Representations of emotion concepts: Comparison across pairwise, appraisal feature-based, and word embedding-based similarity spaces

Mijin Kwon[1], Tor D. Wager[1], Jonathan Phillips[1, 2, 3]
[1] Department of Psychological and Brain Sciences, Dartmouth College
[2] Program in Cognitive Science, Dartmouth College
3 Department of Philosophy, Dartmouth College

Abstract

A question that has long interested cognitive scientists is how to best represent the different emotions we experience and attribute to others. For example, constructionist and appraisal theories propose that differences between emotions can be captured in part by their variation along a set of appraisal dimensions. More recently, researchers have used language models to capture the differences across different emotion terms. Both approaches allow us to represent emotions as occupying different locations in high-dimensional representational spaces. To ask how well these different approaches capture the similarity between emotion concepts, we collected pairwise similarity and appraisal feature ratings for 58 different emotion concepts and then employed representational similarity analysis to investigate the overlap between people's pairwise similarity judgments and emotion similarity in a 14-dimensional appraisal space and three word embedding spaces from two word2vec models (300 dimensions) and the newer GPT-3 model (12288 dimensions). The results indicate that while there is a high correlation between appraisal feature-based similarity and pairwise similarity judgments, word embedding-based similarity exhibits lower correlations, though GPT-3 showed much better performance than the word2vec models. Finally, characterizing the errors made by word embedding models showed that they can be largely attributed to an over-reliance on the valence of emotion concepts.

Keywords: emotion, concepts, representational similarity, appraisal, natural language processing, word2vec, GPT-3

Representation of Emotion

Humans use hundreds of different words to describe emotions. Like many concepts, emotion concepts can be thought of in terms of a network of associations; for example, 'elation' is more closely associated with 'joy' than with 'disgust'. Some of the fundamental questions in emotion research relate to the dimensionality and structure of emotion concepts, and what dimensions or categories might best describe them. For example, can the similarity and differences across concepts be described by a small number of dimensions (i.e., valence and arousal), emotion categories, or situational evaluations of perceived harms, benefits, intentions, and coping resources?

A great deal of research has pursued answers to these questions, with different theorists taking similar positions on some aspects but strongly diverging on others (for reviews of different emotion theories, see Barrett, 2016; Ekman et al., 1983; Ellsworth, 1994; Gendron & Barrett, 2009; Moore et al., 2013). For example, while constructionist and appraisal theories of emotion differ in a number of respects, they share an understanding that differences between emotions are related to differences in the evaluations of the situations that give rise to those emotions.

In one study in this vein, Skerry and Saxe (2015) demonstrated that representations of emotion concepts, at both a behavioral and neural level, are best explained by a multidimensional space constructed by appraisal features. Therefore, these features may provide an important clue in understanding how people represent emotion concepts in relation to one another. Another study demonstrated that our conceptual knowledge of different emotion categories is mirrored by inferences we make about emotions from facial expressions – i.e., the more conceptually different two emotions are, the more different two facial emotions are perceived to be (Brooks & Freeman, 2018).

Representing Emotions using Natural Language Processing

In a separate line of research that has recently been developing in parallel, there has been a growing interest in asking how language models developed by NLP researchers can capture the similarity and differences in the semantics of different emotion terms (see, e.g., Seyeditabari et al., 2019). In one recent paper, for example, Jackson and colleagues (2019) argued for differences in the meaning of emotion terms across different languages on the basis of an analysis of co-lexification of emotion terms.

Present Research

In this paper, we ask how well these different approaches capture the reported similarity between emotion concepts. We collected pairwise similarity and 14 appraisal feature ratings for 58 different emotion concepts and then investigated the relationship between pairwise similarity judgments and pairs of emotion similarity in a 14-dimensional appraisal space and the three word embedding spaces: two word2vec models trained on two independent corpora, and the more recent GPT-3 model (3rd generation Generative Pre-trained Transformer).

