The Language of Interoception: Examining Embodiment and Emotion Through a Corpus of Body Part Mentions

Sophie Wu
McGill University, Canada
sophie.wu@mail.mcgill.ca

Jan Philip Wahle
University of Göttingen, Germany
wahle@uni-goettingen.de

Saif M. Mohammad
National Research Council Canada
saif.mohammad@nrc-cnrc.gc.ca

Abstract

This paper is the first investigation of the connection between emotion, embodiment, and everyday language in a large sample of natural language data. First, we created corpora of body part mentions (BPMs) in online English text (blog posts and tweets). These include a subset featuring human annotations for the emotions of the person whose body part is mentioned in the text. Next, we show that BPMs are common in personal narratives and tweets (~5% to 10% of posts include BPMs) and that their usage patterns vary markedly by time and location. Using word–emotion association lexicons and our annotated data, we show that text containing BPMs tends to be more emotionally charged than text without any BPMs. Finally, we show a strong and statistically significant correlation between body-related language and a variety of negative health outcomes. In sum, we argue that investigating the role of body-part related words in language can open up valuable avenues of future research at the intersection of NLP, the affective sciences, and the study of human wellbeing.

1 Introduction

Embodied cognition—the theory that human cognition is rooted in bodily experiences—has gained significant traction across cognitive science, psychology, linguistics, and philosophy. This framework suggests that our bodily experiences shape not only how we interact with our physical environment, but also how we process, represent, and share abstract concepts. A wide range of disciplines have demonstrated that the intersection of cognition and language is deeply intertwined with sensorimotor experiences. For example, a growing pool of research suggests that: individuals learn new concepts better when they can use their bodies to simulate the concepts (Cook et al., 2008; Johnson-Glenberg et al., 2016); language processing involves mental simulation of physical actions (Pulvermüller, 2005; Glenberg and Kaschak, 2002); and emotions emerge from interpretations of physiological signals through a process known as interoception (Craig, 2002; Barrett, 2017).

In this paper, we ask the question: *to what extent are our embodied experiences encoded in, and reflected through, everyday language?* To this end, we use the concept of **Body Part Mentions (BPMs)**, which we define as *instances of language where words referring to parts of the body are used.*[1] Even though BPMs are a relatively simple method for detecting embodiment-related language, we propose them as a useful initial tool for investigating the relationship between language and our embodied selves in everyday language. As such, we use BPMs in this paper for an exploratory investigation of how words which are semantically related to the human body can provide interesting insights in natural language use.

We introduce two novel BPM corpora, Spinn3r$_{BPM}$ (a corpus of blog posts) and TUSC$_{BPM}$ (a corpus of tweets), and conduct a variety of experiments to show how they can be used to address three areas of inquiry on the connection between embodiment and natural language. First, we investigate the prevalence of body-part related words in everyday language, as well as how frequency of these words differs across different factors such as medium, gendered pronoun usage, time, and place. Second, we look at the relationship between BPMs and affect, motivated by the growing pool of research which posits that emotions originate from interpretations of physiological signals. Lastly, we propose and test the hypothesis that the degree of prevalence of BPMs in social media is indicative of aggregate-level health outcomes.

[1] Some linguistics paper use the term *somatic reference* for such mentions, and the term *somatic unit* or *somatic phraseology* for the vocabulary; however, *somatic reference* and *somatic expression* can be used more generally to express even non-language signals associated with the body.

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 23375–23399

A better understanding of embodiment through language could support work in a range of research areas, especially NLP in health domains (where BPM-heavy corpora are frequent (Chaturvedi et al., 2023)). the growing wave of interest in integrating embodiment within computational models of language (especially text-based LLMs, which rely on BPMs to access information about human embodiment, and struggle with many benchmarks of human cognition due to their lack of embodiment (Chemero, 2023)).

All of our code and data is available at our project repository.[2]

2 Related Work

2.1 Embodiment, Affect, and Health

According to the *theory of constructed emotion* (Barrett, 2017), emotions emerge as interpretations of our bodies' physiological signals. A key supporting argument for this theory is the wide range of research establishing a strong connection between *interoception*—the ability to feel internal bodily sensations (Craig, 2002)—and emotional wellbeing. Better interoceptive awareness has been shown to positively correlate with better emotional regulation (Zamariola et al., 2019), emotional decision-making (Dunn et al., 2010), and emotional granularity (the ability to distinguish different emotions) (Feldman et al., 2024). These results indicate that awareness of our bodily experiences is crucial to our emotional welfare and dysfunctional interoception is a contributor to a variety of mental health conditions (Khalsa et al., 2018).

Embodied experiences manifest frequently in everyday language. For example, descriptions of bodily experiences are frequently found in storytelling contexts (Gallese and Wojciehowski, 2011). But even outside of their role as explicit physical referents, BPMs may reveal a deeper connection to embodied phenomena. The *theory of conceptual metaphor* suggests that metaphors are a fundamental cognitive process (Lakoff and Johnson, 2008). According to this framework, metaphors help us understand abstract concepts by mapping them onto concrete, physical experiences. Treated this way, the widespread usage of body part words in language—even those applied in abstract contexts—are important for understanding how natural language is connected to the physical world from which it originates. Scholars in affective theory have argued that this is why expressions of intensely emotional experiences often use body parts and actions as a metaphor (e.g., *my heart is broken*, *weight lifted off my shoulders*); they are effective ways of grounding subjective experience in a shared embodied reality (Kövecses, 2003).

2.2 Embodiment and NLP

Language plays a key role in understanding the relationship between embodiment and affect. Emotional granularity is measured from the ability to identify and distinguish different emotions using words (Tan et al., 2022), and interoception is often measured through an individual's ability to describe their internal state (Desmedt et al., 2022). A range of NLP projects have taken an interest in body part words for specific applications, such as identifying gendered representations in literature (Silva et al., 2024), mapping bodily sensations for healthcare applications (Wang et al., 2019), or building computational methods for detecting body parts involved in emotional processes to improve machine emotion recognition (Zhuang et al., 2024). We note that other related work in NLP largely uses body-related language to answer adjacent research questions in specialized settings, rather than investigate the significance of BPMs themselves. Since this is the first-ever work investigating the significance of BPMs in *everyday language* using NLP methods, we aim to investigate whether the general relationship between affect and body parts suggested and observed in laboratory environments (i.e., the relationship between affect and body part words suggested by the theory of conceptual metaphor, or the relationship between affect and described bodily experiences suggested by the theory of constructed emotion), can be corroborated using textual corpora.

Many word–emotion association lexicons have been created: mostly for English: e.g., Bradley and Lang (1999), the NRC Emotion Lexicon (Mohammad and Turney, 2013, 2010), Warriner et al. (2013), The NRC VAD lexicon (Mohammad, 2018a); but also for some other languages: e.g., Moors et al. (2013) for Dutch, Võ et al. (2009) for German, and Redondo et al. (2007) for Spanish. The NRC Emotion Lexicon includes entries for whether a words is associated with eight categorical emotions: joy, sadness, fear, anticipation, anger, trust, disgust, and surprise (the Plutchik (1980)

[2] https://www.github.com/sohpei/bpms/

set). It includes entries for about 14,000 words.[3] The NRC VAD lexicon (Mohammad, 2018a) has valence, arousal, and dominance associations for about 20,000 English words.[4] (Version 2 of the NRC VAD Lexicon was released recently, and it includes entries for about 44k words and 10k multiword expressions.)

In this work, we primarily use lexicons of word–emotion associations—which can be used to create accurate emotion arcs in streams of text (Teodorescu and Mohammad, 2023) or effectively distinguish emotional granularity (Vishnubhotla et al., 2024) —as a computationally inexpensive and interpretable method for beginning to look at the relationship between affect, language, and embodiment.

3 BPM Corpora

For this work, we investigate two mediums for online, everyday language: blog posts and tweets. We consider each sentence from a blog post and each individual tweet from the tweet corpora as an instance. We created BPM corpora by extracting all instances that included at least one word referring to a body part from the Spinn3r personal blog datasets (Burton et al., 2009, 2011), and the two TUSC datasets: TUSC$_{ctry}$ (where tweets are geo-located to either the United States of America or Canada) and TUSC$_{city}$ (where tweets are geo-located to cities in North America) (Vishnubhotla and Mohammad, 2022). This results in three final BPM corpora: Spinn3r$_{BPM}$, TUSC$_{ctry-BPM}$, and TUSC$_{city-BPM}$. We will refer to the corpora made up of the rest of the instances as Spinn3r$_{noBPM}$ and TUSC$_{noBPM}$.

We compile a list of body part words by including all the terms in the list used by Zhuang et al. (2024), which extracts BPM samples from the Spinn3r corpus to annotate samples for the presence of explicitly embodied emotion.[5][6]Additionally, we include plural forms of terms (e.g., *hearts, hands, eyes,* etc.). We refer to the body part word forms as *BP word types*. The list of 295 BP word types we used is available on the project webpage.

Two issues need to be addressed when working with BPM corpora to study embodiment. First, some BP word types are ambiguous (some BPM instances may not actually be referring to a body part: e.g, 'I will be *back*'). Second, it is useful to distinguish between the speaker referring to their own body vs. the speaker referring to someone else's body. We are especially interested in mentions of one's own body as a possibly useful indicator for well-being. While we are interested in the insights that all BPMs, including those with a more abstract/metaphorical connections to the human body, can offer on the relationship between embodiment and language, it would also be useful to distinguish instances where the speaker refers directly to their own body parts. and so we use a simple solution that effectively addresses both of the issues raised above. We created three subsets for each BPM corpus that only include the BPM instances preceded by possessive pronouns 'my', 'your', and 'his/her/their', respectively, and call these instances: *possessed BPMs* This helps us create separate corpora for first-, second-, and third-person references to body parts, and also excludes a vast majority of mentions that are not explicit references to body parts. For example, 'I will be *back*' would not be included in any of these corpora, but 'my *back* hurts' will be. We find that this approach delivers a high number of references to an individual's body part (92% of 100 manually inspected instances). While this approach sacrifices recall by excluding possible references to body parts without possessive pronouns, we benefit from higher precision.[7] Additionally, we also conduct some experiments with all BPM instances (the higher-recall and lower-precision corpus).

3.1 Emotion-Annotated BPM Corpus

We also leverage a specialized subset of Spinn3r, previously released by Zhuang et al. (2024). This subset contains sentences with BPMs annotated for the presence of explicitly embodied emotion, which is defined as "the physical experience of an emotion via our body" (i.e., *"Julie pouted and rolled her eyes"* is annotated as containing embodied emotion, but *"Frank breathed heavily through his mouth after his run"* is not). We refer to this subset in the rest of this paper as Spinn3r$_{BPM-Zhuang}$. We extend this work by creating the first human-annotated dataset that explicitly identifies BPM

[3]http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[4]http://saifmohammad.com/WebPages/nrc-vad.html

[5]https://www.collinsdictionary.com/us/word-lists/body-parts-of-the-body

[6]https: //www.enchantedlearning.com/wordlist/body.shtml

[7]For our experiments, we do not need all BPM instances, but rather just a large sample.

ownership (whether the BPM refers to the speaker's body or not) and the emotion of the BPM owner (joy, fear, etc.) as inferred by a human reader.

For full details of all datasets and data collection methods see Appendix A. Further details for the emotion annotation process, which includes a set of quality checks and aggregation methods for final emotion scores, can be found in Appendix B.

4 Research Questions About BPMs

Despite substantial evidence from medical research and psychology that points to connections between the mind and the rest of the body, as well as the connection between interoception and emotional well-being, there is little quantitative work using language to explore this connection. In this section and the next, we make use of large amounts of social media data, massive word–emotion lexicons, and the emotion-annotated corpus described in the previous section, to examine questions on how, when, and in what context we refer to our body parts in text (this section) and whether different body parts mentions tend to be used in different emotional contexts (next section). Since the questions in this section are relevant to the **B**ody, we will index them as B1, B2, etc. The questions in the next section are related to **B**ody and **A**ffect, so we will index them as BA1, BA2, etc.

