optional feedback. After this they are redirected back to the annotation platform Prolific. The full study questionnaire is shown in the following screenshots.



Welcome to the Study: Emotion and Author Perception

Please read the following instructions very carefully.

Dear participant,

Thank you for your interest in our study. This study is part of a series of studies in which we aim to understand how people perceive emotions in text based on who the author of the text is.

In a previous survey, people described events that might have triggered a particular emotion in them. We'll ask you to annotate five texts written by participants in our previous survey. For each of them, you will be asked the same questions that were answered by the event experiencers in the previous survey. Your task is to answer the same way as they did.

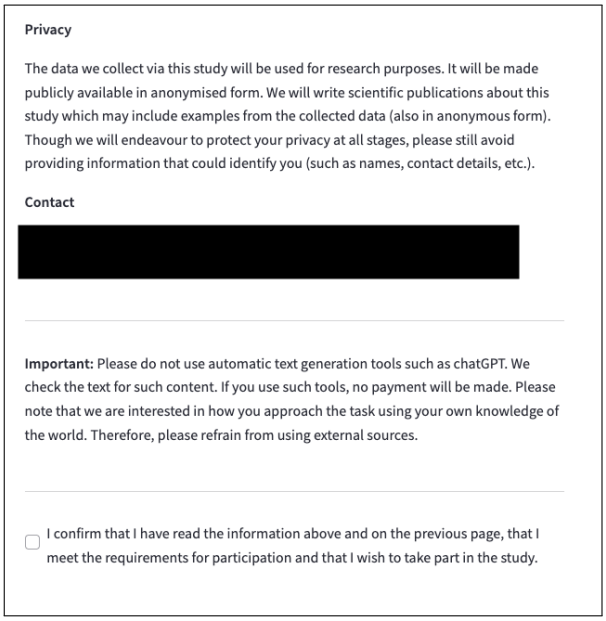
Additionally, we will ask you some demographic and personality-related information about yourself.

Task Conditions

The study should take you 15 minutes, and you will be rewarded with £1.90. Your participation is voluntary.

You must be at least 18 years old and a native speaker of English. Feel free to quit at any time without giving a reason (note that you won't be paid in this case). You cannot use AI to help you complete the study. You will only be paid if you pass all attention checks, and your responses are complete and meaningful.

Page 1



Privacy

The data we collect via this study will be used for research purposes. It will be made publicly available in anonymised form. We will write scientific publications about this study which may include examples from the collected data (also in anonymous form). Though we will endeavour to protect your privacy at all stages, please still avoid providing information that could identify you (such as names, contact details, etc.).

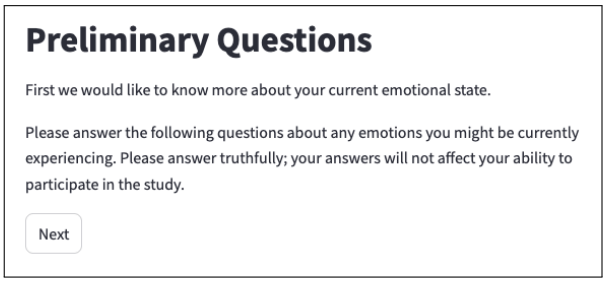
Contact

[redacted]

Important: Please do not use automatic text generation tools such as chatGPT. We check the text for such content. If you use such tools, no payment will be made. Please note that we are interested in how you approach the task using your own knowledge of the world. Therefore, please refrain from using external sources.

☐ I confirm that I have read the information above and on the previous page, that I meet the requirements for participation and that I wish to take part in the study.

Page 2



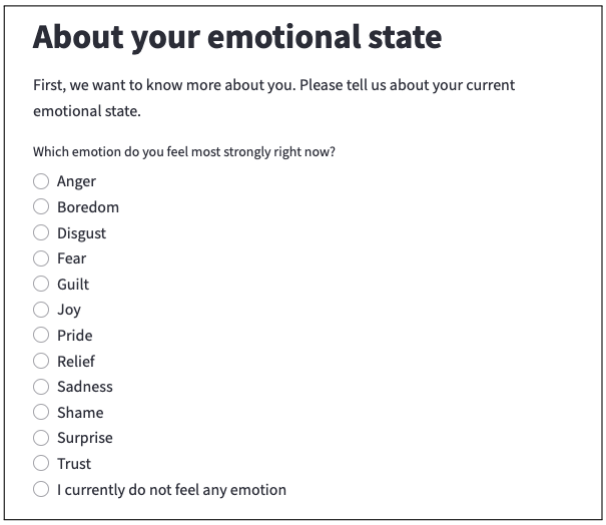
Preliminary Questions

First we would like to know more about your current emotional state.

