Data-driven classification of interpersonal reappraisals reveals eight distinct reappraisal strategies

Alina Herderich[1,2,4,*], Joanna Zun Li[2,3], and Amit Goldenberg[2,3,4]

[1]Graz University of Technology, Institute of Interactive Systems and Data Science

[2]Harvard Business School

[3]Harvard University, Digital, Data and Design Institute

[4]Harvard University, Department of Psychology

Author Note

Alina Herderich https://orcid.org/0000-0002-2940-600X

* The current version of this manuscript was written by Alina Herderich as a documentation of her stay at the Harvard Digital Emotions Lab during the academic year of 2023/2024.

Correspondence concerning this article should be addressed to Alina Herderich, Department of Computer Science and Biomedical Engineering, Graz University of Technology, Sandgasse 36/III, 8010 Graz, Austria. E-mail: alina.herderich@tugraz.at.

Abstract

Reappraisal — rethinking a negative emotional situation to alleviate its impact — is an effective emotion regulation strategy decreasing negative affect, increasing positive affect, social adaptation and well-being. However, the mechanisms behind the effectiveness of reappraisal are not yet well understood. To investigate the determinants of reappraisal success, we developed a reproducible, data-driven classification of reappraisal techniques and linked the identified techniques to reappraisal quality as defined by human raters. We used a novel method, the construct mining pipeline, on a large dataset ($n = 1521$) of interpersonal reappraisal suggestions with respect to six negative emotional vignettes to infer classes of reappraisal. We identified eight distinct techniques, namely growth mindset, positive repurposing, aberration, external justification, normalizing, unwarranted conclusions, validation, and humor. After assigning the presence or absence of each technique to entire reappraisals, we will apply linear regression to establish relationships between reappraisal classes and quality. Using a held-out test set ($n = 1513$), we will replicate our preregistered hypotheses to paint a coherent picture of the relationship between reappraisal techniques and quality. Our findings not only advance the understanding of the mechanisms behind reappraisal, but also enable the development of better targeted reappraisal training both in humans and automatic helper tools such as large language models.

Keywords: reappraisal, quality, emotion regulation, clustering, classification

Data-driven classification of interpersonal reappraisals reveals eight distinct reappraisal strategies

The ability to regulate one's emotions is an essential human skill that can determine a person's mental health (Gross & John, 2003; John & Gross, 2004) or social success (English et al., 2012; K. Wang et al., 2021). Emotion regulation is modifiable: People can learn or unlearn certain emotion regulation techniques (Cohen & Ochsner, 2018), which is crucial considering that most affective disorders such as depression or anxiety can be traced back to an underlying emotion regulation dysfunction (Aldao et al., 2010; Coifman & Summers, 2019; Cui et al., 2024). To leverage its full potential, researchers have set out to identify healthy forms of emotion regulation, as well as ways to teach them.

One emotion regulation technique that has received particular attention is the reevaluation of negative situations in a way that alleviates the impact of the elicited emotion, a technique called cognitive reappraisal (Gross, 1998, 2015; Uusberg et al., 2019). Reappraisal has proven effective in reducing negative emotions in general (i.e., as a habit; Gross & John, 2003), as well as in particular settings (e.g., the global COVID-19 pandemic; K. Wang et al., 2021). People can reappraise for themselves (*intraindividual* reappraisal), but can also share reappraisal suggestions with others (*interindividual* reappraisal). The relative ease with which reappraisal can be taught and shared has inspired the creation of accessible tools supporting people's reappraisal capacities, for example the use of large language models to offer reappraisal suggestions (Sharma et al., 2023, 2024; Zhan et al., 2024; Ziems et al., 2022).

Still, knowledge on the mechanisms behind the success of reappraisal is limited. Recent findings stem from the evaluation of large language models as reappraisal helpers. For example, Sharma et al. (2023) showed that people prefer highly empathic reappraisals that are specific to the situation compared to overly positive ones when generated with GPT-3. When reappraisals incorporate actionable advise, people perceive them as more helpful, relatable, memorable, learnable, and report a significant decrease in negative emotional intensity when provided with such advise (Sharma et al., 2024). In another study, after receiving targeted reappraisal training, GPT-4 provided reappraisal suggestions of higher quality compared to the majority of humans with respect to human judgment. Although reappraisal effectiveness and empathy was associated with situation specificity in GPT-4, for humans the reverse holds: When being semantically similar to the situation, human reappraisals were rated as less effective and less empathic (Li et al., 2024). While studies on the reappraisal ability of large language models provide valuable insights into reappraisal success, they were not designed to understand reappraisal dimensions and their relationship to the quality of reappraisals, especially in humans.

Due to their situation specificity, it is hard to identify overarching components of reappraisals, or reappraisal techniques. A common approach to describe reappraisals stems from psychotherapy, specifically Cognitive Behavioral Therapy (CBT; Beck, 1963). CBT identifies cognitive distortions, also referred to as thinking traps, which can be overcome by shifting perspectives. Such thinking traps include, for example, catastrophizing or overgenalization (Willams & Garland, 2002). Reappraisals can be understood as antagonists of cognitive distortions, where one reappraisal can address several of them. For instance, when designing a reappraisal intervention on a large American mental health platform driven by large language models, Sharma et al. (2024) classified reappraisals by addressing thinking traps, being rational, positive, empathic, actionable, specific, and readable. Other studies follow a similar approach to leverage CBT for teaching large language models reappraisal (e.g., Zhan et al., 2024; Ziems et al., 2022). However, in those studies, boundaries between reappraisal techniques and desirable reappraisal characteristics are not always clear cut and knowing what makes reappraisals desirable or effective does not necessarily imply how to get there.

In the past, only selected studies have tried to distinguish specific reappraisal techniques and relate them to reappraisal effectiveness. McRae et al. (2012) developed a coding system to classify intrapersonal reappraisal attempts with respect to emotional pictures. In this study, people tried to focus on the advantages of the situation, downplay the severity of the situation, challenge the authenticity of the picture, imagine desirable future developments, remind themselves of their sense of agency, distance themselves from the situation, propose practical solutions, or accept the situation. Challenging reality was not effective in increasing positive affect, whereas distancing and agency were effective in decreasing arousal.

Further evidence for the varying effectiveness of reappraisal techniques comes from research on reappraisal inventiveness (Weber et al., 2014). Reappraisal inventiveness describes a person's creativity in generating qualitatively different reappraisals for the same situation. Studies did not find an effect of reappraisal inventiveness on reappraisal success (Weber et al., 2014; Zeier et al., 2020), and that effect was especially pronounced when people were familiar with the situation (Zeier et al., 2024). This might be driven by the fact that only selected techniques are effective and that their effectiveness depends on the situation.

Finally, appraisal theory suggests that people perceive distinct emotions based on their appraisal of the environment (Yeo & Ong, 2024). With reappraisal constituting an intentional change in appraisal, we can theorize that reappraisals need to target an emotion's related appraisal dimensions in order to be successful. That is, if anger is elicited in situations which people perceive to be caused by another person, effective reappraisals would need to target said responsibility appraisal. Uusberg et al. (2019) propose that a shift in appraisal dimensions is achieved either by repurposing, meaning changing a person's goal set, or by reconstruing, meaning changing a person's perception of the situation directly. However, what and how appraisal dimensions are exactly shifted is yet to be defined and there is a limited overlap of categorizations, as well as lacking consensus on what should be called reappraisal (Uusberg et al., 2019).

Understanding different reappraisal techniques can help to create more effective reappraisal training. Reappraisal training has been a fruitful avenue in the past (Kam et al., 2024; K. Wang et al., 2021), but has not yet been targeted at specific reappraisal techniques (except Ranney et al., 2017). Training people how to reappraise is especially crucial since reappraisal success seems to be a matter of skill rather than effort (Li et al., 2024). Furthermore, delving into reappraisal techniques helps us to understand how people reappraise with respect to different modalities (e.g., oral versus visual stimuli), and in different settings (e.g., intrapersonal versus interpersonal reappraisal).

Here, we attempt to discover dimensions of reappraisals in an interpersonal setting using a data-driven approach and relate them to perceived reappraisal quality. We employ a new method, the construct mining pipeline (Herderich et al., 2024), which distills construct dimensions out of a collection of texts describing the construct of interest. Thus, we determine reappraisal to be the suggestions that people generate after being taught and prompted to reappraise. Specifically, after receiving a short reappraisal training, participants created reappraisal suggestions for a stranger with respect to six vignettes, two of which elicited anger, sadness, or anxiety, respectively. Another set of human participants, who received the same training, rated the reappraisals on effectiveness, empathy, novelty, and specificity, which taken together indicate perceived reappraisal quality. Overall, our dataset consisted of 3,034 reappraisals.