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

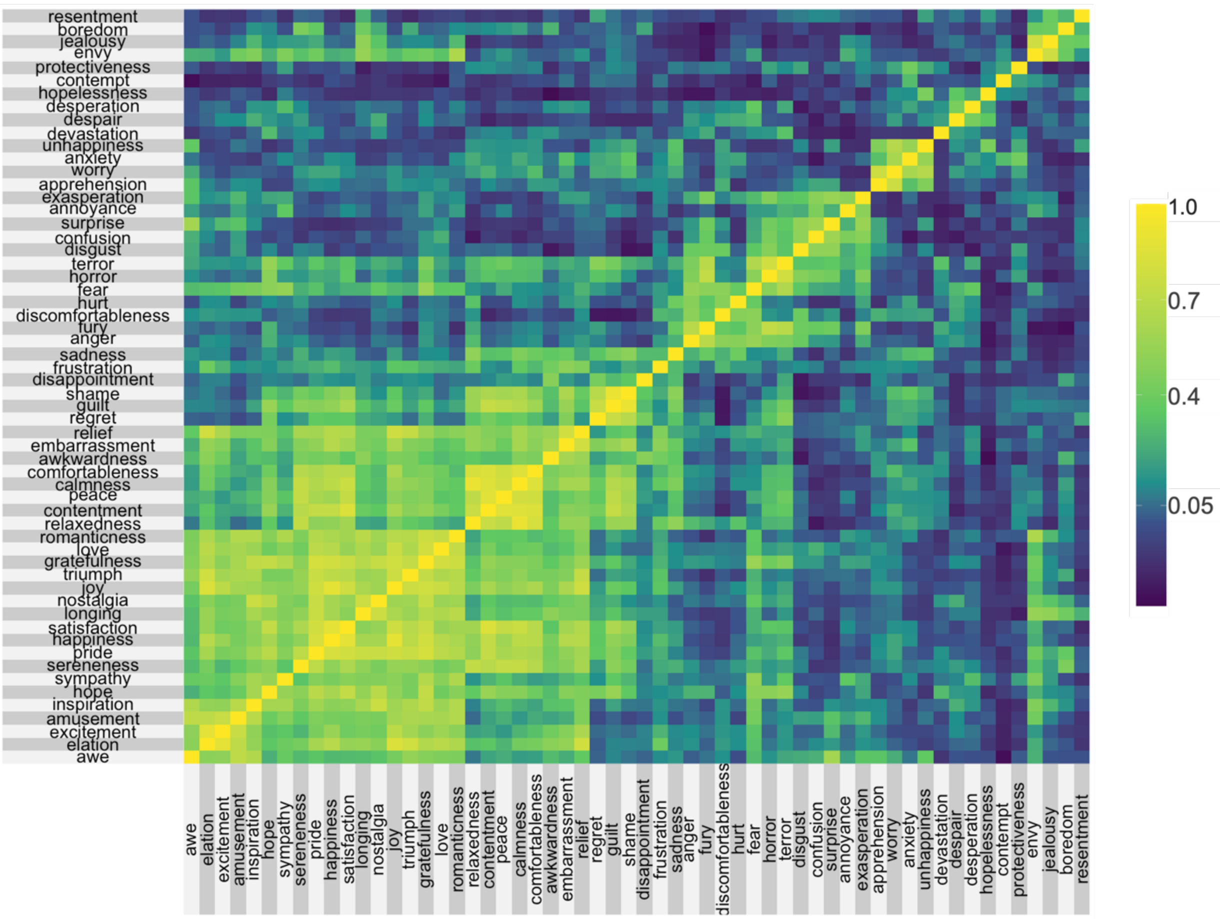

Figure 1. Similarity matrix with hierarchical clustering for all 58 emotions from the pairwise similarity judgment task.

To analyze the relationship between people's pairwise similarity judgments and the similarity of emotions in appraisal and word embedding spaces, we utilized an analytical technique called representational similarity analysis (RSA; see Kriegeskorte et al., 2008 for more on this technique). This approach uses the 2nd-order isomorphic representation (i.e., measures of similarity such as correlation or distance) and therefore, allows us to investigate the relationships between measures in different spaces. This can reveal unique and shared structures across distinct spaces that would otherwise be difficult to uncover using independent analyses within each space. Critically, for our purposes, it allows us to test whether the extent to which any two emotions are similar to one another in one space (e.g., pairwise judgments) is predicted by how related those emotions are to each other in another space (e.g., a word-embedding space). In turn, it also allows to test what's unique in one space and shared across multiple spaces in terms of emotion representation.