B1. To what extent do we use body-related words? While it may be difficult to get natural conversational data for privacy reasons, how often do we mention body parts in social media?
Method: We did not find any past research on how common BPMs are in language. We do not even have a sense of the magnitude: do they occur in 0.01%, 0.0001%, 1%, 10%, 60% of utterances, or something else? To examine the extent to which body-related vocabulary is used in online language, we calculated the number of instances containing at least one BPM.
Results: See Table 1. We find that a substantial proportion of instances contain at least one BPM: 10.4% of blog post sentences in the Spinn3r dataset, 6.4% and 7.3% of tweets in in TUSC_{ctry} and TUSC_{city}, respectively.
Discussion: The consistently high proportion of BPM-containing instances across our corpora emphasizes the ubiquity of body part references in online English text. The markedly larger proportion of samples containing BPMs in blog sentences than tweets implies that usage of BPMs may be more pervasive in personal narratives and longer-form text than in tweets.

Corpus	**S**	$\mathbf{T}_{city}$	$\mathbf{T}_{ctry}$
Instances	*(80,379)*	*(104,575,991)*	*(3,181,879)*
<*BPM*> instances	10.4	6.4	7.3

Table 1: B1 - Percentage of instances in each corpus with at least one BPM. S = Spinn3r. T = TUSC.

Corpus	$\mathbf{S}_{bpm}$	$\mathbf{T}_{city\_bpm}$	$\mathbf{T}_{ctry\_bpm}$
Instances	*(8,371)*	*(6,710,660)*	*(231,577)*
Possessed BPM	28.9	31.9	26.3
“my <BPM>”	16.6	19.2	15.8
your <BPM>	6.5	6.6	5.5
his <BPM>	2.5	3.3	2.7
her <BPM>	2.2	1.6	1.4
their <BPM>	1.2	1.3	1.0

Table 2: B2 - Percentage of instances containing a possessed BPM out of overall BPM samples.

B2. To what extent do we talk about our own body parts (i.e., my BPM) versus others' body parts (i.e., your/his/her/their BPM)?
Method: Although BPMs can be used to study the general significance of body part words in language, we are also interested in examining the extent to which these body parts can be attributed to a particular person (e.g., *her heart*) vs. body part words that are not attributed to a human possessor (e.g., *the heart of the matter*). To do this, we introduce the concept of *possessed BPMs*, which can be attributed to a particular body using a possessive pronoun. We look at three general categories of possessed BPMs: first person instances including *“my <BPM>”*, second person instances including *“your <BPM>”*, and third person instances including *“his/her/their <BPM>”*. In each corpus, we determine the frequency of instances containing at least one instance of a possessed BPM.
Results: Table 2 shows the results. In each corpus, the most common possessed BPM is *“my <BPM>”*, and there are more instances containing *“his <BPM>”* than *“her <BPM”*.
Discussion: By evaluating different distributions of possessed BPM types, we can begin building an understanding of *whose* body parts are most often referred to in conversation. Our results indicate that individuals are more likely to discuss their own bodily experiences in online discourse than that of others. The higher frequencies of *“his <BPM>"* over *“her <BPM>"* instances are also interesting, considering well-established theories that women's bodies are more heavily discussed and scrutinized

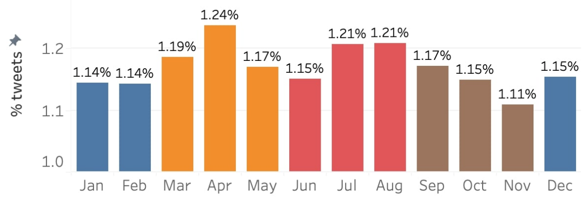

Figure 1: B4 - TUSC$_{ctry}$ - % of tweets with at least one *“my <BPM>”* by month. Colored by season in USA.

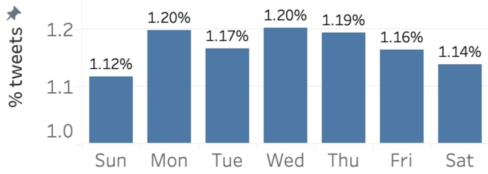

Figure 2: B4 - TUSC$_{ctry}$ - % of tweets with at least one *“my <BPM>”* for different weekdays.

in popular media (Bordo, 2023)—our results indicate that in spite of this, men’s bodies may be referred to more often in everyday speech.

B3. Which of our body parts do we refer to most often? Do we refer to our body differently in different online contexts?

Method: To answer this, we calculate the frequencies of **each individual** BP word type preceded by the possessive pronoun "my" (i.e., *my <BPM>*).

Results: We find that there are certain BP word types that appear very frequently in all corpora—with twelve *“my <BPM>”* word types being shared in the top twenty across all corpora. However, we also observe variation in frequencies across corpora, indicating that we describe our body in different ways across different online contexts. Across all corpora, *my heart* and *my head* are among the most frequently mentioned *“my <BPM>”* word types. These body parts are likely central sources for people’s basic understandings of their embodied experiences, which is reflected in the prevalence of common figurative expressions such as *my heart is broken* and *my head hurts*. We also find that the blog dataset has a much stronger representation of body parts that are strongly related with human senses, such as *my eyes* (10.23% vs. 1.40%), *my ears* (1.19% vs. <0.1%), and *my hands* (3.32% <0.1%). This suggests that the personal narratives in blogs may be more focused on sensory, everyday experiences. Additionally, *my hair* and *my face* appear much more frequently in the TUSC tweet datasets than in the Spinn3r blog datasets, likely a result of personal grooming and appearance being more prevalent in social media updates. This rich divergence between common *“my <BPM>”* word types implies that users refer to their body differently when expressing themselves in different online mediums. (Table 5 in the Appendix shows the top 20 *“my <BPM>”* types in each corpus.)

B4. Does the time of the week/year impact the extent to which we refer to our body?

Method: Each sample in the TUSC$_{ctry}$ dataset has a timestamp indicating the exact time at which it was posted. We use this data to examine whether *“my <BPM>”* usage is higher or lower at different times.

Results: Frequency of instances containing *“my <BPM>”* by day of week and month are shown in Figures 1 and 2. We find that *“my <BPM>”* instances peak during the summer and spring, decline steadily during the fall, and then stay relatively low during the winter. We also find that *"<BPM>"* instance frequency is highly dependent on different days of the week, rising steadily from Sunday to Wednesday and then declining from Wednesday to Saturday. We find a statistically significant decline in BPM usage from April onwards in the year and from Monday onwards throughout the week. More details on statistical tests can be found in Appendix H.

Discussion: The seasonal differences in *“my <BPM>”* usage in warmer months indicates that factors such as temperature, sunlight, and time spent outside could affect awareness and expression of one’s bodily experiences. The weekly rise of referral to one’s own body parts may reflect a renewed engagement with structured activities as the work week begins, while the decline could indicate fatigue or decreased energy as the week progresses, making it difficult to cultivate bodily awareness, consistent with documented patterns of weekly fatigue cycles in organizational research (Zijlstra and Rook, 2008). These results indicate that embodied language use is not static, but responds to environmental and social rhythms.

B5. Do individuals in different regions refer to their bodies at different frequencies?

Method: We take advantage of the geotagged meta data available for the TUSC tweets to evaluate the regional proportion of *“my <BPM>”* tweets.

Results: Figure 3 shows BPM use by city. We find that *“my <BPM>”* instances are used more in

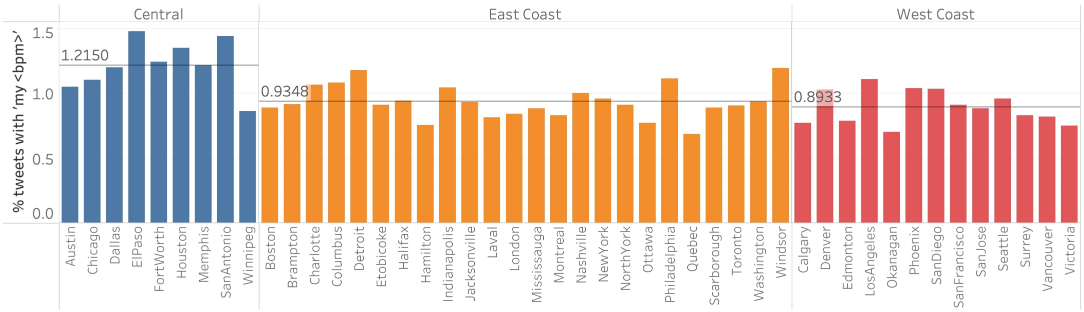

Figure 3: B5 - TUSC$_{city}$ - % of tweets with at least one “my <BPM>” for different cities.

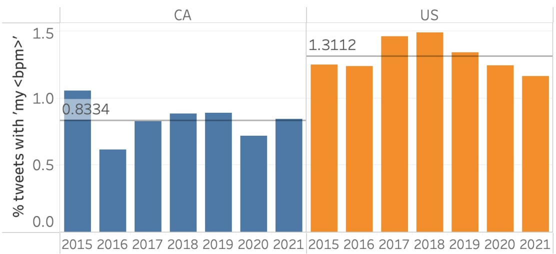

Figure 4: B5 - TUSC$_{ctry}$ - % of tweets with at least one “my <BPM>” for Canada and USA from 2015 to 2021.

central cities than coastal cities. We also find that *“my <BPM>”* instances are more frequent in American tweets than Canadian tweets in the TUSC$_{ctry}$ dataset, as shown in Figure 4 (TUSC$_{city}$ dataset). We find also find that the city of a post has a statistically significant effect on the usage of BPMs in a city. We also find that country has a slightly smaller but still statistically significant effect on usage, but that this pattern also persists across the different months represented in our dataset. Details on statistical tests, as well as a figure showing that usage difference between USA and Canada is consistent across months, can be found in Appendix D.

Discussion: These findings suggest that regional differences influence how individuals refer to their bodies, potentially reflecting broader cultural, social, or environmental factors. Future research could explore how variables such as climate, healthcare access, or local discourse shape how individuals discuss their body in different regions.

5 Research Questions on BPMs–Affect

The primary goal of our work is to explore how language can shed light on the connection between the body, emotion, and well-being. In this section we explore how BPMs are associated with emotions. We explore this question using emotions associated with words that co-occur with BPMs (using large word–emotion association lexicons) as well as perceived emotions of the speaker (using the new human-annotated Spinn3r$_{BPM-Zhuang}$ dataset we introduced earlier). In this section, we take a special interest in samples including *"my <BPM>"*, since our emphasis is rooted in theories of embodied emotion and health psychology, which suggest that references to one’s own body are more likely to reflect internal emotional and physical states.

BA1. Do posts with body part mentions have markedly different emotional associations?

Method: This question aims to shed light on whether the relationship between emotion and embodiment manifests in social media text. In this experiment, we look at the proportion of samples (tweets/blog posts) containing at least one word associated with various emotion categories: *anger, anticipation, disgust, fear, joy, sadness, surprise, and trust* from Plutchik’s set of emotions (Plutchik, 2001), and high or low *valence (positive–negative), arousal (calm–sluggish), and dominance (in control–out of control)*. We obtain the word–emotion associations from the NRC Emotion Lexicon (Mohammad and Turney, 2013, 2010) and the NRC VAD Lexicon (Mohammad, 2018b). We compute these proportions for various BPM categories: *“my <BPM>”*, *"his/her/their <BPM>"*, *"your <BPM>"*, as well as the no BPM corpora.

Results: Figure 5 shows the results for valence, arousal and dominance. We find that instances containing BPMs have higher percentages of emotion-associated co-occurring words than instances not containing BPMs. This is true across 36/42 corpus–dimension pairs. We also observe that the jump in scores from the no BPM corpus to BPM corpora is highest for the low-valence and low-dominance dimensions ($\sim$15 percentage points). We find that the BPM category has a statistically significant effect on the percentage of samples containing emotion-associated terms (more details on these tests can be found in Appendix I).