Please answer the following questions about any emotions you might be currently experiencing. Please answer truthfully; your answers will not affect your ability to participate in the study.

Next

Page 3



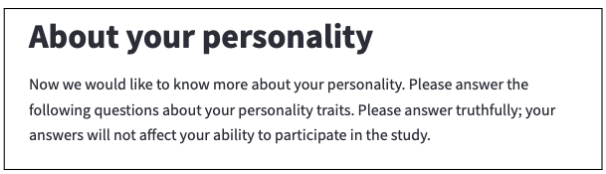
About your emotional state

First, we want to know more about you. Please tell us about your current emotional state.

Which emotion do you feel most strongly right now?

- ○ Anger
- ○ Boredom
- ○ Disgust
- ○ Fear
- ○ Guilt
- ○ Joy
- ○ Pride
- ○ Relief
- ○ Sadness
- ○ Shame
- ○ Surprise
- ○ Trust
- ○ I currently do not feel any emotion

Page 4



About your personality

Now we would like to know more about your personality. Please answer the following questions about your personality traits. Please answer truthfully; your answers will not affect your ability to participate in the study.

Page 5

About your personality

Please select a value below each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

1 = Disagree strongly	2 = Disagree moderately	3 = Disagree a little	4 = Neither agree nor disagree	5 = Agree a little	6 = Agree moderately	7 = Agree strongly

I see myself as extraverted, enthusiastic.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as critical, quarrelsome.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as dependable, self-disciplined.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as anxious, easily upset.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as open to new experiences, complex.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as reserved, quiet.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

Page 6

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as sympathetic, warm.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as disorganized, careless.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as calm, emotionally stable.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

I see myself as conventional, uncreative.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

1 = Disagree strongly	2 = Disagree moderately	3 = Disagree a little	4 = Neither agree nor disagree	5 = Agree a little	6 = Agree moderately	7 = Agree strongly

Page 6 continued

Introduction to the task

You will read five texts. These texts describe events that occurred in the life of their authors. Don't be surprised if they are not perfectly grammatical, or if you find that some words are missing. For each event, you will assess if it provoked an emotion in the experiencer, and if so, what emotion that was.

Next

Page 7

Text 1/5

This is a description of the author

```
A 48-year-old male person who has a Undergraduate
degree (BA/BSc/other) and identifies as European in
terms of ethnicity.
```

This is the event that occurred in the life of the author:

```
we were trying to sell our second home and it seemed
we were not going to actually sell it as we had no
people around in ages. Then someone viewed and offered
asking price, they were a cash buyer and the sale went
through in only 6 weeks
```

Page 8

Tell us more about the emotion.

This is a description of the author

```
A 48-year-old male person who has a Undergraduate
degree (BA/BSc/other) and identifies as European in
terms of ethnicity.
```

This is the event that occurred in the life of the author:

```
we were trying to sell our second home and it seemed
we were not going to actually sell it as we had no
people around in ages. Then someone viewed and offered
asking price, they were a cash buyer and the sale went
through in only 6 weeks
```

What do you think the writer of the text felt when experiencing this event?

- ○ Anger
- ○ Boredom
- ○ Disgust
- ○ Fear
- ○ Guilt
- ○ Joy
- ○ Pride
- ○ Relief
- ○ Sadness
- ○ Shame
- ○ Surprise
- ○ Trust

How confident are you abour your answer? 5 means very confident and 1 means not confident at all.

Page 9

Evaluation of that Experience.

This is a description of the author

```
A 48-year-old male person who has a Undergraduate
degree (BA/BSc/other) and identifies as European in
terms of ethnicity.
```

This is the event that occurred in the life of the author:

```
we were trying to sell our second home and it seemed
we were not going to actually sell it as we had no
people around in ages. Then someone viewed and offered
asking price, they were a cash buyer and the sale went
through in only 6 weeks
```

Put yourself in the shoes of the writer at the time when the event happened, and try to reconstruct how that event was perceived. How much do these statements apply? (1 means "I don't agree at all" and 5 means "I completely agree"). Please read over each of the following statements carefully and answer them using the radio buttons.

