Emotion experience from stories, videos and everyday life: structure and individual differences

Thesis by

Yanting Han

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended May 25, 2022

Yanting Han
ORCID: 0000-0003-3381-2059

ACKNOWLEDGEMENTS

Having spent the past six years at Caltech and in LA has been an amazing experience. Caltech is a small but close community surrounded with nice people. I love the mountains, the beaches, the beautiful sunsets, and numerous authentic Chinese restaurants in SGV (very important!), which together have made my life so much more enjoyable.

First, I would like to express my deepest gratitude to my advisor, Dr. Ralph Adolphs. Ralph is the best advisor one can hope for, who is incredibly knowledgeable, kind, and supportive. Despite his crazy schedule (one that I could never live with), he has always managed to find time to discuss ideas and findings, provide constructive feedback, and offer good advice at moments of doubts and difficulty. He also cares deeply about the mental wellbeing of lab members by constantly reminding us to balance work and life, to spend more time with family and friends, and to prioritize our happiness. We are probably the lab at Caltech which spends most time talking about non-science stuff at lab meetings. Working in academia has always been challenging, but Ralph has managed the lab so well that we've always had each others' back.

I am thankful to members of my thesis committee, Drs. Markus Meister, Frederick Eberhardt, and Shinsuke Shimojo for offering valuable advice on my research. I also thank all of my dear labmates (especially Remya Nair, Lynn Paul, Tessa Rusch, Chujun Lin and Brooks Fu) for their support.

The pandemic over the past two years has made me realize now more than ever how important family and friends are. I am so fortunate to have all of my family members who care for me deeply despite living very far away, including my cutest little niece. I owe many thanks to my dearest friends (Yaoyuan Miao, Hang Yang, Zoey Gu, Ye Jin) whom I don't get to spend a lot of time with but you guys have always been there for me.

I'd like to dedicate this thesis to my parents and my fiancé.

My parents have offered me all of their time, attention, and unconditional love ever since I was born. They have been understanding and supportive of my decisions to pursue dreams in other cities and countries, even though it is not easy for them to have their only child be constantly absent in their lives. Mom and Dad: I love you both so much and thank you for everything.

Finally, I thank my fiancé, Dr. Tianhao Le, for accompanying me on this journey. Tianhao is the smartest and most sincere person I know. I am glad to have found my best friend and the love of my life here at Caltech. Thank you for all that you've done for me and I look forward to spending a lifetime with you.

ABSTRACT

Most studies of emotion have as their subject matter the emotion experiences that people can describe and rate. By contrast to this approach from psychology, studies in animals, and some biological studies in humans, focus on behavior and its adaptive function. These two literatures typically use very different corresponding features by which to characterize emotion: categories or dimensions describing feelings for which we have convenient words, for the former (e.g., happiness, pleasantness), and functional properties for the latter (e.g., persistence, generalizability, approachability). In this thesis I use both sets of ratings, and I ask whether the latter, biologically inspired features could also be used to characterize people's emotion experiences, and might reveal novel dimensions of variability. They also typically use different sets of stimuli to induce the emotions: lexical stimuli in which participants are asked to imagine something hypothetical are common in human studies; ecologically valid stimuli that at least the subjects cannot distinguish from the real world are common in animal studies. Here I used three domains of stimuli: stories, videos, and real-life experiences, in the same set of participants, permitting a unique comparison.

I took advantage of a sample of approximately 1000 Americans who were surveyed longitudinally over the internet during the COVID-19 pandemic. I collected ratings of emotion experiences evoked by three classes of stimuli: a validated set of short stories, a validated set of short videos, and actual experiences in real life across multiple waves. I found that all three types of emotion experiences could be characterized by low dimensional spaces, with the first two factors that accounted for most of the variance in people's ratings corresponding to the dimensions of valence and arousal, in line with prior work. However, I discovered additional novel factors related to generalizability (the extent to which an emotion experience is shared across many different situations and occurrences) or modularity (the extent to which an emotion experience is unique to specific situations). The findings show that emotion features not usually assessed in humans can be recovered from subjective ratings of their experiences. I argue for a revision of current dimensional theories of emotion: they have been incomplete because they were restricted to ratings entrenched in how we think of our conscious experience, and the typical English words we use to describe it. The new dimensions validate some theories of emotion and offer hope for linking psychological studies in humans with behavioral or neurobiological work across species. I also characterized the distributions of the

three types of emotion experiences and found that emotions were distributed along continuous gradients, with no well-separated clusters even for emotions belonging to the six basic emotion categories.

My thesis presents two additional topics that capitalize on my unique sample: the emotions experienced during the COVID pandemic, and individual differences. For example, I also found that resilience buffered individuals against the effect of loneliness on depression, and that people who had tested positive for COVID felt more morally disgusted towards acts of violating social norms. I also explored the association between psychological traits and differences in emotion experiences both in terms of the magnitudes of the ratings and the overall correlation structure across scales. Again, the richness of my dataset reveals a number of associations that are theoretically interesting and that will be of relevance to understanding mood and anxiety disorders as well.

All of the data will be made publicly available, and the core parts of many of the investigations were pre-registered.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Yanting Han and Ralph Adolphs. "Trait resilience protects against depression caused by loneliness during the COVID pandemic". In: *[Under review at Affective Science]* (2022). DOI: `10.31234/osf.io/9dac6`.

Y.H contributed to the conception of the study, data collection, data analysis, drafting, and revising the paper.

[2] Tessa Rusch, Yanting Han, Dehua Liang, Amber Hopkins, Caroline V Lawrence, Uri Maoz, Lynn K Paul, and Damian A Stanley. "COVID-Dynamic: A large-scale multifaceted longitudinal study of socioemotional and behavioral change across the pandemic". In: *[Under review at Scientific Data]* (2021). DOI: `10.31234/osf.io/75eyx`.

Y.H contributed to the conception of the study, data collection, data analysis, drafting, and revising the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

LIST OF TABLES

Chapter 1

THE PSYCHOLOGY OF EMOTION EXPERIENCE: DEBATES AND OPEN QUESTIONS.

1.1 Emotions, concepts, and conscious experience

This dissertation is a psychological investigation of people's emotion experience. It is focused entirely on adult humans, and on self-report. The exclusion of physiological, including neurobiological, measures in my thesis has a good justification: I think it is mostly premature to use these dependent measures to answer fundamental questions about emotion experience, because so much work and clarification is needed at the behavioral level first. For instance, there are specific debates about whether the amygdala (a structure in the vertebrate brain strongly associated with fear in the literature) is involved in emotion experience [1, 2]. There are broader claims that emotion experiences have no specific neural correlates at all [3, 4, 5], and no specific physiological correlates in the body either [6]. As I review below, there is not even consensus on what the word "emotion" means, on how many emotions there are, or on whether animals have them. Given this state of affairs, the general topic of emotion experience first needs to be probed in much more detail at the behavioral level—at least that is a motivation for the focus of my dissertation. Of course this does not exclude the value of carefully designed physiological studies, but, in my view, such studies (especially functional neuroimaging studies, of which there are very many) are mostly uninformative because the conceptual framework for interpretation of the findings (and designing the study) is still insufficiently well developed. Consequently, much of the literature on emotion, especially in humans and especially using neuroimaging, is essentially uninterpretable, since it is unclear what process is supposed to be measured.

At the outset, it is important to clarify the focus and scope of this dissertation. It is focused on people's judgments about the conscious experiences of emotions that they have. While this is the most common domain of inquiry in psychology, it is not a topic that is studied (or indeed can be studied) in nonhuman animals, and it is a topic whose indirect relation to actual emotion states requires cautious claims and inferences. Much of the confusion in the literature on emotion comes from ambiguous usage of terms, so that different studies often talk past one another because they mean different things when they use the word “emotion” [7, 8].

The distinction between emotions, conceived as states of an organism with an objective characterization, and the emotion experienced, the subjective experience of that emotion state (together with whatever else may contribute to the contents of consciousness at that time), is a critical one for this thesis. Throughout, I will simply use the term “emotion” for ease of exposition and when distinctions do not matter too much. However, I will disambiguate when it does matter. Still, it is important to keep in mind that, in general, the main variable of interest of this thesis, and the topic of most psychological theories of emotion, is in fact the conscious experience of emotion, rather than the emotion state itself (which I define functionally, following Adolphs and Anderson [9], see below).

In fact, the dependent measure is even more indirect than that. For the vast majority of psychological studies, and for this thesis, the main dependent measure is the conceptual access that people have to their experiences of emotions, and their ability to map this onto language. If they have a fleeting experience and forget about it, this is not usually measured. If they have an experience so bizarre they cannot describe it, that is also not usually measured. Once words are used, the dependent measure narrows to that aspect of emotion experience that cannot only be remembered and conceptualized, but about which people can think in words, and which they can map onto various scales typically provided by the experimenter. While perhaps obvious, it is worth emphasizing this point: not only do we not have direct access to the emotion state, we also do not have access to unconceptualized or unverbalized subjective experience. We are measuring what participants think about their conscious experiences of emotions, and we are measuring using words and concepts captured in particular rating scales.

This fact makes it critical to pay close attention to the validity and reliability of the words used in ratings. To that end, I constructed brief definitions for each rating scale, ensuring that everybody understood what the term meant, and everybody had the same understanding. I assessed consistency and reliability across these scales, so that unreliable ones could be excluded in analyses. Even after careful selection of scales and provision of definitions for them, some scales produced ratings that were considerably less reliable than others. It will be an important project for the future to determine if low reliability might result from ambiguity in the definition of the words, difficulty in accessing that aspect of one’s emotion experience, high variability in how stimuli elicit that dimension, or other explanations. Ambiguity in the words used, not only to describe specific emotions, but also to make distinctions between emotions, the conscious experiences that we have of them, and what we can think or say about these experiences (roughly: emotion states, feelings, concepts, and words) is ubiquitous across the literature on emotion. Many so-called theories of emotion are completely unclear about which of these domains they intend to address, as already noted.

There are a handful of historical and current emotion theories worth briefly reviewing [10]. However, it is already difficult to evaluate, let alone compare and contrast these, since it is in general unclear what exactly the authors mean by "emotion"—and, in particular, whether they are talking about actual emotion states, the conscious experience accompanying those states, or concepts about them [7]. Perhaps historically the first theory linked to actual biology was Charles Darwin's book, The Expression of the Emotions and Man and Animals [11]. Darwin was impressed by the apparent similarity in behaviors across species, behaviors that he took to be expressions of homologous emotions. We have reason now to believe that this latter inference was incorrect in many cases, but the basic claim that many animals exhibit behaviors caused by emotional states, and that these can be used for social communication, is now very widely accepted [9, 12, 13]. Darwin proposed three so-called principles in his book. The first, the principle of serviceable associated habits, in fact acknowledged that emotional behaviors could be co-opted for social communication. But Darwin insisted that there had to be an ancestral function that served some adaptive purpose. We still do not know why people have tears in their eyes when they are sad, but some co-opted behaviors, such as facial expressions for fear and disgust, have been argued to have plausible ancestral functions (maximizing sensory field-of-view or closing the sensory organs to stimuli, respectively, for these two emotions) [14].

Darwin's second principle was his principle of antithesis, proposing that there was an approach-withdrawal structure to emotions that recruited antagonistic sets of muscles. In essence, this was a precursor to modern dimensional theories of emotion, which all seem to agree on a dimension of approach-withdrawal (or its experiential counterpart, pleasant-unpleasant). In most modern theories of emotion experience this corresponds to a necessary component, a dimension of so-called core affect - valence. Darwin's third and last principle was his principle of the direct action of the nervous system. Although he had nothing to say about this in detail, since the neuroscience of emotion was unknown at the time, he presumed that the expression of emotion was caused by events in the brain. Elaborated and with a few further assumptions, this view is similar to modern views on the relation between brain, behavior, and psychology. A monist ontology proposes that emotions, and emotion experiences (psychological variables) are collections of brain events (they supervene on the brain, in philosophical vocabulary), and in turn cause the various dependent measures from which they can be inferred (behaviors, ratings) [15].

According to one prominent class of emotion theories, conscious experiences of the emotion are a distinct processing step, and correspond to distinct neural correlates in the brain. There are basically two versions of this view. The first is exemplified by William James' somatic feedback theory [16], and by neurological proposals such as the classic Papez circuit. These schemes propose that emotional reactions in the body in turn are perceived by the brain, and that this perception provides the content of the conscious experience. When we experience an emotion, we are thus experiencing something in our body. The intuition behind this view is clear enough: we indeed feel a lump in our throat when sad, or a tension in our stomach when anxious. It is less clear that this is all there is to emotion experience, or that it is a necessary precursor to emotion experience. But William James thought this was the case: "If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find we have nothing left behind, no "mind-stuff" out of which emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains" [16]. Taken literally, James' theory would predict that patients with no afferent input from the body cannot experience emotions. Although there are some arguments from patients with spinal cord lesions [17], the evidence is ultimately inconclusive, since it is impossible to disconnect the brain from all bodily signals (for instance, input from the face enter at the level of the trigeminal nucleus, above the spinal cord, and the vagus nerve provides afferent interoceptive information to brainstem nuclei).

Modern versions of these somatic feedback theories emphasize particular somatic components, such as feedback from the muscles of the face involved in facial expressions [18], or thinly myelinated afferents from the autonomic nervous system [19]. The currently two most prominent theories that propose that somatic feedback generates conscious experiences of emotion are those of Bud Craig [20, 21] and Antonio Damasio [22, 23], and both emphasize central brain structures involved in interoception and representation of the body, notably the insula. While there seems to be agreement in the above theories that representation of bodily states is an essential component of the conscious experience of emotion, the theories are less clear on how this comes about. Damasio, for instance, is explicit in stating that actual somatic feedback is not necessary— it just has to be set up in the right way (in evolution or development), but the actual on-line generation of emotion experience in an adult human could happen entirely in the brain (through his so-called “as-if loop”, which is something like a corollary discharge generating a sensory image of the bodily consequences).

These “internalized” versions of somatic feedback bear a closer resemblance to the second main version of how emotion and emotion experience might be distinct processes: cognitive theories of emotion. Classically, cognitive theories of emotion do not clearly distinguish emotions from emotion experiences [24]. However, a recent theory put forth by the neuroscientist Joe LeDoux is quite explicit in making the distinction [25]. LeDoux begins by famously claiming that we cannot, and should not, study emotions in animals [26]. His reasons are somewhat subtle: he does not deny that animals might have emotions, but he thinks it is not appropriate to study them, and that the studies that have claimed to study them have in fact been studying something else (“survival circuits” in his nomenclature). But the argument he gives hinges entirely on a redefinition of what the word “emotion” means. According to LeDoux “emotion” means “conscious experience of emotion”! Given this starting point, there is some sense to LeDoux’s reasoning: we do not currently have good or accepted methods for evaluating conscious experiences of emotion (or of anything else) in nonhuman animals. So it is methodologically problematic to study conscious experiences of emotions in animals. Some of LeDoux’s most argumentative detractors in fact agree with this statement [9], but they still think it’s perfectly possible to study emotion in animals, as long as we don’t conflate emotions with conscious experiences of emotions.

The logic of LeDoux’s argument aside, his positive contribution is a neurobiological theory of the conscious experience of emotion, which he takes to require working memory and depend on the prefrontal cortex [25]. LeDoux agrees that physiological and behavioral emotional reactions depend in part on subcortical structures like the amygdala, but he thinks we need to have some higher-order representation of these reactions in order to have a conscious experience of the emotion. The view is consonant with general higher-order theories of consciousness, but seems both ad hoc and to suggest a regress (if we need a higher-order representation to make us conscious of any representation, then why don’t we need a second- or third-order representation and so on?). Regardless of these debates, LeDoux, Damasio, and William James all share the idea that emotion states and conscious experiences are distinct both psychologically and neurally. Whatever their merits or demerits, they make a clear distinction and do not conflate emotions with conscious experiences of emotions (Damasio is perhaps the most explicit in this regard [22]).

Another large and complex class of emotion theories with a long history focuses on how emotions are induced in the first place, and how they unfold in time as a function of context and coping mechanisms. These are so-called appraisal theories, which are in a sense at the opposite end of the spectrum from Darwin's biological theory. Appraisal theories seem to discuss the conscious experience of emotion as their primary subject matter (it is not always easy to tell), and propose that how we evaluate, think about, and cope with complex situations is the essential analysis required to understand emotion experience [27]. However, many appraisal theories actually have strong links to ecological and evolutionary considerations that provide adaptive functions for emotions [28]. Nonetheless, appraisal theories generally require a substantial amount of "cognitive" processing before an emotion can be induced — there is a complex set of "stimulus evaluation checks" and other sequenced processes that incorporate context, memory, and thoughts about the self before a stimulus can induce an emotion. In fact, one of the most vigorous debates in the psychology of emotion revolved around exactly this claim: on one side appraisal theorists like Richard Lazarus claimed that cognition has to come before emotion [29]. On the other side of the debate (contra appraisal theories) Bob Zajonc argued that emotions are fast and directly linked to stimuli, and we can evaluate and think about them only subsequently: emotion has to come before cognition [30]. While historically of interest, these extreme views have now been replaced by the realization that both can be right, and that in general emotional and cognitive processing unfold in time in parallel, and continuously influence one another [31]. Both are extended in time, so it makes little sense to try to determine which "comes first".

What do the various theories say specifically about the experience of emotion? Although as already noted most emotion theories are unclear about whether they are talking about emotion states, emotion experiences, or concepts (and in fact some claim that these are the same [5, 32]), there are some explicit treatments of emotion experience. Perhaps the most universal agreement concerns what psychologists call "core affect," which is thought to be a necessary component of all experiences of emotions. According to most psychological theories, core affect consists of two dimensions, valence and arousal, around which additional components can then be added to account for the full richness and diversity of human emotion experience [33]. Lisa Feldman Barrett's theory of constructed emotion incorporates core affect as one essential feature, and then proposes that representations of many of the events and circumstances that accompany any particular episode of emotion experience are added into conscious experience to produce the rich experiences we usually have [32]. As we will see in the rest of this dissertation, a two-dimensional structure of emotion experience is also one of the most prominent findings in my data.

Needless to say, disagreements about emotion theories continue vigorously. In addition to the foundational aspects of emotion theory discussed above, there are disagreements in particular about whether emotion experiences are discrete or dimensional [34], and about how many categories or dimensions there should be. Paul Ekman's seminal studies on the recognition of facial expressions across cultures argued for about 6 so-called basic emotions: happiness, surprise, fear, anger, disgust and sadness (contempt was also sometimes added) [35, 36]. However, those data are based not on trying to measure people's experiences, but rather on their concepts and in general they are restricted to facial expressions [37]. The findings have also been debated, since other cross-cultural studies have produced different results [38], and since there are methodological concerns with Ekman's original studies [39]. However, newer studies propose anywhere from these basic 6 or 7 emotions to possibly more than 30 [40, 41, 42, 43].

My own view does not subscribe necessarily to any of the above theories, but only emphasizes that we need to be very careful in distinguishing emotion states from emotion experiences, emotion concepts, and words for emotions. As noted, my study aims to investigate emotion experiences, and it does so by measuring aspects of people's concepts for their experiences that can be captured by rating scales that have words. The dependent measure I use here cannot be used in nonhuman animals, even though it may well be the case that nonhuman animals also have emotion experiences—but different dependent measures would need to be used there in order to infer them. It is possible that emotion states and their conscious experience depend on distinct brain processes, as LeDoux has proposed. It is also possible that emotion states are themselves intrinsically conscious, or capable of generating conscious experiences. Again, I do not take a stance on this, and my thesis does not address these different possibilities. What my thesis does address is the structure of human emotion experience—how many dimensions or categories might best characterize it? What are those dimensions? How do these look if we compare emotions induced by reading stories, watching videos, or real-life experiences? How do these look across different individuals—males vs. females, extravert vs. introvert? These questions about the structure of emotion experience, and comparisons across stimulus types and people, are what motivated this dissertation (see also further below).

1.2 Ways of inducing emotions

In everyday life, emotions and their conscious experiences are induced by a plethora of stimuli. Some of those stimuli are relatively direct in how they induce emotions, and relatively universal in eliciting emotions across people. Suddenly being chased by a bear will induce a state related to fear, for example. But we also experience emotions that are induced very indirectly, from inferences, presumptions, expectations, and recollections caused by those stimuli. An email notice of a rejected paper induces a negatively valenced emotion in virtue of everything a person is caused to think about by reading the words in that email. Receiving a notice in the mail of an upcoming surgery appointment can induce anxiety whose conscious experience depends on imaging all kinds of events that may happen in the future. Indeed, in many cases it is not even clear what the stimulus was, no matter how indirect — we suddenly remember we are late for an appointment, or remember the death of a loved one, or feel intense panic during a nightmare.

Given this highly indirect and often cognitive fashion in which emotion experiences can be induced in humans, it is perhaps unsurprising that we also have some control over our emotions. The topic of emotion regulation (and its pathological counterpart, emotion dysregulation) have been intensively studied by both psychology and psychiatry [44]. Emotion regulation appears prominently during childhood development, is largely absent in nonhuman animals, and can be compromised by diseases, drugs, or distraction. It is likely one facet of the broader category of cognitive control that is effortful and that has evolved in animals with large brains and complex behaviors, most prominently humans, where it serves a complex social role [45]. Regulation can occur at several stages of processing, all the way from simply avoiding situations that might make one anxious in the first place, to trying to think about or "reappraise" a situation when confronted with it, to effortfully controlling one's own bodily reaction [46, 47].

While emotion regulation operates all the time even in the real world, its role is especially pernicious in laboratory experiments. In many studies, it is clear to subjects that they are supposed to experience certain emotions, and they may answer accordingly. For instance, showing people pictures of distinct facial expressions, or photos of scenes, can elicit ratings with high consensus [48], but it is unclear if subjects really experience emotions, or are simply rating what the intended emotion is supposed to be. Conversely, some classes of stimuli such as music [49] or videos [50] are able to elicit strongly felt emotions, but participants may not want to have those experiences and may thus downregulate their actually experienced emotions. In both cases, weak stimuli or very potent stimuli, what it is that subjects report on may be more related to what it is that they think one should feel, or what they would like to feel, than what they would actually feel if entirely at the mercy of the eliciting stimuli. To some extent, clear instructions, comparisons across different types of stimuli, and comparisons across participants can help address these concerns, but they remain difficult to address.

All of this stands in contrast to emotion elicitation in nonhuman animals. As for humans, emotions in the real world are elicited by complex multi sensory cues in a rich context. In the laboratory, the ways of inducing emotions are considerably more narrow. In animal models, particularly rodents, there are a number of commonly used stimuli, tasks, and environments that are used to assess emotions although the emotions so assessed are usually very specific. By far the largest number of measures have been developed with respect to fear and anxiety, capitalizing on innate behaviors elicited by uncertainty, unconditioned stimulus associations, or specific basic sensory cues, such as those normally provided by predators. Unlike the case for humans, animals do not know that they are in an experiment, and presumably have no concept of what an intended emotion or “correct” behavior should be on a given task. So the ecological validity of the emotional stimuli is improved in that respect in comparison to humans: emotion regulation and explicit knowledge of what other subjects would produce as data are eliminated as sources of confound. It still remains, of course, to interpret the stimuli and conditions as capturing something ecological in the real world.

In humans, there is a large literature that has developed and validated stimuli that the experimenters think should elicit a variety of different emotions [51]. These range from validated images [48] and videos [50, 52] to structured autobiographical recollection of past emotional events [53]. An advantage with humans is that these stimuli are indeed effective in conjuring up real-world equivalents, can certainly be effective in inducing strong emotional experiences (participants can be induced to weep, for instance), and are in fact directly "ecologically valid" in comparison to many situations in the real modern world, such as going to a movie theater or reading a novel or reminiscing about one's past memories. The very flexibility of how emotions can be induced in humans comes to our advantage in strengthening the argument that stimuli in a laboratory setting can be effective and meaningful.

As described further below and as detailed in the General Methods chapter, my study used three very different types of stimuli to induce emotions (in the same set of participants): validated brief stories [40], a large corpus of validated very brief video clips [41], and current real-world emotions sampled longitudinally across the COVID pandemic. The first two had extensive prior work by others, permitting comparisons with those studies (a subset of whose rating scales I also used). The third was a unique opportunity to sample emotions in everyday life during a particularly stressful time and at multiple timepoints. The entire subject sample was recruited from an ongoing study of over 1000 adult Americans, who were also assessed on a rich set of other psychological measures, some of which I use here to probe individual differences [54].

1.3 Measuring emotion experience

A functionalist approach to emotions proposes to characterize, and indeed define, emotions by what they do, rather than how they are implemented [55]. This is a common view in the philosophy of psychology, generally called psychofunctionalism, and it is congenial to neuroscience approaches as well. Roughly, it corresponds to one of the higher "levels of analysis" that David Marr had once proposed [56]. Marr's lowest level he called the "implementation level", which itself could be thought of as multiple levels of scale describing the physical realization of a computationally or psychologically defined process (e.g., actual brain circuits, neurotransmitters). Higher levels were labeled the "algorithmic" and "computational" level, but again these should be seen as a continuum of levels of analysis that abstract from the physical realization and describe the operations, functions, or goals of the process under consideration. Critically, the mapping between lower-to-higher levels is many-to-one, a property in philosophy long emphasized as "multiple realizability." This means that the same algorithmic-level description (e.g., an emotion with particular functional properties, such as fear of predators) could be realized in quite different hardware in the brain of a human, rodent, or octopus—or even a robot of the right kind.

As a functionally defined process, an emotion could thus be inferred from observing its causal relations with sensory stimuli, actions, and other psychological processes. The first would include cases such as presenting highly emotional videos intended to induce particular emotions (in humans), or innately triggering stimuli such as the odor [57] or shadow [58] of a predator (in rodents). The second, actions, would include species-specific behaviors that make sense in light of the stimuli, e.g., increased heart rate, faster breathing, freezing, or EEG arousal in response to a threat stimulus. A large number of specific tasks have been developed for rodents, in particular to assess fear and anxiety; self-report, psychophysiology, and facial EMG are commonly used for humans; and a range of observed behaviors have been catalogued for animals. There is a burgeoning field connected with animal welfare and livestock husbandry that infers emotional states from behaviors such as the ear posture of cows [59] to the whinnying of horses [60]. A recent cross-species study argued for a universal arousal signature in the vocalizations made by all animals [61], and it has been argued that some of the defensive behaviors seen in rodents indeed show strong similarity to such emotional behaviors also in humans, just as Darwin thought [62].

It is important to note that a functionalist view of emotions does not merely anchor them with respect to stimuli and behaviors, but also with respect to many other psychological processes. Indeed, how emotions influence other psychological variables (usually called "cognitive") is a burgeoning field in both animal and human research [63, 64, 65, 66]. The Nobel laureate cognitive scientist Herbert Simon had one of the earliest explicit schemes for emotion-cognition interactions from an engineering perspective. According to Simon, emotions functioned as "interrupt mechanisms" that served to take over ongoing volitional control of goal-directed behavior [67]. Simon proposed that an organism going about its business in everyday search of food and mates needed to have a separate and rather modular controller (or perhaps set of controllers) that could detect features such as potential threats in the environment, interrupt ongoing cognitive processing, and take over behavior to protect the organism. Emotion-cognition interactions are often thought to prominently involve the prefrontal cortex [68, 69], but the examples are ubiquitous. For instance, emotion and attention prominently interact: we can attend to emotionally arousing stimuli even when these are presented subliminally [70]. Emotionally arousing stimuli strongly influence hippocampal-dependent memory consolidation in both humans and animals [71]. Emotionally laden stimuli influence perceptual processing of those stimuli via feedback to visual cortices [72]. There are few if any psychological processes that are not influenced in one way or another by emotion. Characterizing those influences can by itself serve as a basis for inferring a functionally defined emotion state.

A more functional approach to characterizing emotions has been a historical thread in the psychology of emotion, but while purporting to be inspired by biology or evolution, it is typically entrenched in detailed analyses of experience, is invariably strongly focused on humans, and often linked to rather specific theories. The appraisal theories we mentioned earlier analyze the functional role of emotions and aim to move beyond the simple categories for which we happen to have words in English [28]. But there are important differences: although it is often unclear, appraisal theories still focus on experience, whereas my ratings draw from functional features that are in the first instance observed in behavior rather than experienced [9].

To cast a broad net, I included both standard emotion terms (e.g., the names for the so-called basic emotions, such as happiness and sadness), appraisal features (such as valence, self-relevance) as well as novel biologically inspired attributes (such as generalizability and persistence). One immediate question is of course whether people are able to rate such novel attributes. Perhaps they simply do not have access to them in their experience; or perhaps they apply only to observed behavior and not to experience. The question is particularly interesting, because the very same behaviors can be accompanied by quite different experiences, and conversely, the same experience may be accompanied by very different behaviors (or induced by different stimuli). This many-to-many mapping between stimuli, emotion experiences, and behaviors motivates a critical question: is there some coherence among the domains? Intuitively, we think there has to be, but this is far from clear empirically. I tried to provide clear definitions of my scales, but it remains the case that some rating scales produced more consistent ratings than others, and it remains a deeper problem that it is unclear on what exactly it is that subjects are reporting when they produced ratings, issues I take up in the final discussion of this thesis.

The particular words and concepts that experimenters build into questionnaires and rating scales obviously constrain what it is that one can measure about emotion experiences. This can produce biases when participants are forced to choose category labels without sufficient alternatives, producing inflated consensus or validation for stimuli that are in fact better described by alternative emotion concepts that were not offered to the participants [39]. To this end, I also collected free descriptions of the best labels for emotions evoked by stories and videos. For real-life emotions, again, subjects were asked to provide free descriptions of labels and the causes of their emotions in addition to the ratings on the scales.