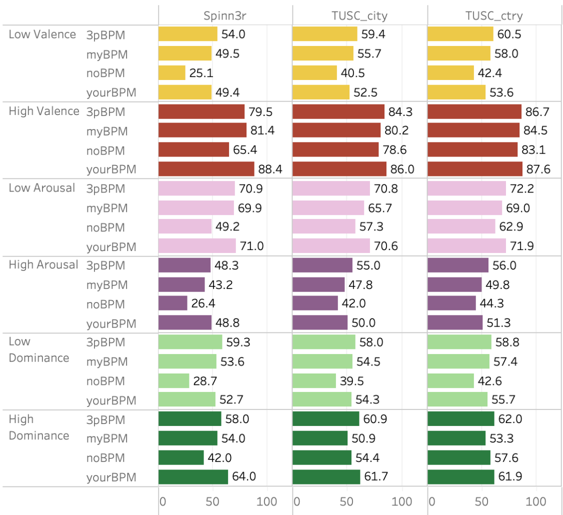

Figure 5: BA1 - Percentage of sentences with at least one high or low valence, arousal, or dominance word (according to the NRC VAD lexicon) in each corpus in myBPM, yourBPM, 3pBPM, and noBPM categories.

Discussion: Referral to one's own body seems to display a strong co-occurrence with emotion-associated language, supporting theories of the connection between embodiment and emotion. When people discuss their own bodies, they tend to use more negative (low valence) emotional language and express less control (low dominance), suggesting these self-references often occur in contexts of pain or powerlessness. In Figure 17 in the Appendix, we demonstrate that the most frequent words associated with BPMs in our corpora support this theory, such as *hurt, sore,* and *sick*.

BA2. What is the impact of explicitly embodied emotion on the emotions expressed through body part mentions?

Method: Body parts are often referenced as physically involved in emotional responses (e.g., *my heart skipped a beat*, *my stomach dropped*). Prior work (i.e. Zhuang et al.) has assumed that such samples in natural language may be responsible for the emotional associations between emotional and body-related language. In BA1, we showed that BPM instances are more likely to have emotion-associated co-terms than their no-BPM counterparts. With this question we explore whether this increase only exists in explicitly embodied emotion. To do this, we analyze samples which are human-annotated by Zhuang et al. (2024) as either containing embodied emotion (where a body part physically participates in expressing the emotion—annotated as *embodied*) or not (annotated as *non-embodied*). Specifically, we looked at the degree of emotion-word co-occurrences in the embodied samples and in the non-embodied samples. As a separate and complementary experiment, we manually annotated these exact instances for whether the speaker was feeling any of the emotion categories (anger, fear, joy, sadness, surprise, and trust), by crowdsourcing on Amazon Mechanical Turk. We note a high inter-annotator agreement rate on our human annotations using Split-Half Class Match Percentage (90.1% on 2 bins to 69.2% on 10 bins). Further details on annotator agreement can be found in Appendix B. This allows us to determine the extent to which explicitly embodied BPMs are more or less emotional than not explicitly embodied BPM instances (non-embodied for short).

Results: Figure 6 shows the percentages of embodied and non-embodied samples where the speaker is experiencing an emotion. Figure 7 shows the percentages of embodied and non-embodied samples that include at least one word associated with an emotion. We find that both methods indicate no notable difference in the percentages of emotional samples across explicitly embodied versus non-embodied (or, more precisely, not *explicitly* embodied) BPM instances (<4 percentage points). In contrast, there is a stark difference between the the emotion percentages of the no BPM samples and the embodied/non-embodied BPM samples) (>30 percentage points).

Discussion: The results show that emotional words appear more frequently in BPM sentences, regardless of whether or not the BPMs are explicitly written as physically connected to emotion in in the text (explicitly embodied). Speakers are also equally likely to be expressing emotion whether the BPM is embodied or not. This unexpected finding supports the theory of constructed emotion and highlights a stronger connection between emotional expression and body-related language than theorized by previous works.

BA3. Do individual body part mentions co-occur with markedly different emotion distributions?

Method: In BA1, we looked at the co-occurrence of posts containing *"my <BPM>"* with emotion-associated words. Here, we are interested in comparing this average score to posts containing specific *"my <BPM>"* types. For the most common *"my <BPM>"* types that occur in at least 100 instances across each corpus, we calculated the aver-

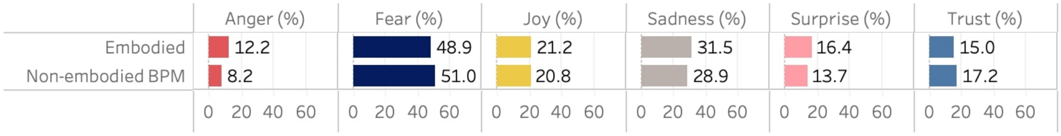

Figure 6: BA2 - Spinn3r$_{BPM-Zhuang}$ - Percentages of embodied and non-embodied samples where the speaker is experiencing an emotion.

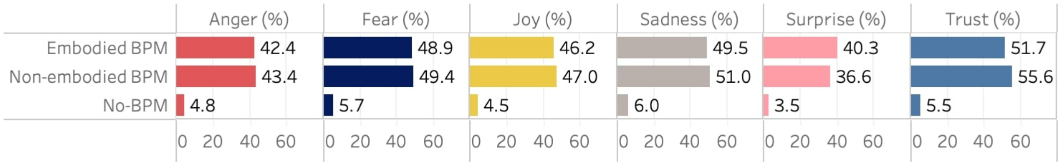

Figure 7: BA2 - Spinn3r$_{BPM-Zhuang}$ - Percentages of embodied and non-embodied samples that include at least one word associated with an emotion.

age proportion of instances that include words associated with specific emotion dimensions. We calculated a mean and standard deviation over these proportions and use these values to find *"my <BPM>"* types which are significantly associated with particular emotional dimensions. We also calculated the standard deviation for each emotional category across all common *"my <BPM>"* types to identify which types are markedly associated with certain emotional dimensions.

Results: We find that different *"my <BPM>"* types are associated with different emotions to markedly different degrees, and that different profiles of associations for the same type can be found in different corpora. However, some *"my <BPM>"* types carry consistent cross-corpus associations, such as *my stomach* being most associated with sadness in both TUSC$_{ctry-BPM}$ and Spinn3r$_{BPM}$, whereas *my chest* is most associated with anger. (Proportions for emotional word co-occurrence across *"my <BPM>"* types for TUSC$_{ctry-BPM}$ and Spinn3r$_{BPM}$ are shown in Figures 13 through 16 and the most associated emotion for each of the BPMs—which are often negative—are shown in Table 8 in the Appendix.)

Discussion: These results indicate that referral to one's body parts are associated with different affective expressions online. Overall trends in $TUSC_{city}$ seem to imply that referral to one's own body parts online often arise from situations of pain, lethargy, and a lack of control.

6 Do BPMs Correlate with Health?

The previous two sections show that BPMs are common in online text and they exhibit many systematic and consistent trends across time and region, as well as w.r.t. co-occurrence with emotion words. These results are consistent with what we would expect if BPMs are linguistic indicators of one's health. In this section, we directly explore whether, at an aggregate level, the degree of BPMs in social media texts correlates with health outcomes. We hypothesize that this occurs because BPMs are frequently used online by individuals to express pain or discomfort in their bodies. If so, regional discrepancies in BPM usage may also be correlated with different health outcomes.

Method: To evaluate this hypothesis, we look at available city-wide health data (City-Health-Dashboard-Dataset, 2025) for all 25 American cities in the TUSC$_{city}$ dataset, and correlations between the proportion of regional tweets containing *"my <BPM>"/"BPM"* and four health measures: *frequent mental distress, frequent physical distress, life expectancy, and physical inactivity*.[8] As points of baseline comparison, we also look at how the health factors investigated are correlated with the number of tweets from each region, and the correlation between the proportion of emotion-associated words (from the NRC Emotion/VAD lexicons) with the health outcomes studied.

Results: Table 3 shows the Spearman rank correlations as well the p-values (we consider the correlations to be statistically significant if the p-value is below 0.05). Observe that the number of tweets per city is not correlated with the health outcomes (See Row A). We find that most emotion–health outcome pairs are also not correlated or only slightly correlated. The highest correlation numbers are for fear–physical activity (See Row B). (Table 9 in the Appendix shows correlations for each of the

[8] https://www.cityhealthdashboard.com

	Freq. Mental Distress		Freq. Phys. Distress		Life Expectancy		Physical Inactivity	
	Spearman's r	p-value	Spearman's r	p-value	Spearman's r	p-value	Spearman's r	p-value
a. Number of tweets	-0.170	0.418	-0.167	0.425	0.290	0.160	-0.243	0.242
b. Prop. of <*Fear word*> tweets	-0.230	0.231	-0.370	0.054	0.160	0.403	**-0.460**	**0.014**
c. Prop. of *"my <BPM>"* tweets	**0.497**	**0.012**	**0.721**	**0.000**	**-0.409**	**0.043**	**0.704**	**0.000**
d. Prop. of *"<BPM>"* tweets	**0.527**	**0.007**	**0.553**	**0.004**	**-0.613**	**0.001**	**0.539**	**0.006**

Table 3: Health Outcomes - TUSC$_{city}$ - Spearman's r correlation and p-values showing the relationship between different health outcomes across cities and various features drawn from tweets from those cities. **Bolded** values indicate statistically significant correlations at $p < 0.05$.

emotion–health outcome pairs.) In contrast, the proportion of *"<BPM>"* and *"my <BPM>"* mentions (rows c and d) are moderately or strongly correlated with all three negative health outcomes and anticorrelated with life expectancy (statistically significant correlations are bolded). Notably, frequent physical distress and physical inactivity are remarkably correlated with higher myBPM usage (Spearman's r = 0.721, Spearman's r = 0.704 respectively), and life expectancy is strongly negatively correlated with BPM use (Spearman's r = -0.613). Overall, these results show that simple metrics capturing the proportion of mentions of body parts in social media can be useful indicators of both physical and mental health. While our findings reveal a statistically significant correlation between body part mentions (BPMs) and regional health outcomes, we do not claim a direct causal relationship. It is likely that both language use and health indicators are shaped by broader social and demographic factors—such as educational attainment, economic status, or regional linguistic norms—which may contribute to the observed patterns. We see this possibility as an exciting result – rather than treating BPMs as independent predictors of health, we interpret these correlations as evidence of shared variance that may offer insight into the sociolinguistic embedding of embodied experience. We highlight these associations as a starting point for future research that more directly models such confounding factors.

7 Conclusion

We created novel corpora designed specifically for the study of Body Part Mentions (BPMs), which includes the first-ever dataset of samples explicitly annotated for the emotions of human entities possessing BPMs. Using these corpora, we answered a series of research questions on the significance of body-related words in everyday language, the relationship between embodiment and emotion, and factors correlated with BPM frequency such as emotional context, time of week/year, and region. We showed that BPMs occur frequently in social media texts and have notable temporal and geographic trends. We also showed that BPM instances have markedly higher emotion associations than non-BPM instances—with an especially marked increase in low valence (negativity) and low dominance (helplessness) instances. Most notably, through experiments on data from 25 US cities, we showed that the degree of BPM usage can is a powerful indicator of aggregate-level wellbeing. Although the connection between language, embodiment, and affect is now well-established, this paper is – to our knowledge – the first-ever approach to understanding this relationship grounded in large amounts of language data. We release our BPM corpus to the public, and hope that our work demonstrates body-related language as a rich and interesting source of material for future NLP research to investigate the deeper connection between language, embodiment, and emotional wellbeing. Although BPMs are a relatively simple tool for investigating the relationship between embodiment and everyday language, they offer a scalable, interpretable signal for understanding this connection empirically. We hope that our initial exploratory work, through showing that body-related language carries diverse and meaningful associations, emphasizes the richness of studying the intersection of embodiment and natural language.