With that, our study aims to address the following questions: First, what techniques do non-experts use to reappraise after receiving a short, general reappraisal training? Second, do the identified techniques overlap with theoretically motivated reappraisal dimensions and if so, in which way? Third, are the identified techniques specific to interpersonal reappraisal and where do they intersect with different settings? Fourth, how do the identified reappraisal techniques relate to perceived reappraisal quality? These questions are particularly suitable for data-driven approaches, and new methods such as the construct mining pipeline render such approaches replicable and scalable. In the following, we will first introduce the construct mining pipeline (Herderich et al., 2024), a method blending computational and psychological techniques to infer construct dimensions from semi-structured text data. We will proceed to describe our dataset of interpersonal reappraisals generated with respect to six different emotional vignettes. Afterwards, we will present our results as intermediate steps along the construct mining pipeline and conclude with a classification of reappraisal techniques before we close with a general discussion and outlook on future steps.

Methods

A Short Introduction to the Construct Mining Pipeline

The construct mining pipeline (Herderich et al., 2024) is a method to establish data-driven classes of psychological constructs such as emotion regulation or reappraisal. By interleaving methods from machine learning and psychology, it aims to make data-driven definitions of psychological constructs robust, replicable, and scalable. As such, it constitutes a viable alternative to methods such as traditional qualitative analysis (Glaser & Strauss, 2010), factor analysis, and computational grounded theory (Nelson, 2020).

The construct mining pipeline comprises nine steps: (1) First, the researcher collects data with semi-structured, open-ended questions (often in the form of scenarios) to obtain a list of texts, where each text reflects a (still unknown) component of the construct of interest. (2) The texts are then embedded into a high-dimensional semantic space using sentence transformers (Reimers & Gurevych, 2019). (3) In order to measure the influence of the data collection on the embedding space an “item bias indicator” is calculated. With that the question is answered of how much the scenarios of the data collection influence the descriptions of the construct. (4) If the indicator reveals a significant “item bias”, the embedding space is debiased by masking relevant key words from the scenarios in the texts and recomputing the embeddings. (5) The debiased sentence embeddings can then be reduced in dimensionality (McInnes et al., 2020) preparing the embedding space for (6) clustering (Campello et al., 2013), which yields the psychologically relevant classes. (7) A robustness check is performed to ensure the independence of the results from stochastic components of the algorithm used for dimensionality reduction. (8) Next, a survey ("intrusion task") is conducted to evaluate the meaningfulness of clusters with respect to human judgement. (9) Finally, clusters remaining from the survey can be interpreted by the researcher as components of the construct of interest.

A Dataset of Reappraisal Attempts

Here, we try to infer a data-driven taxonomy of cognitive reappraisals in interpersonal settings, that is people trying to rethink a negative situation for another person to alleviate the situation's negative emotional impact. We base our analysis on a dataset from experiments evaluating the quality of reappraisals of GPT (Generative Pretrained Transformers) compared to humans (Li et al., 2024).

Specifically, we focus on data from two out of three studies, where we only use human reappraisals for our analysis. In study one, participants completed a training helping them to produce effective reappraisals (Pinus et al., 2023; K. Wang et al., 2021). They then reappraised six vignettes (among others) such as "My roommate laughed at my first attempt at cooking. I feel incompetent." two of which represented situations of sadness, anger, and anxiety, respectively (see Appendix A for details). An example of a reappraisal would be "That was your first attempt and it's okay to make mistakes." Each of the six vignettes received approximately 100 reappraisals, resulting in a set of $n_r = 611$ reappraisals in total.

In study two, participants again received a reappraisal training after which they reappraised the same six vignettes as in study one. The samples from both studies were disjunct. On top of that, participants were assigned to one out of four conditions incentivizing them to produce reappraisals of higher quality ranging from no incentive to 150% bonus based on standard payment. The sample consisted of $n_p = 404$ participants producing $n_r = 2423$ reappraisals in total.

Summing up the data from both studies, our dataset consisted of $n_r = 3023$ reappraisals overall. The studies were conducted on Prolific with residents from the United States or the United Kingdom fluent in English. For a complete description of the experimental set up and applied measures see Li et al. (2024).

Results

In the following, we will describe all intermediate results along the nine steps of the construct mining pipeline (Herderich et al., 2024, and see Section “A Short Introduction to the Construct Mining Pipeline”) to arrive at a taxonomy of interpersonal reappraisal techniques.

Data Preparation

For preparation, we split the data (see Section “A Dataset of Reappraisal Attempts”) in half to conduct exploratory analyses on the first half of the data and replicate our results on the second half of the data at a later stage of the process. The two data partitions are balanced over study one and study two, but are disjunct in human participants. That means, for a stronger replication of our results reappraisals from one person are only assigned to either half of the data. We further tested for differences in potential confounds between data splits: There were no differences in the occurrence of vignettes over data splits ($\chi^2(1) = 0.13$, $p = .99$), no differences in reappraisal length ($U = 1167830.5$, $p = .47$), and no differences in reappraisal quality (measured as a composite score of a 7-point Likert scale of effectiveness, empathy, novelty, and specificity; $U = 1165048.0$, $p = .55$).

Since the reappraisals were considerably long and we had reason to assume that one reappraisal contains several techniques, we split the reappraisals into sentences for the first half of the data on which we planned to fit the construct mining pipeline. Specifically, we split reappraisals by full stops, exclamation marks, question marks, and semicolons. The sentences were then associated with their original vignette. That is, if a reappraisal generated with respect to vignette one contained three sentences, after splitting we obtained three separate data points each associated with vignette one. Accordingly, the first data split included $n_r = 3324$ (split) reappraisals.

Sentence Embeddings

We used Python version 3.11.8 and Sentence Transformers version 2.2.2 (UKP Lab, 2024) for our analysis. We selected the “stsb-mpnet-base-v2” model (https://huggingface.co/sentence-transformers/stsb-mpnet-base-v2) of the Sentence Transformers library to embed the $n_r = 3324$ reappraisals. The model is a fine-tuned version of Microsoft’s “mpnet-base” and was trained on over one billion sentence pairs from Reddit to Wikipedia, as well as a dataset specifically encoding sentence similarity with respect to human ratings (STSB, Semantic Textual Similarity Benchmark, https://huggingface.co/datasets/sentence-transformers/stsb). The model embeds sentences or short paragraphs in a 768 dimensional space. We selected the model, because it is one of Sentence Transformer’s best performing models and explicitly suitable for clustering. We embedded the text of the six vignettes with the same model to estimate the influence of the data collection on the embedding space, which we discuss next.

Item Bias Measurement

When clustering on the original embedding space, clusters might form based on the content of the vignettes rather than different reappraisal techniques. For example, since we used the scenario “My roommate laughed at my first attempt at cooking.”, we could expect clusters forming based on “cooking” defying our goal of identifying reappraisal techniques. Therefore, we use the item bias indicator as proposed in Herderich et al. (2024) to measure how closely reappraisals are associated with their vignettes. Concretely, the indicator measures whether reappraisals created with respect to a certain vignette lie closer to that vignette in the embedding space than to all other vignettes. While a statistic of zero indicates no bias, negative values indicate a bias in the expected direction (i.e., reappraisals are closer to their vignette) and positive values indicate a bias in the unexpected direction (i.e., reappraisals are closer to other vignettes). Item bias is calculated per vignette. Through bootstrapping, we can compute 95% confidence intervals, thus being able to judge whether item bias is significant or not. In Table 1, we show item bias effect sizes (a standardized measure of the statistic) on our embedding space of reappraisals with respect to the six vignettes.

We can see that all vignettes have a negative bias. This means that reappraisals generated with respect to a certain vignette lie closer to that vignette in the embedding space than to all other vignettes. These results are expected, because by definition reappraisals are supposed to be a reinterpretation of a specific situation. However, to later make our reappraisal taxonomy more generalizable, we attempted to reduce item bias in the following step, which we describe next.

Item Bias Reduction

We reduced item bias in the embedding space through a technique termed masking. To this end, we defined keywords from the vignettes and replaced them with "[MASK]" in the reappraisal texts. For example for the vignette "My roommate laughed at my first attempt at cooking.", we determined "roommate*", "flatmate*", "laugh*", "first", "attempt*" and "cook*" as situation-specific key words. Stars indicate regular expressions and account for inflections and declensions of the masked words.