1.4 Motivation for the present study and unique aspects

Three main questions motivated the present dissertation. First, the recent literature makes some strong claims that there are many different dimensions or categories of emotions. According to some studies, there are 18-36 dimensions of emotion [40, 43], or 27 categories that can blend into one another [41]. These claims are related to some emotion theories, like the conceptual act theory of Lisa Feldman Barrett [73], and also to appraisal theories [28], which all emphasize the richness of the conscious experience of emotions. No doubt, we have a large number of words available to describe how we feel, and we can spend a lot of narrative in describing what we feel. But that does not necessarily mean that emotion experiences are best characterized in a high-dimensional space. We also have a large number of words and concepts to describe colors, for instance, but we know that our color experience can be compactly and completely characterized by just three dimensions (hue, saturation, and brightness). As an aside, the analogy between colors and emotion has been taken explicitly in a proposal about the structure of emotions in Drosophila; just like primary colors mix to produce a full spectrum, so do core affects, according to this proposal [74]. So the first question I had was: how many dimensions do we need, and how do we best interpret those dimensions? Relatedly: are there clusters of emotion experiences in this dimensional space that would suggest categories, and if so, which categories of emotions might there be?

A second question concerned the emotions induced by different types of stimuli. Most studies focus on a single type of stimulus, such as viewing pictures or watching videos. This makes it difficult to compare across studies, and makes it difficult to explore the vexing question of whether subjects are reporting just on what they think the intended or "correct" emotion is, versus reporting on the actual contents of their conscious experience of the emotion. To this end I included a type of stimulus where one might reasonably think a clear intended emotion predominates (short verbal scenarios describing something happening to someone), as well as those types of stimuli where one might reasonably think subjects are reporting on their actual emotion experience (potent videos and real-world experience reports). A unique strength of my study was that the very same participants rated their emotions across these three kinds of stimuli, and on the same rating scales. This permitted a comparison across these three stimulus types that, to my knowledge, has not yet been undertaken in the literature.

A third question concerned individual differences. There are a wealth of studies of how people differ in the emotions they experience, ranging from "affective style" in healthy individuals [75, 76] to psychopathology [77, 78, 79]. Capitalizing on the "COVID-Dynamic" study [54], I sampled the same set of participants as those in that study. This provided extensive careful selection of subjects, strong exclusionary criteria and quality control metrics, and very rich assessment of individual differences. I do not examine all of those here, but focus on a small subset of variables of a priori interest. However, the extent and quality of psychological background assessment in this sample of participants is exemplary and provides the basis for Chapter 7 on individual differences.

The first question, how many dimensions or categories of emotion experience there are, of course depends on how you measure it. While there is a long and in many ways unproductive debate in the psychology of emotion on whether emotions are categorical or dimensional, modern views generally acknowledge that both categorical (discrete) and dimensional (continuous) accounts are useful [13, 34]. Both are characterizations that the scientist gives, often on the basis of data that themselves cannot adjudicate between discrete and dimensional accounts: the criterion should then be simply which characterization is most useful for the purposes of the study.

If I give people a single scale on which to rate their experience, it would necessarily be one-dimensional, and everybody would rightly complain that I simply haven't sampled emotion space properly. On the other hand, taking a complete inventory of all different words on which emotions could be rated is infeasible. It is also clearly unnecessary since most words are redundant with one another, and some words are too vague to produce clear ratings. I selected words with relatively clear and non-redundant meanings from three main sources. One is the emotion words classically used, i.e., the words for basic emotions. A second is a selection from modern studies that used specific words. A third are the "features" of emotion proposed by Ralph Adolphs and David Anderson in their book [9].

Another choice that needs to be made is how to induce emotions, and how to induce a sufficient diversity of emotion experiences. Here I felt that directly taking the stimuli from prior studies that claimed to find a large number of dimensions would be advantageous. I thus selected short narratives from [40] and videos from [41].

This might still under sample actual emotions experienced in the real world, so I added a third type of induction: emotions experienced in everyday life during the COVID pandemic. As noted, a unique feature of my dissertation is the ability to compare across these three kinds of stimuli, and to do so in the very same subjects.

Taken together then, the selection both of ratings scales and of stimuli (and of a broad set of participants) all aimed to produce a relatively representative and ideally complete inventory of emotion experiences in my study. Nonetheless, it should be emphasized that there are likely to be rare emotions that were omitted here—particularly intense ones like the devastating loss of a loved one, emotions dependent on a specific culture such as certain emotion concepts unavailable to Western participants, or rare emotions in complex settings such as the feeling of religious awe.

Subject to these caveats, the question then becomes how to characterize emotions. The approach I took focused on the similarity structure of rated emotion experience. The raw data for this are the pairwise correlations between emotion scales across the different types of stimuli. For a given instance of an emotion experience induced by watching a video, for instance, that video would produce a vector of ratings on the scales that I used. Another video would produce a different vector, corresponding to a different emotion experience. Across all the stimuli, we can then ask what is the similarity structure in the ratings, producing a triangle-symmetric correlation matrix. We can visualize the stimuli in a space with dimensions corresponding to the number of rating scales. Some ratings scales would be expected to be highly correlated with one another (e.g., anxiety and fear), whereas others would be expected to be anti correlated (e.g., happiness and sadness). Some emotion stimuli will be located close to one another in the high-dimensional rating scale space (e.g., two almost identical videos) whereas others will be far apart (e.g., two videos evoking very different emotions).

While the similarity relations between scales and between stimuli live in a high-dimensional space, there are numerous techniques available for quantifying the variance in the data in a much lower dimensional space. The main methods that I used were exploratory factor analysis to derive dimensions, and UMAP and clustering techniques to visualize and identify clusters of emotions. In both cases these are tools that need to be applied with a number of criteria in mind; I used both the common criterion of how much variance can be explained, and the additional criterion of interpretability of the results.

1.5 Brief overview of the upcoming Chapters and their relationships

The rest of this dissertation first describes general methods shared across all of the studies, and then separately analyzes the data from each of the three classes of stimuli that I used to induce emotion experiences: stories, videos, and real-life experiences. In each case, the three corresponding chapters have a parallel organization. Then, I devote a chapter to making comparisons across the emotions induced by these three types of stimuli, and in another I explore individual differences. Chapter 8 is a more focused investigation of individual differences in resilience that has been submitted as a separate publication already. I end with a general discussion.

References

[1] Adam K Anderson and Elizabeth A Phelps. "Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions". In: *Journal of cognitive neuroscience* 14.5 (2002), pp. 709–720.

[2] Justin S Feinstein et al. "The human amygdala and the induction and experience of fear". In: *Current biology* 21.1 (2011), pp. 34–38.

[3] Kristen A Lindquist et al. "The brain basis of emotion: a meta-analytic review". In: *The Behavioral and brain sciences* 35.3 (2012), p. 121.

[4] Elizabeth Clark-Polner, Timothy D Johnson, and Lisa Feldman Barrett. "Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions". In: *Cerebral Cortex* 27.3 (2017), pp. 1944–1948.

[5] Lisa Feldman Barrett. "The theory of constructed emotion: an active inference account of interoception and categorization". In: *Social cognitive and affective neuroscience* 12.1 (2017), pp. 1–23.

[6] Erika H Siegel et al. "Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories." In: *Psychological bulletin* 144.4 (2018), p. 343.

[7] Ralph Adolphs. "How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences". In: *Social cognitive and affective neuroscience* 12.1 (2017), pp. 24–31.

[8] Joseph E LeDoux. "Semantics, surplus meaning, and the science of fear". In: *Trends in cognitive sciences* 21.5 (2017), pp. 303–306.

[9] Ralph Adolphs and David J Anderson. *The neuroscience of emotion: A new synthesis*. Princeton University Press, 2018.

[10] Andrew S Fox et al. *The nature of emotion: Fundamental questions*. Oxford University Press, 2018.

[11] Charles Darwin. “The expression of the emotions in man and animals”. In: *The expression of the emotions in man and animals*. University of Chicago press, 2015.

[12] Marc D Hauser and Mark Konishi. *The design of animal communication*. MIT press, 1999.

[13] Michael Mendl, Oliver HP Burman, and Elizabeth S Paul. “An integrative and functional framework for the study of animal emotion and mood”. In: *Proceedings of the Royal Society B: Biological Sciences* 277.1696 (2010), pp. 2895–2904.

[14] Joshua M Susskind et al. “Expressing fear enhances sensory acquisition”. In: *Nature neuroscience* 11.7 (2008), pp. 843–850.

[15] Julien Dubois et al. “Personality beyond taxonomy”. In: *Nature human behaviour* 4.11 (2020), pp. 1110–1117.

[16] William James. *What is an emotion? Mind, os-IX, 188-205*. 1884.

[17] George W Hohmann. “Some effects of spinal cord lesions on experienced emotional feelings”. In: *Psychophysiology* 3.2 (1966), pp. 143–156.

[18] Pamela K Adelmann and Robert B Zajonc. “Facial efference and the experience of emotion”. In: *Annual review of psychology* 40.1 (1989), pp. 249–280.

[19] AD (bud) Craig. “Interoception and emotion: a neuroanatomical perspective”. In: *Handbook of emotions* 3.602 (2008), pp. 272–88.

[20] Arthur D Craig. “How do you feel? Interoception: the sense of the physiological condition of the body”. In: *Nature reviews neuroscience* 3.8 (2002), pp. 655–666.

[21] Arthur D Craig. *How Do You Feel?: An Interoceptive Moment with Your Neurobiological Self*. Princeton University Press, 2015.

[22] Antonio R Damasio. “REVIEW: Toward a Neurobiology of Emotion and Feeling: Operational Concepts and Hypotheses”. In: *The Neuroscientist* 1.1 (1995), pp. 19–25.

[23] Antonio R Damasio. *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Houghton Mifflin Harcourt, 2003.

[24] Philip N Johnson-Laird and Keith Oatley. “Basic emotions, rationality, and folk theory”. In: *Cognition & Emotion* 6.3-4 (1992), pp. 201–223.

[25] Joseph E LeDoux and Richard Brown. “A higher-order theory of emotional consciousness”. In: *Proceedings of the National Academy of Sciences* 114.10 (2017), E2016–E2025.

[26] Joseph LeDoux. “Rethinking the emotional brain”. In: *Neuron* 73.4 (2012), pp. 653–676.

[27] Klaus R Scherer. "The dynamic architecture of emotion: Evidence for the component process model". In: *Cognition and emotion* 23.7 (2009), pp. 1307–1351.

[28] Richard S Lazarus. *Emotion and adaptation*. Oxford University Press, 1991.

[29] Richard S Lazarus. "On the primacy of cognition." In: *American Psychologist* (1984).

[30] Robert B Zajonc. "On the primacy of affect." In: *American Psychologist* (1984).

[31] Klaus R Scherer. "Emotions are emergent processes: they require a dynamic computational architecture". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3459–3474.

[32] Lisa Feldman Barrett. "The conceptual act theory: A précis". In: *Emotion review* 6.4 (2014), pp. 292–297.

[33] Lisa Feldman Barrett et al. "The experience of emotion". In: *Annu. Rev. Psychol.* 58 (2007), pp. 373–403.

[34] Eddie Harmon-Jones, Cindy Harmon-Jones, and Elizabeth Summerell. "On the importance of both dimensional and discrete models of emotion". In: *Behavioral sciences* 7.4 (2017), p. 66.

[35] Paul Ekman. "An argument for basic emotions". In: *Cognition & emotion* 6.3-4 (1992), pp. 169–200.

[36] Paul Ekman. "Basic emotions". In: *Handbook of cognition and emotion* 98.45-60 (1999), p. 16.

[37] Paul Ekman. "Facial expression and emotion." In: *American psychologist* 48.4 (1993), p. 384.

[38] Maria Gendron et al. "Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture." In: *Emotion* 14.2 (2014), p. 251.

[39] Lisa Feldman Barrett et al. "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements". In: *Psychological science in the public interest* 20.1 (2019), pp. 1–68.

[40] Amy E Skerry and Rebecca Saxe. "Neural representations of emotion are organized around abstract event features". In: *Current biology* 25.15 (2015), pp. 1945–1954.

[41] Alan S Cowen and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients". In: *Proceedings of the National Academy of Sciences* 114.38 (2017), E7900–E7909.

[42] Alan S Cowen et al. "Sixteen facial expressions occur in similar contexts worldwide". In: *Nature* 589.7841 (2021), pp. 251–257.

[43] Naoko Koide-Majima, Tomoya Nakai, and Shinji Nishimoto. “Distinct dimensions of emotion in the human brain and their representation on the cortical surface”. In: *NeuroImage* 222 (2020), p. 117258.

[44] James J Gross. “The emerging field of emotion regulation: An integrative review”. In: *Review of general psychology* 2.3 (1998), pp. 271–299.

[45] James J Gross. “Emotion regulation: Affective, cognitive, and social consequences”. In: *Psychophysiology* 39.3 (2002), pp. 281–291.

[46] Kevin N Ochsner and James J Gross. “The cognitive control of emotion”. In: *Trends in cognitive sciences* 9.5 (2005), pp. 242–249.

[47] James J Gross. “Emotion regulation: Current status and future prospects”. In: *Psychological inquiry* 26.1 (2015), pp. 1–26.

[48] Peter J Lang, MM Bradley, and B Cuthbert. “The international affective picture system”. In: *Center for Research in Psychophysiology, University of florida, Gainesville, USA* (1988).

[49] Anne J Blood and Robert J Zatorre. “Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion”. In: *Proceedings of the national academy of sciences* 98.20 (2001), pp. 11818–11823.

[50] James J Gross and Robert W Levenson. “Emotion elicitation using films”. In: *Cognition & emotion* 9.1 (1995), pp. 87–108.

[51] James A Coan, John JB Allen, et al. *Handbook of emotion elicitation and assessment*. Oxford university press, 2007.

[52] Richard J Davidson. “Parsing affective space: Perspectives from neuropsychology and psychophysiology.” In: *Neuropsychology* 7.4 (1993), p. 464.

[53] Antonio R Damasio et al. “Subcortical and cortical brain activity during the feeling of self-generated emotions”. In: *Nature neuroscience* 3.10 (2000), pp. 1049–1056.

[54] Tessa Rusch et al. “COVID-Dynamic: A large-scale multifaceted longitudinal study of socioemotional and behavioral change across the pandemic”. In: (2021).

[55] Ralph Adolphs and Daniel Andler. “Investigating emotions as functional states distinct from feelings”. In: *Emotion Review* 10.3 (2018), pp. 191–201.

[56] David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. MIT press, 2010.

[57] Markus Fendt, Thomas Endres, and Raimund Apfelbach. “Temporary inactivation of the bed nucleus of the stria terminalis but not of the amygdala blocks freezing induced by trimethylthiazoline, a component of fox feces”. In: *Journal of Neuroscience* 23.1 (2003), pp. 23–28.

[58] Melis Yilmaz and Markus Meister. “Rapid innate defensive responses of mice to looming visual stimuli”. In: *Current Biology* 23.20 (2013), pp. 2011–2015.

[59] Helen S Proctor and Gemma Carder. “Can ear postures reliably measure the positive emotional state of cows?” In: *Applied Animal Behaviour Science* 161 (2014), pp. 20–27.

[60] Anne-Laure Maigrot et al. “Vocal expression of emotional valence in Przewalski’s horses (Equus przewalskii)”. In: *Scientific reports* 7.1 (2017), pp. 1–11.

[61] Piera Filippi et al. “Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals”. In: *Proceedings of the Royal Society B: Biological Sciences* 284.1859 (2017), p. 20170990.

[62] D Caroline Blanchard et al. “Human defensive behaviors to threat scenarios show parallels to fear-and anxiety-related defense patterns of non-human mammals”. In: *Neuroscience & Biobehavioral Reviews* 25.7-8 (2001), pp. 761–770.

[63] Mick Power and Tim Dalgleish. *Cognition and emotion: From order to disorder*. Psychology press, 2015.

[64] Raymond J Dolan. “Emotion, cognition, and behavior”. In: *science* 298.5596 (2002), pp. 1191–1194.

[65] Elizabeth A Phelps. “Emotion and cognition: insights from studies of the human amygdala”. In: *Annu. Rev. Psychol.* 57 (2006), pp. 27–53.

[66] Luiz Pessoa. *The cognitive-emotional brain: From interactions to integration*. MIT press, 2013.

[67] Herbert A Simon. “Motivational and emotional controls of cognition.” In: *Psychological review* 74.1 (1967), p. 29.

[68] Helen Barbas. “Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices”. In: *Brain research bulletin* 52.5 (2000), pp. 319–330.

[69] C Daniel Salzman and Stefano Fusi. “Emotion, cognition, and mental state representation in amygdala and prefrontal cortex”. In: *Annual review of neuroscience* 33 (2010), pp. 173–202.

[70] Yi Jiang et al. “A gender-and sexual orientation-dependent spatial attentional effect of invisible images”. In: *Proceedings of the National Academy of Sciences* 103.45 (2006), pp. 17048–17052.

[71] James L McGaugh. “The amygdala modulates the consolidation of memories of emotionally arousing experiences”. In: *Annu. Rev. Neurosci.* 27 (2004), pp. 1–28.

[72] Patrik Vuilleumier et al. “Distant influences of amygdala lesion on visual cortical activation during emotional face processing”. In: *Nature neuroscience* 7.11 (2004), pp. 1271–1278.

[73] Lisa Feldman Barrett. *How emotions are made: The secret life of the brain*. Pan Macmillan, 2017.

[74] Simeng Gu et al. “A model for basic emotions using observations of behavior in Drosophila”. In: *Frontiers in psychology* (2019), p. 781.

[75] Richard J Davidson. “Anterior cerebral asymmetry and the nature of emotion”. In: *Brain and cognition* 20.1 (1992), pp. 125–151.

[76] Richard J Davidson. “Anxiety and affective style: role of prefrontal cortex and amygdala”. In: *Biological psychiatry* 51.1 (2002), pp. 68–80.

[77] Richard J Davidson et al. “Depression: perspectives from affective neuroscience”. In: *Annual review of psychology* 53.1 (2002), pp. 545–574.

[78] Cornelius Gross and Rene Hen. “The developmental origins of anxiety”. In: *Nature Reviews Neuroscience* 5.7 (2004), pp. 545–552.

[79] Lisa M Shin and Israel Liberzon. “The neurocircuitry of fear, stress, and anxiety disorders”. In: *Neuropsychopharmacology* 35.1 (2010), pp. 169–191.

Chapter 2

GENERAL METHODS

2.1 Materials and Procedures

2.1.1 Scale selection

As mentioned earlier, my goal was to have a comprehensive set of scales that describe the properties of emotion experiences. I first assembled an inclusive list from multiple sources: the words for basic emotions; affective scales from modern studies [1, 2, 3] and biologically inspired features proposed by Ralph Adolphs and David Anderson [4]. The initial steps of processing included deleting redundant scales, checking if the scales were indeed describing emotions instead of just contexts and coming up with clear definitions for each scale. All of those steps were finalized through discussion and piloting among members of the Adolphs lab.

To further evaluate the quality of those scales, I recruited 30 subjects from Prolific to obtain ratings of clarity of definition (from very unclear (1) to very clear (5)) and how well each scale applies to human emotional experiences (doesn't apply at all (1) to definitely applies (5)). The results indicated that the final list of 28 scales were considered to be clearly defined and appropriate to describe human emotional experiences (https://osf.io/mqc6b/ and https://osf.io/zyck8/). The definitions of the scales are listed below (Table 2.1).

To assess the readability of my scales (Table 2.1), I calculated the Flesch–Kincaid grade levels [5] using the definitions of the scales. The test is a popular tool based on word length and sentence length and indicates how hard it is to understand my scales. Most of my scales required a grade level of 12 and lower, which corresponds to a high school education.

Table 2.1: Definition of 28 rating scales used in my studies, along with the description of the two ends of the scales and grade level required to understand the definitions.

Labels	Definitions	Lower_end	Higher_end	Grade level
mental_bodily	this scale describes the extent to which an emotion is experienced in the mind or in the body	experienced mostly in the mind	experienced mostly in the body	7.82

controllability	this scale describes how much control you have over an emotion	cannot control this emotion at all	this emotion is easy to control	5.86
valence	this scale describes how pleasant or unpleasant an emotion is	very unpleasant	very pleasant	7.19
approach	this scale describes how much an emotion makes you want to approach or avoid	strongly want to avoid	strongly want to approach	5.88
arousal	this scale describes how physically aroused/stimulated an emotion makes you feel	not aroused at all	highly aroused	10.72
safety_unsafety	this scale describes the degree to which an emotion evokes a sense of safety or unsafety	evokes a strong sense of unsafety	evokes a strong sense of safety	9.82
relief	this scale describes how the feeling turns out to be in the end compared to how you felt in the beginning	felt much worse than it was at the beginning	felt much better than it was at the beginning	7.21
happy	this scale describes how happy this emotion makes you feel	not happy at all	very happy	4.83
sad	this scale describes how sad this emotion makes you feel	not sad at all	very sad	3.65
afraid	this scale describes how afraid (which is immediate and directed towards the present stimulus) this emotion makes you feel	not afraid at all	very afraid	11.07
worried	this scale describes how worried (which is more diffused and longer lasting towards a future threat or risk) this emotion makes you feel	not worried at all	very worried	9.8
surprised	this scale describes how surprised this emotion makes you feel	not surprised at all	very surprised	4.83
angry	this scale describes how angry this emotion makes you feel	not angry at all	very angry	4.83
physical_disgust	this scale describes how physically disgusted (towards things like vomit and spoiled food) this emotion makes you feel	not disgusted at all	very disgusted	9.79
moral_disgust	this scale describes how morally/socially disgusted (towards things like acts of violating social norms) this emotion makes you feel	not disgusted at all	very disgusted	12.27
intensity	this scale describes how strong or weak an emotion is	very faint emotion	very strong, intense emotion	3.65

scalability	this scale describes how much an emotion can scale in intensity. If it does not scale, then it always feels equally strong or weak. Otherwise, it can be either strong or weak.	doesn't scale	can scale	4.8
persistence	this scale describes how long an emotion lasts	emotion is very brief and fleeting	emotion sticks for a long time	3.76
gen_stimuli	this scale describes how many different stimuli can evoke a certain emotion	very rare and specific	very common and often found	10.72
gen_behavior	this scale describes how many different behaviors an emotion can cause	very rare and specific	very common and often found	10.15
consciously_aware	this scale describes how consciously aware you are of an emotion	not aware at all	very aware	8.01
interference	this scale describes the degree to which an emotion disrupts other ongoing activities	not disruptive at all	very disruptive	10.36
common	this scale describes how often you've felt like this	very rare, not often experienced	very common, I experience this emotion on a regular basis	2.34
fairness_unfairness	this scale describes the degree to which an emotion evokes a sense of unfairness or fairness	evokes a strong sense of unfairness	evokes a strong sense of fairness	8.35
intrinsic_extrinsic	this scale describes whether an emotion is primarily a reflection of you (e.g. your personality, your abilities, your past experiences) or a reflection of the surrounding situation (other people, external forces)?	completely intrinsic	completely extrinsic	20.48
future	this scale describes the degree to which an emotion involves anticipation of an event that would or might occur in the future	not related to anticipation of future events at all	totally related to anticipation of future events	11.23
remembering	this scale describes the degree to which an emotion involves remembering events occurred in the past	not related to past events at all	totally related to past events	9.82
self_relevance	this scale describes the level of relevance an emotion has to your life	low relevance	high relevance	6.73

2.1.2 Stimuli selection

I used two types of emotionally evocative stimuli: short stories adapted from [2] and short video clips from [1], in addition to sampling naturally occurring emotions in participants' daily lives.

The original set of stories consisted of 200 stories targeting 20 emotion categories. Each story can be represented using a vector of 46 dimensions (valence/arousal, 6 basic emotions and 38 appraisal dimensions) using data from the original paper. I performed principal component analysis (PCA) on randomly sampled subsets of the whole set of stories (analysis was repeated 100 times for each fixed number of stories) and found that generally, more principal components (PC) were needed to account for 80% of the total variance as the number of stories increased, but the number of PCs required reached a plateau at 150 stories. I therefore determined that 150 stories were enough to represent the majority of the variance for the whole set. The final set of 150 stories were selected using a maximum variation sampling procedure. The procedure sampled the stories by maximizing the sum of Euclidean distances between the story vectors. Specifically, the first story was randomly selected and then the other stories were selected so that each new story had the furthest Euclidean distances from the previously selected stories in the 46 dimensional story vector space. The sampling procedure was repeated until the desired sample size was reached. I repeated the whole process for all possible initializations and selected the specific sample with the maximum sum of Euclidean distances.

Similarly for the video clips, I made use of the ratings collected in the original study and represented each video clip with a vector of 48 dimensions (14 affective and 34 emotion category dimensions). I then carried out a similar PCA procedure as described above to determine that 1000 video clips contained enough variation. The final set of videos were selected in the following ways. First, given the IRB requirements, I deleted the set of extreme videos (the blurred ones in the online map of the original study according to the list provided by the author) and then 1000 videos were sampled according to the maximum variation procedure as described above. Still, I found two videos explicitly sexual and decided to delete them, which left me with a final set of 998 videos.

The list of the final set of stories and videos can be found at `https://osf.io/7594c/`, and interested researchers may contact the authors of the original studies to get the actual stimuli. For session duration considerations, I randomly split the 150 stories into 2 sets, each with 75 stories. Similarly, 998 videos were split into 10 sets, 9 of them with 100 videos and 1 with 98 videos.

2.1.3 Participants

Recruitment

I utilized the same participant pool to study emotion evoked by stories and videos, and real-life emotions (as part of the Covid-Dynamic study). The Covid-Dynamic study was pre-registered before data collection began (`https://osf.io/sb6qx`), and details about the dataset can be found in the data release paper [6]. Here, I provide brief descriptions of the participant recruitment and a subset of psychological measures used in my studies.

The recruitment was done through Prolific (www.prolific.co) and participants were required to be adults 18 or older, fluent in English, and reside in the United States. In addition, they had to have a Prolific approval rating of 98% or higher, and a minimum of 5 Prolific studies completed. In total, 1797 subjects completed Wave 1 of the COVID-Dynamic study.

There was a wide range of psychological assessments administered multiple times throughout the study. I introduced the following ones, specifically related to emotion that I probed to study individual differences. The Connor-Davidson Resilience Scale - 10 Item (CD-RISC) [7] is a self-report questionnaire of coping responses in the past month that is the most common measure of psychological resilience. The NEO Five-Factor Personality Inventory (NEO) [8] is a 60-item self-report questionnaire that assesses an individual on five dimensions of personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism. NIH toolbox: Loneliness scale (NIH-Loneliness) [9] is a 5-item self-report questionnaire of how often an individual felt lonely or alone in the past month. Beck Depression Inventory – II (BDI) [10] is a 21-item self-report questionnaire that examines depressive symptomatology over the past two weeks. State Trait Anxiety Inventory (STAI) [11] is a 20-item self-report questionnaire on the temporary condition of "state anxiety". Perceived Stress Scale (PSS) [12] is a 10-item self-report questionnaire that measures the extent to which a participant perceives personal life events in the past month as stressful.

Determination of sample size for the story and video rating experiments

I determined the sample size to be 15 participants per scale based on a recent study about the point of stability for impression formation from faces [13]. The study introduced a sampling procedure to determine when the average of observations would be stable.I based my sample size estimation on the number of observations required to obtain a stable average because my present research would aggregate ratings across participants for each scale and then take the average rating. Given that the rating scale was ranged from 1 to 7, the corridor of stability (COS) deemed acceptable to me was +/- 0.5 and the level of confidence deemed acceptable to me was 80%. Using the pilot data that I collected which involved three scales of different semantic complexity, I found that 15 participants would be enough for even the most complicated scale to satisfy the above criteria. Also, this sample size was comparable to previous studies which normally had a sample size of about 10 ratings per stimulus. Expecting attrition from data quality exclusions, I decided to collect 20 ratings per scale per stimulus. For each individual study, since there were 28 scales, each recruited 560 participants.

2.1.4 Procedures

As mentioned above, the same participant pool was utilized to study emotion evoked by stories, videos, and in real life (as part of the Covid-Dynamic study). Each wave of the Covid-Dynamic study (16 waves in total, from Apr 2020 to Jan 2021) and each set of the evoked emotion experiments (12 sets in total: 2 sets of stories and 10 sets of videos, from Nov 2020 to March 2021) were posted as separate studies on Prolific.

Covid Dynamic study

As mentioned before, I sampled naturally occurring emotions in participants' daily lives as part of the Covid-Dynamic study. The Covid-Dynamic study is a longitudinal study with at least 18 waves of completed data collection (first 16 waves shown in Fig.2.1). Each wave, participants were asked to finish an approximately hour-long survey that included assessment on multiple psychological domains using standard and custom questionnaires and behavior tasks (see [6] for the full battery of questionnaires and tasks with frequency of administration).

For my experimental measure, I asked participants to rate on 22 scales (a subset from the 28 scales as some were not applicable, collected for every wave since wave 2) regarding their current emotional states. In addition, participants were asked to provide labels (minimum of 1 label and up to five labels, collected for every wave since wave 3) and causes for their emotions using free descriptions (collected for every wave since wave 4) .