While the relationship between any two emotions will likely be sensitive to contextual features, we used simple pairwise comparisons between emotion concepts as a broad estimate of how people represent those emotion concepts in the absence of any particular context (see Brooks & Freeman, 2018; Brooks et al., 2019 for a similar approach). We then asked how the other two types of similarity, derived from appraisal features and word embedding, differ from people's pairwise similarity judgments. Finally, we also investigated whether any of the appraisal features we explored can explain discrepancies between pairwise and word embedding-based similarity to shed light on how word embedding models may represent emotion differently than humans.

Methods

Study materials

Emotion concept words We sought to include as many common English emotion concepts as we could. In total, we included 58 English terms[1] which are relatively frequently used to describe emotional states, including ones retrieved

[1] Following are the 58 emotion concepts used in this study: amusement, anger, annoyance, anxiety, apprehension, awe, awkwardness, boredom, calmness, comfortableness, confusion, contempt, contentment, despair, desperation, devastation, disappointment, discomfortableness, disgust, elation,

from prior studies involving emotion terms (Cowen et al., 2018; O'Reilly & Lundquist, 2016 (EU-Emotion Stimulus Set); Skerry & Saxe, 2014; Tottenham et al., 2009 (NimStim); Gross & Levenson, 1995; Watson, Clark, & Tellegen, 1988 (Positive and Negative Affect Schedule); Scherer, 2005 (Geneva Emotion Wheel)).

Appraisal features 14 appraisal features were sampled from previous research that investigated the impact of a set of appraisal features on emotion experience and perception (Cowen & Keltner, 2017; Ellsworth, 2013; Scherer & Meuleman, 2013; Skerry & Saxe, 2014). These features were selected based on their broad relevance to the emotion concepts included in this study. They include the following appraisal dimensions: valence, seeing vs. imagining, foreground vs. background experience, embodiment, duration, directionality, physical danger, relevance to oneself, attention, disease and contamination, morality, avoidance vs. approach, social appropriateness, and controllability.

Word embeddings Word embeddings of all 58 emotion concepts were extracted from three pre-trained NLP models: two word2vec models trained on Google news (Mikolov et al., 2013) and Wikipedia (Fare et al., 2017), and the recent GPT-3 model, accessed through OpenAI's API (*Davinci*; Brown et al., 2020). Both word2vec models consist of 300 dimensions and the training data corpus was aggregated in the years of 2013 and 2019, respectively. The GPT-3 model includes 12288 dimensions and is trained on a vast range of internet data, including a filtered version of Common Crawl with 410 billion tokens.

Study design

Emotion concept words We recruited 1288 participants (839 for the pairwise similarity judgment task (M_{age} = 40.01, SD_{age} = 13.69, N_{female} = 427), 419 for the appraisal feature rating task (M_{age} = 39.57, SD_{age} = 11.63, N_{female} = 205)) via Amazon Mechanical Turk and Prolific.

Online surveys For pairwise similarity judgments, participants were presented with a pair of two emotion concepts and asked to rate the similarity between them on a scale from 0 (very different) to 100 (very similar). These pairs were randomly selected from all possible pairs of the 58 emotion concepts (1171 pairs including same-concepts pairs). Each participant was asked to rate the similarity of 30 pairs of emotions, randomly sampled from all possible pairs. The similarity value for each pair was calculated by taking the mean of all participants' similarity ratings for that pair.

Ratings of each of the 58 emotion concepts along 14 appraisal features were collected in a separate survey, with each participant randomly assigned to provide relevance ratings ("based on your experience, how much is [description of appraisal feature] relevant for whether you experience [emotion]?"; see Online Supplements for full list) of 1 of the 14 appraisal features and asked to rate 30 different emotion concepts, randomly sampled from the full set of 58 emotion concepts. The relevance score for each emotion-appraisal feature pair was again calculated by simply taking the mean of all participants' ratings.