To check for additional situation-specific keywords beyond the vignette texts, we calculated a class based TF-IDF (term frequency inverse document frequency; Grootendorst, 2020), which allows to determine the most informative words — or "top words" — per vignette. With that, we found seven additional words, although most of the identified words stemmed from inspecting the vignette text directly. We list the keywords for all vignettes in Table A1 in the Appendix.

After masking the reappraisal texts with the identified keywords, we recomputed the sentence embeddings with our previous model and applied the item bias statistic to quantify the effect of debiasing. As shown in Table 1, we achieved a reduction of 28% to 37% for all vignettes, except vignette "Father", for which we were only able to reduce bias by 16%. Still, bias reduction was significant for all vignettes as indicated by the non-overlapping confidence intervals before and after. Given the successful debiasing, we

Table 1

Item Bias Effect Sizes Before and After Item Bias Reduction with 95% Confidence Intervals

Vignette	Before	After	Reduction
Cooking	-1.45 [-1.50; -1.41]	-0.92 [-1.00; -0.84]	37%
Hair	-1.26 [-1.32; -1.21]	-0.86 [-0.94; -0.78]	32%
Cans	-1.47 [-1.52; -1.43]	-1.03 [-1.11; -0.94]	30%
Father	-1.59 [-1.63; -1.55]	-1.33 [-1.38; -1.28]	16%
Gift	-1.46 [-1.52; -1.41]	-0.95 [-1.03; -0.87]	35%
Friend	-1.49 [-1.54; -1.44]	-1.08 [-1.16; -1.01]	28%

Note. For detailed vignette texts see Appendix A.

used the masked reappraisal embeddings going forward.

Dimensionality Reduction and Clustering

To infer reappraisal classes, we reduced the dimensionality of and clustered on the masked embedding space using UMAP (version 0.5.3) and HDBSCAN (version 0.8.33) as implemented in Python (McInnes, 2018; McInnes et al., 2018). UMAP and HDBSCAN have a set of hyperparameters that need to be chosen to control the fitting process. Specifically, for UMAP, we need to decide on the minimum distance that points in the reduced embedding space can be apart (d_{min}), the number of neighbors considered for density estimation (k_{UMAP}), and the number of dimensions to project into (n_{UMAP}). For HDBSCAN, we need to set the minimum number of points to form a cluster (c_{min}), as well as the number of neighbors for density estimation ($k_{HDBSCAN}$).

We chose $d_{min} = 0.05$ and $c_{min} = 10$ based on theoretical considerations (Herderich et al., 2024), and determined the remaining three parameters through hyperparameter tuning. To this end, we employed a grid search with $k_{UMAP} = [15, 30, 40, 45, 50]$, $n_{UMAP} = [10, 30, 50]$, and $k_{HDBSCAN} = [10, 15, 20, 25, 30, 35, 40, 45, 50]$. Since UMAP utilizes a random seed, we tried four different random seeds with the above-mentioned hyperparameter set, which we generated using Python's `random.randint` function. Desirable solutions are characterized by (1) being stable across different random seeds, (2) yielding the expected numbers of clusters (or more), (3) leaving as little points unclustered as possible (also called noise), and (4) reducing the number of dimensions as little as possible.

For the number of resulting clusters, we expected to find at least eight based on McRae et al. (2012) and around ten based on our own qualitative analyses. Considering that reappraisals are situation-specific and we could expect each reappraisal type to manifest differently over all six vignettes, we planned to settle on a solution between 10 and 60 clusters. Figure 1 displays the number of resulting clusters over all combinations of hyperparameters. We can see that for $n_{UMAP} = 50$ and and $k_{UMAP} = [15, 30]$, solutions are unstable, that is resulting in trivial solutions with no to very little clusters, especially for small values of $k_{HDBSCAN}$, which would otherwise yield numbers of clusters in our expected range. We therefore settled on $n_{UMAP} = 30$, $k_{UMAP} = 45$ and $k_{HDBSCAN} = 15$ resulting in solutions with roughly 30 clusters. Importantly, we aimed for a solution with more number of clusters, since discarding and merging clusters is still possible in upcoming steps of the pipeline as noted in Herderich et al. (2024). For the chosen set of hyperparameters the number of unclustered points revolved around 1500 to 1750, but noise was mostly proportional to the number of clusters, which is why we did not consider it further (see Figure B1 for details).

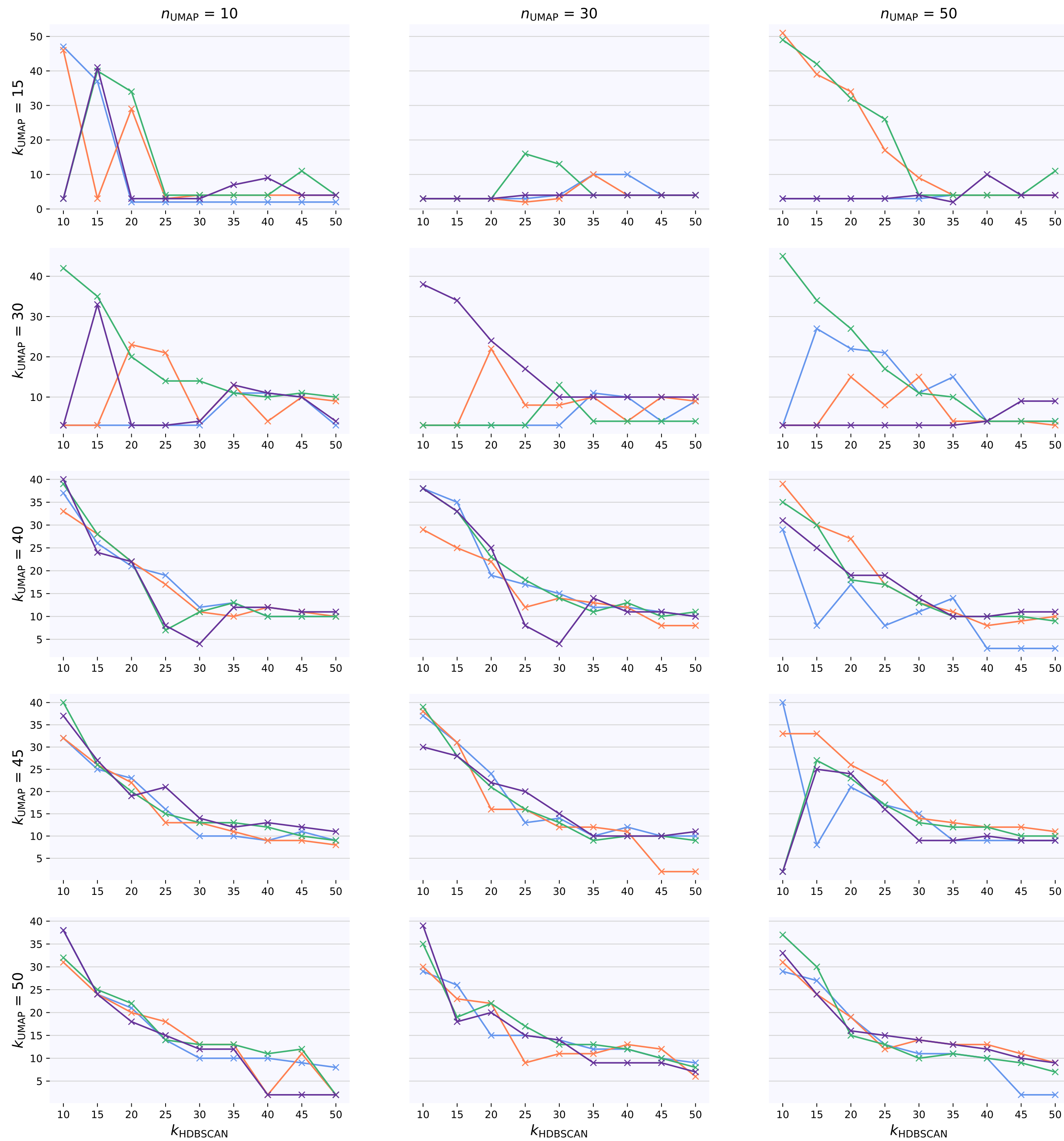
Robustness Check

To ensure that the number of clusters for our chosen hyperparameter combination ($n_{UMAP} = 30$, $k_{UMAP} = 45$, $k_{HDBSCAN} = 15$) is not the result of chance, we ran UMAP and HDBSCAN with 5000 different random seeds keeping all other parameters constant. Furthermore, to determine the final clustering, out of the 5000 runs, we chose one random seed yielding the number of clusters of the majority of solutions.