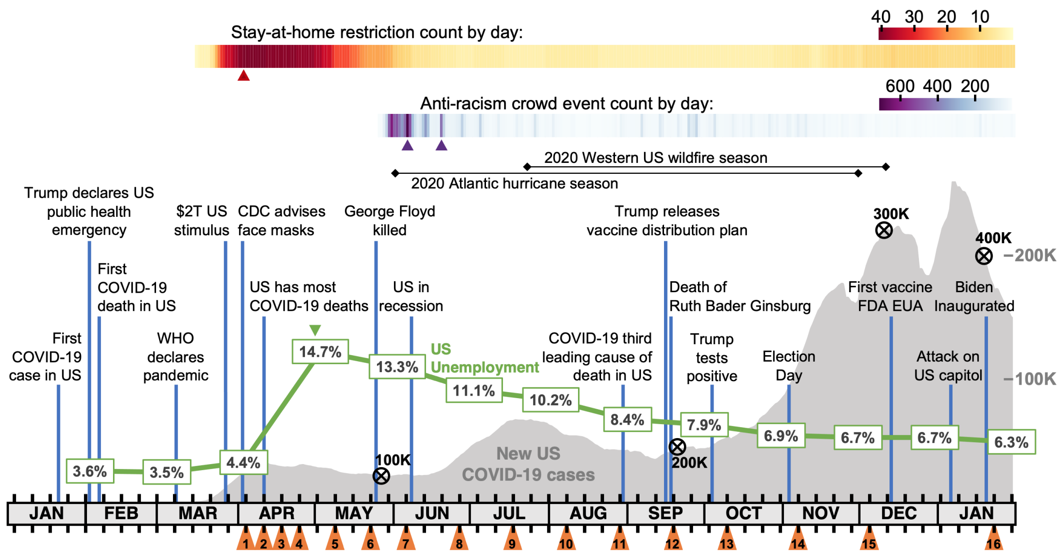

Figure 2.1: Timeline of Real-World Events and COVID-Dynamic Wave Administrations. Visualization of the COVID-Dynamic data collection schedule in the context of the events of January 2020 to January 2021. Orange triangles denote each wave administration (black tick marks depict weekly intervals). The gray curve indicates the daily 7-day average of new, confirmed COVID-19 cases in the U.S., black encircled X's on top of the curve mark grim U.S. COVID-19-related death milestones (100,000 to 400,000 dead). The green line shows the monthly unemployment rate. The upper gradient (yellow-red) indicates the daily count of states with active stay-at-home restrictions (peak=41). The lower gradient (blue-purple) shows the daily count of U.S. anti-racism crowd events. Colored triangles below the gradients indicate local maxima for the various measures. All these external data (aligned to our data collection) are included in the dataset. Events of interest are indicated with vertical blue lines. (credit: Covid Dynamic Study, [6]

Evoked emotion experiments

Participants first signed up for my studies on Prolific and were then directed to Qualtrics for informed consent and brief surveys. Note that as described above (in 2.1.3), for each of the evoked emotion experiments, the number of participants needed was 560. All participants from the Covid-Dynamic study were eligible, but they were accommodated on a first come first serve basis. That is, the evoked emotions studies were stopped as soon as the number of responses reached 560.

The Toronto Alexithymia Scale (TAS) [14] and emotion regulation questionnaire (ERQ) [15] questionnaires were administered at the very first session and the Positive and Negative Affect Schedule (PANAS) [16] was administered at the first session of the day for each participant. Then, participants were randomly assigned one of 28 affective scales and were directed to Pavlovia where the rating experiment was hosted.

Story rating experiment

The story rating experiment consisted of evaluating emotions evoked by 75 stories on the assigned scale (Fig.2.2).

In the practice trial, participants were shown an example story and were asked to move the slider to the middle of the scale (which would be 4 for a scale from 1 to 7) as close as possible. The scale would appear below the story once the participants clicked the 'finished' button. This was designed to get accurate response time that's not contaminated by the time spent on reading the stories. There would be a warning message for clicking the 'finished' button too soon within five seconds of the start of the trial. If no rating was made after 50 seconds, there would be a message to urge for a response or to contact the researcher for assistance. The experiment would end if the participant failed to proceed after 5 minutes.

In the main trials, stories were presented in random order. Similarly as in the practice trial, participants would read the story, click the 'finished' button, and then move the slider to rate. Messages would be shown if clicking the 'finished' button too soon (within 5 seconds), or if responding too slow (after 50 seconds). The trial would be timed out if no response was made within 60 seconds. And the experiment would end if approximately 10% (8) trials were skipped.

In the retest trials where a random set of 8 stimuli previously shown in the main trials would be presented again, participants were asked to rate them again and to also provide the best emotion labels as free text responses. Similarly as before, slow responses would be warned and timed out if necessary.

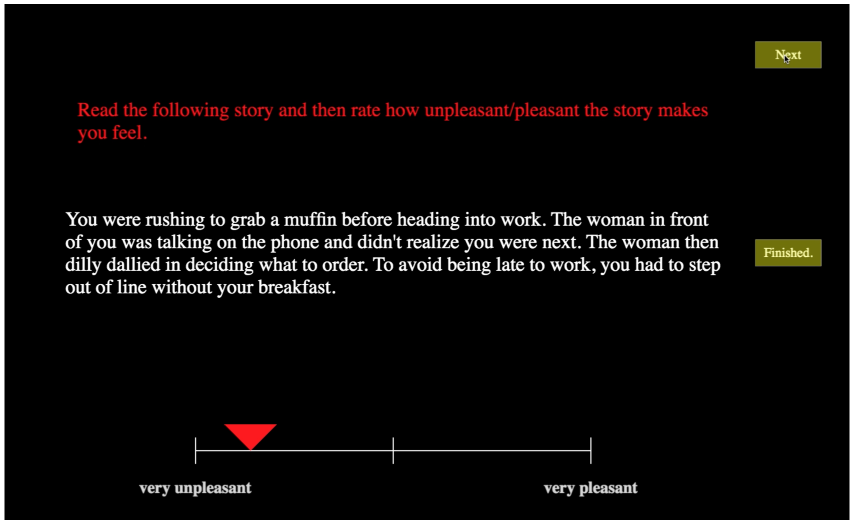

Figure 2.2: Example of a main trial for the story rating experiment.

Video rating experiment

The video rating task consisted of evaluating about 100 video clips on the assigned scale (Fig.2.3).

In the practice trial, participants were shown an example video (which could be replayed) and were asked to move the slider to the middle of the scale (which would be 4 for a scale from 1 to 7) as close as possible. Once the participants clicked the ‘finished’ button, the scale would appear below and video would be stopped (no replay allowed at this point). This was designed to get accurate response time that’s not contaminated by the time spent on watching the videos. There would be a warning message for clicking the ‘finished’ button too soon within the duration of the video (t) since the start of the trial. If no rating was made after 2t + 50 seconds, there would be a message to urge for a response or to contact the researcher for assistance. The experiment would end if the participant failed to proceed after 2t + 5 minutes.

In the main trials, videos were presented in random order. Similarly as in the practice trial, participants would watch the video, click the ‘finished’ button and then move the slider to rate. Messages would be shown if clicking the ‘finished’ button too soon (within t), or if responding too slowly (after 2t +50 seconds). The trial would be timed out if no response was made within 2t +60 seconds, and the experiment would end if approximately 10% (10) trials were skipped.

In the retest trials where a random set of 8 stimuli previously shown in the main trials would be presented again, participants were asked to rate them again and to also provide the best emotion labels as free text responses. Similarly as before, slow responses would be warned and timed out if necessary.

I posted 12 individual studies on Prolific with 12 different sets of stimuli (2 sets of stories and 10 sets of video clips). Each experiment session lasted about 20 to 30 minutes. Participants were allowed to finish multiple sessions with different sets of stimuli as they wished (each session had a different set of stimuli, and was posted as an individual study on Prolific), but were not allowed to participate more than once for the same session (viewing the same stimuli more than once would be potentially problematic, especially for scales such as surprised).

Figure 2.3: Example of a main trial for the video rating experiment.

2.1.5 Exclusion

Evoked emotion experiments

I pre-registered the exclusion criteria for the evoked emotion experiments before data collection began (`https://osf.io/vprz8`). The exclusions were applied to several levels (see counts after each level of exclusion in Fig.2.4).

Trial-wise deletions were done if responses were timed-out or if reaction time was extremely short (<400 ms). Session-wise deletions were done if any of the following conditions was met: failing more than 1 attention checks (out of a total of 3 checks); extremely low test-retest reliability estimated from the retest session (below 3 standard deviations from the mean reliability compared to all participants who rated the same scale on the same set of stimuli) or having more than 10% of invalid trials. Participant-wise deletions were done if they had more than 3 invalid sessions out of all the sessions that they did.

I included additional inclusion at the participant level from the Covid Dynamic study (as I describe below) because of the shared participant pool, which was not planned at the time of pre-registration.

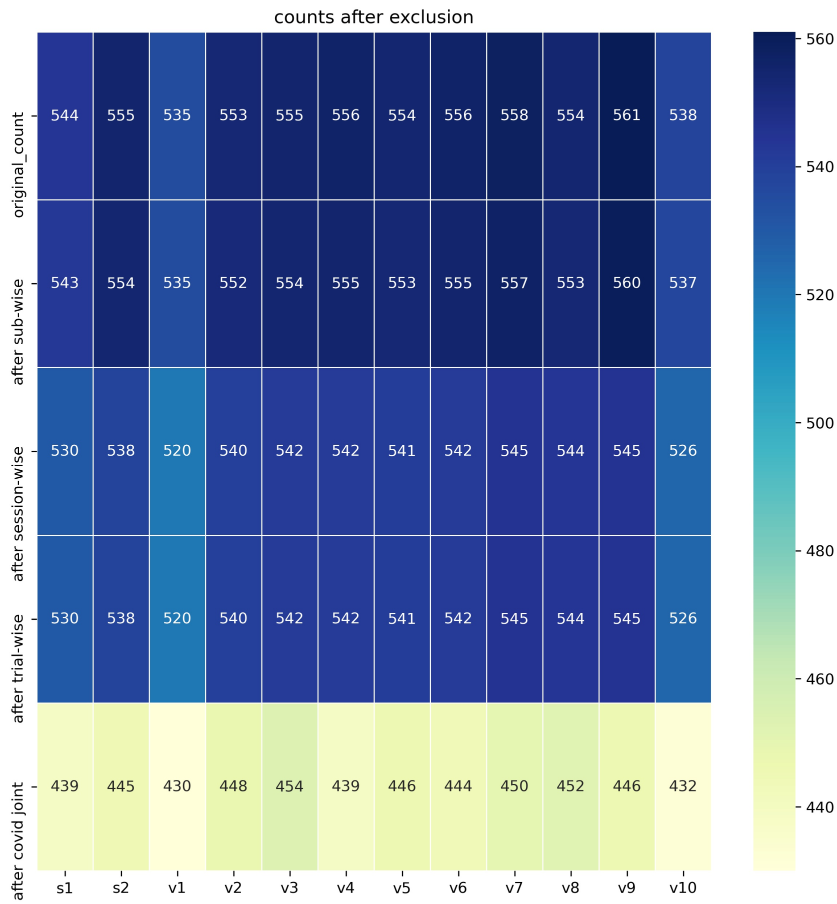

Figure 2.4: The number of remaining participants after each level of exclusion (each row) for different experiment sessions (each column: story: s1, s2; video: v1 to v10).

Covid Dynamic study

The exclusion criteria were chosen based on pilot analysis conducted using data from 50 randomly selected participants (from the total sample of 1797 participants) across 16 waves of data collections and pre-registered (`https://osf.io/y78mz`) before applying to the entire dataset. Again, exclusions were applied at multiple levels as described below (see counts after each level of exclusion in Fig.2.5).

Wave-wise deletion: I deleted participants' data from a wave based on the data quality metrics (for detailed info, please see the data quality section in the data release paper, [6]). More specifically, data were deleted if they failed 2 or more attention questions or the percentage of quality checks failed was equal or higher than 20%.

Participant-wise deletion: I deleted a participant's data for all waves if any of the following conditions were met:

1) I asked about pre-existing mental conditions at wave1 and also newly-diagnosed mental conditions in the past month (both questions allowed for multiple selections) several times at wave 4,7,9,11,13,15,16. I only included participants who self-reported to be Autism Spectrum Disorder only or Major depressive disorder only or Anxiety disorder only or Major depressive disorder and Anxiety disorder or None or Prefer not to disclose across all waves.

Participants reported otherwise were excluded, which included: those who reported to be Schizophrenia only or Bipolar only or Others (with text input for specification) only or PTSD only or multiple conditions other than Depression and Anxiety.

2) Participants were excluded if they had 3 or more waves where their data were deemed as low quality.

3) Participants were excluded if they had completed less than 50% of all waves, i.e., 8 waves.

4) Given that the Covid Dynamic study and the evoked emotion experiments share the same participant pool. A participant was excluded in both datasets if they met the exclusion criteria from either of the studies.

After exclusion as described above, the number of unique participants for the story rating experiments, the video rating experiments and the Covid-Dynamic study were 554, 638 and 1000 respectively (see Fig.2.6 for an overview of data collection and Table 2.2 for a characterization of the sample).

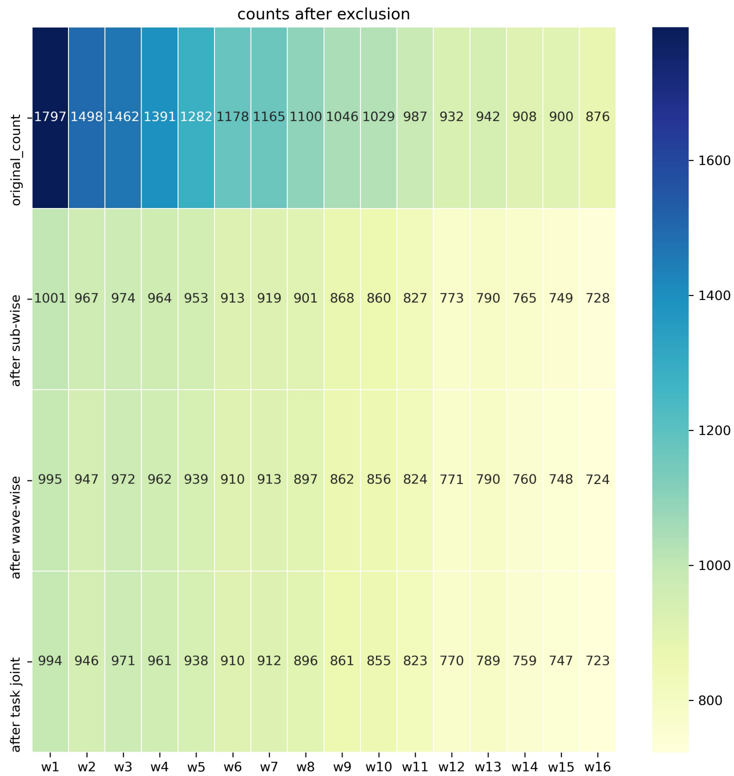

Figure 2.5: The number of remaining participants after each level of exclusion (each row) for different waves (each column, from wave 1 to wave 16).

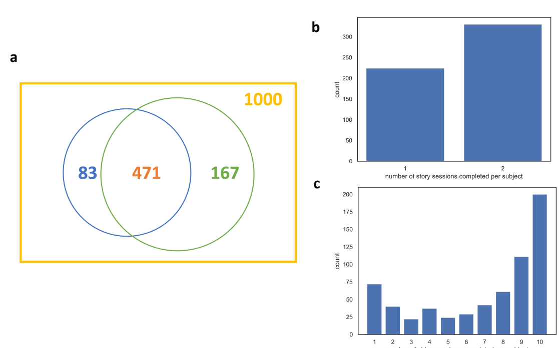

Figure 2.6: Overview of data collection. (a) Venn diagram showing the overlap of participants (completed at least one session or one wave) for each type of experiments (yellow rectangle: the Covid-Dynamic study/real-life emotions, blue circle: story rating experiment, and green circle: video rating experiment). (b) histograms for the number of story sessions completed by each participant and (c) histograms for the number of video sessions completed by each participant.

Table 2.2: Demographic characteristics, means, and standard deviations of all measures used for the final sample.

	All (N = 1000)	Female (N=507)	Male (N=493)
Age (in years) (mean; sd)	39.65 (14.22)	40.39 (14.47)	38.89 (13.92)
Education: below Bachelor (n; %)	442 (44.2%)	220 (43.39%)	222 (45.03%)
Education: Bachelor and above (n; %)	558 (55.8%)	287 (56.61%)	271 (54.97%)
Religious level (mean; sd)	1.73 (0.92)	1.79 (0.94)	1.66 (0.9)
CD-RISC 10 (mean; sd)	26.38 (7.3)	25.59 (7.4)	27.18 (7.12)
NEO Openness (mean; sd)	30.75 (6.51)	31.69 (6.47)	29.79 (6.42)
NEO Conscientiousness (mean; sd)	33.34 (7.54)	33.26 (7.38)	33.43 (7.7)
NEO Extraversion (mean; sd)	22.68 (8.26)	22.34 (8.1)	23.04 (8.42)
NEO Agreeableness (mean; sd)	32.75 (6.26)	33.82 (5.89)	31.66 (6.45)
NEO Neuroticism (mean; sd)	20.4 (10.55)	22.18 (10.61)	18.58 (10.19)
BDI (mean; sd)	11.2 (10.12)	12.16 (10.28)	10.21 (9.88)
STAI (mean; sd)	41.82 (12.4)	43.12 (12.65)	40.49 (12.0)
PSS (mean; sd)	15.71 (7.24)	17.01 (7.2)	14.38 (7.03)
TAS (mean; sd)	44.21 (10.99)	42.94 (10.97)	45.52 (10.87)

2.2 Analytical methods

2.2.1 Evaluation of scale quality

Test-retest reliability

For the evoked emotion experiments, I had retest trials where I collected ratings for a random set of 8 stimuli previously shown in the main trials, which allowed me to assess within-subject test-retest reliability. I therefore calculated Pearson's correlation between the ratings at these two time points for each experiment session (Fig.2.7). Scales varied in their test-retest reliability, but the pattern of relative performance among scales was consistent across experiment sessions.

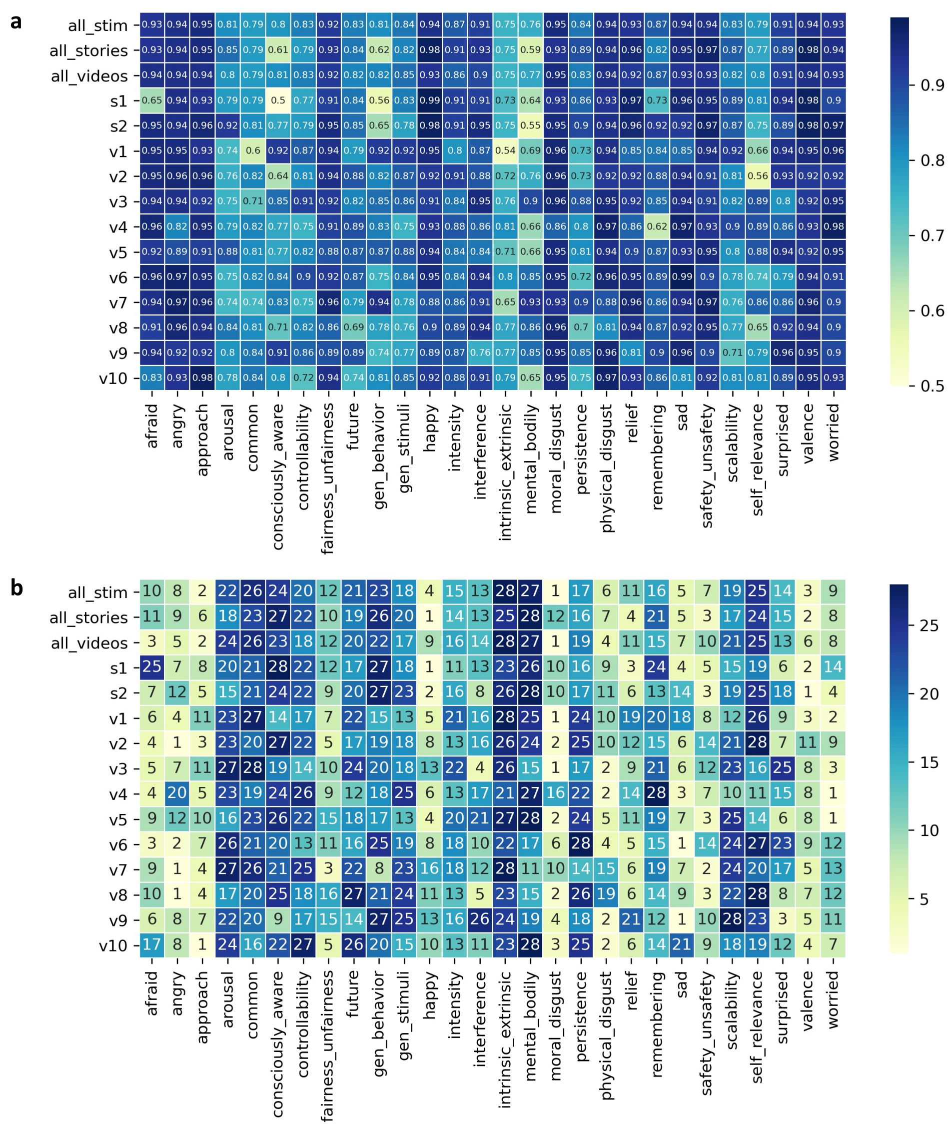

Figure 2.7: Test-retest reliability for each scale. (a) median value, and (b) median rank among 28 scales for each scale calculated using different data (all experiment sessions, all story sessions, all video sessions, and individual sessions alone).

Split-half reliability

I assessed between-subject consensus using split-half reliability. Specifically, for subjects who rated the same set of stimuli on the same scale, I randomly split their data into two halves, and correlated the average ratings derived from the two halves. Since the split was random, I repeated the process for 50 times to get a more reliable estimate. (Fig.2.8)

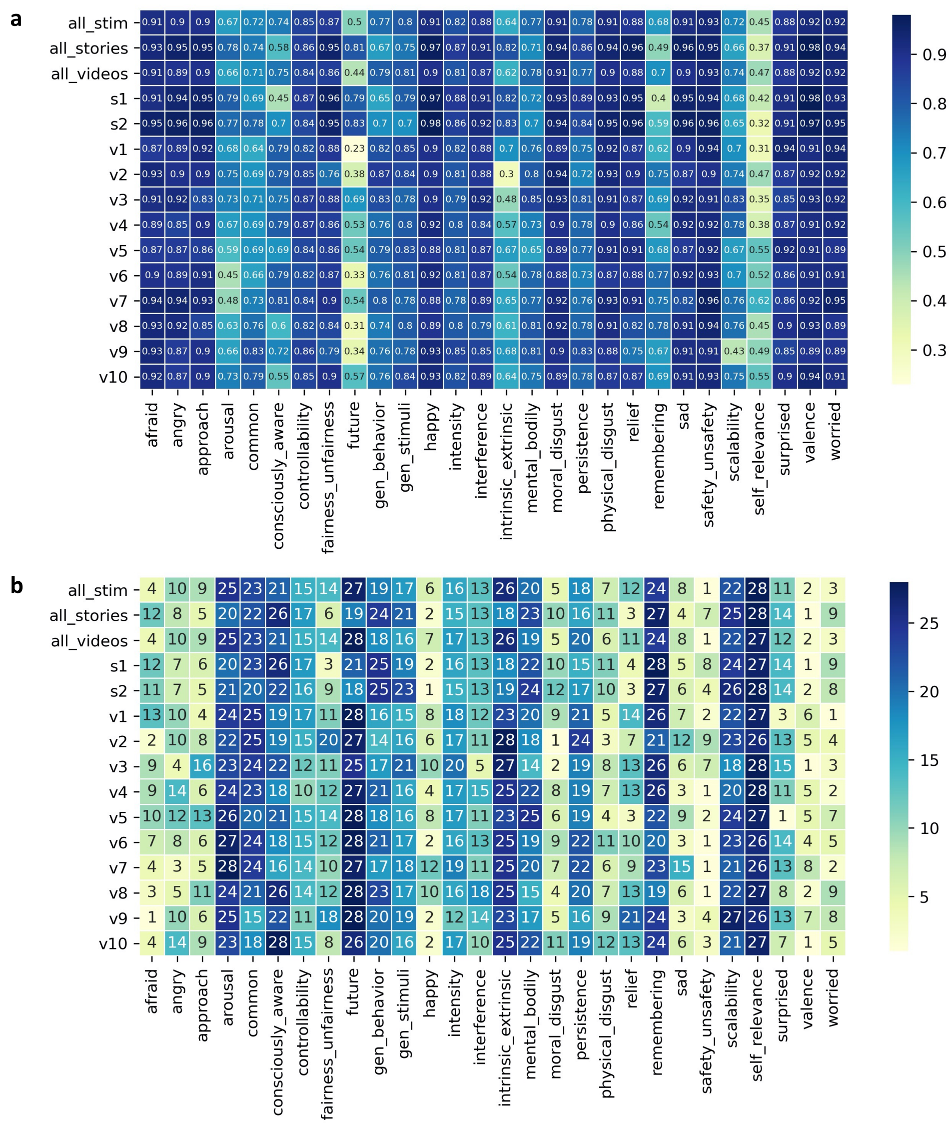

Figure 2.8: Split-half reliability for each scale. (a) median value and (b) median rank among 28 scales for each scale calculated using different data (all experiment sessions, all story sessions, all video sessions, and individual sessions alone).

Scales varied in their split-half reliability, but the pattern of relative performance among scales was consistent across experiment sessions. It's also worth noting that the relative performance pattern was robust across these two evaluation metrics, that is, scales with high test-retest reliability also had high split-half reliability and vice versa.

As can be seen from Fig.2.7 and Fig.2.8, those scales with the clearest semantic meaning in the first place were also the ones with the greatest reliability. Notably, scales with low reliability were those describing complex constructs such as intrinsic-extrinsic emotion experience, or whether the emotion was self-relevant. Somewhat surprisingly, the commonly used scale of arousal was also of relatively low reliability. Overall, readability as evaluated using grade levels did not correlate significantly with scale quality as evaluated using test-retest reliability($r = -0.19$, $p = 0.324$) and split-half reliability ($r = -0.21$, $p = 0.295$).

Scale exclusion

Based on the quality metrics that I described above, I decided to exclude five scales of low quality (self_relevance, Intrinsic_extrinsic, remembering, mental_bodily, future), which left me with 23 scales for the story/video rating experiments and 18 scales for the real-life emotions.

2.2.2 Representational similarity analysis (RSA)

RSA provides a framework for comparing the similarity structure across scales derived from different stimuli types [17]. To create a correlation matrix across scales, I first averaged ratings across participants who rated the same set of stimuli on the same scale for the evoked emotion experiments. For real-life emotions from the Covid Dynamic study, since emotions for each individual were idiosyncratic, raw ratings were used without averaging. I then calculated the scale-by-scale Pearson correlation matrices for emotions evoked by stories, for emotions evoked by videos and for real-life emotions. For easier visualization, the matrices were sorted using hierarchical clustering to intuitively depict the underlying structure.

To assess how similar the correlation matrices were across stimuli types, I calculated second-order similarity. Because the correlation matrices were symmetric about a diagonal of ones, I vectorized the lower triangle of each matrix and calculated the Spearman rank correlations (without assuming a strict linear match) for each pair of matrices.

To assess the relatedness of matrices, I followed the randomization procedure as outlined in [17]. More specifically, I reordered rows and columns of one of the two correlation matrices according to a random permutation, and computed the correlations between the two matrices. A distribution of correlations simulating the null hypothesis that the two correlation matrices were unrelated can be derived by repeating this step 10000 times, which can be combined with the actual correlation to derive the p-value.

2.2.3 Factor analysis

Determining the optimal number of factors

A number of standard statistical methods have been proposed for determining the appropriate number of factors to retain, but no single method is considered to be optimal [18]. In the hope of finding converging evidence, I tried out the following methods which are commonly used.

Parallel analysis compares the eigenvalues of the actual data with eigenvalues of random data with the same size and only retains factors that are not due to chance [19]. Both the optimal coordinate (OC) and the acceleration factor (AF) attempt to provide non graphical solutions to the scree plot [20]. OC measures the gradients associated with eigenvalues and their preceding coordinates and finds the elbow based on a series of linear extrapolations. AF tries to identify where the slope of the curve changes most abruptly. The Very Simple Structure (VSS) simplifies the pattern matrix by only keeping the greatest loadings for each item and examines how well the original correlation matrix is reproduced [21]. Velicer's Minimum Average Partial test tries to identify factors that represent systematic variances, as opposed to residual or error variance [22]. Empirical BIC evaluates models with different numbers of factors, taking both the model fit and the parsimony into account.

Parallel analysis, the acceleration factor and the optimal coordinate were computed using the nScree function in the "nFactors" package in R. Very Simple Structure, Empirical BIC and Velicer's MAP were computed using the nfactors function in the "psych" package.

In addition to the statistical procedures described above, I also used a cross validation procedure to choose how many factors (tested a reasonable range of n = 1 to 8 factors) to retain for my data. More specifically, for each n, I randomly split data into two halves (repeated for 20 iterations). I applied exploratory factor analysis (EFA) to the first half of the data which would result in a factor loading matrix. I then assigned each item to a factor if the absolute loading was higher than a cutoff value of 0.2, and then fitted a confirmatory factor analysis (CFA) model to the other half of the data. To evaluate the performance of different factor solutions, I derived the percentage of explained variance from the EFA and root mean square error of approximation (RMSEA) fit index from the CFA.

Assessing the robustness of factor solutions

I quantified the robustness of my factor solutions both across different numbers of stimuli and across different numbers of scales.