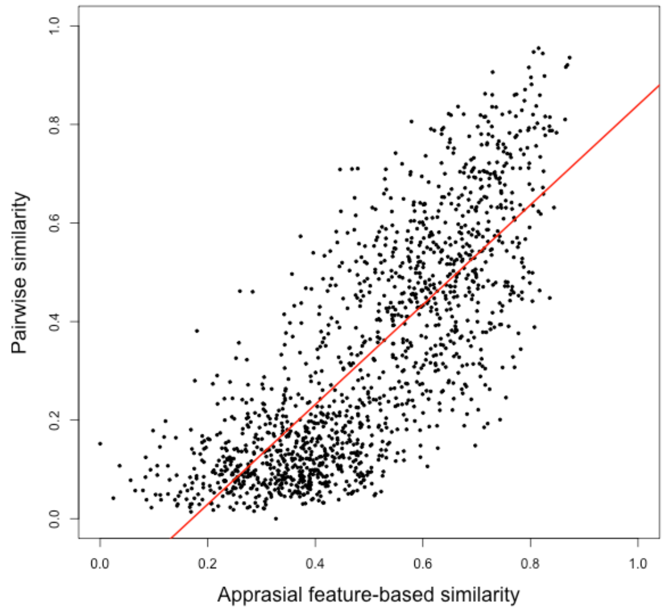

Figure 2. Correlation between pairwise similarity and appraisal feature-based similarity.

Results

Common and distinct representation of emotion concepts across three representational spaces

We investigated the similarity structure uncovered by each measure and the overlap between them. Our approach was to construct similarity matrices of emotion concepts using (1) the direct pairwise judgments, (2) the 14 appraisal feature ratings, and (3-5) word embeddings from the two word2vec models and embeddings from GPT-3. To measure similarity, we used Euclidean distance for the 14 appraisal feature ratings and used cosine similarity for word embeddings between emotion concepts. The pairs between the same concepts were excluded in the analysis and all measures were rescaled to range from 0 to 1. See Figure 1 for an illustration of the similarity matrix for pairwise similarity judgments. We then calculated the correlations between these five similarity matrices and tested how well the feature-based and word embedding-based similarity predicts the pairwise similarity as well as how similar they are to each other. Spearman's rank correlations were used, and significance was assessed

embarrassment, envy, exasperation, excitement, fear, frustration, fury, gratefulness, guilt, happiness, hope, hopelessness, horror, hurt, inspiration, jealousy, joy, longing, love, nostalgia, peace, pride, protectiveness, regret, relaxedness, relief, resentment, romanticness, sadness, satisfaction, sereneness, shame, surprise, sympathy, terror, triumph, unhappiness, worry.

with permutation tests (5000 permutations of randomly shuffling the labels of one of the similarity matrices).

Pairwise and appraisal representational similarity We first asked how much common structure there was between emotion concepts in the 14-dimensional appraisal feature space and our baseline pairwise similarity judgments. The results indicate a large overlap between the similarity structures revealed by the pairwise and feature-based similarity ($r = .79$, $p < .001$), see Figure 2. This notable amount of overlap suggests that both approaches recover a common underlying structure in the way that the 58 emotion concepts are represented.

Representational similarity within NLP models Next, we compared the similarity of all of the 58 emotion concepts across the three NLP models: two word2vec models trained on different corpora and GPT-3. This analysis revealed a high correlation between the two word2vec models ($r = .69$, $p < .001$) and a lower correlation between the two word2vec and GPT-3 (Google news: $r = .33$, $p < .001$; Wikipedia: $r = 0.37$, $p < .001$). These results suggest that while the two word2vec models converge on a common underlying structure recovered from the different corpora they were trained on, GPT-3's representation of emotion concepts differed substantially.

Pairwise and word embedding representational similarity Lastly, we asked how much common structure was shared between these NLP models and people's pairwise similarity judgments (see Figure 3). The word2vec models had a relatively low correlation with the pairwise similarity (Google news: $r = .34$, $p < .001$; Wikipedia: $r = .35$, $p < .001$). In contrast, GPT-3 similarity showed a quite high correlation, $r = 0.74$, $p < .001$), though it was still lower than the 14-dimension appraisal feature similarity. Given the varying room for improvement across these three word-embedding models, we next asked whether we could uncover any systematicity in the errors made by these models.