Figure 2 shows that the number of clusters over different random seeds ranges

Figure 1

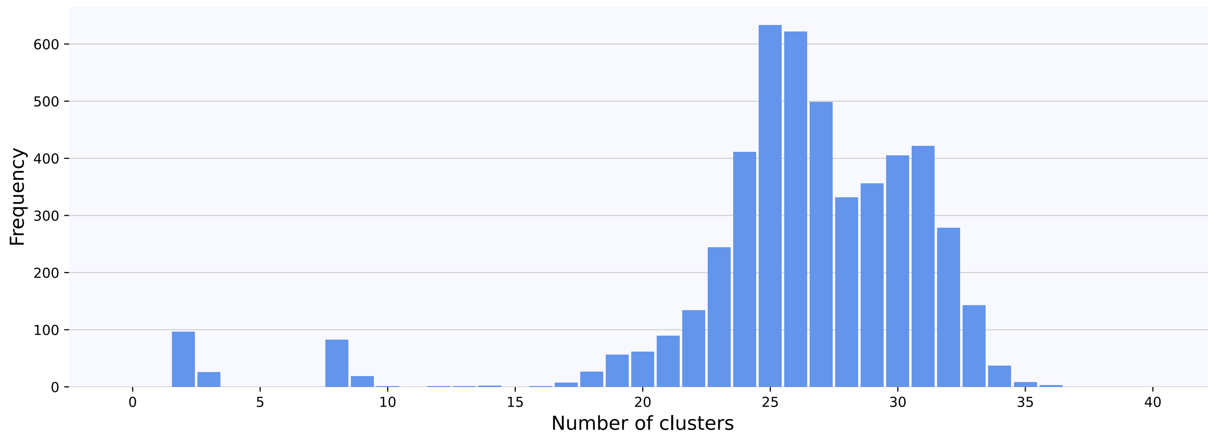
Number of Resulting Clusters Over a Set of Different UMAP and HDBSCAN Hyperparameters



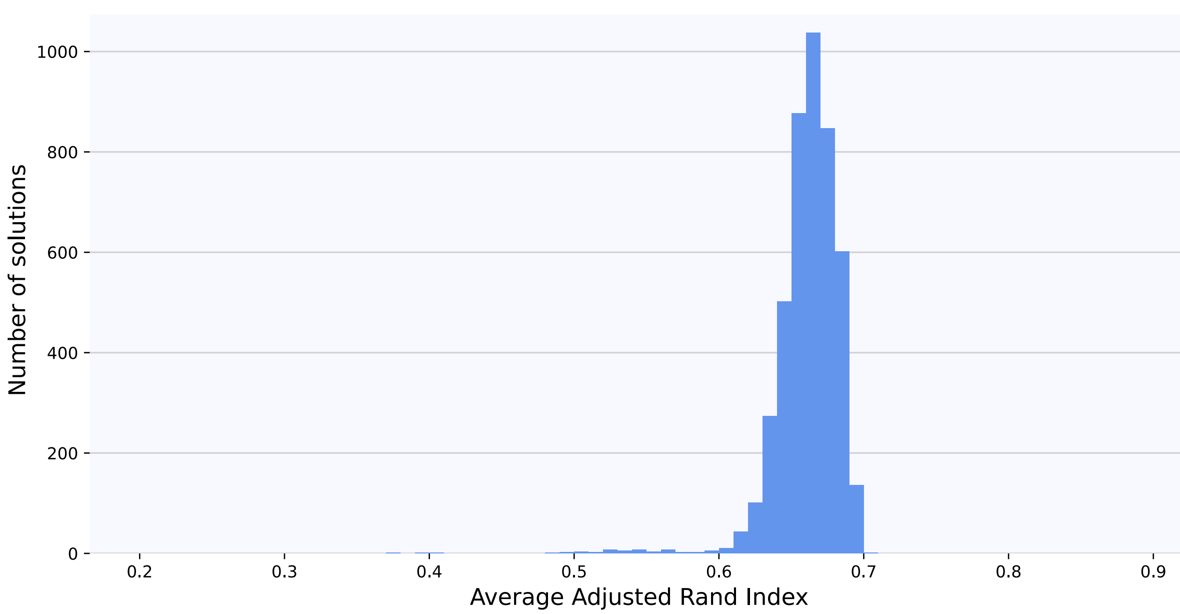
Note. The four different lines in each subplot denote the runs over four different random seeds.

Figure 2

Robustness Check



(a)



(b)

Note. (a) Histogram of number of resulting clusters over 5000 different UMAP and HDBSCAN runs. (b) Histogram of average Adjusted Rand Index per random seed over 5000 different UMAP and HDBSCAN runs.

considerably between 20 and 35, but rarely results in solutions with very little clusters. The distribution is bimodal, however, solutions with 25 clusters are more frequent

($n = 633$) than solutions with 31 clusters ($n = 422$). We thus chose a random seed yielding 25 clusters.

To examine the overlap of clusters across different solutions, we applied the adjusted rand index (ARI; Hubert & Arabie, 1985; Steinley et al., 2016). The ARI can range between 0 and 1 and is high if points are grouped similarly across two different clusterings. We defined "trivial solutions" to be solutions with less than four clusters (Figure 2A) and excluded them for this analysis. Effectively, ARIs ranged between .60 and .70 (Figure 2B), indicating moderate recovery (Steinley, 2004). Importantly, the ARI correlated with the difference in number of clusters between pairs of solutions ($r_{spearman} = -.73$), meaning that ARIs were higher for solutions with similar number of clusters.

Validating Clusters

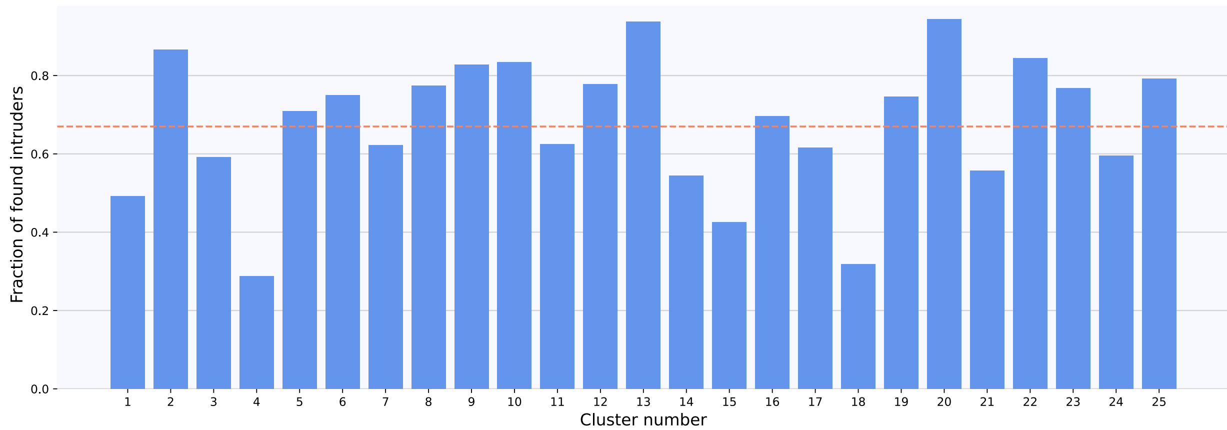
With the final clustering solution, we conducted an "intrusion survey" to validate the clusters beyond expert judgement. In the survey, participants are presented with four sentences at a time, three being from the same and one from a different cluster. They are supposed to identify which one of the sentences does not belong to the others. If people are able to identify the majority of "intruders" for a cluster on average, we declare it coherent and therefore interpretable.

We aimed at evaluating 30% of sentences for each cluster resulting in 205 survey questions in total. To limit the strain for participants, we presented each with a random set of 51 questions out of the entire question pool. We conducted the survey with a sample of $n = 81$ on Prolific with adult residents of the United States, who were fluent in English. We further added four attention checks to the questionnaire and terminated participation if two attention checks were failed. We obtained IRB approval for this study previous to conducting the survey.

For data analysis, we excluded four participants, because their completion time was three standard deviations below the mean ($\bar{x} = 22.7min$, $SD = 3.9min$). Our remaining sample of $n = 77$ was 71.4% female, 27.3% male, and 1.3% non-binary. People were 39.9

Figure 3

Fraction of Found Intruders per Cluster



Note. The dashed, orange line denotes our chosen cut-off score of .67 (two third majority).

years on average ($SD = 13.0$). We obtained 15 to 22 votings per question. Figure 3 depicts the fraction of found intruders for each of the clusters. Out of 25 clusters, 11 fall below our chosen cut-off score of .67 representing a two third majority. We defined said cut-off score, because a two third majority still represents a strong majority while taking the difficulty of the intrusion survey into account. In summary, we kept 14 clusters to interpret moving forward.