To test the robustness of my results against the number of stimuli, I systematically reduced the number of stimuli and computed factor congruences between factors derived using the reduced set and the original set. For the story rating data, I started with the full set of 150 stories, and then removed 5 random stories (20 randomizations each) at each step, until I was left with 5 stories for the last step. At each step, I used the new aggregated ratings for EFA and calculated Tucker indices of factor congruence for all sub-datasets (with orthogonal Procrustes rotation). Video rating data was assessed in a similar way that I started with 998 videos, removed 25 each time (20 randomizations each) until I was left with 23 videos for the last step. For real-life emotions, I started with 12861 instances, and removed 250 each time (20 randomizations each) until I was left with 111 instances for the last step.

To test the robustness of my results against the number of scales, I systematically reduced the number of scales and quantified the relatedness of the original factors from the full set and the ones from the subset by correlating the factor scores.

The order of removal was determined based on the redundancy of each scale with the rest of the scales. More specifically, I started with the full correlation matrix (23 by 23, the average of the two matrices using story and video rating data) across all scales after excluding scales of low quality. I quantified the global redundancy of each scale by computing the means of each row (or column), that is, the means of correlations of a scale with all the rest of the scales. The scale with the highest mean correlation and thus the highest redundancy was removed and the correlation matrix was updated. I repeated the same process with the updated matrices until I was left with two scales.

After determining the order of removal of scales, I removed scales one by one as specified and reperformed EFA to extract the same number of factors as before until I was left with 5 scales. I quantified the relatedness of the original factors from the full set and the ones from the subset by correlating the corresponding factor scores.

Exploratory factor analysis procedure

I determined the optimal number of factors to retain by taking multiple aspects into consideration: the suggestions from the standard statistical methods, results from the cross validation approach, results from the robustness assessment and the interpretations of the factor loadings. The number of factors to retain for story, video and real-life emotions was determined to be 3, 3, and 4 respectively. Exploratory factor analysis was then performed to extract the factors using the minimal residual method, and the solutions were rotated with oblimin rotation for interpretability. The Tenberge method was used to obtain oblique factors scores (using the "‘fa" function in the "psych" package in R).

2.2.4 Uniform Manifold Approximation and Projection (UMAP)

UMAP is a nonlinear dimensionality reduction technique that tries to learn the manifold structure and find a low dimensional embedding that preserves the structure [23]. I used UMAP to reduce the original high dimensional space to a two-dimensional space to visualize the distribution of emotional experiences.

The most important parameter of UMAP is the size of the local neighborhood, which controls the tradeoff between preserving global and local structure. Specifically, larger values lead to a better preservation of global structure while the local structure gets worse. I used a cross validation procedure to choose this parameter with the goal of preserving both global and local structures. Two metrics were used to evaluate the preservation of local and global structure respectively: trustworthiness which is based on the change of nearest neighbors [24] and the Spearman rank correlation of the pairwise Euclidean distances. Data was randomly split into two halves (repeated for 10 times) where one half was used for training and the other half for testing for a range of neighborhood sizes. The size of the local neighborhood was determined to be 10, 15, and 25 for story, video and real-life ratings respectively (Fig.2.9, Fig.2.10, Fig.2.11).

To further elucidate the structure, I used several color coding schemes. For story and video ratings, I color-coded using both the categorical labels from the original studies and factor scores derived from my own data. For real-life emotions, only factor scores were used.

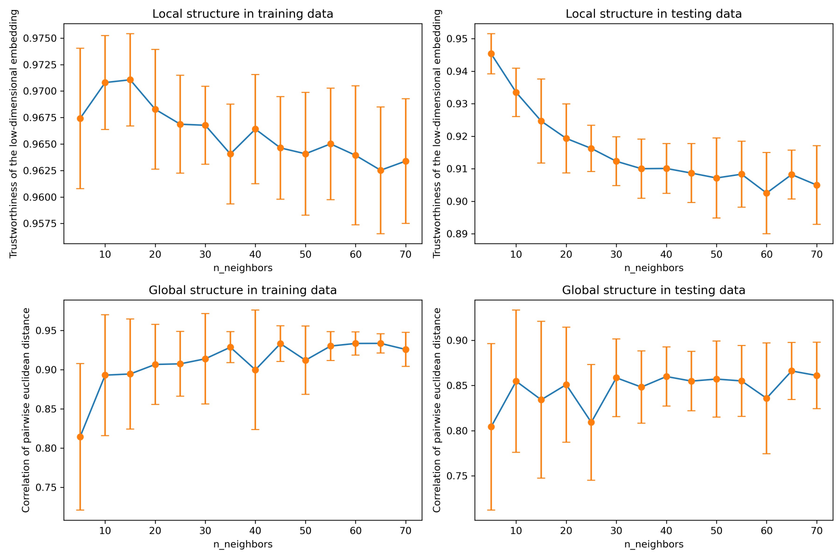

Figure 2.9: The means (points) and standard deviations (error bars, n = 10 iterations) of trustworthiness (upper) and rank correlation of pairwise distances (lower) of the training data and testing data (left to right) for the story ratings, with different sizes of the local neighborhood for UMAP.

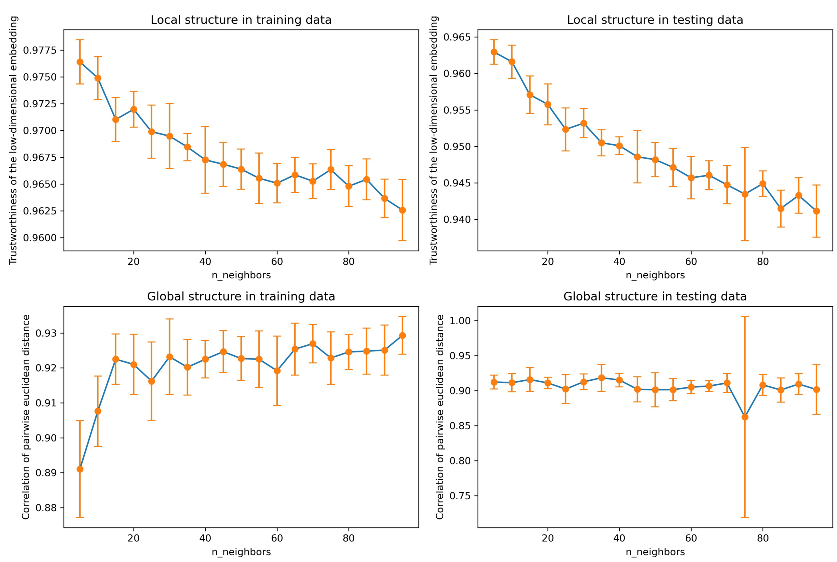

Figure 2.10: The means (points) and standard deviations (error bars, n = 10 iterations) of trustworthiness (upper) and rank correlation of pairwise distances (lower) of the training data and testing data (left to right) for the video ratings, with different sizes of the local neighborhood for UMAP.

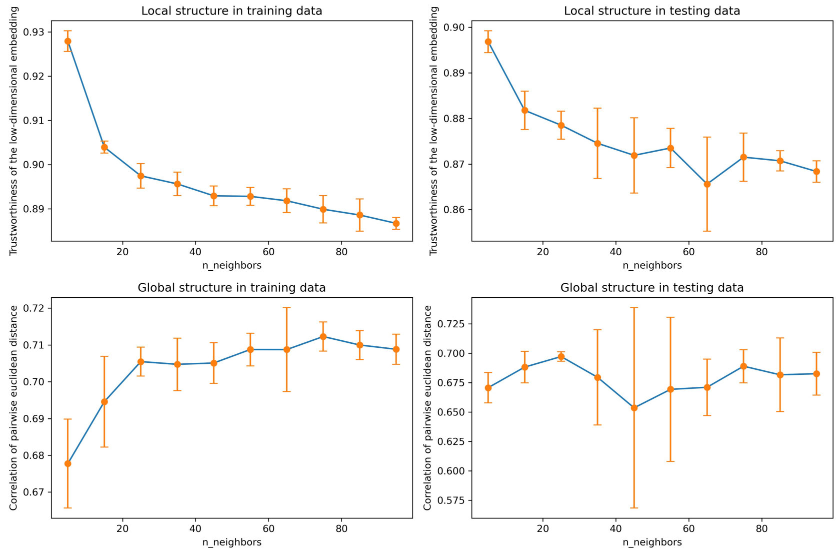

Figure 2.11: The means (points) and standard deviations (error bars, n = 10 iterations) of trustworthiness (upper) and rank correlation of pairwise distances (lower) of the training data and testing data (left to right) for real-life emotion ratings, with different sizes of the local neighborhood for UMAP.

2.2.5 Clustering

Clustering algorithms

I review some of the commonly used clustering algorithms below.

K-means clustering (implemented using sklearn.cluster.KMeans in Python) is a centroid-based clustering algorithm that tries to find the k clusters (as specified) that minimize the inertia, that is, the within-cluster sum-of-squares [25] . The number of clusters was chosen either to directly compare with previous studies or determined using the evaluation metrics as described below.

Mean-shift (implemented using sklearn.cluster.MeanShift in Python) is a clustering algorithm that assigns the data points to the clusters iteratively by shifting points towards the mode, that is, the highest density of data points in the region. Unlike K-means, the algorithm does not require setting the number of clusters in advance [26].

DBSCAN (implemented using sklearn.cluster.DBSCAN in Python) is a popular density-based clustering algorithm where clusters are detected according to density drop [27]. The algorithm divides points into core points, neighbors of core points, and outliers. The two important parameters of DBSCAN that need to be specified are min_samples (number of samples in a neighborhood for a point to be considered as a core point excluding the point itself) and eps (the maximum distance between two samples for one to be considered as in the neighborhood of the other) as named in the sklearn implementation. A core point is therefore defined as a sample in the data where there exist min_samples other samples within a distance of eps. Heuristics [28] suggest that min_samples can be set at 2 times the number of dimensions of the data and eps can be chosen based on the k (min_samples -1) nearest distance plot.

To probe a hierarchy of clusters, I carried out hierarchical agglomerative clustering (HAC, implemented using sklearn.cluster.AgglomerativeClustering in Python) which is a bottom-up approach where each observation starts in its own cluster, and similar clusters are successively merged. I used Euclidean distance as the distance metric and variance-minimizing Ward linkage which is similar to the objective of k-means [29].

For all of the clustering algorithms described above, I used Euclidean distance as the distance measure. Since all ratings on the scales were in the same range of 1 to 7, and I believed that the actual distributions for each scale contained intrinsic information on how informative the scale was at describing emotional experiences, I didn't perform standardization to force the same variance across scales.

Evaluation metrics for clustering

Evaluation of clustering results is difficult with no agreed-upon standards, and can be divided into two types of approaches. When ground truth is available, one can assess the agreement between the true labels and the clustering results. Otherwise, the evaluation is based on the data itself, with the general idea of assessing whether members of the same cluster are more similar than members of different clusters. I review some commonly used evaluation metrics below.

Inertia is the within-cluster sum-of-squares, lower values indicate that the clusters are more internally coherent. The Silhouette coefficient (calculated using sklearn.metrics.silhouette_score) compares the distances of a point to other points in the same cluster with its distances to points in the nearest cluster [30]. The value ranges from -1 to 1, with higher values indicating better clustering. The Silhouette coefficient of a set of points is given as the mean of the Silhouette coefficient of the individual points. The Davies-Bouldin index (calculated using sklearn.metrics.davies_bouldin_score) measures the average 'similarity' between clusters, where the similarity is a measure that compares the intra-cluster distances with the inter-cluster distances. Lower values indicate better clustering with 0 being the lowest possible value [31]. The Calinski-Harabasz index (calculated using sklearn.metrics.calinski_harabasz_score) is defined as the ratio of the sum of between-cluster variance and of within-cluster variance for all clusters with higher values indicating better clustering [32].

The adjusted Rand index (calculated using sklearn.metrics.adjusted_rand_score) is a version of the Rand index corrected for chance, ranging from -1 to 1 with higher values indicating better match [33]. The unadjusted Rand index indicates how many pairs are in agreement between the clustering result and true labels out of the total number of pairs [34]. The adjusted mutual information score (calculated using sklearn.metrics.adjusted_mutual_info_score) is a version of the mutual information corrected for chance with higher values indicating better match. The mutual information measures how much information is shared between the clustering result and the ground truth [35]. A contingency matrix (calculated using sklearn.metrics.cluster.contingency_matrix) can be used to visualize the agreement for every true/predicted cluster pair, allowing one to examine the spread of each true cluster across predicted clusters and vice versa.

References

[1] Alan S Cowen and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients". In: *Proceedings of the National Academy of Sciences* 114.38 (2017), E7900–E7909.

[2] Amy E Skerry and Rebecca Saxe. "Neural representations of emotion are organized around abstract event features". In: *Current biology* 25.15 (2015), pp. 1945–1954.

[3] Lauri Nummenmaa et al. "Maps of subjective feelings". In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9198–9203.

[4] Ralph Adolphs and David J Anderson. *The neuroscience of emotion: A new synthesis*. Princeton University Press, 2018.

[5] J Peter Kincaid et al. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Tech. rep. Naval Technical Training Command Millington TN Research Branch, 1975.

[6] Tessa Rusch et al. "COVID-Dynamic: A large-scale multifaceted longitudinal study of socioemotional and behavioral change across the pandemic". In: (2021).

[7] Kathryn M Connor and Jonathan RT Davidson. "Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC)". In: *Depression and anxiety* 18.2 (2003), pp. 76–82.

[8] PT Costa and RR McCrae. "The NEO Five Factor Inventory: Manual". In: *New York, NY: Psychological Assessment Resources* (1991).

[9] Richard C Gershon et al. "NIH toolbox for assessment of neurological and behavioral function". In: *Neurology* 80.11 Supplement 3 (2013), S2–S6.

[10] Aaron T Beck, Robert A Steer, Gregory K Brown, et al. *Manual for the beck depression inventory-II*. 1996.

[11] CD Spielberger et al. "Manual for the State-Trait Anxiety Inventory (STAI)." In: *Palo Alto: Consulting Psychologists Press* (1983).

[12] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. "A global measure of perceived stress". In: *Journal of health and social behavior* (1983), pp. 385–396.

[13] Eric Hehman et al. "Assessing the point at which averages are stable: A tool illustrated in the context of person perception". In: (2018).

[14] R Michael Bagby, James DA Parker, and Graeme J Taylor. "The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure". In: *Journal of psychosomatic research* 38.1 (1994), pp. 23–32.

[15] James J Gross and Oliver P John. "Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being." In: *Journal of personality and social psychology* 85.2 (2003), p. 348.

[16] John R Crawford and Julie D Henry. "The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample". In: *British journal of clinical psychology* 43.3 (2004), pp. 245–265.

[17] Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. "Representational similarity analysis-connecting the branches of systems neuroscience". In: *Frontiers in systems neuroscience* 2 (2008), p. 4.

[18] Matthew Courtney and Ray Gordon. "Determining the number of factors to retain in EFA: using the SPSS R-menu v2 0 to make more judicious estimations". In: *Practical Assessment, Research, and Evaluation* 18.1 (2013), p. 8.

[19] John L Horn. "A rationale and test for the number of factors in factor analysis". In: *Psychometrika* 30.2 (1965), pp. 179–185.

[20] Gilles Raiche et al. "Non-graphical solutions for Cattell's scree test". In: *Methodology* (2013).

[21] William Revelle and Thomas Rocklin. "Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors". In: *Multivariate Behavioral Research* 14.4 (1979), pp. 403–414.

[22] Wayne F Velicer. “Determining the number of components from the matrix of partial correlations”. In: *Psychometrika* 41.3 (1976), pp. 321–327.

[23] Leland McInnes, John Healy, and James Melville. “Umap: Uniform manifold approximation and projection for dimension reduction”. In: *arXiv preprint arXiv:1802.03426* (2018).

[24] Jarkko Venna and Samuel Kaski. “Neighborhood preservation in nonlinear projection methods: An experimental study”. In: *International Conference on Artificial Neural Networks*. Springer. 2001, pp. 485–491.

[25] Stuart Lloyd. “Least squares quantization in PCM”. In: *IEEE transactions on information theory* 28.2 (1982), pp. 129–137.

[26] Keinosuke Fukunaga and Larry Hostetler. “The estimation of the gradient of a density function, with applications in pattern recognition”. In: *IEEE Transactions on information theory* 21.1 (1975), pp. 32–40.

[27] Martin Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *kdd*. Vol. 96. 34. 1996, pp. 226–231.

[28] Erich Schubert et al. “DBSCAN revisited, revisited: why and how you should (still) use DBSCAN”. In: *ACM Transactions on Database Systems (TODS)* 42.3 (2017), pp. 1–21.

[29] Joe H Ward Jr. “Hierarchical grouping to optimize an objective function”. In: *Journal of the American statistical association* 58.301 (1963), pp. 236–244.

[30] Peter J Rousseeuw. “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.

[31] David L Davies and Donald W Bouldin. “A cluster separation measure”. In: *IEEE transactions on pattern analysis and machine intelligence* 2 (1979), pp. 224–227.

[32] Tadeusz Caliński and Jerzy Harabasz. “A dendrite method for cluster analysis”. In: *Communications in Statistics-theory and Methods* 3.1 (1974), pp. 1–27.

[33] Douglas Steinley. “Properties of the Hubert-Arable Adjusted Rand Index.” In: *Psychological methods* 9.3 (2004), p. 386.

[34] William M Rand. “Objective criteria for the evaluation of clustering methods”. In: *Journal of the American Statistical association* 66.336 (1971), pp. 846–850.

[35] Nguyen Xuan Vinh, Julien Epps, and James Bailey. “Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance”. In: *The Journal of Machine Learning Research* 11 (2010), pp. 2837–2854.

Chapter 3

EMOTIONS EVOKED BY STORIES

3.1 Introduction

Lexical stimuli are commonly used in human psychological studies to study emotions. The simplest form is to use a set of emotional words [1, 2], but this approach comes with major drawbacks. Words without context are ambiguous in meaning and thus not ecologically valid. It's also likely that when presented with words, subjects simply rate the semantic meaning of those words, without experiencing any emotions at all. For example, rating the word "happy" would presumably elicit high ratings on a scale of "happiness"—but this could be done simply by matching the semantic meaning of the word with the semantic meaning of the scale, in a very shallow way that does not require the induction of any emotion experience. Sentences or vignettes are better than words alone as they provide a more vivid description of an emotional situation and thus are more specific and effective at eliciting emotions [3, 4, 5].

On the other hand, it is certainly possible to use words to induce strong emotions—provided that they are not rated directly, but used as triggers to elicit rich experiences that can then be rated. For instance, autobiographical recollection of past emotional events can be studied this way, and can be quite effective at eliciting strong emotions. For instance, subjects can be induced to cry when given the cue "sad", when this instructs them to recall sad autobiographical memories [6]. However, the limitation of this approach is that researchers can't fully control for the differences in the content of the recalls across individuals, since idiosyncratic memories would be evoked in each individual.

Taking multiple factors into consideration, I decided to use the short vignettes developed by [4], as they featured a large number of scenarios and were verified to be able to reliably elicit 20 fine-grained emotions. Considerable effort was put into designing and validating these stimuli. The authors used the 200 vignettes to study people's ability to infer other people's emotions (not first-person emotion experiences) and found that an appraisal model resembled neural patterns evoked in brain regions thought to be involved in emotion, a result that was compared with the basic emotion model and the circumplex model. Given the extensive prior assessment of these stimuli, and their link to brain activation in a neuroimaging study, it is plausible that the stimuli indeed elicited the emotions that they were designed to evoke.

Across the many studies that have used lexical stimuli, valence and arousal were most consistently identified as the two dimensions that characterized emotions [1, 7, 2, 3, 5]. In Jim Russell's seminal work, emotion words form a circumplex (a circular arrangement) in this 2-D space. The two dimensions of valence and arousal are also universally acknowledged to constitute what psychologists refer to as "core affect", that aspect of emotion experience that underlies all affective conscious states. Disagreements across studies then focus more on the variety of specific emotions that can be added, elaborated, or constructed on top of core affect—disagreements that in many cases can be largely attributed to the use of different scales and words used.

3.2 Results

3.2.1 Three dimensions underlying emotion experiences evoked by stories

Correlation structure across scales

After excluding five scales due to their low reliability (as explained in 2.2.1), I derived a pairwise Pearson correlation matrix across the 23 remaining scales (Fig.3.1), which I interpret as a (likely incomplete) representation of the similarity structure in the underlying psychological space of emotion experiences evoked by stories.

First, I observed strong correlations across scales as expected, suggesting that the dimensionality can be reduced to more efficiently represent the psychological space. Closer inspection revealed that the scales can be grouped into several groups. The first group of scales (afraid, worried, interference, physical disgust, angry, moral disgust) are the ones at the very top, all describing how negative the emotions are while the other group of scales (safety, approach, fairness, valence, happy, relief) describe how positive the emotions are. I also noted that in the middle, there's a group of scales that characterize the intensity and persistence of emotions (arousal to scalability). And finally, the scales at the bottom try to describe how generalized an emotion can be.

Looking at specific correlations suggests that more negative emotions are felt more intensively and last longer. They are also less controlled and less generalizable over both stimuli and behavior. Some of the scales didn't have strong correlations with the generalizability-related scales possibly because of the content of the stories.

For instance, most sad stories featured in my study describe scenarios related to death which are uncommon and not representative of the richness of the sadness experienced in real life.

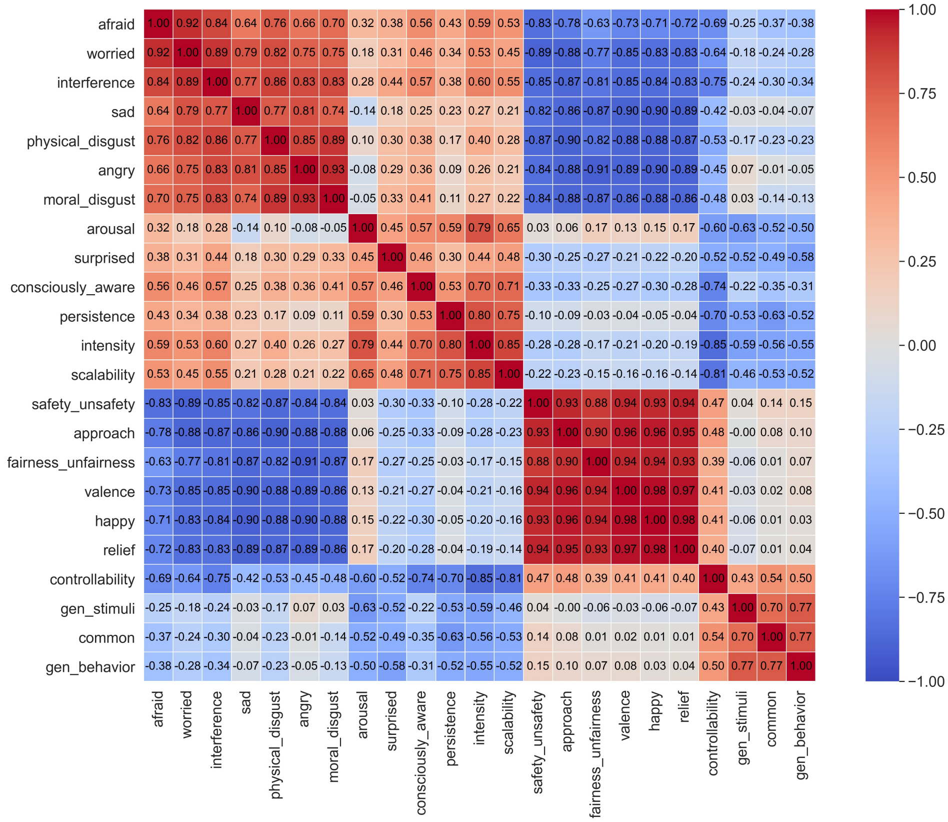

Figure 3.1: Correlation matrix across scales for emotions evoked by stories (sorted using hierarchical clustering to intuitively depict the underlying structure).

Factor analysis

As mentioned above, the correlation matrix across scales suggested that the psychological space of emotion experiences evoked by stories can be characterized using a smaller number of underlying factors, but the number of factors to retain in an exploratory factor analysis needs to be determined first.

Determining the optimal number of factors to extract

Ultimately, it's a subjective decision for how many factors to retain, and there's no universal agreement on the exact procedure. I determined the optimal number of factors to retain by taking multiple aspects into consideration: the suggestions from the standard statistical methods, results from the cross validation approach, results from the robustness assessment, and the interpretations of the factor loadings, which I describe below.

First, I used the six commonly used statistical tests (see details in 2.2.3). The results (Fig.S3.1) did not converge to a single number, but suggested either 2 or 3 factors to retain. Very Simple Structure, Empirical BIC, and Velicer's MAP suggested 2, 3 and 3 factors respectively. Parallel analysis, the acceleration factor and the optimal coordinate suggested 3, 2 and 3 factors respectively.

In addition to the standard statistical procedures described above, I also used a cross validation procedure (see details in 2.2.3) to choose how many factors to retain for the story rating data. The general idea was to extract factors by applying exploratory factor analysis (EFA) on one half of the data and then test the factor structure using confirmatory factor analysis (CFA) on the other held-out half. Visual inspection of the result (Fig.S3.2) suggested that two factors were most appropriate. As the number of factors increased from 1 to 2, there's a significant increase in explained variance from EFA and model fit for CFA while adding more factors subsequently showed marginal improvement.

I then applied EFA to further assess the interpretability of different factor solutions. When extracting 2 factors, the factors each explained 49% and 28% of the common variance in the data (77% in total). I interpreted the factors as "valence" and "arousal" (Fig.S3.3a). When extracting 3 factors, the factors each explained 48%, 21% and 13% of the common variance in the data (82% in total). I interpreted the factors as "valence", "arousal" and "generalizability" (Fig.3.3). I also attempted the 4 factor solution, the factors each explained 48%, 20%,13% and 2% of the common variance in the data (84% in total) which showed a marginal improvement and the last factor was uninterpretable (Fig.S3.3b).

Evidence so far suggested both the 2 and 3 factor solutions seemed reasonable, so I assessed the robustness of both with regard to the number of stimuli and number of scales.

First, I systematically reduced the number of stories starting from the whole set of 150 stories. At each step, I used the new aggregated ratings for EFA and calculated Tucker indices of factor congruence between factors derived using the reduced set and the original set of stimuli (with orthogonal Procrustes rotation). Both the 2 factor and 3 factor solutions were robust to the number of stories (Fig.3.2 a.b). For the 2 factor solution, all mean factor congruences were higher than 0.9 with no fewer than 10 stories (roughly 6.7% of the whole set). For the 3 factor solution, all mean factor congruences were higher than 0.9 with no fewer than 25 stories (roughly 16.7% of the whole set).

Second, I removed scales one by one (the order of removal was based on global redundancy, as explained in 2.2.3) and quantified the relatedness of the original factors from the full set and the ones from the subset by correlating the factor scores. Both the 2 factor and 3 factor solutions were robust to the number of scales as the factors derived from the full set versus the subsets of scales were highly correlated (Fig.3.2 c.d).

All evidence taken together, it was still difficult to choose decisively between the 2 and 3 factor solutions as both were interpretable and robust. I decided to retain 3 factors for completeness. However, I make no strong claims that the 3 factor solution is superior to the 2 factor solution.

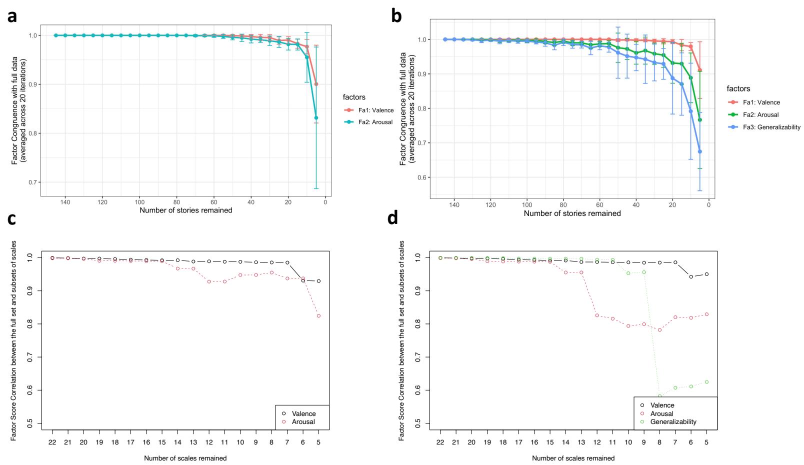

Figure 3.2: Robustness of factor solutions with respect to the number of stories and number of scales. (a,b) Points indicate the means and error bars indicate standard deviations of Tucker indices of factor congruence (with orthogonal Procrustes rotation) between the full set versus subsets of stories across 20 iterations, and are color-coded for different factors for (a) the 2 factor solution and (b) the 3 factor solution. (c, d) Pearson's correlations between factor scores from the full set versus subsets of scales, color-coded for different factors for (c) the 2 factor solution and (d) the 3 factor solution.

Interpretation of the three factors

Exploratory factor analysis was then performed to extract 3 factors using the minimal residual method, and the solutions were rotated with oblimin for interpretability. The Tenberge method was used to obtain oblique factors scores (all using the "'fa" function in the "psych" package in R). I interpreted the three factors as "valence", "arousal", and "generalizability" (see Fig.3.3 for factor loadings). The first two factors are in line with previous literature [8] while the third factor is novel, and will be discussed in more detail in Chapter 6.

The two groups of scales that I identified earlier in the correlation matrix which describe positive and negative emotions respectively load strongly onto the “valence” factor. The group of scales that characterize the intensity and persistence of emotions load strongly onto the “arousal” factor. And lastly, the three scales with the highest loadings on the “generalizability” factor are “common”, “generalizability over stimuli”, and “generalizability over behavior”.