Limitations

Our work introduces the relevance of BPMs to NLP, and we argue for BPMs as a source of interesting research by demonstrating that their usage is correlated to the presence of emotional expression on social media as well as certain indicators of physical health and emotional wellbeing. But since we focus on BPMs occurring in a specific medium (online social media, specifically blog posts and tweets), much remains to be discovered about how body part words – and their relationship to everyday language and affect – manifest differently in other contexts.

Cultural and linguistic backgrounds significantly influence how people express emotions. Additionally, social media platforms and other digital communication channels produce unique language use patterns that may not reflect everyday language use in other environments (i.e., spoken conversation). We hope that in the future, other researchers can consider the relevance and limitations of producing BPM lists and conducting similar experiments in other languages and with other datasets. This can both extend our general knowledge of embodiment within language as well as help us consider the ways in which our results may differ in other linguistic contexts.

Since the primary aim of this paper is descriptive rather than explanatory to highlight that there are diverse and meaningful associations which BPMs carry, warranting future study, we do not isolate any specific mechanisms (i.e. effects of social variables, classification of BPM usage types, or deeper linguistic analysis) which could explain these associations. We encourage future work which can further probe the exact causes and explanations for the associations which we have discovered in this paper.

For this reason, we chose to use lexicon-based approaches to studying affect in our corpus, since although they cannot capture the full nuance of emotional expression in a single sample, especially in figurative or context-dependent language, they remain valuable tools for capturing broad trends across large corpora, and their interpretability makes them suitable for an exploratory, large-scale study like ours. such as ML-based classifiers, could possibly help provide more fine-grained and context-sensitive emotion detection for more specific and contextual research questions (compared to the more general and exploratory research questions we attempt to answer in this paper).

Ethics Statement

Our approach, as with any other data-driven approach to affective science/emotional wellbeing, should be considered an *aggregate-level indicator* rather than a biomarker for individual's affective states (Guntuku et al., 2017). The measures we introduce for evaluating body part related words in everyday language, as well as their relationships to aspects of emotional and physical health should not be used as standalone indicators of these factors. Instead, they should be an additional metric that can be used in conjunction with a myriad of other investigative tools. This is especially important considering the diverse ways in which different individuals use words in everyday speech. Further best practices for ethical applications of emotional lexicons can be seen here: (Mohammad, 2022).

We also note that conceptions of emotion and wellbeing, especially as expressed through language, are heavily influenced by culture and linguistic variance (Barrett and Lindquist, 2008). Interpretations of affective language may differ not only across languages but also within communities and individuals, shaped by socio-cultural norms, lived experiences, and context. As such, any claims or insights drawn from our analysis should be situated within a broader understanding of cultural and linguistic diversity, and we caution against universalizing interpretations without further cross-cultural validation.

Acknowledgments

Many thanks to Tara Small for helpful discussions and comments. We also thank Lisa Pennel for her feedback and support on this paper.

References

Lisa Feldman Barrett. 2017. The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1):1–23.

Lisa Feldman Barrett and Kristen A Lindquist. 2008. The embodiment of emotion. *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*, pages 237–262.

Susan Bordo. 2023. *Unbearable weight: Feminism, Western culture, and the body*. Univ of California Press.

Margaret M Bradley and Peter J Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida.

Kevin Burton, Akshay Java, and Ian Soboroff. 2009. The icwsm 2009 spinn3r dataset. In *Third Annual Conference on Weblogs and Social Media (ICWSM 2009)*, San Jose, CA. AAAI.

Kevin Burton, Akshay Java, and Ian Soboroff. 2011. The icwsm 2011 spinn3r dataset. In *Proceedings of the Annual Conference on Weblogs and Social Media (ICWSM 2011)*. AAAI.

Jaya Chaturvedi, Natalia Chance, Luwaiza Mirza, Veshalee Vernugopan, Sumithra Velupillai, Robert Stewart, Angus Roberts, et al. 2023. Development of a corpus annotated with mentions of pain in mental health records: Natural language processing approach. *JMIR Formative Research*, 7(1):e45849.

Anthony Chemero. 2023. Llms differ from human cognition because they are not embodied. *Nature Human Behaviour*, 7(11):1828–1829.

City-Health-Dashboard-Dataset. 2025. Nyu langone health, v12.03.2024. *NYU School of Medicine*.

Susan Wagner Cook, Zachary Mitchell, and Susan Goldin-Meadow. 2008. Gesturing makes learning last. *Cognition*, 106(2):1047–1058.

Arthur D Craig. 2002. How do you feel? interoception: the sense of the physiological condition of the body. *Nature reviews neuroscience*, 3(8):655–666.

Olivier Desmedt, Alexandre Heeren, Olivier Corneille, and Olivier Luminet. 2022. What do measures of self-report interoception measure? insights from a systematic review, latent factor analysis, and network approach. *Biological Psychology*, 169:108289.

Barnaby D Dunn, Hannah C Galton, Ruth Morgan, Davy Evans, Clare Oliver, Marcel Meyer, Rhodri Cusack, Andrew D Lawrence, and Tim Dalgleish. 2010. Listening to your heart: How interoception shapes emotion experience and intuitive decision making. *Psychological science*, 21(12):1835–1844.

MJ Feldman, E Bliss-Moreau, and KA Lindquist. 2024. The neurobiology of interoception and affect. *Trends in Cognitive Sciences*.

Vittorio Gallese and Hannah Wojciehowski. 2011. How stories make us feel: Toward an embodied narratology. *California Italian Studies*, 2(1).

Arthur M Glenberg and Michael P Kaschak. 2002. Grounding language in action. *Psychonomic bulletin & review*, 9(3):558–565.

Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49.

Mina C Johnson-Glenberg, Colleen Megowan-Romanowicz, David A Birchfield, and Caroline Savio-Ramos. 2016. Effects of embodied learning and digital platform on the retention of physics content: Centripetal force. *Frontiers in psychology*, 7:1819.

Sahib S Khalsa, Ralph Adolphs, Oliver G Cameron, Hugo D Critchley, Paul W Davenport, Justin S Feinstein, Jamie D Feusner, Sarah N Garfinkel, Richard D Lane, Wolf E Mehling, et al. 2018. Interoception and mental health: a roadmap. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 3(6):501–513.

Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.

George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

Saif Mohammad. 2018a. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia. Association for Computational Linguistics.

Saif Mohammad. 2018b. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 174–184.

Saif Mohammad and Peter Turney. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Los Angeles, CA. Association for Computational Linguistics.

Saif M Mohammad. 2022. Best practices in the creation and use of emotion lexicons. *arXiv preprint arXiv:2210.07206*.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Agnes Moors, Jan De Houwer, Dirk Hermans, Sabine Wanmaker, Kevin Van Schie, Anne-Laura Van Harmelen, Maarten De Schryver, Jeffrey De Winne, and Marc Brysbaert. 2013. Norms of valence, arousal, dominance, and age of acquisition for 4,300 dutch words. *Behavior research methods*, 45(1):169–177.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*, 1(3):3–33.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350.

Friedemann Pulvermüller. 2005. Brain mechanisms linking language and action. *Nature reviews neuroscience*, 6(7):576–582.

Jaime Redondo, Isabel Fraga, Isabel Padrón, and Montserrat Comesaña. 2007. The spanish adaptation of anew (affective norms for english words). *Behavior research methods*, 39(3):600–605.

Mariana O Silva, Luiza de Melo-Gomes, and Mirella M Moro. 2024. From words to gender: Quantitative analysis of body part descriptions within literature in portuguese. *Information Processing & Management*, 61(3):103647.

Tse Yen Tan, Louise Wachsmuth, and Michele M Tugade. 2022. Emotional nuance: Examining positive emotional granularity and well-being. *Frontiers in psychology*, 13:715966.

Daniela Teodorescu and Saif M Mohammad. 2023. Evaluating emotion arcs across languages: Bridging the global divide in sentiment analysis. *arXiv preprint arXiv:2306.02213*.

Krishnapriya Vishnubhotla and Saif M. Mohammad. 2022. Tweet Emotion Dynamics: Emotion word usage in tweets from US and Canada. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4162–4176, Marseille, France. European Language Resources Association.

Krishnapriya Vishnubhotla, Daniela Teodorescu, Mallory J Feldman, Kristen A Lindquist, and Saif M Mohammad. 2024. Emotion granularity from text: An aggregate-level indicator of mental health. *arXiv preprint arXiv:2403.02281*.

Melissa LH Võ, Markus Conrad, Lars Kuchinke, Karolina Urton, Markus J Hofmann, and Arthur M Jacobs. 2009. The berlin affective word list reloaded (bawl-r). *Behavior research methods*, 41(2):534–538.

Yipei Wang, Xingyu Fan, Luoxin Chen, Eric I-Chao Chang, Sophia Ananiadou, Junichi Tsujii, and Yan Xu. 2019. Mapping anatomical related entities to human body parts based on wikipedia in discharge summaries. *BMC bioinformatics*, 20:1–11.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4):1191–1207.

Giorgia Zamariola, Nollaig Frost, Alice Van Oost, Olivier Corneille, and Olivier Luminet. 2019. Relationship between interoception and emotion regulation: New evidence from mixed methods. *Journal of Affective Disorders*, 246:480–485.

Yuan Zhuang, Tianyu Jiang, and Ellen Riloff. 2024. My heart skipped a beat! recognizing expressions of embodied emotion in natural language. In *Proceedings of NAACL 2024 (Volume 1: Long Papers)*, pages 3525–3537, Mexico City, Mexico. Association for Computational Linguistics.

Fred RH Zijlstra and John W Rook. 2008. The weekly cycle of work and rest: A diary study. In *Time in Organizational Research*, pages 84–101. Routledge.

Appendix

A Full Dataset Descriptions

All relevant datasets and subsets to this paper can be viewed in Table 4.

B Obtaining human ratings of emotion for Spinn3r$_{BPM-Zhuang}$ corpus

The crowd-sourced annotations presented in this paper were approved by our Institutional Research Ethics Board. About 52% of the annotators were male and about 48% female, with average age of annotators being 39. Our final data collection process stored no information about annotator identity and as such there is no privacy risk to them. The annotators were free to do as many word annotations as they wished. The instructions included a brief description of the purpose of the task as well.

The key steps in producing the emotion annotation for this are:

1. developing the questionnaire for emotion annotation
2. developing measures for quality control (QC)
3. annotating instances on the crowdsource platform (Amazon Mechanical Turk)
4. discarding data from outlier annotations
5. aggregating data from multiple annotators to determine final scores for each emotion

We annotate the Spinn3r$_{BPM-Zhuang}$ corpus, taken from (Zhuang et al., 2024), for the presence of six emotions. – *joy, fear/anxiety, sadness, anger, disgust, and trust* from Plutchik's wheel of basic emotions (Plutchik, 2001).

For each instance, we identify all possible BPMs. For each sample presented, we ask the crowd-worker to identify whether a BPM in the sample belongs to the "speaker" or the "non-speaker". We also present them a description of the emotions we will want annotations for (See Figure 8). For example, in the sentence "Robin placed her hand on Kevin's shoulder", we would tell the annotator to identify the owner of "hand" or "shoulder", which would lead them to annotate for the emotion of Robin or Kevin respectively. We also note that all samples from the Spinn3r$_{BPM-Zhuang}$ include BPMs that are preceded by the possessive pronoun "my", "his", or "her", guaranteeing that there is always an entity whose emotional state can be inferred from the BPM.

We then present six emotional categories that they can annotate from, along with descriptions of these emotional categories (Figure 9). For each emotional category, they are five ranked categories they can choose from to indicate the severity of the emotion (no/slight/moderate/high/very high) (Figure 10).