Capturing the difference between pairwise similarity and word embedding similarity

To understand how word embedding models are differing from human emotion concept representation, we next conducted a linear regression with the pairwise similarity as dependent variable and the word embedding-based similarity as predictor and extracted the signed residuals (i.e., the extent to which the word embedding-based similarity over- or under-estimated the similarity between pairs of emotion concepts). The absolute value of these residuals indicates the extent to which the NLP models were failing to align with human emotion concept representation, and the signed error

Table 1. Contribution of appraisal features to 5 principal components

indicates whether the models were over- or under-estimating the similarity between two emotion concepts compared to human judgments.

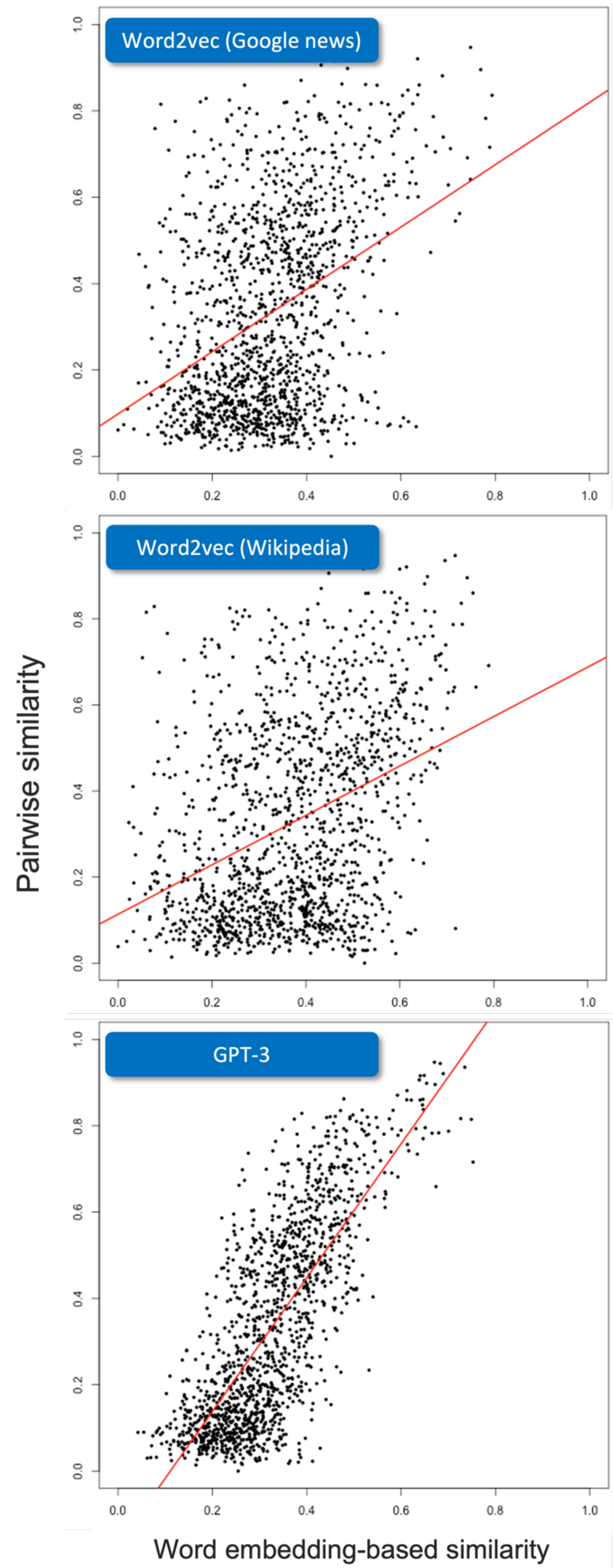

Figure 3. Correlations between pairwise similarity and two word embedding-based similarity (From top: Word2vec (Google news), Word2vec (Wikipedia), GPT-3).

Table 1. Contribution of appraisal features to 5 principal components.