Examining the factor scores for individual stories also allowed me to verify my interpretation of the scales. For instance, stories describing accomplishments evoked the most positive emotions while stories featuring death evoked the most negative emotions. Intense emotions were evoked by both positive and negative events, finding out about pregnancy brought intense joy while killing children while driving drunk resulted in intense guilt. Feeling content after a long day at work or feeling annoyed listening to gossip on a long train ride were examples of the most generalized emotions, of positive and negative valence respectively. On the other hand, feeling terrified after getting lost in the woods or feeling grateful for receiving free medication for cystic fibrosis from an altruistic doctor were examples of least generalized emotions.

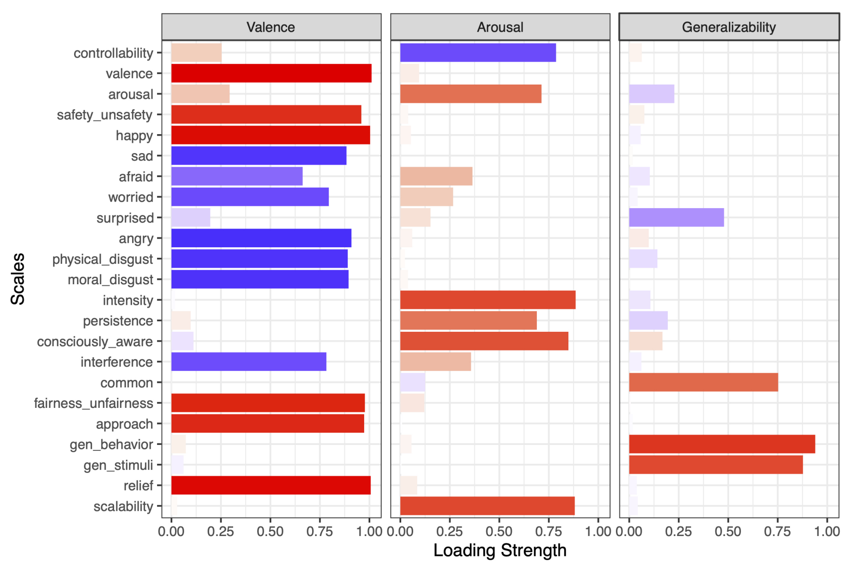

Figure 3.3: Factor loadings of scales on the three factors from EFA. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

3.2.2 Distribution of emotion experiences evoked by stories

The stories that I adapted from [4] featured a wide range of content and were designed to evoke 20 kinds of emotions. Among the three types of stimuli that I used, stories were the most manipulated as they were constructed from scratch by the authors. The number of stories intended to evoke positive and negative emotions were roughly balanced. Therefore, the distributions for many scales related to valence (for instance, valence, relief, fairness) were bimodal while ratings on other scales were more normally distributed (Fig.3.4).

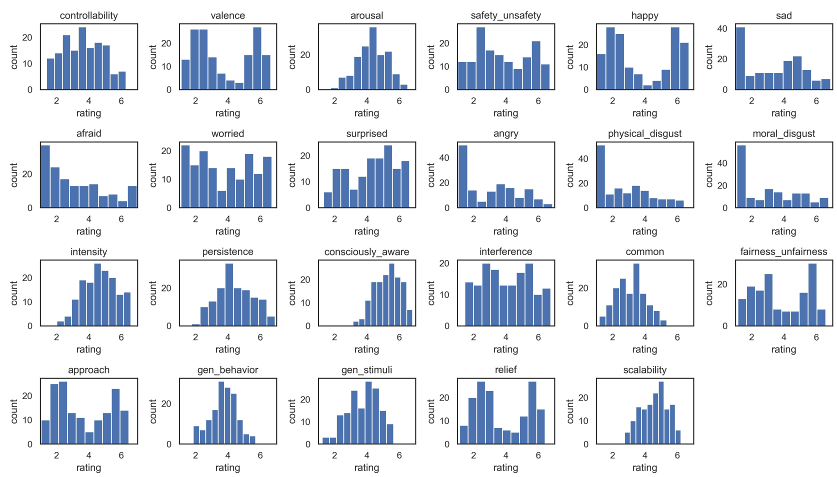

Figure 3.4: Distributions of aggregated ratings (across subjects) on the 23 scales for emotions evoked by 150 stories.

Uniform Manifold Approximation and Projection (UMAP)

The histograms on individual scales gave me some hints on the distributions of emotions in the high dimensional space as defined by the 23 scales of mine. For instance, I might find two large groups of emotions defined by their valence.

Still, I need tools to project data from the original high dimensional space to a lower dimensional space (ideally 2d) that's easier for visualization and interpretation. To that end, I applied UMAP which is a nonlinear dimension reduction technique, aimed at preserving structure [9]. Since three factors accounted for most of the total variance in my data, representing the emotions using a two dimensional UMAP plot should be a reasonable approximation. I further combined two sources of information which are the labels for the intended emotion categories and the factor scores with the UMAP plots, trying to address a key question of whether emotion experiences are discrete or dimensional.

Color code UMAP by intended categories (20 categories)

As I have noted and suspected earlier, the UMAP plot revealed two large groups of emotions (of positive and negative valence respectively) that seemed to be well separated (Fig.3.5). This pattern of two clusters was obvious even without the categorical labels, what the labels did do was to reveal the meaning of the clusters.

However, if I were to look at the embedding alone without the categorical labels, I wouldn't necessarily recognize additional clusters nested within the two overall clusters, that is, the 20 emotion categories didn't appear well separated.

With the labeling for the intended emotion categories, for some emotions, instances belonging to the same categories were located in closer proximity to one another (such as apprehensive and content) compared to other categories where instances were more scattered (such as surprised).

It is possible that the discrepancy among categories is related to the content of the stories; for some categories, the content was more similar and therefore the evoked emotions were more unified. It could also be related to the effectiveness and specificity of the stories at evoking the intended emotions. For instance, on the top right side, a story described a situation where a driver hit a boy because of texting while driving, but the boy turned out to be ok. This was intended to evoke gratitude but failed as subjects reported feeling guilty and scared but not grateful.

Another possible explanation is that for some categories, the intra-category variance is intrinsically larger than the inter-category variance. Take experiences of surprise for example (as indicated by brown dots in Fig.3.5 a), these were located at various positions, associated with neutral (such as a turnaround game), positive (such as a surprise party after PhD defense), and negative (such as sudden book ending with characters all killed) valences.

Even for categories where instances are located in relatively close proximity, it is a separate question of whether the categories as defined semantically actually form clusters in the high dimensional space with clear boundaries. Since loss of information is inevitable for any dimensionality reduction tool, it is possible that well separated clusters in the high dimensional space overlap in the low dimensional embeddings. I addressed this question more directly using clustering analysis as UMAP alone can't answer that.

Color code UMAP by factor scores

In addition to testing whether emotions form discrete clusters, I also tested whether emotions varied along continuous gradients, possibly encoded by the factors that I identified.

Color coding using the "valence" factor revealed a continuous global gradient linking positive emotions on the left to negative emotions on the right (Fig.3.5 b), in line with the dimensional view of emotions [10]. I also color-coded for "arousal" and "generalizability", and found that emotions varied along those two dimensions smoothly as well, but not in a single global direction as "valence". Extremely positive or negative emotions are more intense and arousing compared to neutral ones (Fig.3.5 c). Roughly, I found the opposite pattern for "generalizability": extremely positive or negative emotions do not generalize well (Fig.3.5 d).

Clustering

As already noted, UMAP has revealed continuous gradients, most notably the "valence" factor, in the dimensional space of emotions evoked by stories. Here, I applied clustering analysis to probe whether emotions form discrete clusters in the high dimensional space with clear boundaries.

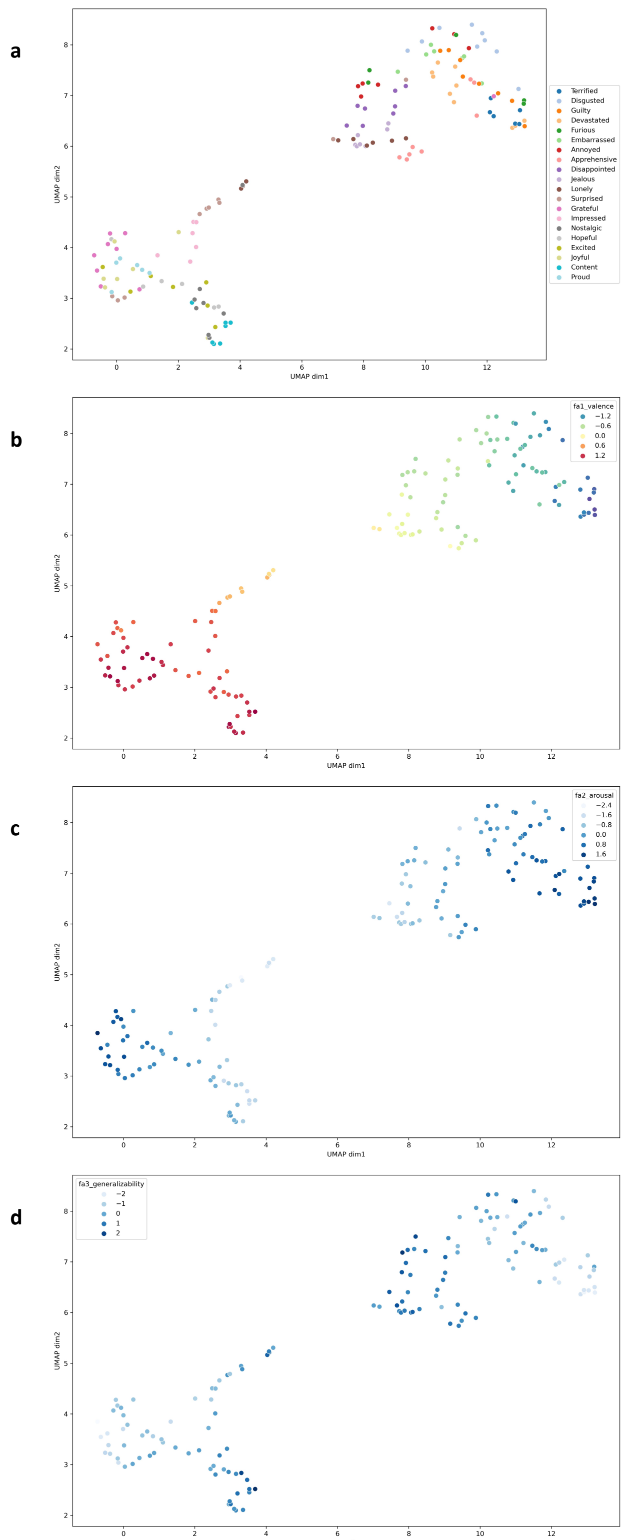

Figure 3.5: UMAP plots color-coded for (a) intended categories, (b) the "valence" factor, (c) the "arousal" factor, and (d) the "generalizability" factor.

Recovering the 20 intended categories

The stories that I used consisted of 2 to 3 sentences with an average length of 50 words, carefully engineered to elicit 20 different emotion categories. So the first question I asked was whether I can re-discover the same categories as clusters if I set the number of clusters to be 20.

I therefore applied the K-means clustering algorithm to derive 20 clusters and assessed the agreement between my results and the intended categories. Overall, I found a low level of agreement between the two with an adjusted rand score of 0.244, and an adjusted mutual info score of 0.395. The contingency matrix (Fig.3.6) showed the intersection of every intended/predicted cluster pair.

The clustering result largely confirmed the intuition I got from the UMAP result. Emotions of the same intended category shared some level of similarity, but they were not necessarily more similar than emotions of different intended categories. This varied across categories. The categories with all instances located in close proximity in UMAP, for instance, apprehensive and content, spread less across predicted clusters. On the contrary, instances of surprise located far from each other in UMAP, were spread out more across predicted clusters.

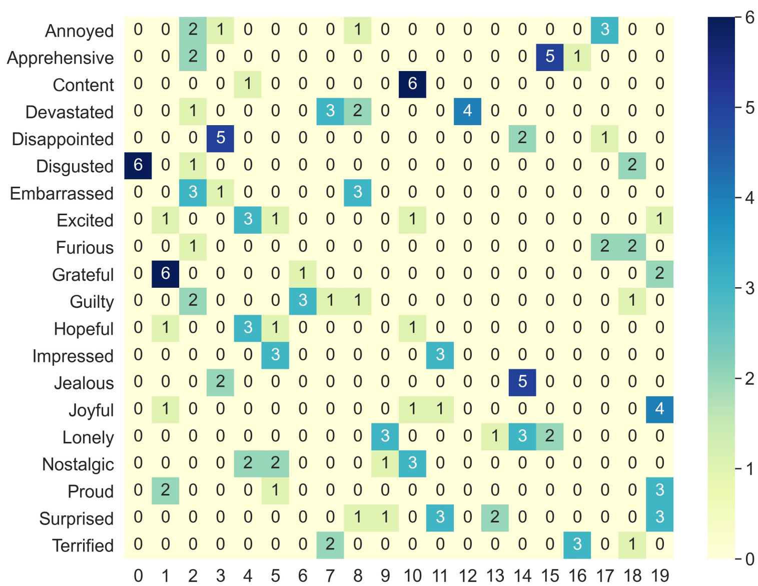

Figure 3.6: Contingency matrix between the 20 discovered categories (columns) and the ones intended (rows).

Data driven clustering

The fact that I failed to recover the 20 intended categories suggested that at least the

20 category structure was not valid given my data. However, this didn't rule out the possibility of emotions forming clusters in the high dimensional space. So I tried to determine the optimal number of clusters in a data-driven manner, ignoring the intended labels.

I tried both centroid-based and density-based algorithms, hoping to find converging evidence.

First, I used K-means, which is a centroid-based clustering algorithm that requires specification for the number of clusters to extract. I tried out a range of possible number of clusters to extract, the evaluation metrics (see details in 2.2.5) suggested that the optimal number of clusters to extract should be 2 (Fig.S3.4). Mean-shift, an algorithm that automatically determines the number of clusters, also suggested 2 clusters. DBSCAN, which detects the boundary of clusters by a density drop in the high dimensional space, also found 2 clusters. I chose the hyperparameters for DBSCAN based on prior knowledge. Specifically, the number of samples in a neighborhood for a point to be considered as a core point was set to 5 since there were approximately 5 instances per intended emotion categories, and subsequently eps which represents the maximum distance between two samples for one to be considered as being in the neighborhood of the other was determined to be 4 (Fig.S3.5).

Three different clustering algorithms with different assumptions all suggested a two-cluster structure in my data. Visualizing the solutions along the two UMAP dimensions (Fig.3.7) revealed two clusters of emotions with positive and negative valence respectively, with minor disagreement across models. The clustering analysis thus confirmed my previous observations of the bimodal distributions of the valence-related scales and the UMAP result.

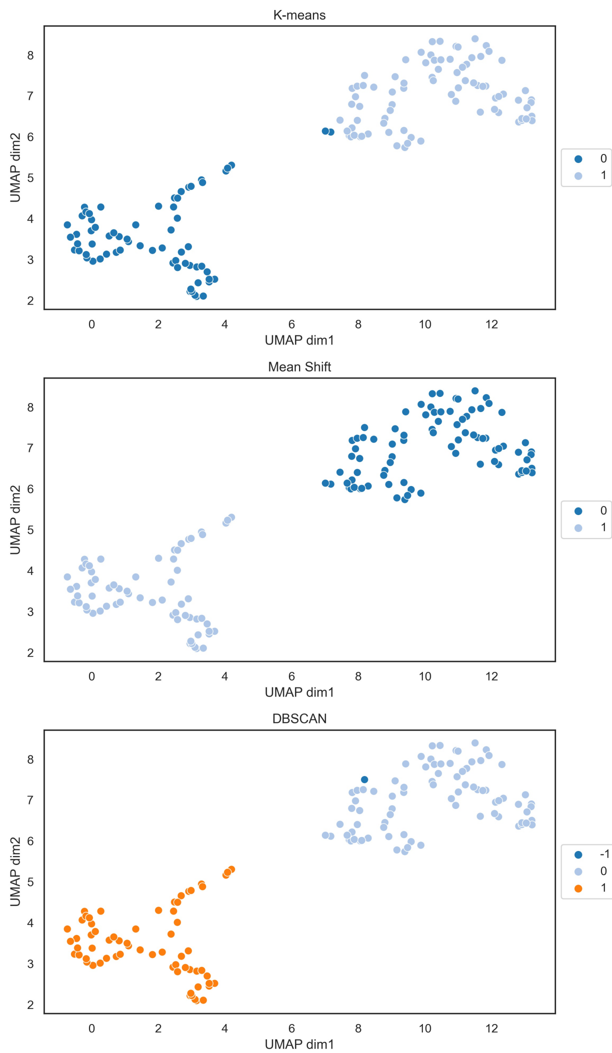

Figure 3.7: Visualization of two-cluster structure as determined by different algorithms; points were color coded for cluster membership (for DBSCAN, points labeled as -1 were outliers). Location was based on UMAP coordinates.

So far, I have established that a two-cluster structure was most appropriate for a single partitioning of the data. In addition, I probed an extensive hierarchy of clusters, using hierarchical agglomerative clustering (HAC) with Euclidean distance and Ward linkage (Fig.3.8). I combined UMAP coordinates and free descriptions of the emotion labels to interpret the meaning of the clusters (Fig.S3.6, Fig.S3.7).

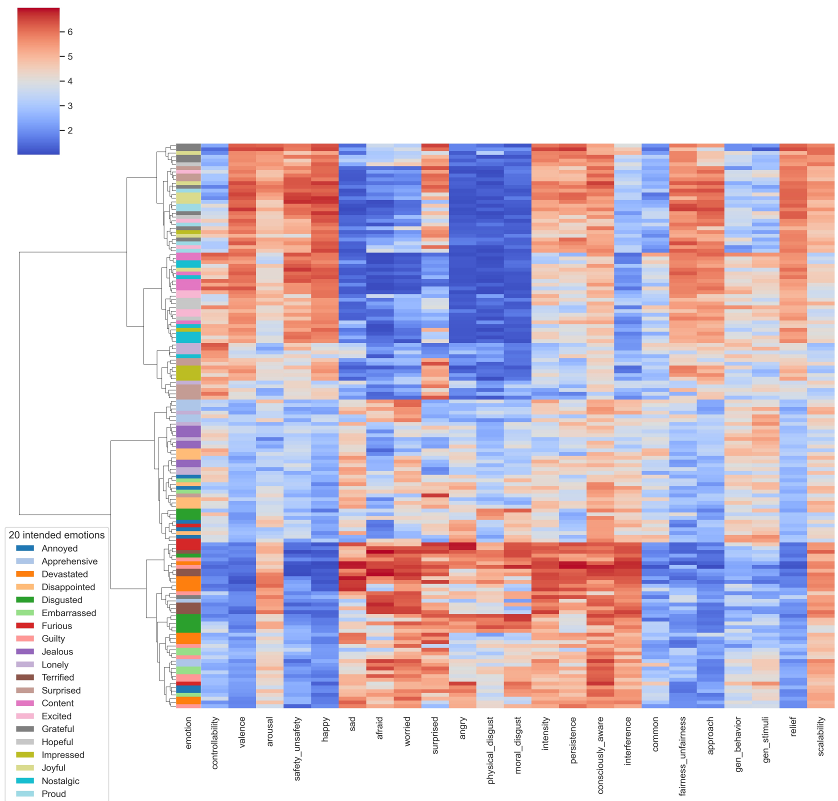

Figure 3.8: Visualization of the hierarchical clustering results where the first column indicates the intended emotion categories and the subsequent columns indicate ratings on the 23 scales (color indicates rating magnitude: blue for lower ratings and red for higher ratings). Each row represents one emotion evoked by one story, 150 rows in total.

The two clusters corresponding to the positive and negative emotions that I identified before are at the top of the hierarchy. The freely-generated words that subjects used the most to label the negative emotions were scared, sad, fear, nervous, and annoyed while words for positive emotions included happy, excited, grateful, content, and proud.

Further down the dendrogram, four clusters emerge which can be interpreted as strong positive emotions (with words like proud, happy, and excited), weak positive emotions (with words like content and happy), weak negative emotions (with words like annoyed, sad and jealous), and strong negative emotions (with words like scared, sad and fear) from top to bottom respectively. The basis of partitioning at these two levels corresponds to the "valence" and "arousal" factors that I have identified, again justifying the dimensional account of emotions.

In principle, the hierarchy of clusters can be probed at all possible levels, with the bottom level of having a single instance as its own cluster. However, it should be noted that each solution represents a different level of fit and having more clusters doesn't necessarily imply better interpretation. For instance, for the six cluster solution, the strong negative emotion cluster is further separated into two clusters but the word descriptions didn't reveal a clear conceptual difference.

Inspecting the row labels which encodes the intended categories, the result is again somewhat mixed. Intra-category similarity does exist as indicated by some sequential rows of the same color, but I also observed a mixing of colors indicating inter-category similarity.

3.3 Summary and discussion

Using 150 well validated stories that are rich in content, I characterized the evoked emotions in a high dimensional space using 23 affective scales. Exploratory factor analysis was performed to extract three robust factors that captured the majority of the variance. The dimensions were interpreted as "valence" (pleasantness-unpleasantness), "arousal" (intensity), and "generalizability" (extent to which the experience applied commonly across many situations, or was specific to only one or a few). The first two factors that I identified are in line with the literature (constituting core affect) while the third factor emerges given my novel scales. I further characterized the distribution of emotions within this three-dimensional space, and found that emotions evoked by stories varied along continuous gradients (most notably, the "valence" dimension). I didn't find evidence for well-separated clusters in the dimensional space, contrary to theories postulating discrete emotion categories.

My study of emotions evoked by stories has several limitations that I discuss below.

First, the stimulus set can be further improved. On average, the stories used 2-3 sentences to describe a scenario, which is probably shorter than lexical stimuli of emotional relevance commonly encountered in real life. Longer stimuli such as novels would probably be better at eliciting strong emotions. However, the question of how to balance the effectiveness and the quantity of stimuli is a difficult one. The content of the stories can be improved as well. For instance, a story described holocaust liberation which is uncommon and probably difficult to imagine. The similarity of content for stories belonging to the same intended categories will likely affect the conclusion of whether emotions formed well-separated clusters. For instance, if all sad stories feature death, then the evoked emotions would be more unified and likely to form a closer cluster than if the stories feature different topics.

Second, my study collected behavioral ratings only and therefore the question of to what extent the stories actually evoked emotions remains unknown. The comparison across domains allowed me to address this question to some extent, but a more direct approach would be to collect physiological measures in addition to behavioral ratings.

Besides the limitation with the stories themselves, I was also limited by the sparseness of my data. Ideally, I would like to collect ratings on all of my scales for all of the stimuli from each subject, which would require a substantial amount of testing time (around 20 hours per subject for stories alone). Because of that, the analyses in this chapter were based on aggregate ratings across subjects, and I can't address whether the conclusions hold for data from a single subject. Relatedly, it limits my ability to investigate individual differences, for example, it's impossible to construct a correlation matrix across scales for each subject.

Finally, it is worth noting that the lexical character of the story stimuli, as well as of the rating scales, limits the conclusions about emotion experience that can be drawn, as I already noted in the general introduction. I am limiting myself to a subset of the words/narratives in English that convey emotion concepts. Other languages and cultures would no doubt offer additional, and different, words and concepts, and even English provides a plethora of emotion words that go unexplored here. On the other hand, it seems likely that much of the variability, at least in English emotion words, can be captured with relatively few words, since many are closely related and are synonyms or antonyms. Nonetheless, given the limitations of lexical stimuli, a larger set of richer and more diverse stimuli that do not require language would be an important comparison. I thus next turn to the elicitation of emotions from videos, in the next chapter.

3.4 Supplementary information

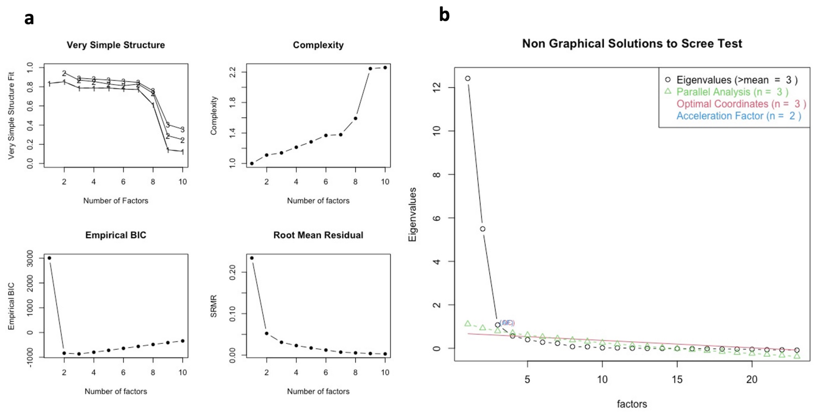

Figure S3.1: Results for the various statistical methods. (a) Very Simple Structure and Empirical BIC (Velicer's MAP is not plotted), (b) Parallel analysis, the acceleration factor and the optimal coordinate.

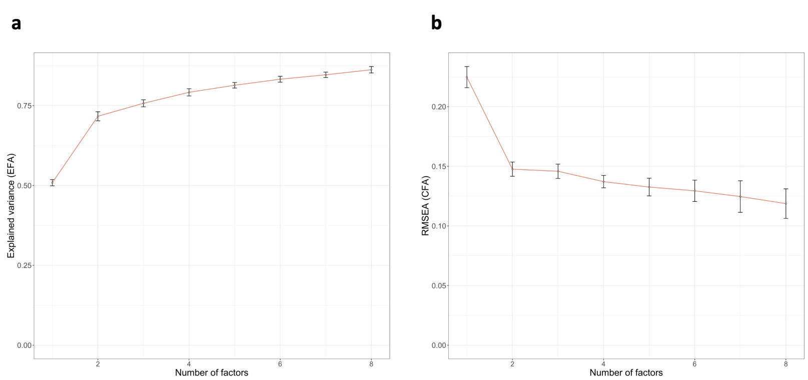

Figure S3.2: Results for the cross validation procedure. The means (points) and standard deviations (error bars, n = 20 iterations) of (a) explained variance from the EFA on training data and (b) root mean square error of approximation (RMSEA) fit index from the CFA on testing data.

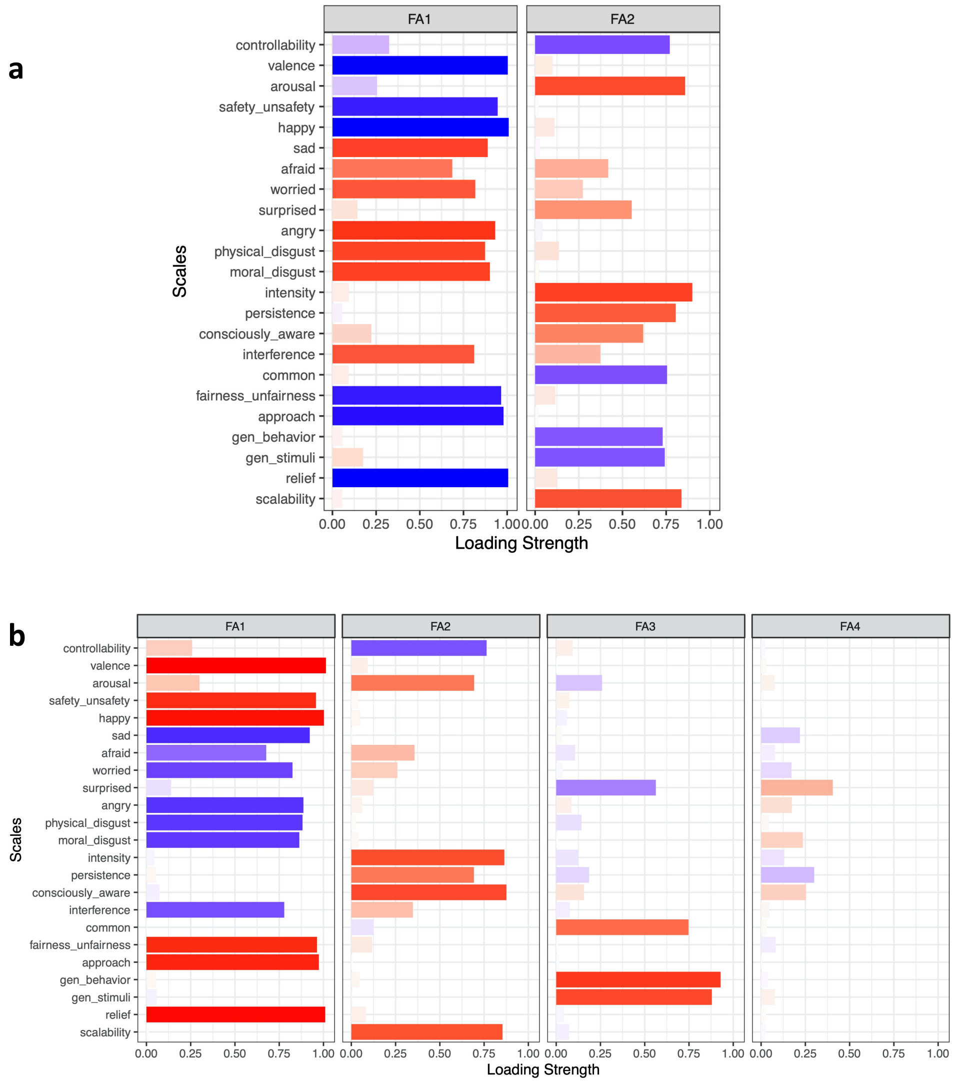

Figure S3.3: Factor loadings of scales on (a) the two factors from EFA and (b) the four factors from EFA. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

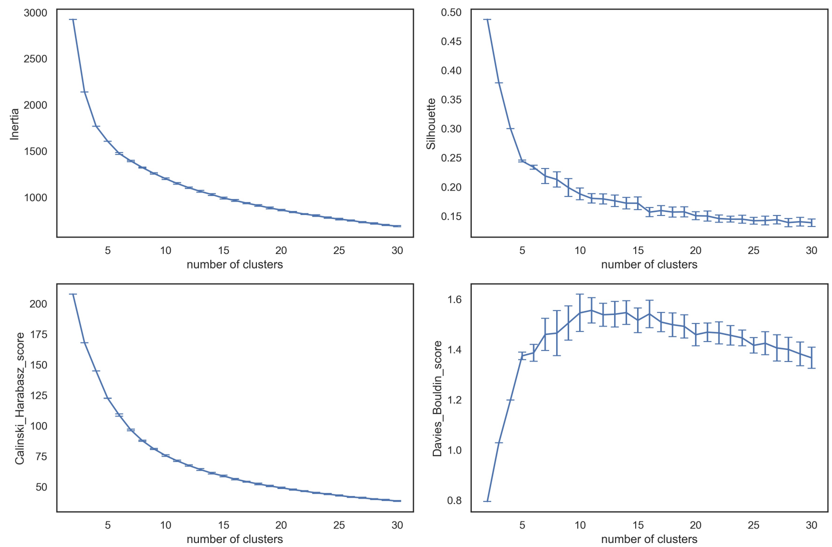

Figure S3.4: Determining the number of clusters for K-means. The means (points) and standard deviations (error bars, n = 20 iterations) of inertia, the Silhouette coefficient, the Calinski-Harabasz index, and the Davies-Bouldin index for K-means results with different number of clusters (2 to 30).