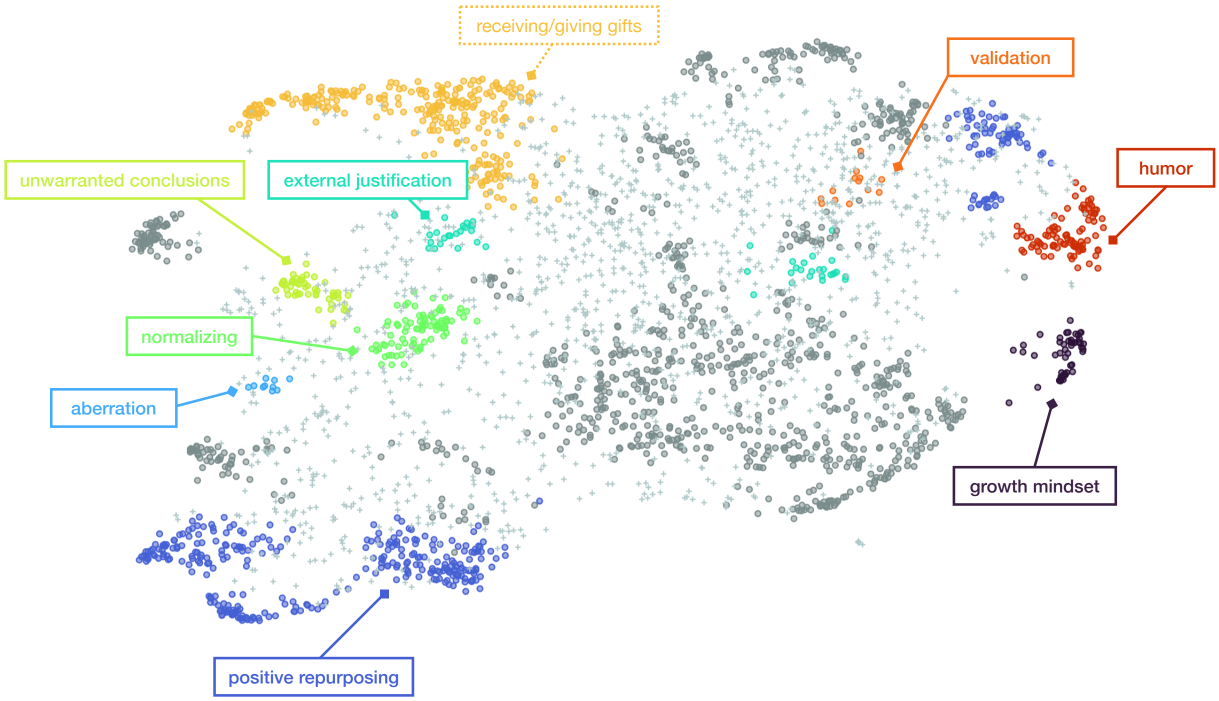
Interpreting and Consolidating Clusters

Finally, we interpreted and consolidated the clusters withstanding the intrusion survey through deep reading. Figure 4 shows a two dimensional representation of the sentence embedding space with annotated clusters. Table 2 summarizes the consolidated reappraisal classes including example sentences.

We found eight reappraisal classes in total: growth mindset, positive repurposing, aberration, external justification, normalizing, unwarranted conclusions, validation, and humor. We were able to interpret almost all of the clusters as reappraisal strategies, except for one (cluster 18), which revolved around receiving and giving gifts, but contained a mixture of strategies. The strategy to repurpose the situation by reevaluating the

Figure 4

Annotated Map of Reappraisal Clusters



Note. The figure displays a two dimensional representation of the sentence embedding space of reappraisals obtained with UMAP. Grey crosses denote unclustered points. Dark-grey circles denote clusters that did not withstand the intrusion survey. Colored circles denote interpretable clusters. Colors were chosen based on consolidated reappraisal classes. The orange cluster revolved around the topic "receiving/giving gifts" instead of an overall reappraisal strategy.

intentions of the other person was by far the most prevalent strategy, which is reflected in "positive repurposing" consisting of five clusters.

Despite our efforts to make the embedding space on which we clustered more independent from the vignettes used for data collection, the reappraisal classes were still highly situation-specific. We considered this a possibility, because reappraisals are by definition situation-specific. Furthermore, we used a variety of situations to capture reappraisal attempts theorizing that different situations might necessitate different strategies. Therefore, although situation-specific, the found reappraisal classes were still

Table 2

Description and Examples of Consolidated Reappraisal Classes

Reappraisal Class	**Related Clusters**	**Description**	**Example Sentence**
Growth mindset	2	understanding the situation as a learning opportunity	"Cooking is a skill, and every skill takes practice."
Positive repurposing	5, 9, 10, 20, 25	reevaluating the intentions of the other person(s) involved	"It shows how much he loves you."
Aberration	6	thinking of the situation as an exception rather than the norm	"She's actually looking forward to catching up with you next time."
External justification	8, 16	understanding that the cause of the situation might lie outside a person's control	"Perhaps your boss had other pressures in their life that made them lash out."
Normalizing	12	understanding that the situation is normal and advantageous	"Your friend has other friends and that's okay."
Unwarranted conclusions	13	questioning a person's conclusions from the situation	"Just because she hung out with other people doesn't mean she will abandon you."
Validation	22	validating a person's perception of the situation	"Your mother-in-law made a comment that was not very kind."
Humor	23	giving the situation a humorous twist	"You need to fire your hairstylist and laugh."

generalizable.

Discussion

The goal of the current investigation was to identify reappraisal dimensions, or specific techniques that people can use to rethink negative situations to alleviate their emotional impact, in an interpersonal setting. To this end, we had people reappraise negative emotional vignettes for strangers after receiving a short, general reappraisal training. We used the construct mining pipeline (Herderich et al., 2024), a novel method blending psychological and computational techniques, to infer a data-driven classification of reappraisals. We found eight distinct reappraisal techniques, namely growth mindset, positive repurposing, aberration, external justification, normalizing, unwarranted conclusions, validation, and humor.

The reappraisal techniques we identified demonstrate how negative thinking traps, or cognitive distortions (Beck, 1963; Burns, 1980; Yurica & DiTomasso, 2005), can be targeted concretely. For example, "positive repurposing", that is rethinking another person's intentions, is a response to disqualifying positive evidence and mind reading (i.e., interpreting another person's actions to be negatively motivated). "Aberration", that is declaring a situation a coincidence rather than the norm, can target overgeneralization and magnifying negative experiences. Imagining "external justifications" can prevent a person to attribute the cause of negative experiences to themselves. Finally, being reminded that certain implications might not be justified (a technique we termed "unwarranted conclusions") can answer to selective abstraction of negative details of a situation, and jumping to conclusions. Still, reappraisal techniques must not only target specific thinking traps, but can correct negative thinking more broadly as well.

Until today, only selected studies have explicitly investigated the effectiveness of different reappraisal tactics. When teaching people either one of two broad, theoretical reappraisal techniques, both reconstrual and repurposing (Uusberg et al., 2019) had similar positive effects compared to a control condition (K. Wang et al., 2021). McRae et al. (2012) identified eight reappraisal tactics with respect to intrapersonal reappraisal of emotional pictures. Although investigating a different setting and modality, many of the identified themes reflect in our classes. In their sample the authors, too, found the idea that the situation has explicit advantages or is not as bad as it seems ("positive repurposing"), that the situation isn't real ("unwarranted conclusions"), that the situation will improve on its own ("aberration") or due to a person's intervention ("growth mindset"), that the situation is unrelated to the self ("external justification"), and that the occurrence of the negative situation is normal at least to some extent (“normalizing”).

When reappraising, people exploit the ambiguity of the situation and its social signals. This is evidenced by the fact that rethinking a person’s intentions is the most prevalent technique in our clustered reappraisals, and supported by techniques such as external justification and unwarranted conclusions. Other techniques are taking the situation as a given, but try to manipulate its importance. For instance, growth mindset, aberration, and normalizing all do not contest the situation but portray it as normal or easily modifiable. Validation — the nonjudgmental acceptance of another person’s feelings — could be a technique specific to interpersonal reappraisal. Parental validation is an important factor in the emotional development of children (Jeon & Park, 2024; Lambie & Lindberg, 2016). In interpersonal reappraisal, validation could function as a pathway to helping other’s access their own emotion regulation skills by reducing initial stress (Shenk & Fruzetti, 2011), or buffering against a decrease in positive affect (Benitez et al., 2022), thus freeing cognitive resources.