Dataset	Type	Description	# Instances
1. Spinn3r	Blogs	English subset of ICWSM 2009 Spinn3r Blog Dataset.	80,379
2. Spinn3r$_{BPM}$	Blogs	Subset of Spinn3r containing only posts with at least one BPM	8,371
3. Spinn3r$_{noBPM}$	Blogs	Subset of Spinn3r not containing any instances with a BPM	72,008
4. Spinn3r$_{myBPM}$	Blogs	Subset of Spinn3r$_{BPM}$ containing only instances including BPMs preceded by 'my'.	1,391
6. Spinn3r$_{yourBPM}$	Blogs	Subset of Spinn3r$_{BPM}$ containing only instances including BPMs preceded by 'your'.	541
7. Spinn3r$_{3pBPM}$	Blogs	Subset of Spinn3r$_{BPM}$ containing only instances including BPMs preceded by 'his'/'her'/'their'.	474
8. Spinn3r$_{BPM-Zhuang}$	Blogs	Subset of Spinn3r$_{BPM}$ where BPM mentions are annotated for embodied emotion by (Zhuang et al., 2024).	6,359
9. TUSC$_{city}$	Tweets	The TUSC$_{city}$ dataset.	104,575,991
10. TUSC$_{city-BPM}$	Tweets	The TUSC$_{city}$ dataset contatining only posts with at least one BPM.	6,710,660
11. TUSC$_{city-myBPM}$	Tweets	Subset of TUSC$_{city-BPM}$ containing only instances including BPMs preceded by 'my'.	1,060,507
12. TUSC$_{city-yourBPM}$	Tweets	Subset of TUSC$_{city-BPM}$ containing only instances including BPMs preceded by 'your'.	363,860
13. TUSC$_{city-3pBPM}$	Tweets	Subset of TUSC$_{city-BPM}$ containing only instances including BPMs preceded by 'his'/'her'/'their'.	338,510
14. TUSC$_{ctry}$	Tweets	The TUSC$_{ctry}$ dataset.	3,181,879
15. TUSC$_{ctry-BPM}$	Tweets	The TUSC$_{ctry}$ dataset contatining only posts with at least one BPM.	231,577
16. TUSC$_{ctry-myBPM}$	Tweets	Subset of TUSC$_{ctry-BPM}$ containing only instances including BPMs preceded by 'my'.	37,183
17. TUSC$_{ctry-yourBPM}$	Tweets	Subset of TUSC$_{ctry-BPM}$ containing only instances including BPMs preceded by 'your'.	12,936
18. TUSC$_{ctry-3pBPM}$	Tweets	Subset of TUSC$_{ctry-BPM}$ containing only instances including BPMs preceded by 'his'/'her'/'their'.	18,492

Table 4: Datasets used in this work.

Finally, we aggregate emotions to produce binary scores of an emotion being present/not present.

We assess annotation reliability using the Split-Half Class Match Percentage (SHCMP) as reported in the paper, a method adapted from traditional split-half reliability to handle categorical labels like those used for emotion intensity. SHCMP evaluates how consistently items are classified across multiple random groupings of the dataset. Specifically, the data is divided into n random subsets (with n = 2 representing a typical half-split) 1,000 times, and the average proportion of items that receive the same label across these splits is computed. A higher SHCMP score reflects greater reliability, indicating that the labels are likely to remain stable across repeated annotations.

C Most Frequent myBPMs by Corpus (Supplementary Table, B3)

The top 20 most frequent myBPMs in each corpus, along with their frequency relative to all myBPMs present in their respective corpus, can be viewed in Table 5.

D BPM scores by region (Supplementary Figures and Statistical Tests, B5)

Figure 11 displays the monthly percentage of tweets with at least one *"my <BPM>"* from Canada and the USA between 2015 and 2021.

To assess whether the city of a post influences the probability of including a body-part mention (BPM), we modeled the number of posts containing BPMs out of the total posts per city using a binomial logistic regression, with city included as a categorical predictor. This approach models the log-odds of a post containing a BPM as a function of city. We then performed a likelihood ratio test comparing this model to a null model containing only an intercept, which indicated that cities overall have a highly significant effect on BPM usage (LR = 357.61, $p < 10^{-67}$). When performing the same test using country as a categorical predictor, we find statistically significant results as well (LR = 9.8, p = 0.002), although its effect is considerably weaker than that of individual cities.

Summary Instructions

You will be given an English sentence taken randomly from a blog post. Each of these sentences includes the mention of a body part (face, eyes, gut, etc.). We will refer to the person whose body part is mentioned as the **target person.** Your task is to:

1. Identify whether the body part mentioned belongs to the speaker of the sentence or to someone else. In other words, is the target person the same as the speaker?
2. Identify the emotion that the target person is likely feeling through the multiple choice options. Select all emotions that apply and the level of emotion intensity that applies.

The options correspond to six emotions (anger, fear, sadness, fear/anxiousness, happiness, social warmth, and surprise) on a scale from no emotion, slight emotion, moderate emotion, to high emotion.

You will be asked to infer the emotion felt by the target person, even if it is not explicitly stated. For example:

- Speaker says: In the late afternoon I am lying in the sun, my music plugged firmly into my ears, dancing .

- Mention of body part: ears

Here, the target person (who the body part belongs to) is the speaker. The passage does not explicitly describe happiness, but through readig this passage we can intuitively imagine that the target person is experiencing at least slight amounts of happiness.

Figure 8: Summary instructions for crowdworkers annotating Spinner$_{BPM-Zhuang}$ on how to identify the body part 'target person' and their emotion.

Emotion Descriptions

- **Anger:** irritated, annoyed, aggravated, indignant, resentful, offended, exasperated, livid, irate, etc.
- **Fear/Anxiousness:** frightened, apprehensive, intimidated, panicky, wary, dreadful, jittery, antsy, insecure, nervous, tense, worried, apprehensive, fretful, etc.
- **Sadness:** melancholic, despondent, gloomy, heartbroken, longing, mourning, dejected, downcast, disheartened, dismayed, etc.

- **Happiness:** joyful, elated, content, cheerful, blissful, delighted, gleeful, satisfied, ecstatic, upbeat, pleased, etc.
- **Social warmth:** warmness, sociableness, generosity, helpfulness, tolerance, understanding, thoughtfulness, etc.

- **Surprise:** taken aback, bewildered, astonished, amazed, startled, stunned, shocked, dumbstruck, confounded, stupefied, etc.

Quality Control

Some questions have pre-determined correct answers. If you mark these questions incorrectly, we will often give you immediate feedback in a pop-up box. An occasional misanswer is okay. However, if the rate of misanswering is high (e.g., >20%), then all of one's HITs may be rejected.
In addition, for some questions, we record the misanswers, but do not show a popup. Rejection decisions are based on one's full set of annotations.

Demographics

Provide your age, country, and gender in the first HIT (or at least one HIT) that you do. You can leave the text boxes blank in subsequent HITs. This information will be used to study trends of age, location, etc. with emotion.

Figure 9: Instructions for crowdworkers annotating Spinner$_{BPM-Zhuang}$ on the various emotional categories to annotate.

Q3-Target. Does the body part mentioned in the sentence belong to the speaker or to someone else? In other words, is the target person the same as the speaker?
The speaker often refers to themselves by first-person pronouns such as I, my, mine, etc. (i.e. *I felt my heart racing*), while others are usually referred to in the second or third-person with pronouns such as he, she, you, they, xe, xem, etc. (i.e. *She clenched her fists*).

○ speaker ○ other

Q3-Target's-Emotion. Based on the sentence above, which of the options below best describe the feelings of the target person (the one you identified in the above question)?
Select one option for each of the rows below:

○ no anger	○ slight anger	○ moderate anger	○ high anger
○ no sadness	○ slight sadness	○ moderate sadness	○ high sadness
○ no fear/anxiousness	○ slight fear/anxiousness	○ moderate fear/anxiousness	○ high fear/anxiousness
○ no happiness	○ slight happiness	○ moderate happiness	○ high happiness
○ no surprise	○ slight surprise	○ moderate surprise	○ high surprise
○ no social warmth	○ slight social warmth	○ moderate social warmth	○ high social warmth

Feedback (optional):

Figure 10: Questionnaire for annotating Spinner$_{BPM-Zhuang}$ for emotion felt by the BPM owner.

BPM	Spinn3r$_{myBPM}$ *Blog sentences (%)*	TUSC$_{ctry-myBPM}$ *Tweets (%)*	TUSC$_{city-myBPM}$ *Tweets (%)*
my arm	-	1.08	1.28
my arms	1.51	-	-
my back	1.88	4.61	3.95
my blood	-	1.14	1.07
my body	4.20	6.54	6.61
my brain	2.57	4.43	6.55
my chest	1.88	1.25	1.19
my ears	1.19	-	-
my eye	1.07	-	-
my eyes	10.23	1.40	1.21
my face	4.14	7.31	6.61
my feet	1.19	2.30	1.92
my fingers	1.07	-	-
my hair	3.26	11.18	10.08
my hand	2.82	2.21	2.07
my hands	3.32	-	-
my head	12.11	12.24	13.58
my heart	20.39	17.46	16.70
my legs	-	-	-
my lips	-	-	-
my mouth	1.82	2.89	2.94
my neck	-	1.34	1.29
my nerves	-	-	0.91
my nose	-	1.78	1.59
my side	1.51	-	-
my skin	2.13	1.83	1.87
my stomach	1.25	3.00	3.01
my teeth	-	1.21	1.13
my throat	-	1.13	-
Total	79.54	86.36	85.58

Table 5: B3 - Top 20 most common BPMs preceded by ‘my‘ throughout the Spinn3r$_{BPM}$, TUSC$_{ctry}$, and TUSC$_{city}$ corpora with the frequency of appearance relative to total BPM distribution. The list shows the union of the top 20 unique BPMs for each dataset. Empty entry means that the BPM was not in the dataset’s top 20.

E Are mentions of the body associated with longer utterances: longer sentences/tweets?

Method: We investigate whether BPMs are associated with more descriptive language by comparing samples with and without BPMs by length.
Results: In Spinn3r, sentences containing BPMs are substantially longer than sentences without BPMs, with an average length of 482.07 characters compared to 146.53 characters for non-BPM sentences, making the average BPM sentence length 3.29x the length of the average non-BPM sentence. In the TUSC datasets, a similar trend holds but with a smaller magnitude: BPM tweets average 130.99 characters in TUSC$city$ (vs. 93.98 for non-BPM) and 133.42 characters in TUSC$ctry$ (vs. 101.45 for non-BPM), corresponding to a 1.39x and 1.32x increase, respectively.
Discussion: These results suggest that body part mentions are consistently associated with longer sentences in a variety of online domains. The larger difference in the Spinn3r corpus may reflect the affordances of long-form narrative text, where BPMs may often occur in detailed narratives or reflective writing. Tweets often include just one sentence, but may at times include more; however, the total number of characters cannot exceed 280. It is interesting that even such character-limited conditions, tweets with BPMs are markedly longer than those without BPMs.

F Emotion associations between myBPMs, yourBPMs, 3pBPMs, noBPMs (Supplementary Figures, BA1)

Figure 12 shows the percentage of sentences containing at least one positive word associated with Plutchik’s eight emotional categories (NRC Emotion Lexicon) across the same BPM categories in each corpus.