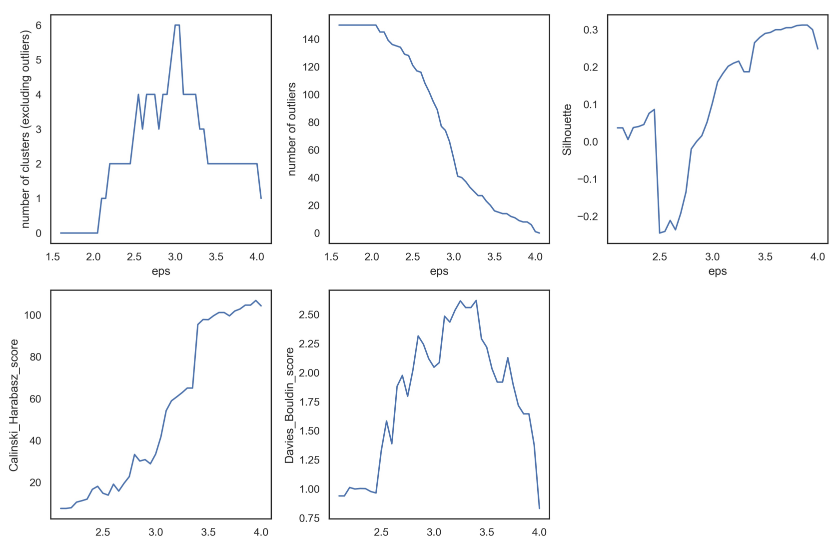

Figure S3.5: Determining hyperparameters for DBSCAN. The number of clusters, the number of outliers, the Silhouette coefficient, the Calinski-Harabasz index, and the Davies-Bouldin index for DBSCAN with different values of eps.

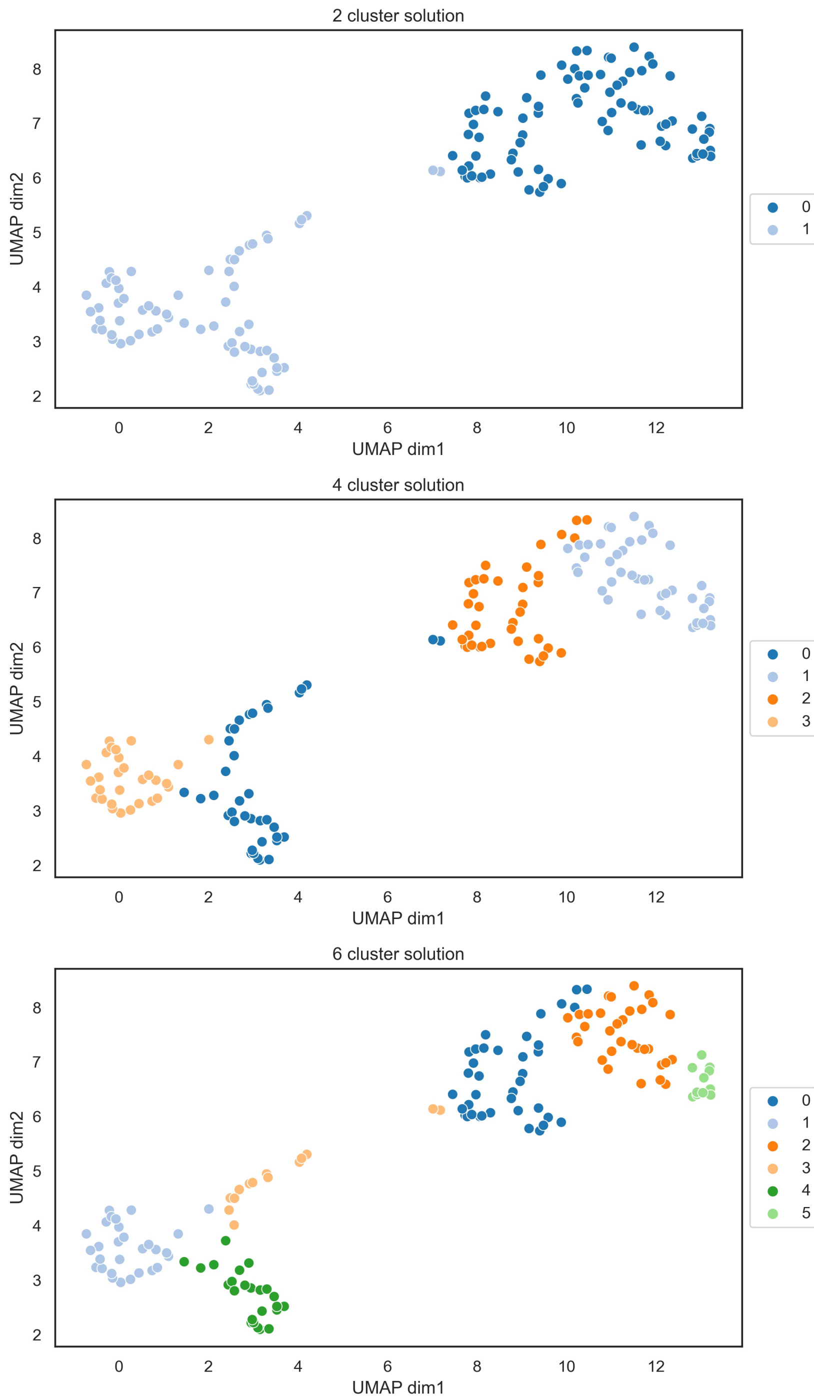

Figure S3.6: Visualization of the different cluster solutions (2, 4, 6 clusters) as determined by HAC. Points were color coded for cluster membership and location was based on UMAP coordinates.

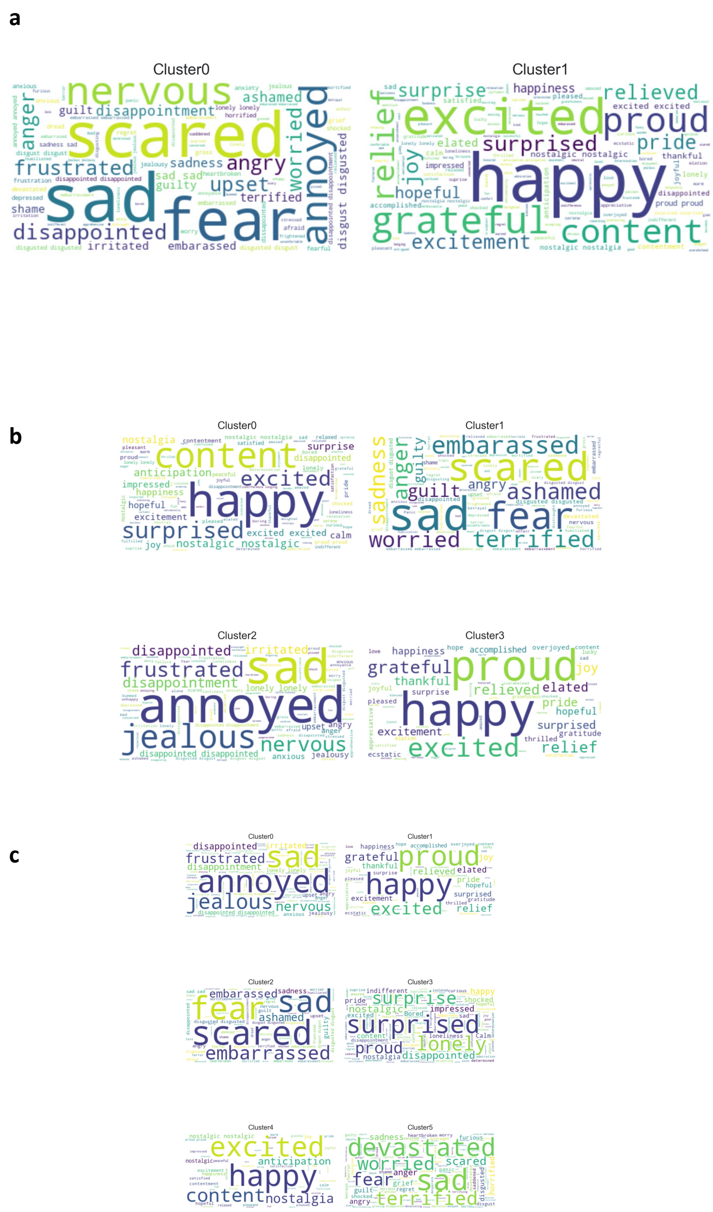

Figure S3.7: Word Clouds of the free descriptions for each cluster of the different cluster solutions as determined by HAC: (a) 2 clusters, (b) 4 clusters, and (c) 6 clusters.

References

[1] Johnny RJ Fontaine et al. "The world of emotions is not two-dimensional". In: *Psychological science* 18.12 (2007), pp. 1050–1057.

[2] Lauri Nummenmaa et al. "Maps of subjective feelings". In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9198–9203.

[3] Michael Argyle and Jill Crossland. "The dimensions of positive emotions". In: *British Journal of Social Psychology* 26.2 (1987), pp. 127–137.

[4] Amy E Skerry and Rebecca Saxe. "Neural representations of emotion are organized around abstract event features". In: *Current biology* 25.15 (2015), pp. 1945–1954.

[5] CHARLOTTE VanOYEN WITVLIET and Scott R Vrana. "Psychophysiological responses as indices of affective dimensions". In: *Psychophysiology* 32.5 (1995), pp. 436–443.

[6] Antonio R Damasio et al. "Subcortical and cortical brain activity during the feeling of self-generated emotions". In: *Nature neuroscience* 3.10 (2000), pp. 1049–1056.

[7] James A Russell. "Pancultural aspects of the human conceptual organization of emotions." In: *Journal of personality and social psychology* 45.6 (1983), p. 1281.

[8] Lisa Feldman Barrett et al. "The experience of emotion". In: *Annu. Rev. Psychol.* 58 (2007), pp. 373–403.

[9] Leland McInnes, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction". In: *arXiv preprint arXiv:1802.03426* (2018).

[10] Eddie Harmon-Jones, Cindy Harmon-Jones, and Elizabeth Summerell. "On the importance of both dimensional and discrete models of emotion". In: *Behavioral sciences* 7.4 (2017), p. 66.

Chapter 4

EMOTIONS EVOKED BY VIDEOS

4.1 Introduction

In the previous chapter, I characterized the dimensional space of emotions evoked by stories. Here, I present results for emotions evoked by short video clips.

Visual stimuli are generally considered better than lexical stimuli at eliciting emotions because of richer contextual information and thus higher ecological validity. Emotional images (including emotional faces) and emotional videos are two types of commonly used visual stimuli. The International affective picture system (IAPS) is perhaps the most widely used image set with 956 images, which includes a wide range of everyday scenes and objects [1, 2, 3]. Main issues with the IAPS include image quality and outdated contexts, and alternative image sets have been developed, for example, the Geneva Affective Picture Database (GAPED) and the Nencki Affective Picture System (NAPS) [4]. But images, being static stimuli, are fundamentally limited in their ability to depict context. In contrast, videos are capable of eliciting emotions dynamically, similar to real life emotional episodes, and are therefore more preferred than images. 16 film clips developed by [5] were widely used to elicit eight discrete emotion categories of amusement, anger, contentment, disgust, fear, neutral, sadness, and surprise. Studies using these film clips claimed to find discrete autonomic representations for each emotion category [6, 7], but it's likely that the conclusions were heavily biased by their choice of such a small set of highly specific stimuli. My goal was to elicit a wide range of emotions, not restricted to certain categories. Ambiguous emotions should not be excluded, in fact, they are needed to properly sample the emotion space. I therefore used videos developed by [8], whose original array included 2185 short video clips (lasting about 5 seconds on average) depicting an exceptionally wide range of emotional situations (the complete list of videos can be viewed at `https://s3-us-west-1.amazonaws.com/emogifs/map.html#`).

Using the 2185 videos to elicit emotions, the authors claimed that 27 dimensions were needed to explain the variance of the emotion experiences. Specifically, for categorical judgements that they collected, subjects were asked to select at least 1, but as many as desired, out of the 34 intended categories for each video. Aggregate ratings across subjects which essentially represented the percentage of selection for each of 34 categories were then used for a split-half canonical correlations analysis, a specific method devised by the authors. In my view, the conclusion is likely an overestimation as the authors performed the dimensionality reduction analysis using the categorical judgements data which biased towards linear independency among categories, instead of the continuous ratings on the affective scales which they also collected.

4.2 Results

4.2.1 Three dimensions underlying emotion experiences evoked by videos

Correlation structure across scales

I derived a pairwise correlation matrix across 23 scales (Fig.4.1), which is a representation for the underlying psychological space for emotion experiences evoked by videos. Similarly as for stories, I observed strong correlations across scales, suggesting that the dimensionality of the space can be further reduced.

For stories, I observed that the scales can be grouped into four groups. In contrast, correlation structure for videos showed clearly that the scales can be partitioned into two groups. The first group at the top includes not only scales that characterize how positive the emotions are, but also those describing how generalized an emotion can be. On the other hand, the other group of scales includes both scales describing how negative the emotions are and scales that characterize the intensity and persistence of emotions. One possible explanation for the clear superordinate structure of videos could be that videos are more diverse and realistic than the stories. It's also possible that videos evoke intense emotions, helping to sharpen the correlation structure.

The specific correlations are qualitatively similar to the story ones. Previously, generalizability-related scales correlated weakly with the scales describing how positive or negative emotions are, this relationship is more clear with the video data. Specifically, positive emotions generalize well while negative ones seem more modular.

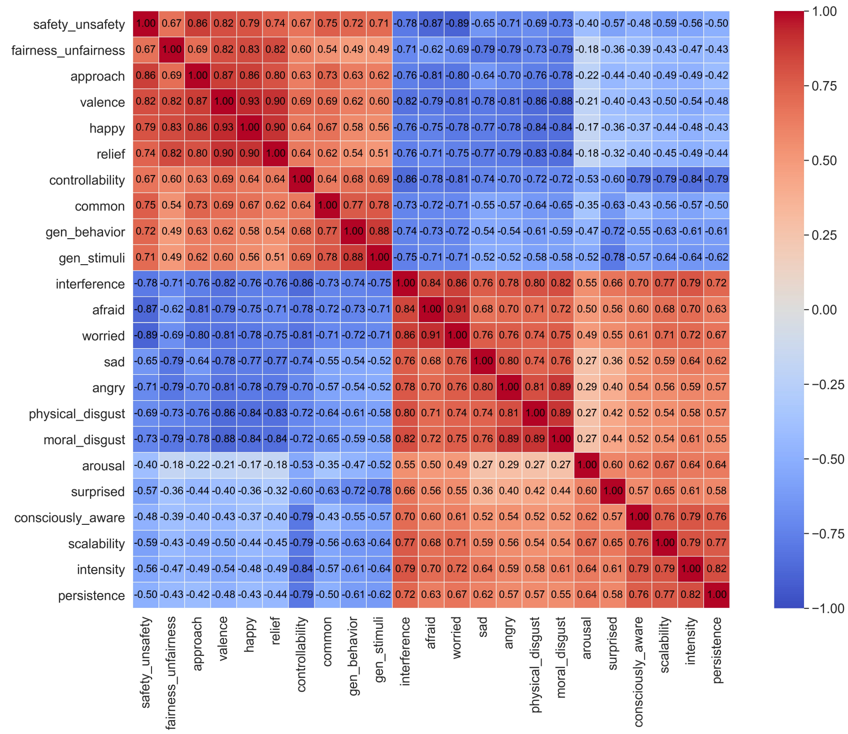

Figure 4.1: Correlation matrix across scales for emotions evoked by videos (sorted using hierarchical clustering to intuitively depict the underlying structure).

Factor analysis

As mentioned above, the correlation matrix across scales suggested that the psychological space of emotion experiences evoked by videos can be characterized using a smaller number of underlying factors, but the number of factors to retain in an exploratory factor analysis needs to be determined first.

Determining the optimal number of factors to extract

I determined the optimal number of factors to retain by taking multiple aspects into consideration: the suggestions from the standard statistical methods, results from the cross validation approach, results from the robustness assessment and the interpretations of the factor loadings, which I describe below.

First, I used the six commonly used statistical tests (see details in 2.2.3). The results (Fig.S4.1) did not converge to a single number. Very Simple Structure, Empirical BIC, and Velicer's MAP suggested 1, 4, and 4 factors respectively. Parallel analysis, the acceleration factor, and the optimal coordinate suggested 6, 1, and 6 factors respectively.

In addition to the standard statistical procedures described above, I also used the cross validation procedure (see details in 2.2.3) to choose how many factors to retain for the video rating data. Visual inspection of the result (Fig.S4.2) is less clear compared to the story rating data, where 2 or 3 or 4 factors all seemed reasonable.

The various methods have suggested multiple values for the number of factors to retain. I therefore applied EFA to further assess the interpretability of different factor solutions. When extracting 2 factors, the factors each explained 44% and 32% of the common variance in the data (76% in total). I interpreted the factors as "valence" and "arousal" (Fig.S4.3 a). When extracting 3 factors, the factors each explained 41%, 24% and 15% of the common variance in the data (81% in total). I interpreted the factors as "valence", "arousal" and "generalizability" (Fig.4.3). When extracting 4 factors, the factors each explained 32%, 20%, 18% and 14% of the common variance in the data (83% in total). I interpreted the factors as "valence", "arousal" , "generalizability" and "safety" (Fig.s4.3 b). I also attempted the 5 factor solution, which cumulatively explained 84% of the total variance representing a marginal improvement and the factors were uninterpretable (Fig.S4.3 c).

Evidence so far suggested both the 3 and 4 factor solutions seemed reasonable. I don't argue against the 2 factor solution, but just decided to retain more factors for completeness. I further assessed the robustness of both the 3 and 4 factor solutions with regard to the number of stimuli and number of scales.

First, I systematically reduced the number of videos starting from the whole set of 998 videos. At each step, I used the new aggregated ratings for EFA and calculated Tucker indices of factor congruence between factors derived using the reduced set and the original set of stimuli (with orthogonal Procrustes rotation). Both the 3 factor and 4 factor solutions were robust to the number of videos (Fig.4.2 a,b). For the 3 factor solution, all mean factor congruences were higher than 0.9 with no fewer than 23 videos (roughly 2.3 % of the whole set). For the 4 factor solution, all mean factor congruences were higher than 0.9 with no fewer than 73 videos (roughly 7.3 % of the whole set).

Second, I removed scales one by one (same order of removal as I did with stories based on global redundancy, as explained in 2.2.3) and quantified the relatedness of the original factors from the full set and the ones from the subset by correlating the factor scores (Fig.4.2 c,d). The 3 factor solution was robust to the number of scales as the factors derived from the full set versus the subsets of scales were highly correlated. The 4 factor solution didn't behave as well, most notably the 4th factor, "safety" was unstable. All evidence taken together, I decided to retain 3 factors which are highly interpretable and robust.

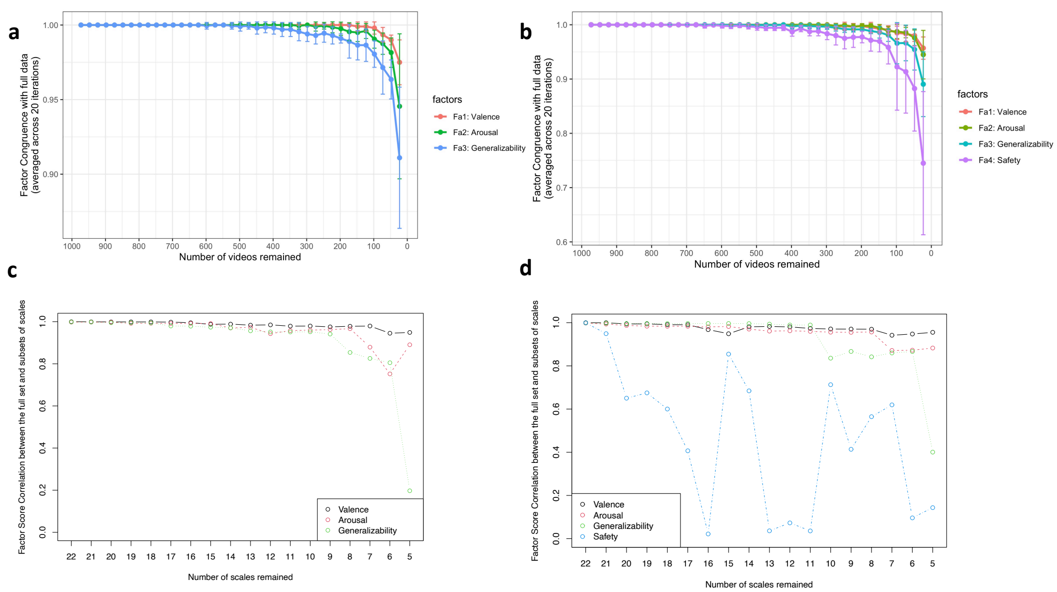

Figure 4.2: Robustness of factor solutions with respect to the number of videos and number of scales. (a,b) Points indicate the means and error bars indicate standard deviations of Tucker indices of factor congruence (with orthogonal Procrustes rotation) between the full set versus subsets of videos across 20 iterations, and are color-coded for different factors for (a) the 3 factor solution and (b) the 4 factor solution. (c, d) Pearson's correlations between factor scores from the full set versus subsets of scales, color-coded for different factors for (c) the 3 factor solution and (d) the 4 factor solution.

Interpretation of the three factors

Exploratory factor analysis was then performed to extract 3 factors using the minimal residual method, and the solutions were rotated with oblimin for interpretability. The Tenberge method was used to obtain oblique factors scores (all using the "'fa" function in the "psych" package in R). I interpreted the three factors as "valence", "arousal", and "generalizability" (see Fig.4.3 for factor loadings).

The factor loadings again show that scales which describe how positive and negative emotions load strongly onto the "valence" factor. The group of scales that characterize the intensity and persistence of emotions load strongly onto the "arousal" factor. Lastly, three scales ("common", "generalizability over stimuli", and "generalizability over behavior") load strongly onto the "generalizability" factor. Overall, the loading pattern closely resembles that of the stories.

Examining the factor scores for individual videos also allowed me to verify my interpretation of the scales. For instance, some of the most positive emotions involved watching cute animals while some of the most negative emotions involved watching horrifying scenes. Intense emotions were evoked by both positive and negative videos, watching family reunions brought intense joy while watching tragic scenes evoked intense horror and shock. Watching mellow scenes that happened in everyday life evoked most generalized emotions while extremely negative and intense emotions were the least generalized.

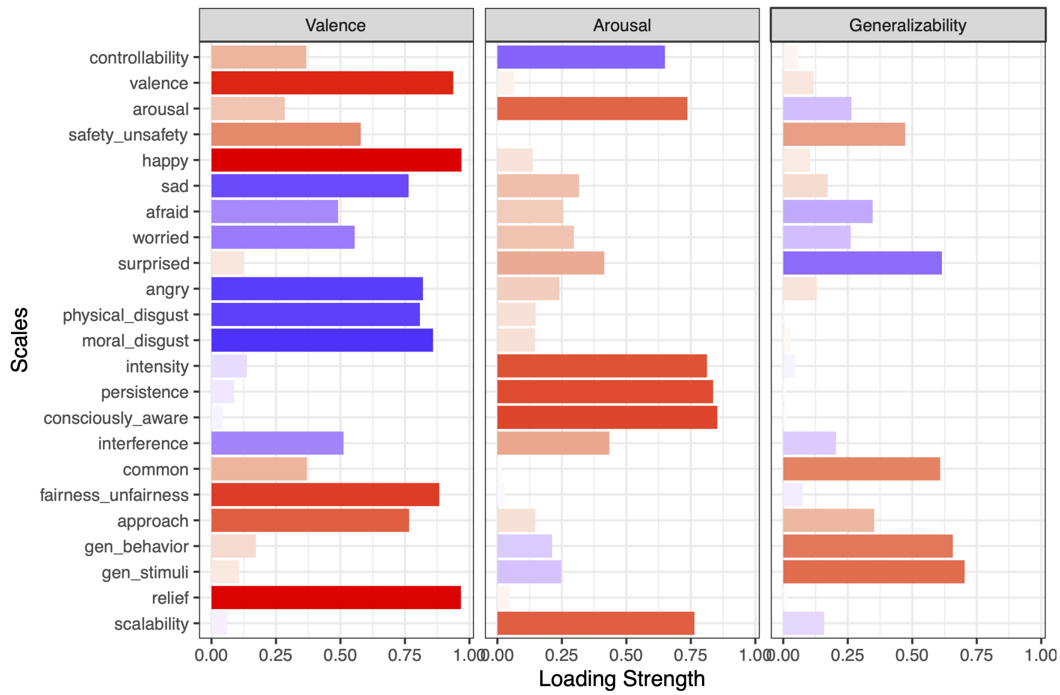

Figure 4.3: Factor loadings of scales on the three factors from EFA. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

4.2.2 Distribution of emotion experiences evoked by videos

The videos that I used are published in [8] (the complete list of videos can be viewed at `https://s3-us-west-1.amazonaws.com/emogifs/map.html#`). The authors intended to compile a list of videos targeting 34 different emotions by coming up with keywords for each emotion category and querying the search engines and other websites. Therefore the authors had relatively less control over the videos, in contrast to the stories which were completely made up. Still, they were meant to be relatively good examples of the intended emotion categories.

Part of the data collected in the original study was categorical judgment where subjects were required to select at least one category but could select as many as desired among the 34 options for each video. And averaging those judgements across subjects would result in percentages of selection for all the 34 categories for each video. I simplified the data by assigning a single dominant emotion out of the 34 categories with the highest percentage of selection for each video. Four categories (contempt, disappointment, envy, and guilt) were never selected as dominant emotions for any of the videos, indicating that either they were difficult to elicit or the videos intended to evoke those emotions were poorly chosen. It's also worth noting that the emotions evoked by videos were not as carefully balanced as in the case of stories. In particular, there's an over-representation of amusement (selected as the dominant emotion for 200 out of 998 videos) while certain emotions were severely under-represented (such as pride, selected for 2 out of 998 videos).

Looking at the distributions, unlike stories, the rating distributions for most scales resemble a normal distribution, except for those describing basic emotion ones (Fig.4.4).

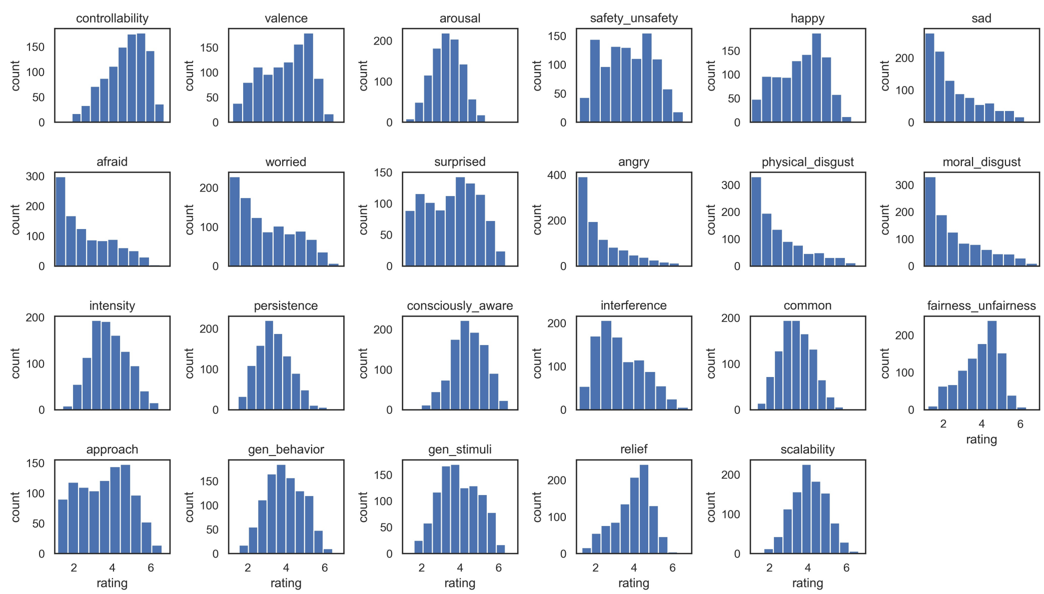

Figure 4.4: Distributions of aggregated ratings (across subjects) on the 23 scales for emotions evoked by 998 videos.