Compared to serious reappraisals, humorous reappraisals have been shown to be an effective strategy, at least when people were coping for themselves (Kugler & Kuhbander, 2015; Samson et al., 2013). In an interpersonal setting, however, there’s a flip side: People have different humor styles and not everybody will consider the same things funny. In general, the key to a successful humorous reappraisal appears to be the use of benign instead of aggressive or self-defeating humor (Amjad & Dasti, 2022; Samson & Gross, 2012). On the one hand, humor could be beneficial in distant social settings, since people preferred social sharing over other emotion regulation strategies in close relationships (Tanna & MacCann, 2023). On the other hand, using humor to reappraise can have beneficial effects in close relationships by increasing psychological intimacy or relationship satisfaction (Horn et al., 2019; Walker et al., 2024).

Independent of the applied technique, reappraisals are always highly situation-specific. When interpreting the extracted clusters, we found that although representing generalizable reappraisal techniques, each cluster largely referred to one single vignette. Furthermore, within each cluster specific reappraisal suggestions were fairly common. For example, people often attributed the fact that the father stops by unannounced as him being lonely. This suggests that some reappraisals come more easily and the question is whether obvious or unobvious suggestions are perceived as more helpful. Research has shown that people with similar emotional responses to a stressful laboratory task also indicated similar appraisals (Siemer et al., 2007). That said, people with specific appraisals might prefer matching reappraisals, and the challenge in interpersonal settings might be to find this sweet spot. This is supported by the fact that generating more reappraisals is not indicative of reappraisal effectiveness (Zeier et al., 2020) and that this is especially true for people with high situational familiarity (Zeier et al., 2024).

Interpersonal and intrapersonal emotion regulation can be substantially different (Zaki & Williams, 2013), and so can reappraisal. For example, a recent study of interpersonal emotion regulation showed that both expressive suppression and cognitive reappraisal reduced negative and increased positive affect in targets, while increasing negative effect in regulators (Y. Wang & Shi, 2024). Intrapersonal reappraisal might be characterized through unique techniques not represented in our dataset. Mapping reappraisal tactics over different settings and modalities is a ripe avenue for future research. Another question concerns relevant pathways (e.g., empathy) over which different interpersonal techniques function (Levy-Gigi & Shamay-Tsoory, 2017). Finally, the success of interpersonal reappraisal depends on the intensity of the negative emotion (Nozaki & Mikolajczak, 2023) raising the question of further potential constraining factors.

Still, understanding reappraisal techniques will help improving reappraisal capacity more purposefully. On the one hand, reappraisal techniques will allow us to differentiate high quality from low quality reappraisals and reappraisers. On the other hand, knowledge about high quality reappraisals can be utilized to improve automatic tools such as large language models specialized in reappraisal (Li et al., 2024). Incorporating domain knowledge in machine learning to improve performance has been proven extremely effective in other fields to date (Jumper et al., 2021). Furthermore, understanding reappraisal techniques can help to develop more targeted and systematic reappraisal training. Research indicates that several reappraisal techniques could be equally effective (Ranney et al., 2017), although this has not been investigated exhaustively. Other findings point to the variation in reappraisal effectiveness based on the fit of technique and emotion, and people will choose techniques based on perceived and not actual effort and effectiveness (Vishkin et al., 2020). In general, reappraisal quality, which is supposed to differ between tactics, has been shown to mediate the relationship between the use of reappraisal and its effects on positive, as well as negative affect (Southward et al., 2022).

Our study constitutes the first attempt in classifying reappraisal techniques data-driven, yet standardized. However, it is not without limitations. First, reappraisals were generated with respect to only six scenarios. It is not clear whether those vignettes evoke all potential reappraisal tactics. People might be able to identify more or less with the presented scenarios and therefore, their reappraisals might be more or less diverse. Future research needs to replicate the current classification of reappraisal techniques. The construct mining pipeline is repeatable and extensible, and in principle suitable to do so. Second, the construct mining pipeline is suitable for discovering, not assigning classes. It has been shown that identifying high-density regions in the embedding space, or in other words identifying the most representative samples, can paint an exhaustive picture of classes at least within a given sample (Herderich et al., 2024). However, the method leaves a considerable amount of reappraisals not clustered and even if clusters are coherent, they might not be entirely pure. Therefore, identified classes ideally need to be reassigned to each sample either manually or with a separate machine learning model trained specifically for classification before statistical analysis. Third, we used effectiveness, novelty, empathy, and specificity to rate reappraisals, but there might be other indicators to assess reappraisal quality. Furthermore, we used an independent sample to rate reappraisal quality rather than people receiving the reappraisals directly. Future research will need to investigate the relationship between reappraisal techniques and these or other quality indicators.

Outlook

Identifying reappraisal dimensions was only the first step in our endeavor to link different reappraisal techniques to effectiveness. We now plan to undertake the following steps to round off the current project: First, we will manually assign each of the eight techniques to full reappraisals in the training dataset. That is, for the classification, we split reappraisals into smaller entities (see Section "Data Preparation") to separate reappraisal tactics more neatly. For the analysis of reappraisal techniques on effectiveness, we would like to analyze reappraisals as a whole, first and foremost, because reappraisal quality was assigned to entire reappraisals. To this end, for each of the eight techniques two raters will assign a one to an entire reappraisal if said technique is present and a zero if not. After calculating interrater reliability, and only if interrater reliability is sufficient, we will apply linear regression to model the relationship of reappraisal techniques on each of the four quality measures (effectiveness, novelty, empathy, and specificity). This approach will also give us the opportunity to factor in further predictors such as the amount of techniques used in a single reappraisal or interactions between techniques. Based on this exploratory analyses, we will generate hypotheses, which we will preregister and test on the remaining half of our data, the test set. With that, we hope to paint a more coherent picture of what specific reappraisal techniques are beneficial and which are not.

Acknowledgments

We thank Diane Sun for joining us as an annotator in the upcoming part of the project.

Declarations

Funding

AH was supported by the ERP Program of the German Academic Scholarship Foundation, the Marshall Plan Scholarship of the Austrian Marshall Plan Foundation, and the Marietta-Blau Fellowship of Austria's Agency for Education and Internationalization (Award No. MPC-2023- 00677).

Competing Interests

The authors declare no competing interests.

Inclusion and Ethics

The current research was approved by the IRB of Harvard under application number IRB23-0216 and IRB23-1383 (original studies), as well as IRB24-0943 (intrusion survey). The intrusion survey was further evaluated and approved by Graz University of Technology (EK-39/2024).

Data Availability

The data of the original studies is available on OSF under accession code https://osf.io/z9e48/.

Code Availability

Analysis code will be shared publicly once the project is released as a preprint.

Author's Contributions

AH contributed to the conceptualization, methodology, formal analysis, software, data curation, visualization, and funding acquisition of this project and wrote the initial draft of the manuscript.

JZL contributed to the conceptualization, methodology, investigation, data curation, and project administration.

AG contributed to the conceptualization, methodology, resources, and funding acquisition and was the supervisor of this project.

References

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237. https://doi.org/https://doi.org/10.1016/j.cpr.2009.11.004

Amjad, A., & Dasti, R. (2022). Humor styles, emotion regulation and subjective well-being in young adults. *Current Psychology*, *41*, 6326–6335. https://doi.org/https://doi.org/10.1007/s12144-020-01127-y

Beck, A. T. (1963). Thinking and depression. *Archives of General Psychiatry*, *9*, 324–333. https://doi.org/10.1001/archpsyc.1963.01720160014002

Benitez, C., Howard, K. P., & Cheavens, J. S. (2022). The effect of validation and invalidation on positive and negative affective experiences. *The Journal of Positive Psychology*, *17*(1), 46–58. https://doi.org/https://doi.org/10.1080/17439760.2020.1832243

Burns, D. D. (1980). *Feeling good*. Morrow.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates [Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160–172, Vol. 7819). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14

Cohen, N., & Ochsner, K. N. (2018). From surviving to thriving in the face of threats: The emerging science of emotion regulation training. *Current Opinion in Behavioral Sciences*, *24*, 143–155. https://doi.org/https://doi.org/10.1016/j.cobeha.2018.08.007

Coifman, K. G., & Summers, C. B. (2019). Understanding emotion inflexibility in risk for affective disease: Integrating current research and finding a path forward. *Frontiers in Psychology*, *10*, 392. https://doi.org/https://doi.org/10.3389/fpsyg.2019.00392

Cui, X., Ding, Q., Yu, S., Zhang, S., & Li, X. (2024). The deficit in cognitive reappraisal capacity in individuals with anxiety or depressive disorders: Meta-analyses of behavioral and neuroimaging studies. *Clinical Psychology Review*, *114*, 102480. https://doi.org/https://doi.org/10.1016/j.cpr.2024.102480

English, T., John, O. P., Srivastava, S., & Gross, J. J. (2012). Emotion regulation and peer-rated social functioning: A 4-year longitudinal study. *Journal of Research in Personality*, *46*(6), 780–784. https://doi.org/https://doi.org/10.1016/j.jrp.2012.09.006

Glaser, B. G., & Strauss, A. L. (2010). *The discovery of grounded theory: Strategies for qualitative research* (Second Edition). Aldine Transaction.