Appraisal features	PC1	PC2	PC3	PC4	PC5
Attention	1.030	0.350	6.083	1.317	62.214
Avoidance vs. approach	12.486	14.426	0.768	0.634	1.527
Controllability	2.853	0.595	8.687	0.401	23.933
Physical danger	10.009	20.447	6.218	0.070	0.137
Disease and contamination	9.366	16.152	8.624	1.464	0.220
Morality	5.506	16.255	0.131	7.120	1.363
Relevance	13.319	0.812	0.004	6.329	0.030
Social appropriateness	13.838	16.201	0.523	0.233	0.243
Directionality	0.894	0.868	26.309	0.849	0.993
Duration	4.050	0.207	18.929	0.104	0.105
Embodiment	1.051	0.839	0.486	41.063	2.936
Foreground vs. background	6.034	6.126	18.695	0.243	1.498
Seeing vs. imagining	0.733	6.581	0.309	35.280	4.801
Valence	18.833	0.140	4.234	4.893	0.000

Having extracted these residuals, we could then leverage participants' prior ratings of emotion concepts along the 14 appraisal features to ask whether any of these features may help explain how and why NLP models are misrepresenting the similarity of emotion concepts. To pursue this question, we took each appraisal feature and each pair of emotion concepts and calculated the absolute difference between the appraisal feature ratings of each pair of emotion concepts, so that each pair had a difference score between each appraisal feature rating (referred to as "difference scores" from here on). Then, for simplicity in interpretation, we used principal component analysis to reduce the number of appraisal features down to 5 principal components that can explain a large amount of the variance in our data (60.84%; see Online Supplements for full results). Respectively, each component has relatively a high contribution from the following appraisal features: PC1 ("valence", "social appropriateness", and "relevance"), PC2 ("physical danger", "morality, and "social appropriateness"), PC3 ("directionality", "duration", and "foreground vs. background"), PC4 ("embodiment" and "seeing vs. imagining"), and PC5 ("attention"). Based on the loadings of each appraisal feature on these 5 principal components, we labeled 5 components as "valence", "physical and social danger", "directionality", "embodiment", and "attention" (Table 1).

Then we conducted a multiple regression analysis with the residuals from the first linear regression model as the dependent variable and the 5 principal components as predictors. This analysis allowed us to test if any of the appraisal features or their combination can explain the discrepancy between the emotion concept similarity judged by human raters and the three NLP models. This analysis revealed significant effects for all 5 components in all three models, controlling for the effects of other components in the regression model (see Table 2 for the results of the word2vec model trained on Google news; see Online Supplements for the results of the other two models).

To illustrate the relationship between these principal components and the residuals between the pairwise similarity word embedding similarity, we categorized the emotion concept pairs into four categories based on their residuals: i.e., 1) highly under-estimated pairs, 2) slightly under-estimated pairs, 3) slightly over-estimated pairs, and 4) highly over-estimated pairs. These categories were simply mapped to the four quartiles of the residual distribution arranged from most under-estimated to most over-estimated. We can then plot the relationship between the residual quartiles and the principal components we derived for the NLP models, see Figure 4.

Table 2. Multiple regression summary for 5 principal components predicting the residuals between pairwise and word-embedding similarity (Word2vec; Google news).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0019	0.0043	0.44	0.6590
Valence (PC1)	-0.0741	0.0025	-29.13	0.0000
Physical and social danger (PC2)	-0.0067	0.0032	-2.08	0.0380
Directionality (PC3)	0.0328	0.0034	9.64	0.0000
Embodiment (PC4)	0.0369	0.0039	9.58	0.0000
Attention (PC5)	-0.0228	0.0043	-5.27	0.0000

To give a sense for the pattern, we can illustrate the kinds of terms in each quartile. For the Google news, examples of highly under-estimated pairs (1st quartile) included <"elation", "sadness">, and <"despair", "joy>. Slightly under-estimated pairs (2nd quartile) included <"hope", "sadness"> and <"fear", "longing">. Slightly over-estimate pairs (3rd quartile) included <"exasperation", "shame"> and <"awe", "terror>. Highly over-estimated pairs (4th quartile) included <"hurt", "sadness">, and <"horror", "terror">. As these examples illustrate and as can be clearly seen in Figure 4, all three NLP models underestimate the similarity between two emotion concepts when the difference in their loadings on the valence component is bigger and over-estimate it when the difference is smaller. In other words, the similarity of a pair of positive and negative emotions (such as "happiness" and "sadness") is more likely to be underestimated by the NLP models compared to human raters' judgment on their similarity, and correspondingly the similarity of a pair of similarly valenced emotions (such as "happiness" and "love") is likely to be over-estimated. The other four principal components did not show strong patterns and had comparatively similar values across all four quartiles.