Umap

Unlike stories where valence-related scales had a bimodal distribution, the univariate distributions of videos didn't seem to suggest a good separation of positive and negative emotions. Similarly as I did before, I applied UMAP to better visualize the distributions in the embedded two dimensional space. Since three factors accounted for most of the total variance in the video data, representing the emotions using a two-dimensional UMAP plot should be a reasonable approximation. I further combined the dominant emotion category labels and the factor scores with the UMAP plots, trying to address the key question of whether emotion experiences evoked by videos are discrete or dimensional.

Color code UMAP by intended categories (30 dominant categories)

Unlike stories where the two-cluster structure was prominent even without categorical labels, the emotions evoked by videos formed just one cluster in the UMAP space (Fig.4.5 a).

With the labeling for the dominant emotion categories, for some categories, instances belonging to the same categories were located relatively closer to one another (such as disgust at the top left corner) compared to other categories where instances were way more scattered (such as amusement).

To some extent, this is expected because the videos were quite short in general, lasting 5 seconds on average, and the quality varied with some being very brief and blurred and thus of low quality. As mentioned, the videos were compiled from online resources, so presumably the availability was also a limiting factor for finding good videos. These caveats limit the effectiveness and specificity of the videos at evoking the intended emotions, which would naturally result in a more scattered distribution because the evoked emotions would be more ambiguous than desired.

However, it should be noted that the emotions evoked by videos were more diverse than the emotions evoked by stories because of the sheer number of stimuli. And therefore, it is also possible that the continuous distribution was driven primarily by the diversity of the emotion experiences.

Color code UMAP by factor scores

Color coding using the factors revealed qualitatively similar patterns as the story data. "Valence" again emerged as a continuous global gradient linking negative emotions on the left to positive emotions on the right (Fig.4.5 b), in line with the dimensional view of emotions. Emotions varied along the "arousal" and "generalizability" dimensions smoothly as well, but not in a single global direction as "valence". Extremely positive or negative emotions are more intense compared to neutral ones (Fig.4.5 c). Positive emotions generalize better than negative ones in general. (Fig.4.5 d).

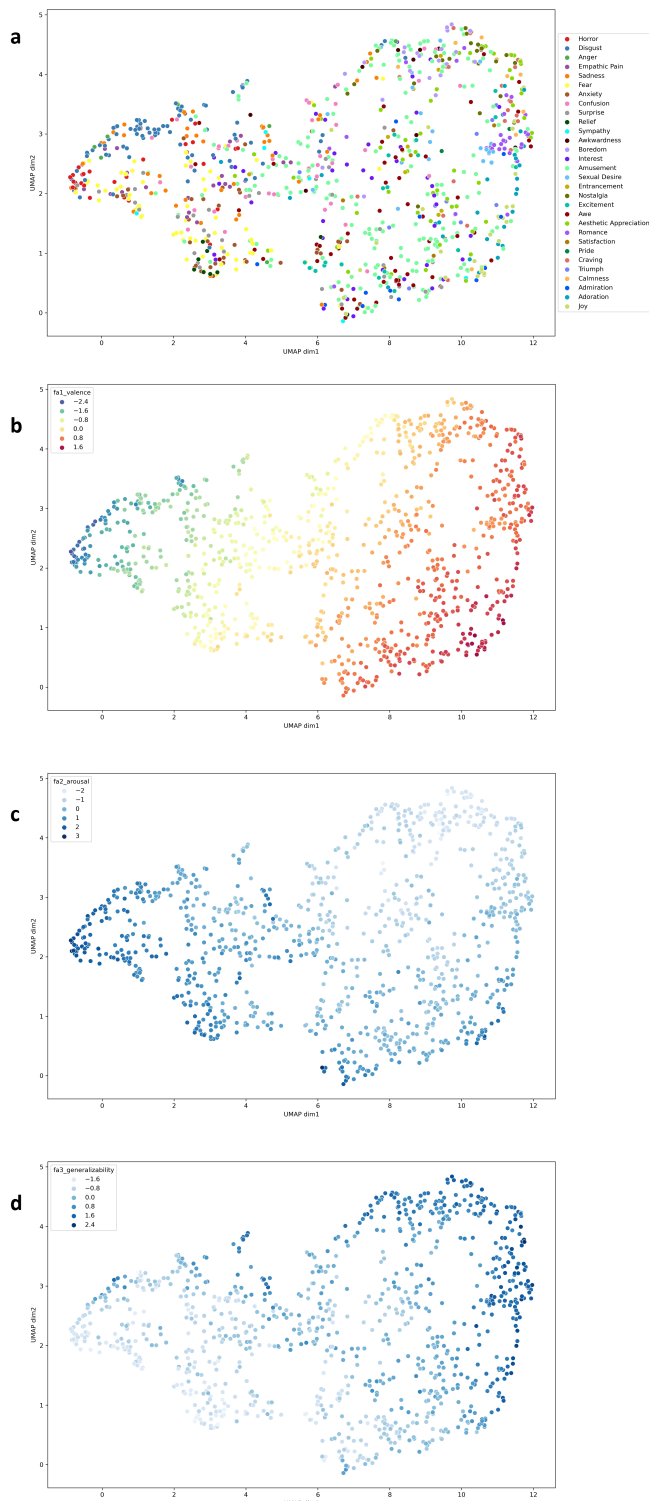

Figure 4.5: UMAP plots color-coded for (a) dominant emotion categories, (b) the "valence" factor, (c) the "arousal" factor, and (d) the "generalizability" factor.

Clustering

As already noted, UMAP has revealed continuous gradients, most notably the "valence" factor, in the dimensional space of emotions evoked by videos. Here, I applied clustering analysis to probe whether emotions form discrete clusters in the high-dimensional space with clear boundaries.

Recovering the 30 intended categories

So the first question I asked was whether I can re-discover the same categories as clusters if I set the number of clusters to be 30.

I therefore applied the K-means clustering algorithm to extract 30 clusters and assessed the agreement between my results and the intended categories. Overall, I found a low level of agreement between the two with an adjusted rand score of 0.08, and an adjusted mutual info score of 0.254. The contingency matrix (Fig.4.6) showed the intersection of every intended/predicted cluster pair.

The clustering result was in line with the intuition that I got from the UMAP result. Closer inspection of the contingency matrix revealed that some categories (for instance, craving, sexual desire, romantic love, and nolstalgia) clustered relatively better than other categories. This finding was consistent with what the authors reported in the original paper, albeit using a completely different dataset with different scales.

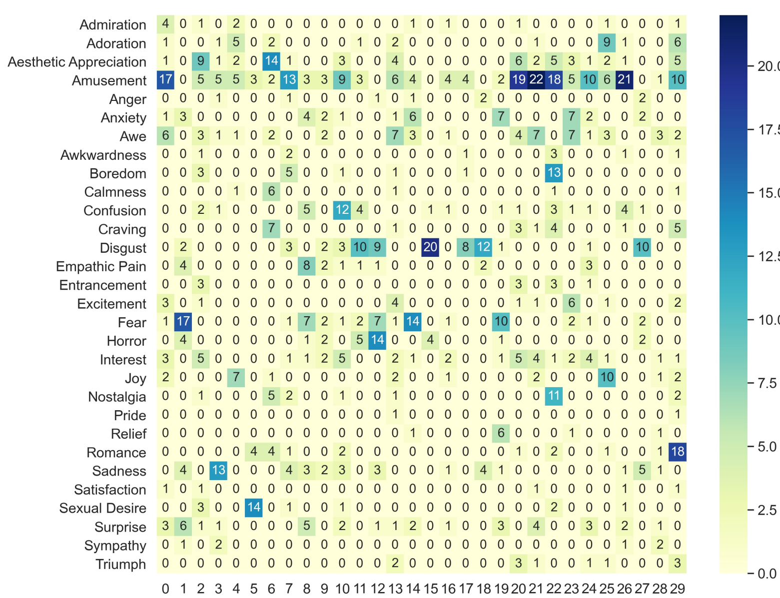

Figure 4.6: Contingency matrix between the 30 discovered categories (columns) and the ones intended (rows).

Data driven clustering

The fact that I failed to recover the 30 intended categories suggested that at least the 30 category structure was not valid given my ratings. However, this didn't rule out the possibility of emotions forming clusters in the high dimensional space, albeit ones that might not correspond well with the 30 originally intended emotion categories. So I tried to determine the optimal number of clusters in a data-driven manner, ignoring the intended labels.

I tried both centroid-based and density-based algorithms, hoping to find converging evidence.

First, I used K-means, and tried out a range of possible number of clusters to extract. The evaluation metrics suggested that the optimal number of clusters to extract should be 2 (Fig.S4.4). it should be noted that this didn't exclude the possibility of one cluster being the optimal choice which wasn't evaluated because most of the evaluation metrics require at least 2 clusters. Mean-shift suggested 1 cluster.

As mentioned, the videos were not balanced across categories so it's not possible to set the min_samples parameter for DBSCAN based on prior knowledge as I did with stories. Therefore, I used heuristics as suggested in literature [9] and set min_samples to be 46 and eps to be 4 which suggested one cluster again (Fig.S4.5).

The clustering analysis has confirmed my previous observations based on the univariate distributions and from the UMAP result that the emotion experiences evoked by videos are best characterized by just one cluster. Still, I think an extensive hierarchy of clusters would be informative. So I applied hierarchical agglomerative clustering (HAC) with Euclidean distance and Ward linkage (Fig.4.7) and combined UMAP coordinates and free descriptions of the emotion labels to interpret the meaning of the clusters (Fig.S4.6, Fig.S4.7).

The two clusters correspond to the positive and negative emotions at the top of the hierarchy. The freely-generated words that subjects used the most to label the negative emotions were scared and fear while words for positive emotions included happy, awe and bored. Further down the dendrogram, four clusters emerge which can be interpreted as weak negative emotions (with words like scared and fear, but also confused and worried), strong negative emotions (with words like scared and fear, but also sad and horrified), less generalized positive emotions (with words like funny, impressed and amazed), and more generalized positive emotions (with words like happy, joy and calm) from top to bottom respectively. The basis of partitioning at these two levels corresponds to the three factors that I have identified, again justifying the dimensional account of emotions.

In principle, the hierarchy of clusters can be probed at all possible levels, with the bottom level of having a single instance as its own cluster. However, it should be noted that each solution represents a different level of fit and having more clusters doesn't necessarily imply better interpretation. For instance, the five or nine cluster solutions were hard to interpret (Fig.S4.6, Fig.S4.7).

Inspecting the row labels which encode the intended categories, again, I observed some level of intra-category similarity but also a mixing of colors indicating inter-category similarity.

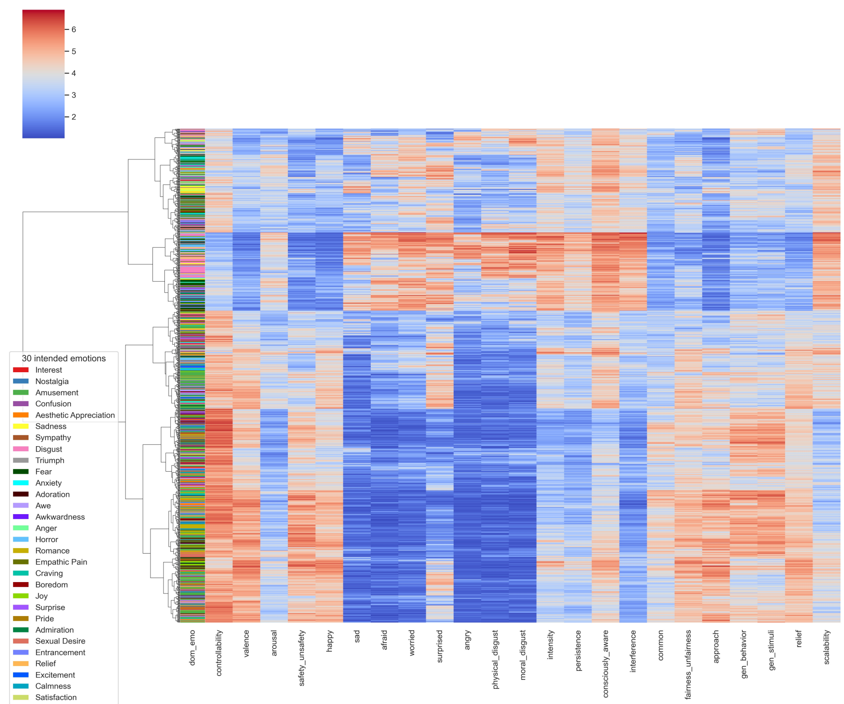

Figure 4.7: Visualization of the hierarchical clustering results where the first column indicates the intended emotion categories and the subsequent columns indicate ratings on the 23 scales (color indicates rating magnitude: blue for lower ratings and red for higher ratings). Each row represents one emotion evoked by one video, 998 rows in total.

4.3 Summary and discussion

Using 998 well validated videos that are rich in content, I characterized the evoked emotions in a high-dimensional space using 23 affective scales. Exploratory factor analysis was performed to extract three robust factors that captured the majority of the variances. The dimensions were interpreted as "valence", "arousal", and "generalizability". The first two factors that I identified are in line with the literature while the third factor emerges given my novel scales. I further characterized the distribution of emotions, and found that emotions evoked by videos varied along continuous gradients (most notably, the “valence” dimension). I didn’t find evidence for well-separated clusters in the dimensional space. The findings are qualitatively the same as the ones discovered using the story data.

My study of emotions evoked by videos has several limitations that I discuss below.

First, the stimuli set can be further improved. The length and quality varied across videos, with some being very short or having low resolution. The specificity of videos was much poorer than stories, indicated by the categorical judgment data from the original paper. Whether the specificity of the videos is an issue or not depends on the research questions. One main issue was that the number of videos were not balanced across intended categories, for instance, there’s an over-representation for videos evoking amusement. Having imbalanced classes is a major issue for analyses such as clustering and classification. The exact content of the videos can also be optimized. For videos evoking negative emotions, for instance, many featured horrifying scenes, which were not common in real life. It also negatively impacted the data collection process because some subjects, after finishing some sessions of the video experiments, indicated that they were unwilling to view such extremely negative videos and would not participate in the later sessions (through messaging on the Prolific platform).

As discussed before, it would be good to collect physiological measures in addition to behavioral ratings to further characterize the emotion experiences. Unlike stories where I am concerned that emotions could be weak, the issue with videos is more related to the possibility of subjects down-regulating their emotions, especially when viewing the negative videos. One future direction is to examine individual differences using the emotion regulation questionnaire that I collected.

The data sparseness issue is the same as I discussed with stories, but it would be even more difficult to resolve with videos as the number of videos is five times that of stories.

4.4 Supplementary information

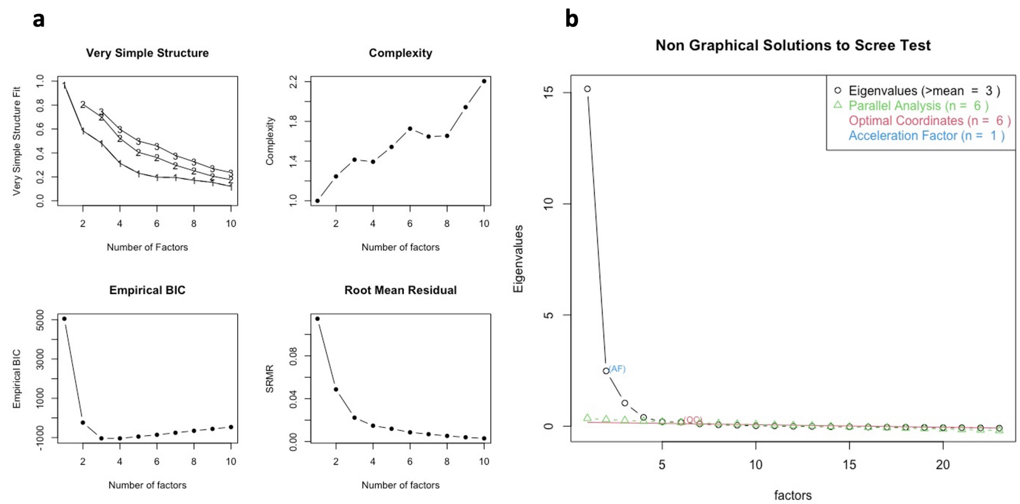

Figure S4.1: Results for the various statistical methods. (a) Very Simple Structure and Empirical BIC (Velicer's MAP is not plotted), (b) Parallel analysis, the acceleration factor and the optimal coordinate.

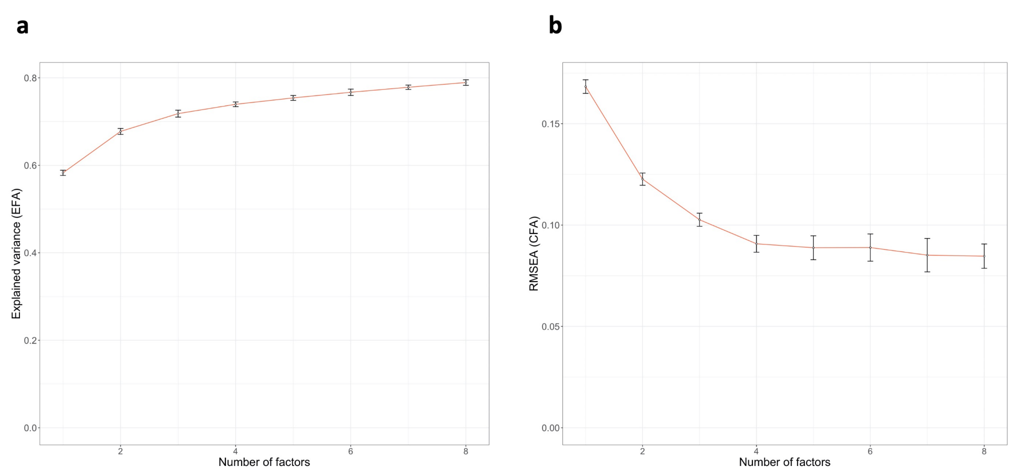

Figure S4.2: Results for the cross validation procedure. The means (points) and standard deviations (error bars, n = 20 iterations) of (a) explained variance from the EFfiguA on training data and (b) root mean square error of approximation (RMSEA) fit index from the CFA on testing data.

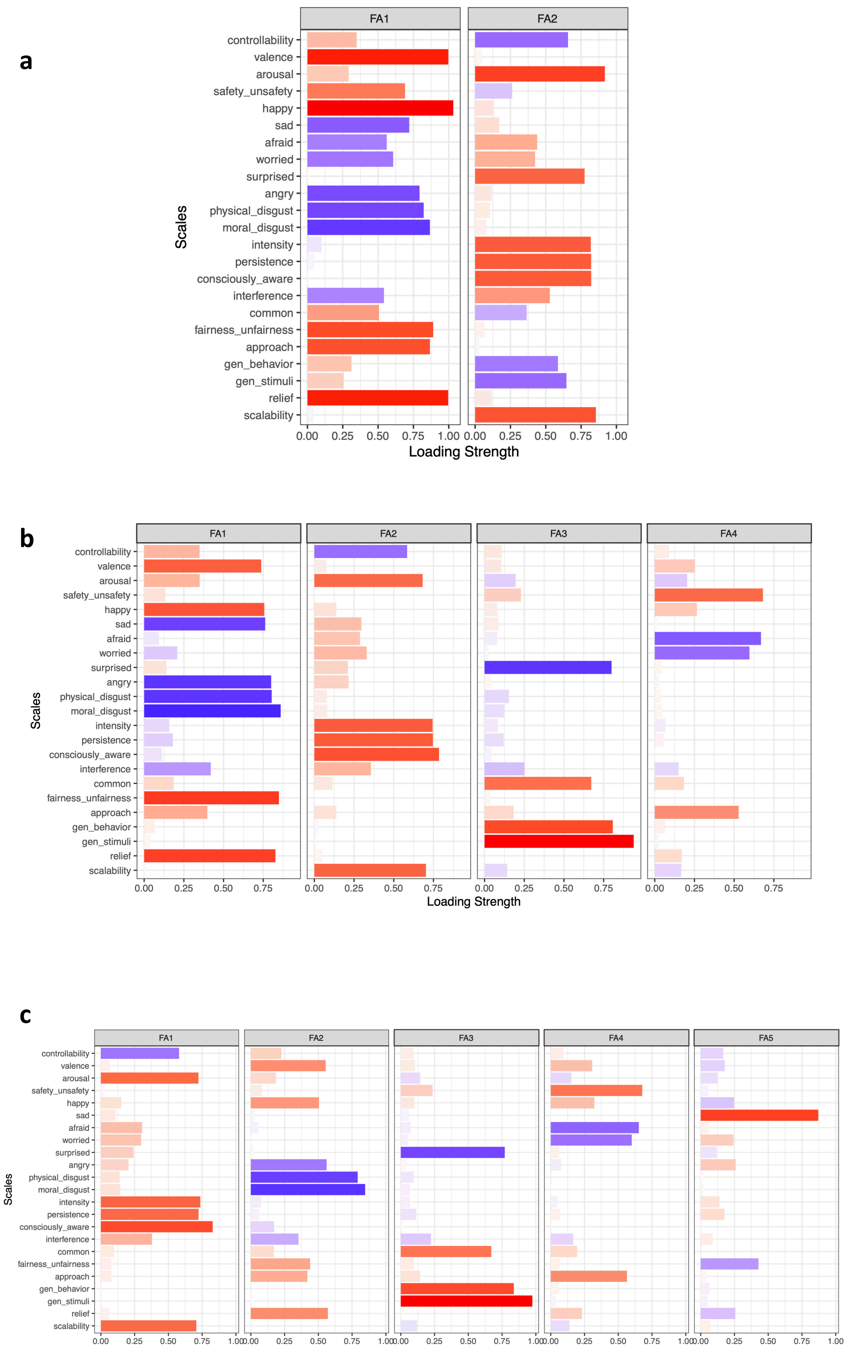

Figure S4.3: Factor loadings of scales on (a) the 2 factors from EFA and (b) the 4 factors from EFA and (c) the 5 factors from EFA. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

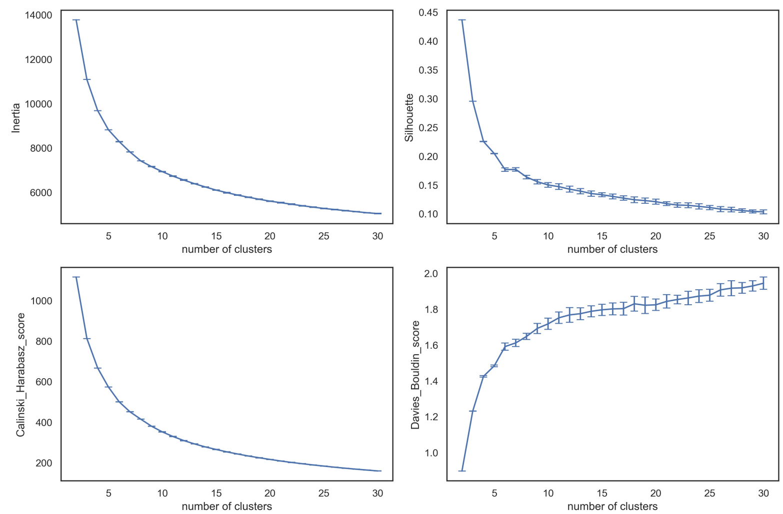

Figure S4.4: Determining the number of clusters for K-means. The means (points) and standard deviations (error bars, n = 20 iterations) of inertia, the Silhouette coefficient, the Calinski-Harabasz index, and the Davies-Bouldin index for K-means results with different number of clusters (2 to 30).

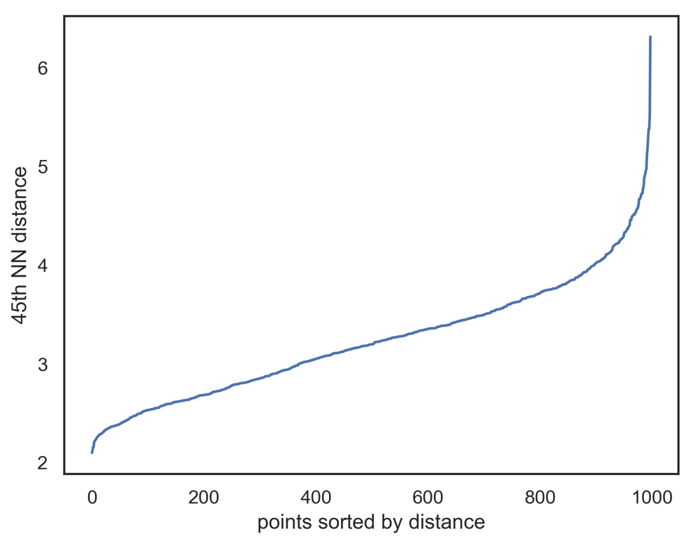

Figure S4.5: The 45th nearest distance plot for DBSCAN.

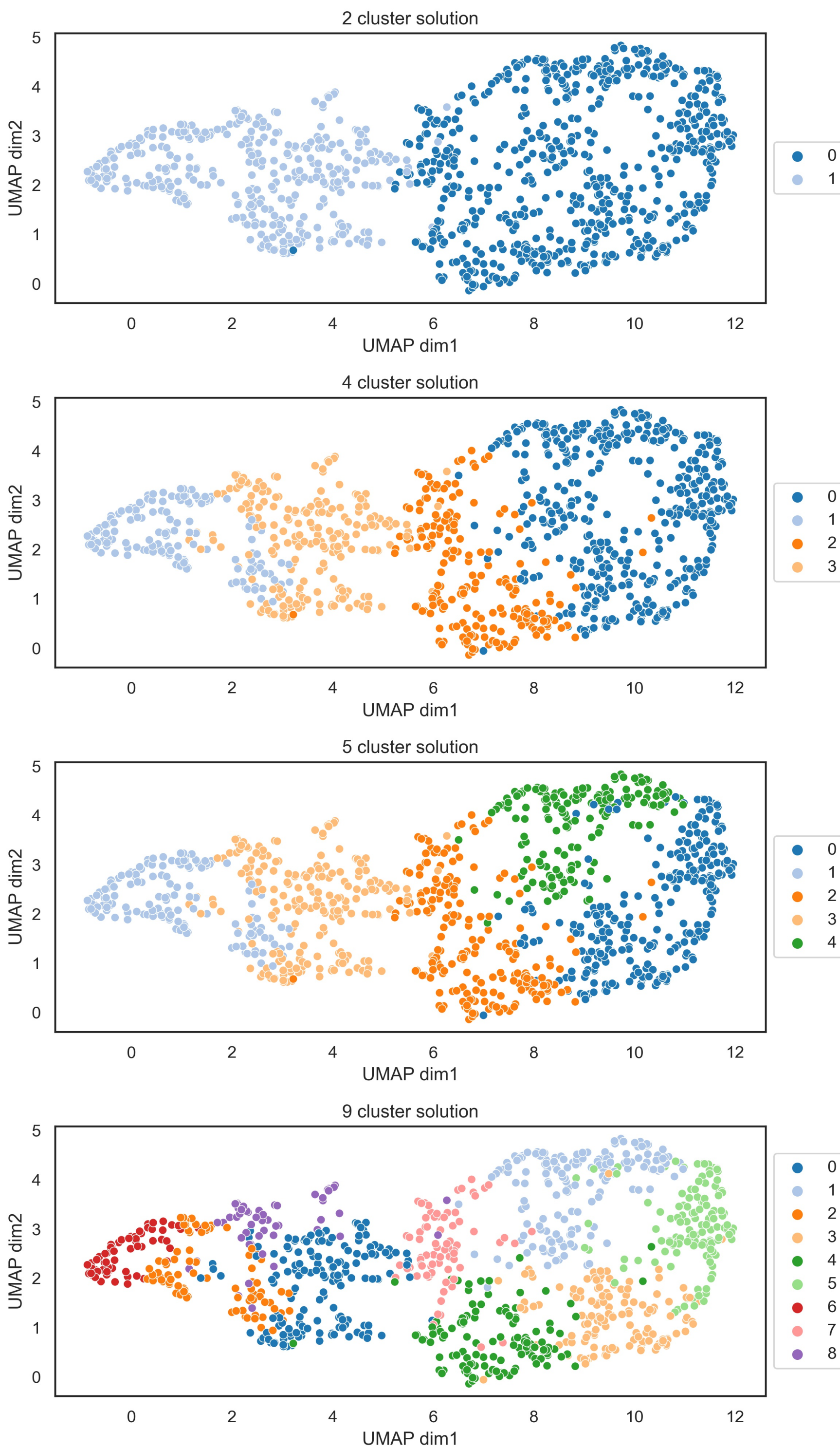

Figure S4.6: Visualization of the different cluster solutions (2, 4, 5, 9 clusters) as determined by HAC. Points were color coded for cluster membership and location was based on UMAP coordinates.