Grootendorst, M. (2020). Ctfidf [Accessed: 2024-10-18].

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*, 1–26.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Herderich, A., Freudenthaler, H. H., & Garcia, D. (2024). A computational method to reveal psychological constructs from text data. *Psychological Methods*, Advance Online Publication.

Horn, A. B., Samson, A. C., Debrot, A., & Perrez, M. (2019). Positive humor in couples as interpersonal emotion regulation: A dyadic study in everyday life on the mediating role of psychological intimacy. *Journal of Social and Personal Relationships*, *36*(8), 2376–2396. https://doi.org/https://doi.org/10.1177/0265407518788197

Hubert, L., & Arabie, P. (1985). Coparing partitions. *Journal of Classification*, *2*, 193–218. https://doi.org/10.1007/BF01908075

Jeon, J., & Park, D. (2024). Your feelings are reasonable: Emotional validation promotes persistence among preschoolers. *Developmental Science*, *27*(5), Article e13523. https://doi.org/https://doi.org/10.1111/desc.13523

John, O. P., & Gross, J. J. (2004). Healthy and Unhealthy Emotion Regulation: Personality Processes, Individual Differences, and Life Span Development. *Journal of Personality*, *72*(6), 1301–1334. https://doi.org/10.1111/j.1467-6494.2004.00298.x

Jumper, J., Evans, R., Pritzel, A., & et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*, 583–589. https://doi.org/https://doi.org/10.1038/s41586-021-03819-2

Kam, J. W. Y., Wan-Sai-Cheong, L., Zuk, A. A. O., Mehta, A., Dixon, M. L., & Gross, J. J. (2024). A brief reappraisal intervention leads to durable affective benefits. *Emotion*, *24*(7), 1676–1688. https://doi.org/https://doi.org/10.1037/emo0001391

Kugler, L., & Kuhbander, C. (2015). That's not funny! – but it should be: Effects of humorous emotion regulation on emotional experience and memory. *Frontiers in Psychology*, *6*, Article 1296. https://doi.org/https://doi.org/10.3389/fpsyg.2015.01296

Lambie, J. A., & Lindberg, A. (2016). The role of maternal emotional validation and invalidation on children's emotional awareness. *Merrill-Palmer Quarterly*, *62*(2), 129–157. https://doi.org/https://doi.org/10.13110/merrpalmquar1982.62.2.0129

Levy-Gigi, E., & Shamay-Tsoory, S. G. (2017). Help me if you can: Evaluating the effectiveness of interpersonal compared to intrapersonal emotion regulation in reducing distress. *Journal of Behavior Therapy and Experimental Psychiatry*, *55*, 33–40. https://doi.org/https://doi.org/10.1016/j.jbtep.2016.11.008

Li, J. Z., Herderich, A., & Goldenberg, A. (2024, April). Skill but not effort drive gpt overperformance over humans in cognitive reframing of negative scenarios. https://doi.org/10.31234/osf.io/fzvd8

McInnes, L. (2018). Umap: Uniform manifold approximation and projection for dimension reduction [Accessed: 2024-10-18].

McInnes, L., Healy, J., & Astels, S. (2018). Docs: The hdbscan clustering library [Accessed: 2024-10-18].

McInnes, L., Healy, J., & Melville, J. (2020, September). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Number: arXiv:1802.03426]. Retrieved December 14, 2022, from http://arxiv.org/abs/1802.03426
Comment: Reference implementation available at http://github.com/lmcinnes/umap.

McRae, K., Ciesielski, B., & Gross, J. J. (2012). Unpacking cognitive reappraisal: Goals, tactics, and outcomes. *Emotion*, *12*(2), 250–255. https://doi.org/10.1037/a0026351

Nelson, L. K. (2020). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*, *49*(1), 3–42. https://doi.org/10.1177/0049124117729703

Nozaki, Y., & Mikolajczak, M. (2023). Effectiveness of extrinsic emotion regulation strategies in text-based online communication. *Emotion*, *23*(6), 1714–1725. https://doi.org/https://doi.org/10.1037/emo0001186

Pinus, M., Halperin, E., Cao, Y., Coman, A., Gross, J., & Goldenberg, A. (2023, October). Emotion regulation contagion. https://doi.org/10.31219/osf.io/km6r4

Ranney, R. M., Bruehlman-Senecal, E., & Ayduk, O. (2017). Comparing the effects of three online cognitive reappraisal trainings on well-being. *Journal of Happiness Studies*, *18*, 1319–1338. https://doi.org/https://doi.org/10.1007/s10902-016-9779-0

Reimers, N., & Gurevych, I. (2019, August). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks [Number: arXiv:1908.10084]. Retrieved December 14, 2022, from http://arxiv.org/abs/1908.10084
Comment: Published at EMNLP 2019.

Samson, A. C., Glasscoa, A. L., Leea, I. A., & Gross, J. J. (2013). Humorous coping and serious reappraisal: Short-term and longer-term effects. *Europe's Journal of Psychology*, *10*(3), 571–581. https://doi.org/10.5964/ejop.v10i3.730

Samson, A. C., & Gross, J. J. (2012). Humour as emotion regulation: The differential consequences of negative versus positive humour. *Cognition and Emotion*, *26*(2), 375–384. https://doi.org/https://doi.org/10.1080/02699931.2011.585069

Sharma, A., Rushton, K., Lin, I. W., Nguyen, T., & Althoff, T. (2024). Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring. https://arxiv.org/abs/2310.15461

Sharma, A., Rushton, K., Lin, I. W., Wadden, D., Lucas, K. G., Miner, A. S., Nguyen, T., & Althoff, T. (2023). Cognitive reframing of negative thoughts through human-language model interaction. https://arxiv.org/abs/2305.02466

Shenk, C. E., & Fruzetti, A. E. (2011). The impact of validating and invalidating responses on emotional reactivity. *Journal of Social and Clinical Psychology*, *30*(2), 163–183. https://doi.org/https://doi.org/10.1521/jscp.2011.30.2.163

Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation–different emotions: How appraisals shape our emotions. *Emotion*, *7*(3), 592–600. https://doi.org/https://doi.org/10.1037/1528-3542.7.3.592

Southward, M. W., Holmes, A. C., Strunk, D. R., & Cheavens, J. S. (2022). More and better: Reappraisal quality partially explains the effect of reappraisal use on changes in positive and negative affect. *Cognitive Therapy and Research*, *46*, 73–85. https://doi.org/https://doi.org/10.1007/s10608-021-10255-z

Steinley, D. (2004). Properties of the Hubert-Arable adjusted rand index. *Psychological Methods*, *9*(3), 386–396. https://doi.org/10.1037/1082-989X.9.3.386

Steinley, D., Brusco, M. J., & Hubert, L. (2016). The variance of the adjusted rand index. *Psychological Methods*, *21*(2), 261–272. https://doi.org/10.1037/met0000049

Tanna, V. J., & MacCann, C. (2023). I know you so I will regulate you: Closeness but not target's emotion type affects all stages of extrinsic emotion regulation. *Emotion*, *23*(5), 1501–1505. https://doi.org/https://doi.org/10.1037/emo0001073

UKP Lab. (2024). Sentence transformers documentation [Accessed: 2024-10-17].