Discussion

Many different accounts have been proposed to explain how we represent emotions and what can best explain the representational structures underlying emotion concepts. Here, using representational similarity analysis, we compared three different ways of representing the similarity of emotion concepts: pairwise similarity judgments, ratings along 14 different appraisal features, and word embeddings from three NLP models. This approach allowed us to ask how well appraisal features and word embeddings align with the representational structure of emotion concepts uncovered by the direct pairwise similarity judgments of emotions. The results show that there is a high correlation between appraisal feature similarity and pairwise similarity judgments. In contrast, word embedding models showed mixed results, likely as a result of differences in training methods and the scope of data they are trained on. Two word2vec models trained only either on Google news or Wikipedia show a much lower correlation with pairwise similarity compared to feature-based similarity. On the other hand, the GPT-3 model,

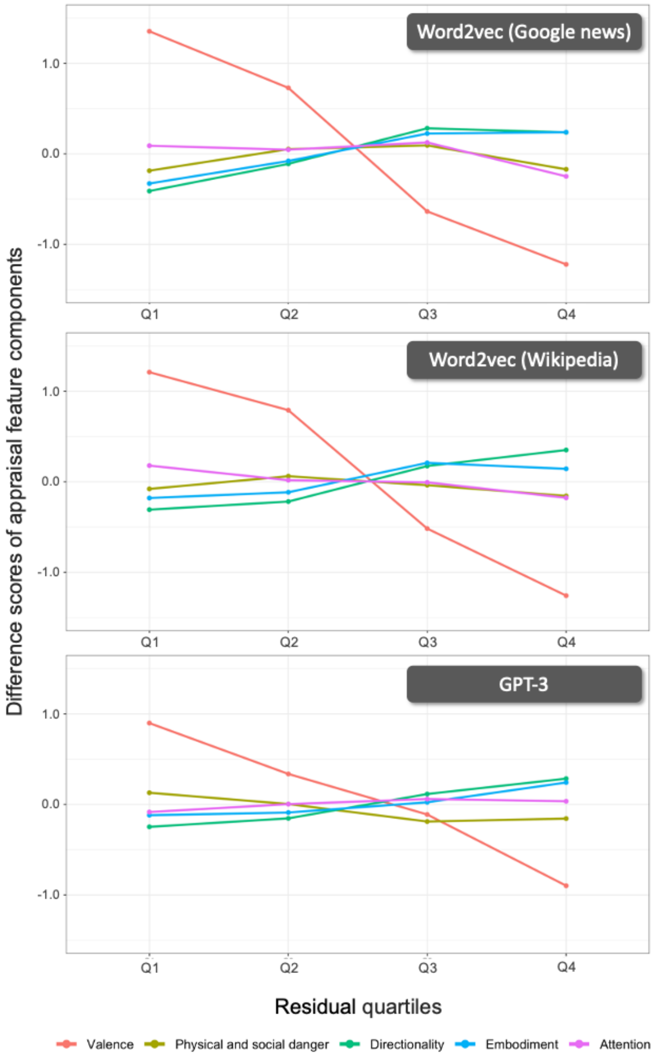

Figure 4. Relationship between the residual quartiles and the 5 principal components derived from the appraisal features.

which is sensitive to context, is trained on a much wider range of data, and has many more parameters showed a correlation much closer to appraisal feature-based similarity.

Finally, we turned to why word embedding similarity was over- or under-estimating the similarity between some emotion concepts. We took appraisal features that are known to help describe human inference about emotion and asked if they helped explain the discrepancy between pairwise and word embedding-based similarity. We found that the errors made by word embedding models are best attributed to the changes in how different two emotion concepts are in the valence dimension (in contrast to other appraisal feature components).