Country	Month	2015	2016	2017	2018	2019	2020	2021
CA	Jan	0.839	0.801	1.083	1.081	0.913	1.006	0.765
	Feb	1.034	0.818	0.910	0.831	0.964	0.951	0.987
	Mar	1.007	0.869	0.950	0.893	0.842	1.006	0.918
	Apr	0.907	0.993	1.032	1.224	0.866	1.077	0.918
	May	1.166	0.740	0.985	0.792	1.033	1.009	0.879
	Jun	0.965	1.114	0.985	0.792	1.034	0.883	0.774
	Jul	1.069	1.064	1.054	0.969	0.958	0.955	1.026
	Aug	1.090	0.969	1.139	1.058	0.968	0.898	0.892
	Sep	0.922	0.831	1.369	0.958	0.968	0.894	0.851
	Oct	1.050	0.909	1.005	0.998	0.893	0.897	0.851
	Nov	1.120	0.803	0.945	1.152	0.917	0.844	0.846
	Dec	0.859	0.970	1.189	1.010	0.919	0.833	0.845
USA	Jan	1.344	1.061	1.178	1.596	1.381	1.374	1.236
	Feb	1.307	1.385	1.364	1.578	1.283	1.271	1.271
	Mar	1.372	1.312	1.462	1.578	1.283	1.209	1.299
	Apr	1.411	1.342	1.224	1.287	1.323	1.496	1.496
	May	1.367	1.433	1.607	1.394	1.394	1.308	1.266
	Jun	1.221	1.328	1.361	1.461	1.285	1.229	1.306
	Jul	1.312	1.352	1.673	1.286	1.335	1.307	1.265
	Aug	1.340	1.354	1.531	1.318	1.318	1.228	1.182
	Sep	1.640	1.346	1.531	1.418	1.418	1.239	1.112
	Oct	1.406	1.300	1.301	1.428	1.233	1.339	1.118
	Nov	1.342	1.254	1.749	1.193	1.178	1.128	1.182
	Dec	1.193	1.570	1.678	1.497	1.365	1.279	1.122

Figure 11: B5 - TUSC$_{ctry}$ - % of tweets with at least one “my <BPM>“ for Canada and USA from 2015 to 2021 for each month.

G Controlling for Post Length (Supplementary Tables, BA2)

Note that tweets (which we use as individual posts for our tweet dataset, TUSC) are limited to 280 characters. However, blog sentences (which we use as individual posts for our blog dataset, Spinn3r) can be longer. Tables 6 and 7 in the Appendix show the percentage of samples co-occuring with emotion-associated words when controlling for blog post length for VAD and emotion categories, respectively.

H Effect of time on BPM usage (Supplementary statistical tests, B4)

To assess the statistical significance of seasonal variation in BPM usage, we modeled the log-odds of a post containing a BPM as a function of month using a binomial logistic regression. We treat month as a continuous variable, starting from April, since this is the month where we can start to observe the trend of decreasing BPM use throughout the year. The results indicate a statistically significant decline in BPM usage over the months following April (coefficient = -0.0036, p < 0.001). When performing the same test with weekday, starting from Monday, we find a statistically significant decline in BPM usage from Monday as well (coefficient = -0.0104, p < 0.001).

I Emotion associations between specific *“my <BPM>”* types (Supplementary Figures and Statistical Tests, BA3)

In this section, we include exact values for the differences in the percentage of sentences with emotion-associated words in samples containing "*my <BPM>*" types in. These are shown in Figures 13 (TUSC$_{ctry-BPM}$ and VAD), 14 (TUSC$_{ctry-BPM}$ and emotion categories), 15 (Spinn3r$_{BPM}$ and emotion categories), and 16 (Spinn3r$_{BPM}$ and VAD). We focus on the top 30 most common *“my <BPM>”* types in TUSC$_{city-BPM}$. and the and top 15 most common body parts in Spinn3r$_{BPM}$. The mean and standard deviation is also calculated over all common *“my <BPM>”* types analyzed for each corpus for each emotional dimension, and for each *“my <BPM>”* type we display the ’delta’ as the proportion of the *“my <BPM>”* type sample co-occuring with an emotion-associated words subtracted from the mean. All word-emotion associations are from the NRC VAD Lexicon and the NRC Emotion Lexicon.

We also use a two-way ANOVA test to examine the independent and interactive effects of BPM category and emotional category on VAD (Valence-Arousal-Dominance) values, which revealed significant main effects for both category ($F(3,48) = 26.94$, $p < 0.001$) and emotion ($F(5,48) = 60.72$, $p < 0.001$), with emotion showing the stronger effect, but no significant interaction between factors ($F(15,48) = 1.25$, $p = 0.267$). This indicates that while values vary across the different emotions (as expected, since some emotion-associated words naturally occur in certain corpora more frequently), the different BPM instance categories independently have a statistically significant effect emotion-associated word co-occurences. We also evaluate this test for the different emotional categories, again showing significant effects of the BPM categories ($F(3,64) = 54.33$, $p < 0.001$) and emotion ($F(7,64) = 44.68$, $p < 0.001$), and in this case also a significant interaction between category and emotion ($F(21,64) = 2.03$, $p = 0.016$) – suggesting that the impact of BPM category is statistically significant, and also varies in strength depending on the specific emotion dimension.

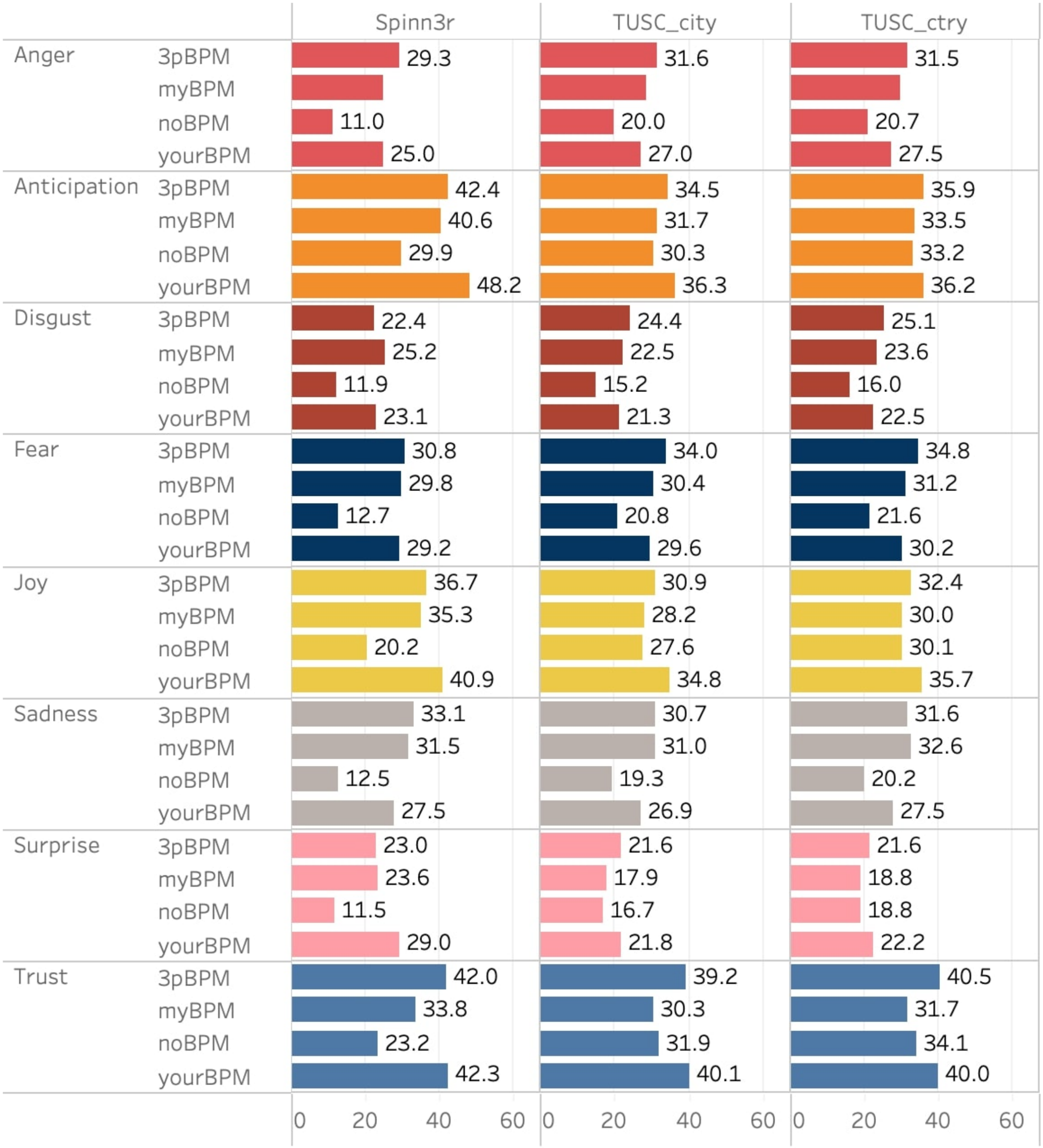

Figure 12: BA1 - Emotional categories. Percentage of sentences with at least one positive word in the eight emotions from Plutchik's emotion wheel (according to the NRC Emotion Lexicon) in each corpus in myBPM, yourBPM, 3pBPM, and noBPM categories.

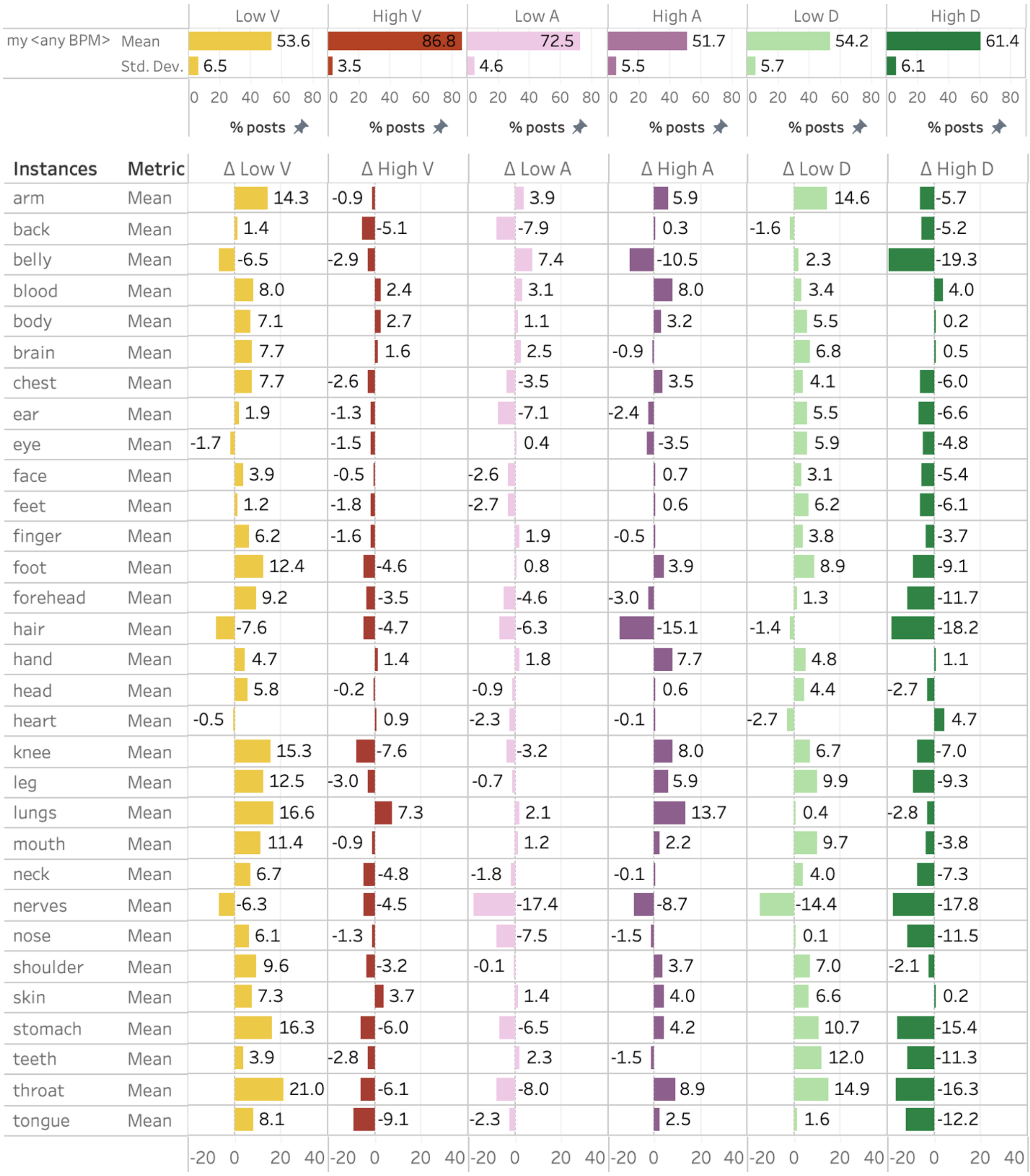

Figure 13: BA2 - TUSC$_{ctry-myBPM}$ - Variance in emotion-associated term co-occurrence for top 30 most common *"my <BPM>"* types present in the dataset. For each type, we display the delta in ("my <BPM>" type minus *"my <BPM>"* mean) in the percentage of tweets with at least one word that is associated with high/low valence, arousal, and dominance (according to the NRC VAD lexicon). Mean and standard deviation are calculated over all body parts considered (top 30 most common *"my <BPM>"* types present in the dataset).