Uusberg, A., Taxer, J. L., Yih, J., Uusberg, H., & Gross, J. J. (2019). Reappraising reappraisal. *Emotion Review*, *11*(4), 267–282. https://doi.org/https://doi.org/10.1177/1754073919862617

Vishkin, A., Hasson, Y., Millgram, Y., & Tamir, M. (2020). One size does not fit all: Tailoring cognitive reappraisal to different emotions. *Personality and Social Psychology Bulletin*, *46*(3), 469–484. https://doi.org/10.1177/0146167219861432

Walker, S., Pinkus, R., Olderbak, S., & MacCann, C. (2024). People with higher relationship satisfaction use more humor, valuing, and receptive listening to regulate their partners' emotions. *Current Psychology*, *43*, 2348–2356. https://doi.org/https://doi.org/10.1007/s12144-023-04432-4

Wang, K., Goldenberg, A., Dorison, C. A., & et al. (2021). A multi-country test of brief reappraisal interventions on emotions during the covid-19 pandemic. *Nature Human Behaviour*, *5*, 1089–1110. https://doi.org/10.1038/s41562-021-01173-x

Wang, Y., & Shi, Y. (2024). Interpersonal emotion regulation and physiological synchrony: Cognitive reappraisal versus expressive suppression. *Cognition and Emotion*, Advance online publication. https://doi.org/https://doi.org/10.1080/02699931.2024.2371092

Weber, H., Loureiro de Assunção, V., Martin, C., Westmeyer, H., & Geisler, F. C. (2014). Reappraisal inventiveness: The ability to create different reappraisals of critical situations. *Cognition and Emotion*, *28*(2), 345–360. https://doi.org/10.1080/02699931.2013.832152

Willams, C., & Garland, A. (2002). Identifying and challenging unhelpful thinking. *Advances in PSychiatric Treatment*, *8*, 377–386. https://doi.org/10.1192/apt.8.5.377

Yeo, G. C., & Ong, D. C. (2024). Associations between cognitive appraisals and emotions: A meta-analytic review. *Psychological Bulletin*, Advance Online Publication. https://doi.org/https://doi.org/10.1037/bul0000452

Yurica, C. L., & DiTomasso, R. A. (2005). Cognitive distortions. In A. Freeman, S. H. Felgoise, C. M. Nezu, A. M. Nezu, & M. A. Reinecke (Eds.), *Encyclopedia of cognitive behavior therapy* (pp. 117–122). Springer US. https://doi.org/10.1007/0-306-48581-8_36

Zaki, J., & Williams, W. C. (2013). Interpersonal emotion regulation. *Emotion*, *13*(5), 803–810. https://doi.org/https://doi.org/10.1037/a0033839

Zeier, P., Sandner, M., & Wessa, M. (2020). Script-based reappraisal test introducing a new paradigm to investigate the effect of reappraisal inventiveness on reappraisal effectiveness. *Cognition and Emotion*, *34*(4), 793–799. https://doi.org/10.1080/02699931.2019.1663153

Zeier, P., Sandner, M., & Wessa, M. (2024). Regulating emotions with experience – the effectiveness of reappraisal inventiveness depends on situational familiarity. *Anxiety, Stress, & Coping*, *37*(1), 77–85. https://doi.org/10.1080/10615806.2023.2200999

Zhan, H., Zheng, A., Lee, Y. K., Suh, J., Li, J. J., & Ong, D. C. (2024). Large language models are capable of offering cognitive reappraisal, if guided. https://arxiv.org/abs/2404.01288

Ziems, C., Li, M., Zhang, A., & Yang, D. (2022). Inducing positive perspectives with text reframing. https://arxiv.org/abs/2204.02952

Appendix A

Vignette Texts

Table A1

Vignettes and Their Respective Masked Words

Emotional Vignette	Masked Words
My roommate laughed at my first attempt at cooking. I feel incompetent.	roommate*, flatmate*, laugh*, first, attempt*, cook*
My mother-in-law said my hair looks disheveled in front of everyone at our family dinner. I feel miserable.	mother-in-law*, mother in law*, hair, disheveled, family, dinner
My boss made me clean the bathroom all week after I knocked over some cans during my first shift. I feel offended.	boss*, clean*, bathroom*, week*, knock*, can*, shift*, rota*, task*[1], work*[1], job*[1]
My father comes over to my house unannounced to "check on me" even though I am married and have kids. I feel irritated.	father*, dad, house, unannounced, check*, married, kids, child*[1]
My new friend rejected my gift even though it was supposed to be a show of thanks for their help on a task last week. I worry I'm weird.	friend*, reject*, gift*, help*, task*
My friend hung out with others without me and posted a lot of pictures together on Instagram. I worry she is going to abandon me.	friend*, hang*, out, hung out, without, post*, picture*, Instagram, Insta, IG, group*[1], together[1], social media[1]

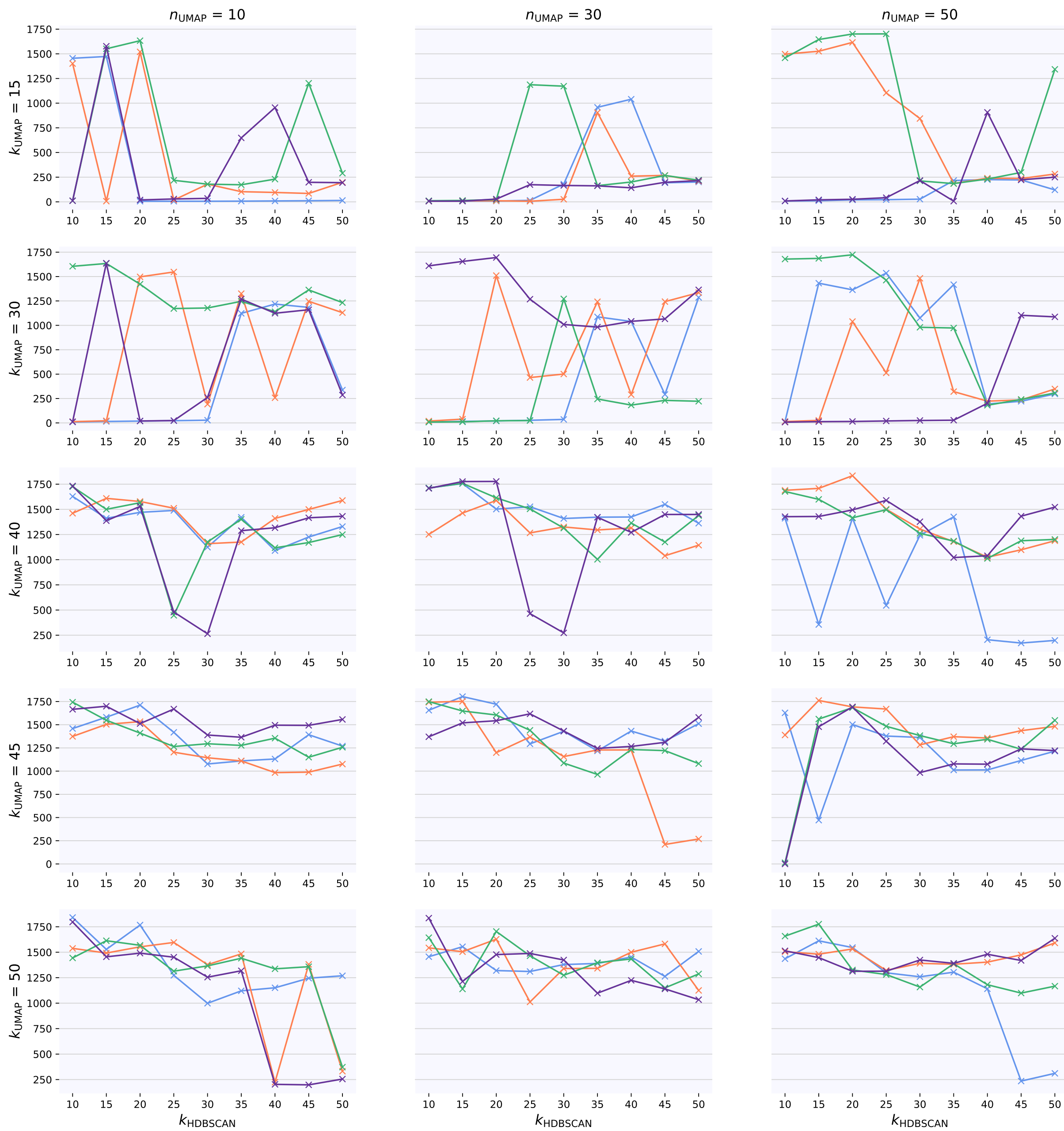
Note. [1] Words identified with ctfidf.

Appendix B

Number of Unclustered Points

Figure B1

Number of Unclustered Points Over a Set of Different UMAP and HDBSCAN Hyperparameters



Note. The four different lines in each subplot denote the runs over four different random seeds.