Our results broadly support the previous findings that situational evaluations based on a set of appraisal features can do a good job of capturing general emotion representation (Skerry & Saxe, 2015), as we found a large overlap between the pairwise and appraisal feature spaces, but for a much larger set of emotion concepts. However, as we only used 14 appraisal dimensions, the resulting appraisal space we used was relatively impoverished compared to that in prior work (Skerry & Saxe, 2015), and a higher-dimensional appraisal space may do a still better job of mirroring the similarity structure of emotion concept representation.

Our results also point to a potential limitation of large language models in capturing the meaning of different emotion terms. Large language models derive the similarity of terms by exploiting co-occurrence statistics in the training data, whether they are sensitive to context (GPT-3) or not (word2vec). Some of the appraisal features that humans use to determine the similarity of emotion concepts are likely to have a systematic relationship with the words that occur around them in corpora. An emotion term's valence, for example, is likely to have a systematic relationship to whether it co-occurs with terms like "puppy" or "murder." This kind of systematic relationship can easily be exploited by language models, essentially allowing them to capture that valence is a relevant dimension along which emotion terms differ. However, many of the other appraisal features that humans use to determine the similarity of emotion concepts are unlikely to have any such systematic relationship to co-occurrence statistics. For example, it is much harder to see what kind of systematic effect the extent to which an emotion is embodied will have on the words that co-occur with more embodied vs. less-embodied emotion terms. Features like embodiment, however, may still have a central role in the way that humans understand different emotions, since bodily feelings play a significant role in actually experiencing these emotions. This fact suggests a potential limitation in the extent to which large language models will be able to represent emotion concepts in a way that is similar to how humans do.

Future directions

There are a number of ways that the current study may be expanded or improved. First, future work may want to incorporate a wider scope of language models that use different training data, are fine-tuned on emotion-relevant text, employ different training methods, or are even more sensitive to contextual information. This would allow us to systematically investigate the impact of different characteristics of language models in their ability to capture the nuances of how humans represent and express emotion concepts. Second, as stated in the Discussion, future work should incorporate more appraisal features known to play a role in emotion representation. This can further inform how language models may differ from humans in the weight they place on different dimensions in understanding different emotion concepts.

Appendix: Online Supplements

Data, code, and supplementary results are available at https://doi.org/10.7910/DVN/6DPPKH

References

Barrett, L. F. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*.

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour*, 2(8), 581–591.

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences*, 116(32), 15861–15870.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. *ArXiv*:2005.14165.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, 114(38), E7900–E7909.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616), 1208–1210.

Ellsworth, P. C. (1994). William James and emotion: is a century of fame worth a century of misunderstanding?. *Psychological review*, 101(2), 222.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, 5(2), 125-131.

Fares, M., Kutuzov, A., Oepen, S., & Velldal, E. (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. *Proceedings of the 21st Nordic Conference on Computational Linguistics*, 271–276.

Gendron, M., & Feldman Barrett, L. (2009). Reconstructing the Past: A Century of Ideas About Emotion in Psychology. *Emotion Review*, 1(4), 316–339.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9(1), 87–108.

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472), 1517–1522.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 26.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review,* 5(2), 119–124.

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., Bölte, S., Baron-Cohen, S., & Lundqvist, D. (2016). The EU-Emotion Stimulus Set: A validation study. *Behavior Research Methods*, 48(2), 567–576.

Seyeditabari, A., Tabari, N., Gholizade, S., & Zadrozny, W. (2019). Emotional Embeddings: Refining Word Embeddings to Capture Emotional Content of Words. *ArXiv*:1906.00112

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 693-727.

Scherer, K. R., & Meuleman, B. (2013). Human Emotion Experiences Can Be Predicted on Theoretical Grounds: Evidence from Verbal Labeling. *PLoS ONE*, 8(3).

Skerry, A. E., & Saxe, R. (2015). Neural Representations of Emotion Are Organized around Abstract Event Features. *Current Biology*, 25(15), 1945–1954.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3), 242–249.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.