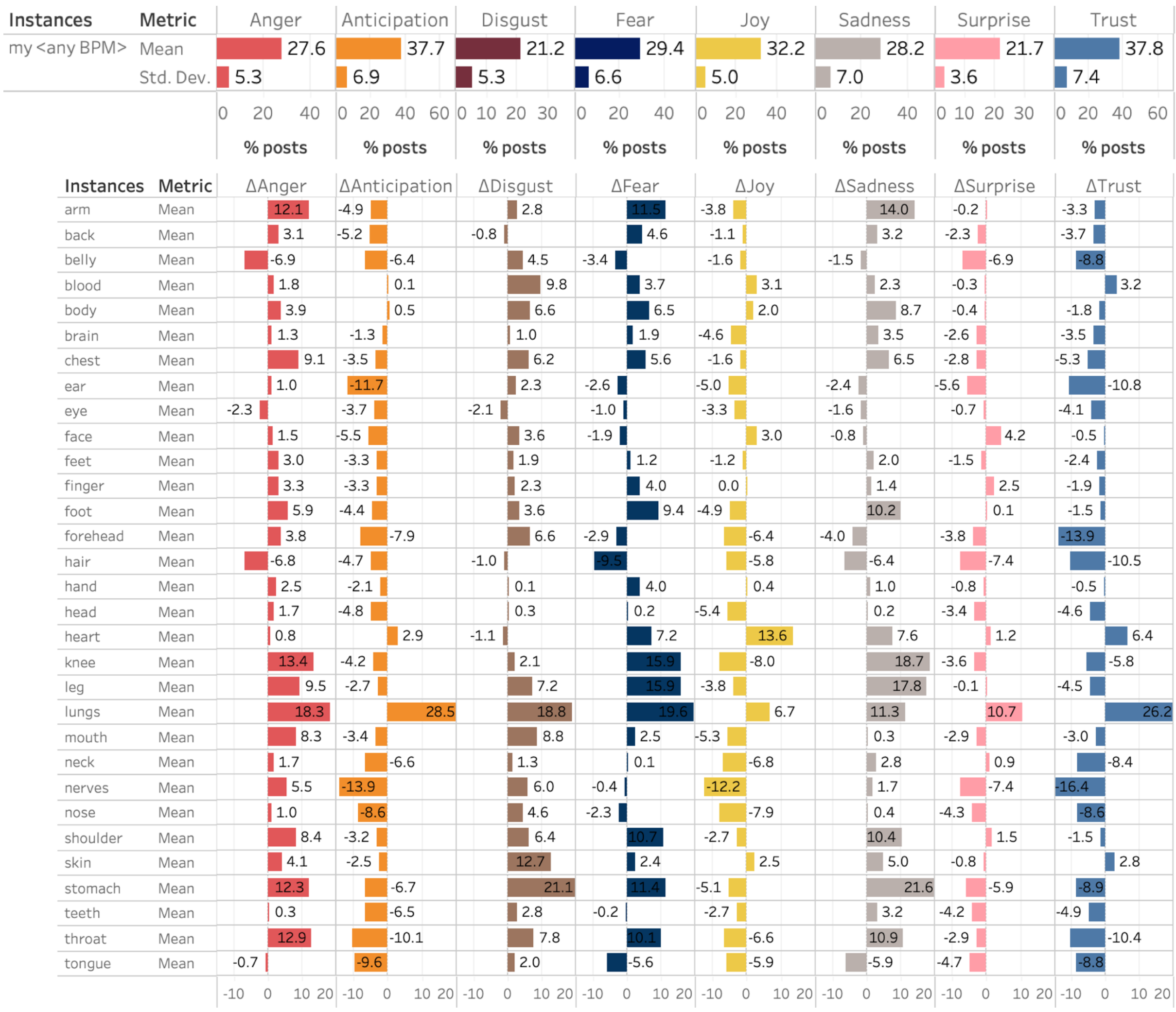

Figure 14: BA2 - TUSC$_{ctry-myBPM}$ - Variance in emotion-associated term co-occurrence for top 30 most common *"my <BPM>"* types present in the dataset. For each type, we display the delta in ("my <BPM>" type minus *"my <BPM>"* mean) in the percentage of tweets with at least one word that is associated with each emotional category (according to the NRC Emotion Lexicon). Mean and standard deviation are calculated over all body parts considered (top 30 most common *"my <BPM>"* types present in the dataset).

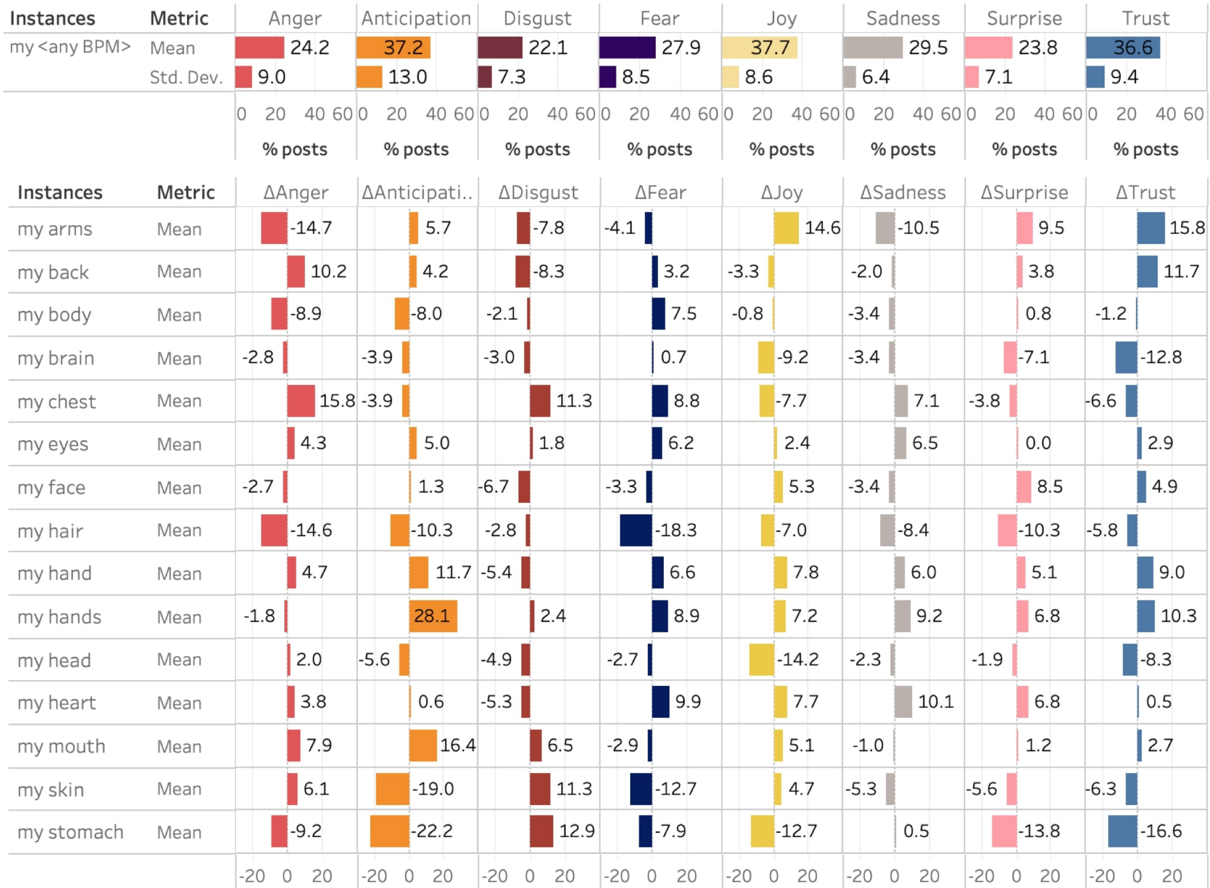

Figure 15: BA2 - Spinn3r$_{myBPM}$ - Variance in emotion-associated term co-occurrence for top 30 most common *"my <BPM>"* types present in the dataset. For each type, we display the delta in ("my <BPM>" type minus *"my <BPM>"* mean) in the percentage of blog sentences with at least one word that is associated with each emotional category (according to the NRC Emotion Lexicon). Mean and standard deviation are calculated over all body parts considered (top 30 most common *"my <BPM>"* types present in the dataset).