1 = Not at all 3 = Moderately 5 = Extremely

The event was sudden or abrupt.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The event was familiar to its experiencer.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer could have predicted the occurrence of the event.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Page 10

The event required an immediate response.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer was able to influence what was going on during the event.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Someone other than the experiencer was influencing what was going on.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The situation was the result of outside influences of which nobody had control.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer anticipated that he/she could live with the unavoidable consequences of the event.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The event clashed with her/his standards and ideals.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The actions that produced the event violated laws or socially accepted norms.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer had to pay attention to the situation.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer wanted to shut the situation out of her/his mind.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Page 10 continued

The event was pleasant for the experiencer.

◉ 1 ○ 2 ○ 3 ○ 4 ○ 5

The event was unpleasant for the experiencer.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

This is an attention check. Please select the value 1.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer expected the event to have important consequences for him/herself.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The event was caused by chance, special circumstances, or natural forces.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The event was caused by the experiencer's own behavior.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The event was caused by somebody else's behavior.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer anticipated the consequences of the event.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The experiencer expected positive consequences for her/himself.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Page 10 continued

This is an attention check. Please select the value 2.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

The situation required her/him a great deal of energy to deal with it.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

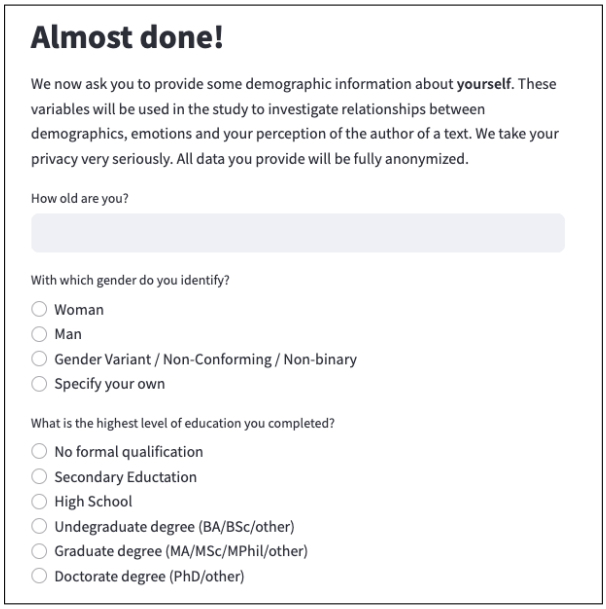
1 = Not at all 3 = Moderately 5 = Extremely

Page 10 continued

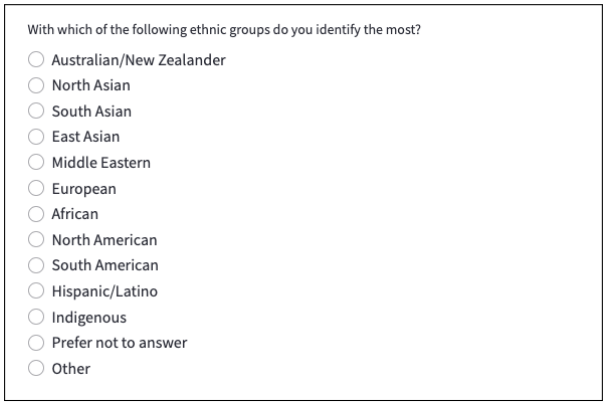
Introduction to the next text

Now, you will see a new text. Keep in mind that the text might be written by **the same author or by a different author**.

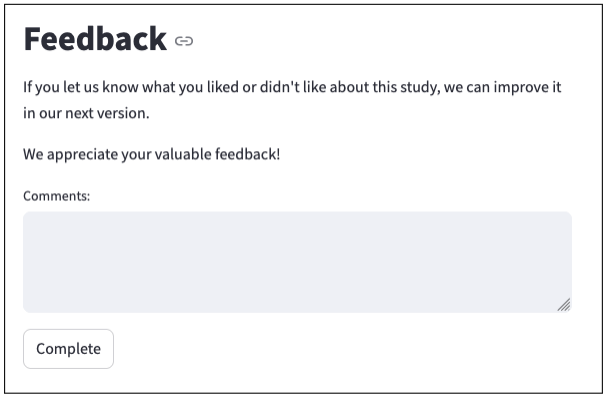
Page 11 (go to Page 8, next text)



Page 12



Page 12 continued



Page 13