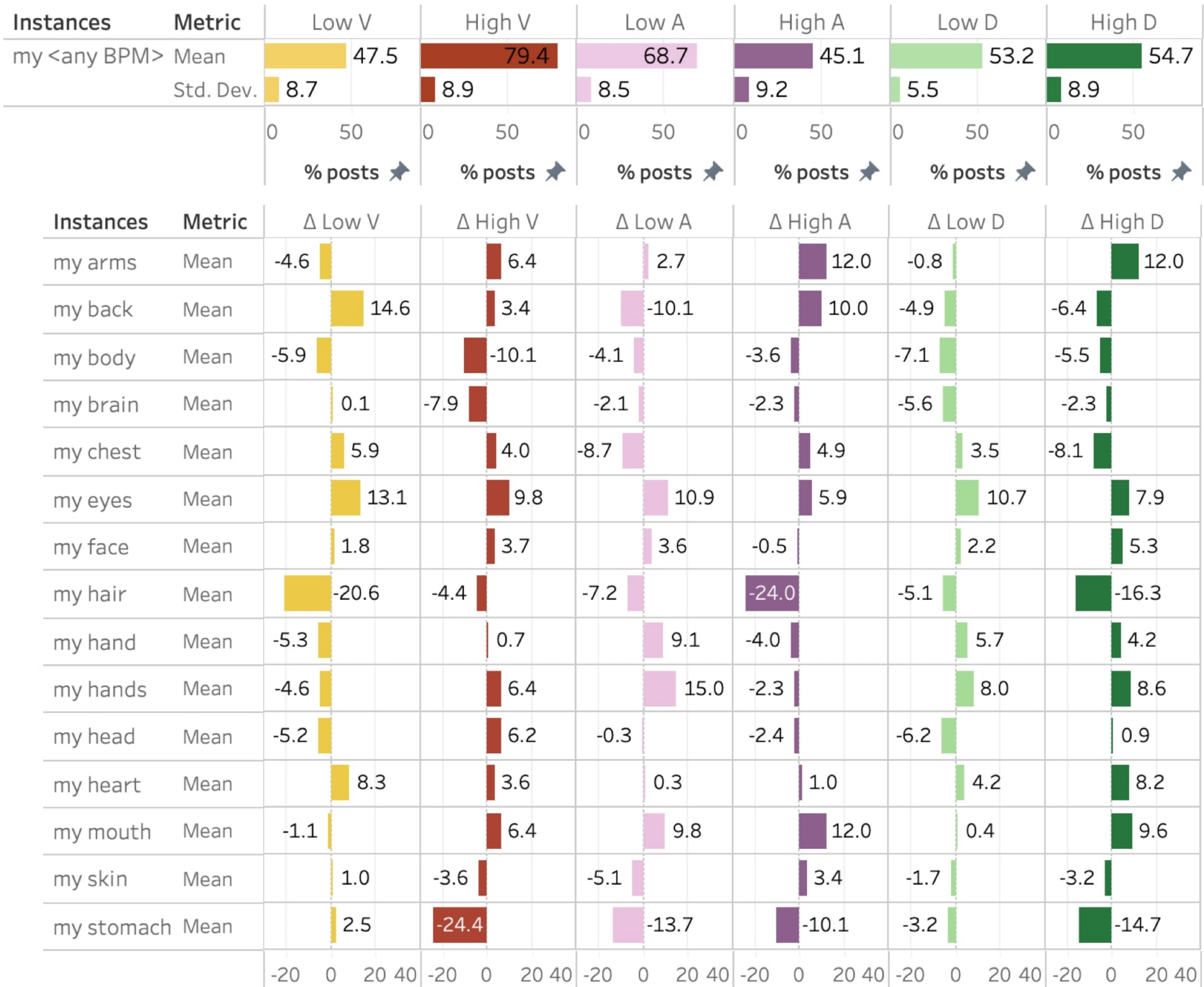

Figure 16: BA2 - Spinn3r$_{myBPM}$ - Variance in emotion-associated term co-occurrence for top 30 most common *"my <BPM>"* types present in the dataset. For each type, we display the delta in ("my <BPM>" type minus *"my <BPM>"* mean) in the percentage of blog sentences with at least one word that is associated with high/low valence, arousal, and dominance (according to the NRC VAD lexicon). Mean and standard deviation are calculated over all body parts considered (top 30 most common *"my <BPM>"* types present in the dataset).

J How are body part words represented in word-emotion association lexicons? (Additional Experiment)

Method: 144 BPMs from our list are found in the NRC Emotion Lexicon, and 200 are found in the arousal, dominance, and valence lexicon. All of the BPMs represented in the top 20 myBPMs across our corpora are represented in our lexicons (except for plural versions of the same BPM). We compare average scores for VAD and emotional categories using the NRC VAD lexicon and the NRC Emotion Lexicon respectively, for high frequency BPM words (defined as a word found in the top 20 myBPM list in any of our corpora), words that are in our BPM list, and words that are not in our BPM list. *Results:* We find that high frequency BPMs exhibit high changes in associations with valence, arousal, and dominance from non-BPMs (significantly than the average BPM score). However, frequent BPMs are rarely ranked as positive instances for any of the seven emotional categories (less than the non-BPM baseline for all categories except for Surprise and Disgust). *Discussion:* Although common BPMs seem to have a different emotional signature than non-BPM words, they seem to have little everyday association to particular emotional categories. This corroborates the theory that bodily interpretations form the basis of our most basic emotional categories, but that more specific emotional categories are produced from contextual interpretations of these bodily signals.

K How does referring to one's own body change the emotional signature in personal narratives online (BA1: Supplementary Figures)

Table 6 displays the percentage of BPM vs no BPM sentences across High/Low VAD categories containing at least one word associated with each emotional category according to NRC VAD lexicon in the Spinn3r dataset across different bins of sentence length for high/low VAD, and Table 7 displays this data for emotional categories.

L What are the emotions most commonly associated with the most frequently discussed body parts?

In Table 8, we display the top emotion (with associated increase in emotion-associated word from *"my <BPM>"* average) for top *"my <BPM>"* types across Spinn3r$_{BPM}$ and TUSC$_{BPM}$ datasets. A dash indicates the body part is not present in top *"my <BPM>"* types considered for the dataset (top 15 for Spinn3r$_{BPM}$ and top 30 for TUSC$_{BPM}$).

M Is physical wellbeing correlated with emotional word use? (Additional Experiment)

We also look at whether the physical wellbeing indicators we examine in other experiments are correlated with emotion-related words according to the NRC lexicon. See Table 9.

N Why/when do we refer to our own bodies? (Additional Experiment)

Method: We evaluate the context of words that tend to surround myBPMs by looking at word clouds which visualize the words which are most likely to appear within the context window of particular myBPMs (See Figure 17).
Results: We find that there are significantly more 3pBPM types with $> 0.1\%$ occurrence compared to *"my <BPM>"* types with $> 0.1\%$ (131 vs 57 in the Spinn3r$_{BPM}$ dataset, and 108 vs 56 in the TUSC$_{ctry-BPM}$ dataset). We also find that, although *"my <BPM>"* types exhibit a rich diversity in associated contexts, that some *"my <BPM>"* types share common contexts as well, especially "hurt", "pain", and "sick", which frequently co-occur with several frequent myBPMs such as "my head", "my back", "my neck", and" my stomach".
Discussion: The analysis reveals that third-person BPM types (3pBPM) in are significantly more diverse than *"my <BPM>"* types at the 0.1% occurrence threshold in the Spinn3r dataset, indicating a more limited and concentrated vocabulary when people refer to their own body than the bodies of others. The words with negative associations with health frequently accompanying some of the most common *"my <BPM>"* types also highlight health concerns and physical pain as central themes for myBPM usage.

Figure 17: B6 - Wordclouds for the twenty most frequent *"my <BPM>"* types in the TUSC$_{city}$ dataset with the most frequent co-occurring words.

Table 6: BA2 - Spinn3r - Percentage of BPM vs no BPM sentences across High/Low VAD categories containing at least one word associated with each emotional category according to NRC VAD lexicon in the Spinn3r dataset.

Bin	High Valence		Low Valence		High Arousal		Low Arousal		High Dominance		Low Dominance	
(# of words)	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM
(0,10]	0.38	0.40	0.19	0.12	0.14	0.11	0.27	0.24	0.15	0.20	0.17	0.13
(10,20]	0.77	0.80	0.39	0.28	0.32	0.29	0.57	0.60	0.36	0.50	0.37	0.33
(20,30]	0.89	0.93	0.47	0.40	0.44	0.44	0.78	0.79	0.58	0.69	0.62	0.47
(30,40]	0.98	0.97	0.66	0.53	0.55	0.56	0.92	0.88	0.73	0.79	0.71	0.60
(40,50]	0.98	0.97	0.69	0.57	0.65	0.60	0.91	0.90	0.83	0.83	0.79	0.65

Table 7: Percentage of BPM vs no BPM sentences across bins containing at least one word associated with each emotional category (anger, fear, joy, sadness, surprise, trust) according to the NRC VAD Lexicon in Spinn3r.

Bin	Anger		Fear		Joy		Sadness		Surprise		Trust	
(# of words)	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM	*"my <BPM>"*	no BPM
(0,10]	0.07	0.04	0.10	0.05	0.11	0.09	0.10	0.05	0.06	0.05	0.09	0.10
(10,20]	0.20	0.11	0.21	0.13	0.19	0.22	0.24	0.13	0.12	0.12	0.18	0.25
(20,30]	0.19	0.18	0.24	0.22	0.32	0.32	0.27	0.21	0.23	0.19	0.35	0.39
(30,40]	0.30	0.26	0.39	0.30	0.50	0.44	0.40	0.29	0.28	0.27	0.46	0.51
(40,50]	0.44	0.31	0.51	0.33	0.56	0.48	0.48	0.33	0.44	0.29	0.52	0.54
(50,60]	0.36	0.19	0.41	0.23	0.68	0.41	0.36	0.21	0.45	0.26	0.68	0.48

Body Part	**Spinn3r (Top Emotion)**	**TUSC (Top Emotion)**
arms	Trust (15.8)	-
arm	-	Anger (12.1)
back	Trust (11.7)	Fear (4.6)
belly	-	Disust (4.5)
blood	-	Disgust (9.8)
body	Surprise (0.8)	Disgust (6.6)
brain	Fear (0.7)	Sadness (3.5)
chest	Anger (15.8)	Anger (9.1)
ear	-	Disgust (2.3)
eye	-	Surprise (-0.7)
face	Surprise (8.5)	Surprise (4.2)
feet	-	Anger (3.0)
finger	-	Fear (4.0)
foot	-	Sadness (10.2)
forehead	-	Disgust (6.6)
hair	Disgust (-2.8)	Disgust (-1.0)
hand	Anticipation (11.7)	Fear (4.0)
hands	Anticipation (28.1)	Anger (1.7)
head	Anger (2.0)	Disgust (0.3)
heart	Sadness (10.1)	Joy (13.6)
knee	-	Sadness (18.7)
leg	-	Sadness (17.8)
lungs	-	Anticipation (28.5)
mouth	Anticipation (16.4)	Disgust (8.8)
neck	-	Sadness (2.8)
nerves	-	Disgust (6.0)
nose	-	Disgust (4.6)
shoulder	-	Anger (8.4)
skin	Disgust (11.7)	Disgust (12.7)
stomach	Disgust (12.9)	Sadness (21.6)
teeth	-	Sadness (3.2)
throat	-	Anger (12.9)
tongue	-	Disgust (2.0)

Table 8: Most associated emotion (with associated increase in emotion-associated word from *"my <BPM>"* average) for *"my <BPM>"* types across Spinn3r$_{BPM}$ and TUSC$_{BPM}$ datasets (top 15 for Spinn3r$_{BPM}$ and top 30 for TUSC$_{BPM}$). A dash indicates the body part is not present in top *"my <BPM>"* types considered for the dataset.

Emotional Category	**Mental Distress**	**Physical Distress**	**Life Expectancy**	**Physical Inactivity**
Anger	-0.05 (p=0.8170)	-0.12 (p=0.5520)	-0.16 (p=0.4030)	-0.12 (p=0.5480)
Anticipation	-0.10 (p=0.6110)	-0.24 (p=0.2090)	0.07 (p=0.7260)	-0.33 (p=0.0870)
Disgust	0.07 (p=0.7090)	0.07 (p=0.7200)	-0.23 (p=0.2340)	0.09 (p=0.6610)
Fear	-0.23 (p=0.2310)	-0.37 (p=0.0540)	0.16 (p=0.4030)	**-0.46 (p=0.0140)**
High Arousal	-0.12 (p=0.5350)	-0.25 (p=0.2070)	0.03 (p=0.8770)	-0.34 (p=0.0760)
High Dominance	-0.18 (p=0.3570)	-0.31 (p=0.1090)	0.16 (p=0.4210)	**-0.38 (p=0.0440)**
High Valence	-0.13 (p=0.4950)	-0.24 (p=0.2110)	0.12 (p=0.5390)	-0.33 (p=0.0900)
Joy	-0.05 (p=0.7960)	-0.15 (p=0.4580)	0.09 (p=0.6490)	-0.24 (p=0.2260)
Low Arousal	-0.11 (p=0.5710)	-0.27 (p=0.1590)	0.05 (p=0.8090)	-0.35 (p=0.0700)
Low Dominance	-0.13 (p=0.5050)	-0.26 (p=0.1830)	0.06 (p=0.7710)	-0.37 (p=0.0530)
Low Valence	-0.19 (p=0.3400)	-0.26 (p=0.1790)	0.07 (p=0.7110)	**-0.38 (p=0.0460)**
Sadness	-0.15 (p=0.4430)	-0.27 (p=0.1600)	0.08 (p=0.6710)	**-0.39 (p=0.0410)**
Surprise	-0.10 (p=0.6300)	-0.23 (p=0.2490)	0.09 (p=0.6380)	-0.33 (p=0.0820)
Trust	-0.16 (p=0.4050)	-0.31 (p=0.1120)	0.13 (p=0.5190)	**-0.38 (p=0.0480)**

Table 9: Spearman's ρ and p-values between proportion of emotional words and city-level health outcomes. Bolded values are statistically significant at $p < 0.05$.