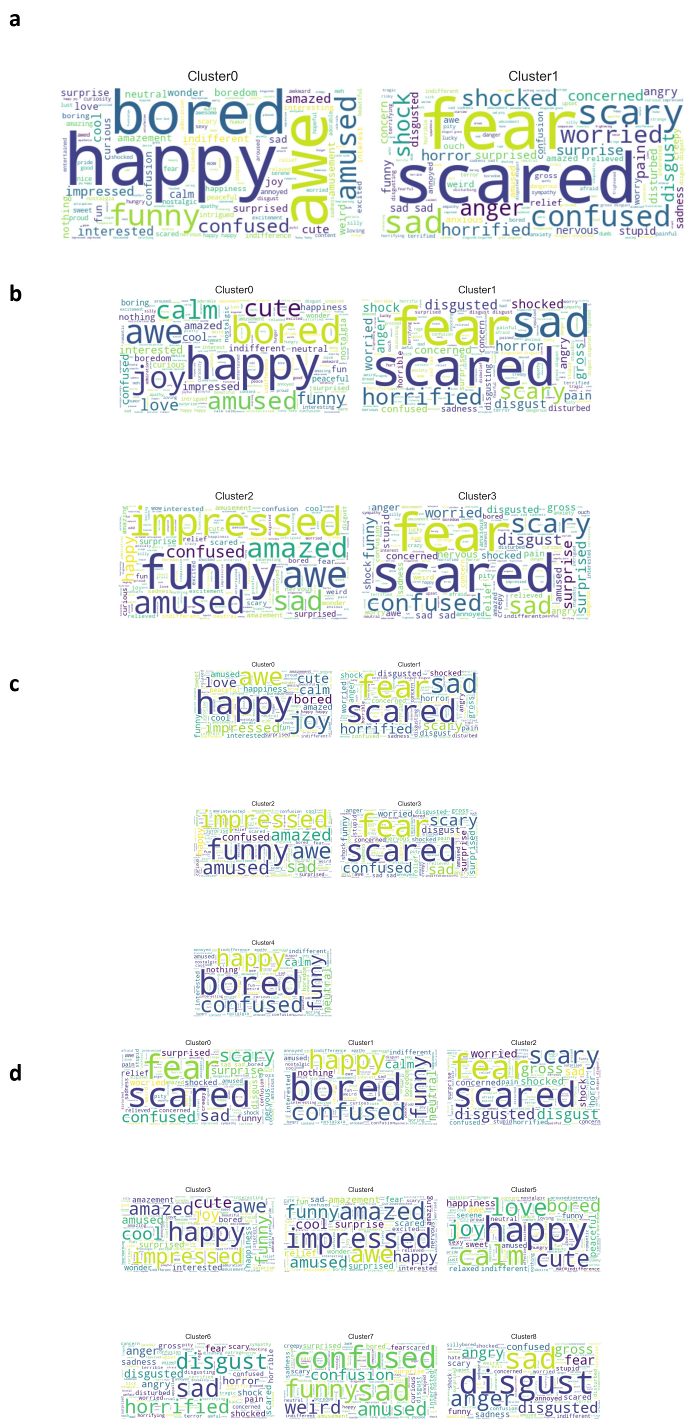

Figure S4.7: Word Clouds of the free descriptions for each cluster of the different cluster solutions as determined by HAC: (a) 2 clusters, (b) 4 clusters, (c) 5 clusters, and (d) 9 clusters.

References

[1] Peter J Lang, MM Bradley, and B Cuthbert. "The international affective picture system". In: *Center for Research in Psychophysiology, University of florida, Gainesville, USA* (1988).

[2] Ahmad R Hariri et al. "The amygdala response to emotional stimuli: a comparison of faces and scenes". In: *Neuroimage* 17.1 (2002), pp. 317–323.

[3] Andrew H Kemp et al. "Gender differences in the cortical electrophysiological processing of visual emotional stimuli". In: *NeuroImage* 21.2 (2004), pp. 632–646.

[4] Daniel Grühn and Neika Sharifian. "Lists of emotional stimuli". In: *Emotion measurement*. Elsevier, 2016, pp. 145–164.

[5] James J Gross and Robert W Levenson. "Emotion elicitation using films". In: *Cognition & emotion* 9.1 (1995), pp. 87–108.

[6] Philip A Kragel and Kevin S LaBar. "Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions." In: *Emotion* 13.4 (2013), p. 681.

[7] Israel C Christie and Bruce H Friedman. "Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach". In: *International journal of psychophysiology* 51.2 (2004), pp. 143–153.

[8] Alan S Cowen and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients". In: *Proceedings of the National Academy of Sciences* 114.38 (2017), E7900–E7909.

[9] Erich Schubert et al. "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN". In: *ACM Transactions on Database Systems (TODS)* 42.3 (2017), pp. 1–21.

Chapter 5

REAL-LIFE EMOTIONS DURING THE COVID PANDEMIC

5.1 Introduction

In the previous chapters, I discussed emotions evoked by laboratory stimuli (stories and videos). In this chapter, I present results for real-life emotions, sampled longitudinally, during the COVID pandemic.

Previous studies on emotions in everyday life have largely focused on how frequently different emotions are experienced (especially the basic emotions) using mostly experience sampling [1, 2, 3, 4, 5] and sometimes diaries [6]. The studies suggested that happiness and anger are the most frequently experienced basic emotions. And in general, positive emotions are more common, while intense negative emotions are rare. Because these studies often used emotion terms (for example, the basic emotions categories) instead of affective scales, and data was often collected in a dichotomous format to encourage more responses, the dimensional space for real-life emotions is less probed compared to the distribution of these emotions.

There is substantial variability in the emotions people experience in everyday life. In particular, there is variability in the extent to which they differentiate different emotions from one another: some make many subtle distinctions, whereas others tend to lump their emotion experiences into a single category. This discriminative ability has been called "emotional granularity", and greater emotional granularity (that is, making more distinctions) is associated with better emotional health [7, 8]. This aspect could also be explored further in my COVID-Dynamic dataset, which among other scales includes the Toronto Alexithymia scale: a measure of the ability to feel, conceptualize, and verbalize emotions.

As mentioned before, I sampled real-life emotions across multiple waves as part of the Covid Dynamic study using the same set of affective scales as the ones used for stories and videos. There were similar attempts during the pandemic, for instance, a study in Spain collected ratings on valence and arousal for emotions experienced from March to June in 2020 [9]. Using these data, I tried to not only characterize the dimensional space, but also the distribution of real-life emotions.

5.2 Results

5.2.1 Four dimensions underlying real life emotion experiences

Correlation structure across scales

18 scales remained after exclusion that can be used to characterize real-life emotions experienced during the COVID pandemic, which were a subset of the 23 scales used for stories and videos (five absent scales were approach, relief, scalability, and generalizability over stimuli and behavior). Some scales (approach, relief) were taken out because I didn't think they would apply in this scenario. For instance, the relief scale contrasts how one feels in the end compared to the beginning, but I didn't expect this kind of temporal evolution for the real-life emotions that I was sampling. The other scales (scalability, and generalizability over stimuli and behavior) were omitted because of the concern of them being too complicated to rate, as the real-life emotion ratings were collected as part of an hour-long survey which already poses considerable cognitive demands.

I derived a pairwise correlation matrix across 18 scales (Fig.5.1), which is a representation of the underlying psychological space for real-life emotion experiences. I observed a similar broad correlation structure as for stories and videos, but in general, the correlations were weaker. This is probably expected for at least two reasons. One is that real-life emotions were probably weaker and more ambiguous than the ones evoked using carefully selected stories and videos. And since the stimuli were all different and individual, I was not able to aggregate ratings across subjects as I did with stories and videos, so the data was noisier than the stimuli-evoked emotions.

Judging from the correlation structure, the scales can be grouped into three groups. At the top, I have the same group of scales as seen in stories and videos, describing how negative the emotions are. At the bottom, I again have a group of scales describing how positive the emotions are. The scales at the middle describe general properties of emotions that are less tightly correlated with the positive and negative related scales. I discuss more about the similarities and dissimilarities of correlation structures across stimuli types in Chapter 6.

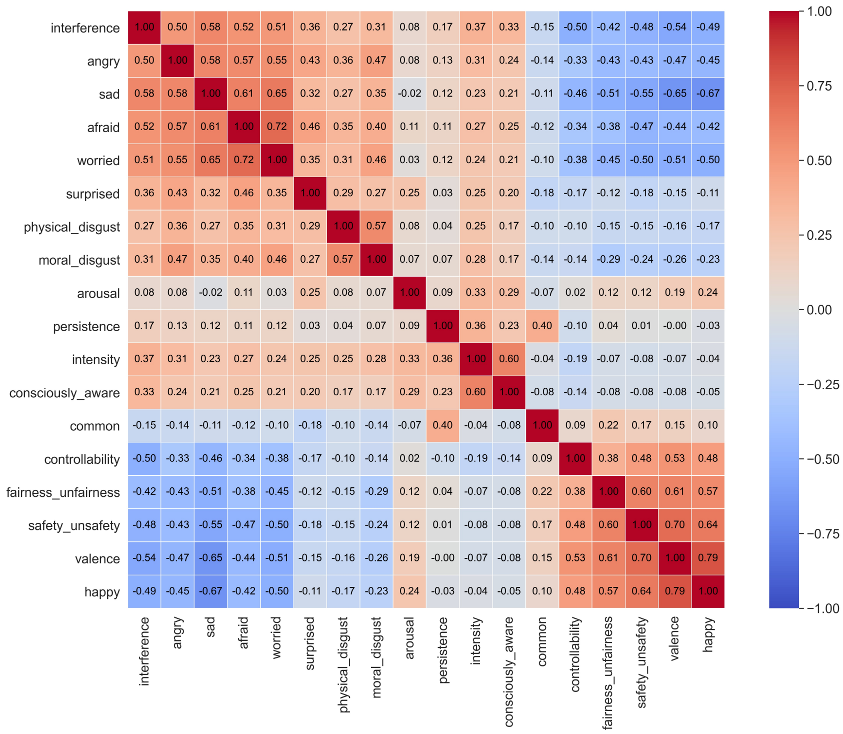

Figure 5.1: Correlation matrix across scales for real-life emotions (sorted using hierarchical clustering to intuitively depict the underlying structure).

Factor analysis

The correlation matrix across scales suggested that the psychological space of real-life emotion experiences can be characterized using a smaller number of underlying factors, but the number of factors to retain in an exploratory factor analysis needs to be determined first.

Determining the optimal number of factors to extract

I determined the optimal number of factors to retain by taking multiple aspects into consideration: the suggestions from the standard statistical methods, results from the cross validation approach, results from the robustness assessment, and the interpretations of the factor loadings, which I describe below.

First, I used the six commonly used statistical tests (see details in 2.2.3). The results (Fig.S5.1) varied across methods. Very Simple Structure, Empirical BIC and Velicer's MAP suggested 1, 8 and 2 factors respectively. Parallel analysis, the acceleration factor and the optimal coordinate suggested 8, 1 and 6 factors respectively.

In addition to the standard statistical procedures described above, I also used the cross validation procedure (see details in 2.2.3) to choose how many factors to retain for real-life emotions. Visual inspection of the result (Fig.S5.2) again didn't suggest a clear value to use, and 2 to 5 factors all seemed reasonable.

The various methods have suggested multiple values for the number of factors to retain. I therefore applied EFA to further assess the interpretability of different factor solutions. When extracting 2 factors, the factors each explained 28% and 15% of the common variance in the data (43% in total). I interpreted the factors as "valence" and "arousal" (Fig.S5.3 a). When extracting 3 factors, the factors each explained 26%, 13%, and 10% of the common variance in the data (49% in total). I interpreted the factors as "valence", "negative affect", and "arousal" (Fig.S5.3 b). When extracting 4 factors, the factors each explained 25%, 13%, 10%, and 5% of the common variance in the data (54% in total). I interpreted the factors as "valence", "negative affect", "arousal", and "common" (Fig.5.3). I also attempted the 5 factor solution, which cumulatively explained 58% of the total variance but the factors were uninterpretable (Fig.S5.3 c).

Evidence so far suggested both the 3 and 4 factor solutions seemed reasonable. I don't argue against the 2 factor solution, but just decided to retain more factors for completeness. I further assessed the robustness of both the 3 and 4 factor solutions with regard to the number of stimuli and number of scales.

First, I systematically reduced the number of real-life emotion instances from the whole set of 12861 instances. At each step, I used the remaining data for EFA and calculated Tucker indices of factor congruence between factors derived using the reduced set and the original set of stimuli (with orthogonal Procrustes rotation). Both the 3 factor and 4 factor solutions were robust to the number of instances (Fig.5.2 a,b). For the 3 factor solution, all mean factor congruences were higher than 0.9 with no fewer than 111 instances (roughly 0.86 % of the whole set). For the 4 factor solution, all mean factor congruences were higher than 0.9 with no fewer than 361 instances (roughly 2.8 % of the whole set).

Second, I removed scales one by one (same order of removal based on global redundancy as explained in 2.2.3) and quantified the relatedness of the original factors from the full set and the ones from the subset by correlating the factor scores (Fig.5.2 c,d). The 4 factor solution was robust to the number of scales as the factors derived from the full set versus the subsets of scales were highly correlated. The 3 factor solution didn't behave as well, most notably the "negative affect" factor was unstable. All evidence taken together, I decided to retain 4 factors which are highly interpretable and robust.

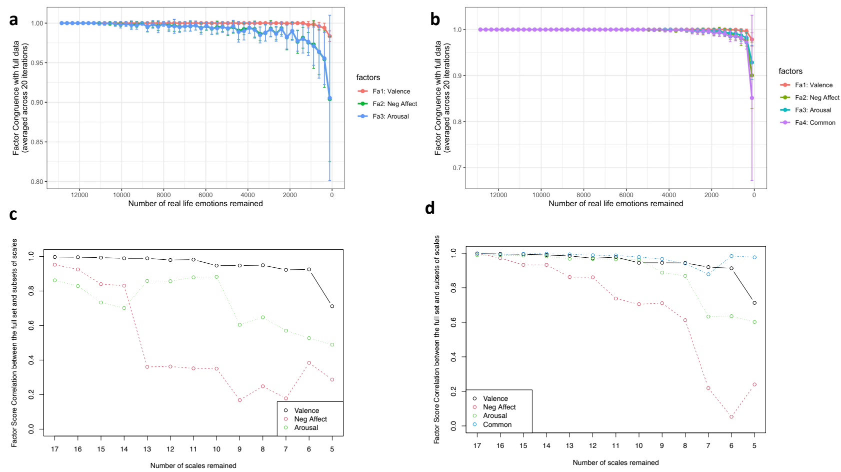

Figure 5.2: Robustness of factor solutions with respect to the number of instances and number of scales. (a,b) Points indicate the means and error bars indicate standard deviations of Tucker indices of factor congruence (with orthogonal Procrustes rotation) between the full set versus subsets of real-life emotions across 20 iterations, and are color-coded for different factors for (a) the 3 factor solution and (b) the 4 factor solution. (c, d) Pearson's correlations between factor scores from the full set versus subsets of scales, color-coded for different factors for (c) the 3 factor solution and (d) the 4 factor solution.

Interpretation of the four factors

Exploratory factor analysis was then performed to extract 4 factors using the minimal residual method, and the solutions were rotated with oblimin for interpretability. The Tenberge method was used to obtain oblique factors scores (all using the "'fa" function in the "psych" package in R). I interpreted the 4 factors as "valence", "negative affect", "arousal", and "common" (see Fig.5.3 for factor loadings).

For "valence" and "arousal", the factor loadings closely resemble those of stories and videos. The "common" factor describes how common and persistent emotions are and is somewhat related to the "generalizability" factor that I found for stories and videos. The "negative affect" factor is unique with the real-life emotions and not found with the story or video data. A specific set of scales (disgust, fear, surprise, and anger) load strongly onto this factor. One reasonable interpretation is that this factor is specific to the COVID pandemic (and all the other stressors associated with it).

Examining the factor scores for individual emotions (together with the descriptions of the cause of emotion from subjects) also allowed me to verify my interpretation of the scales. Most real-life emotions are experienced commonly as expected and not necessarily associated with certain levels of valence or intensity.

People generally feel positive when they are at home, relaxing with family. Some of the most negative emotions were associated with financial concerns as a consequence of the pandemic while for some people, they were feeling depressed constantly with no specific cause.

It's also interesting that the same real-world event can evoke completely different emotions in different people. For instance, the data collection for wave 14 happened just around the election, and for people supporting Joe Biden, the result brought intense joy and hope while for others, it evoked intense anger.

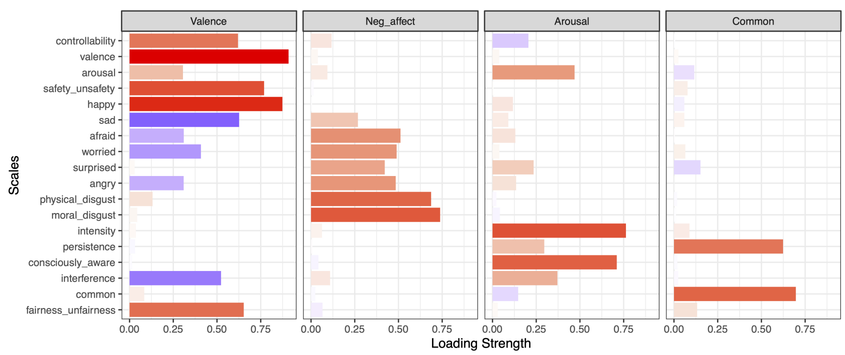

Figure 5.3: Factor loadings of scales on the four factors from EFA. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

5.2.2 Distribution of real life emotion experiences

The data included here corresponds to wave 2 to 16 of the COVID-Dynamic project (from April 2020 to January 2021) during which time the US faced multiple COVID surges. In addition, this period included multiple national events (e.g., BLM protests, US-presidential elections, the attack on the US Capitol, etc.) (Fig.2.1).

In such a changeable environment and over such a long time period, I believe that the sampled real-life emotions would be particularly rich in content. It's worth noting that unlike the previous studies where certain emotion categories were intended with the stories and videos, subjects were simply asked to rate and describe whatever emotions they were feeling at the moment. This is not to claim that there's no bias in my data. Admittedly, I am probably missing out on extremely positive or negative emotions as it's unlikely that subjects experiencing those emotions would sit down for an hour-long survey.

From the distributions, I note that real-life emotions are more common and easier to control, compared with the ones evoked by stories and videos (Fig.5.4). Distributions for many other scales resemble a normal distribution, except for those basic emotion ones.

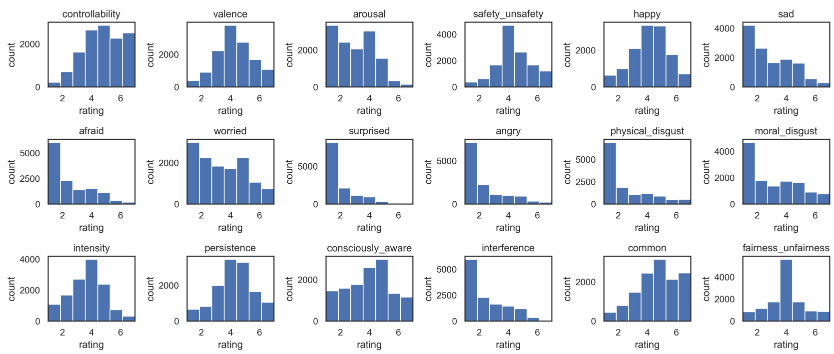

Figure 5.4: Distributions of ratings on the 18 scales for 12861 real-life emotions.

Umap

The univariate distributions on valence-related scales for real-life emotions resemble those of videos, but not those of stories, again suggesting that I might not see a good separation of positive and negative emotions.

Similarly as before, I applied UMAP to better visualize the distributions in the embedded two dimensional space. Since four factors accounted for most of the total variance, representing the emotions using a two dimensional UMAP plot should be a reasonable approximation. I further combined the factor scores with the UMAP plots to probe whether real-life emotion experiences are discrete or dimensional.

The first observation was that in general, real-life emotions formed just one cluster with some outliers near the boundaries. Interestingly, at the very top, there's a cluster, well-separated from the rest of the emotions, that encodes apathy.

Color coding using the factors again revealed "valence" as a continuous global gradient linking positive emotions on the left to negative emotions on the right (Fig.5.5 a). Emotions varied along the "arousal" dimension smoothly as well, but not in a single global direction as valence. Extremely positive or negative emotions are more intense and arousing compared to neutral ones (Fig.5.5 c). Emotions with high scores on the “negative affect” factor were located on the right side only (Fig.5.5 b). The “common” factor didn’t have a prominent direction of gradient (Fig.5.5 d).

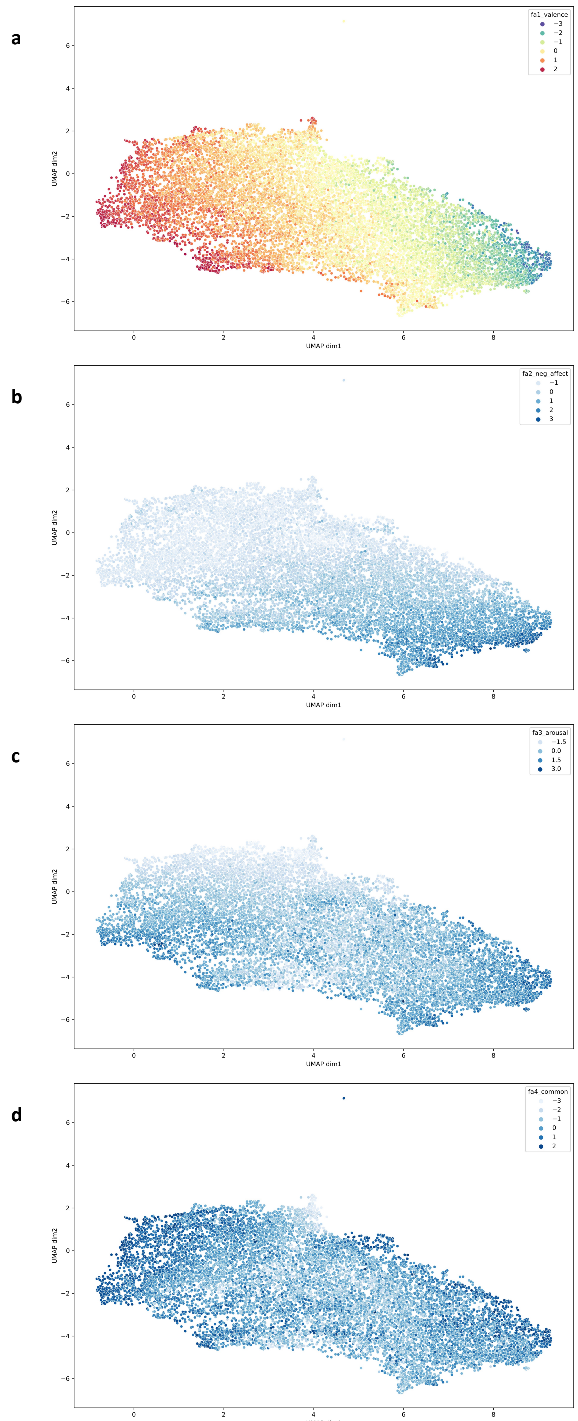

Figure 5.5: UMAP plots color-coded for (a) the “valence” factor, (b) the “negative affect” factor, (c) the “arousal” factor, and (d) the “common” factor.

Clustering

As already noted, UMAP revealed continuous gradients, most notably the "valence" factor, in the dimensional space of real-life emotions. Here, I applied clustering analysis to probe whether emotions form discrete clusters in the high dimensional space with clear boundaries.

I used several clustering algorithms to determine the optimal number of clusters for a single partitioning of the data. First, I used K-means, and tried out a range of possible number of clusters to extract, the evaluation metrics suggested that the optimal number of clusters to extract to be 2 (Fig.S5.4). It should be noted that this didn't exclude the possibility of one cluster being the optimal choice which wasn't evaluated because most of the evaluation metrics require at least 2 clusters. Mean-shift suggested 1 cluster. Min_samples and eps were set to be 36 and 6 respectively for DBSCAN (using heuristics), which again suggested one cluster (Fig.S5.5).

These results confirmed my previous observations based on the univariate distributions and from the UMAP result that the real-life emotion experiences are best characterized by just one cluster. Still, I think an extensive hierarchy of clusters would be informative. So I applied hierarchical agglomerative clustering (HAC) with Euclidean distance and Ward linkage (Fig.5.6) and combined UMAP coordinates and free descriptions of the emotion labels to interpret the meaning of the clusters (Fig.S5.6, Fig.S5.7).

The two clusters correspond to the positive and negative emotions at the top of the hierarchy. The words that subjects used the most to label the negative emotions were anxious, worried, and tired while words for positive emotions included happy, content, and calm. Note that these words differed from the ones used for stories and videos, especially for describing negative emotions.

Further down the dendrogram, four clusters emerge which can be interpreted as weak positive emotions(with words like calm, content and neutral), strong positive emotions(with words like calm, content and happy), strong negative emotions (with words like anxious, worried, frustrated and depressed), and weak negative emotions (with words like anxious, tired, and worried) from top to bottom respectively. The basis of partitioning at these two levels corresponds to the "valence" and "arousal" factors that I have identified, again justifying the dimensional account of emotions.

In principle, the hierarchy of clusters can be probed at all possible levels, with the bottom level of having a single instance as its own cluster. But, the five cluster solution was already hard to interpret (Fig.S5.6, Fig.S5.7).

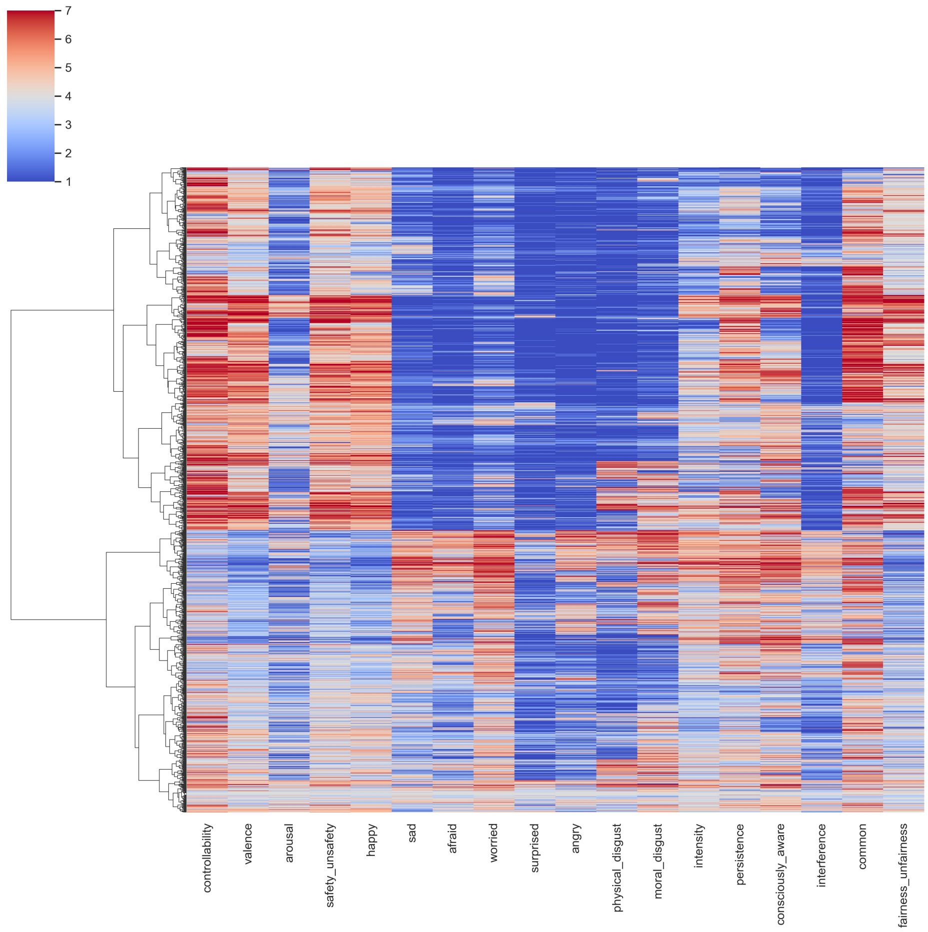

Figure 5.6: Visualization of the hierarchical clustering results where columns indicate ratings on the 18 scales (color indicates rating magnitude: blue for lower ratings and red for higher ratings). Each row represents one real-life emotion, 12861 rows in total.

5.3 Summary and discussion

I characterized over 10,000 real-life emotions in a high dimensional space using 18 affective scales. Exploratory factor analysis was performed to extract four robust factors that captured the majority of the variances. The dimensions were interpreted as “valence”, “negative affect”, “arousal” and “common”. “Valence” and “arousal” are in agreement with what I found with stories and videos. I believed that the “negative affect” factor was COVID-specific. The “common” factor might be separating general and ongoing emotions, from emotions specifically triggered by events. I further characterized the distribution of emotions, and again found that emotions varied along continuous gradients with no well-separated clusters.

My study of real-life emotions has several limitations that I discuss below.

First, there are some limitations for my sampling. As noted, I am probably missing out on extremely positive or negative emotions of high intensity. For some individuals, I noted that even with 15 waves of data collection, there was little variance on some of the scales which made it impossible to construct the full correlation matrices across scales for those subjects. In addition, since all emotions were sampled during the pandemic, they are likely biased to be more negative in general. It's also worth noting that unlike traditional emotion sampling, my sampling was embedded in the hour-long survey which itself can be viewed as a stimulus.

Second, I didn't use the full set of scales for real-life emotions which limited my ability to compare across domains for some questions. For instance, it might have resulted in the absence of the generalizability scale in real-life emotions.

I again only collected self-report ratings for real-life emotions, without physiological measures. Wearables devices may be utilized to collect measures such as heart rate for future studies for a more comprehensive investigation of real-life emotions.

5.4 Supplementary information

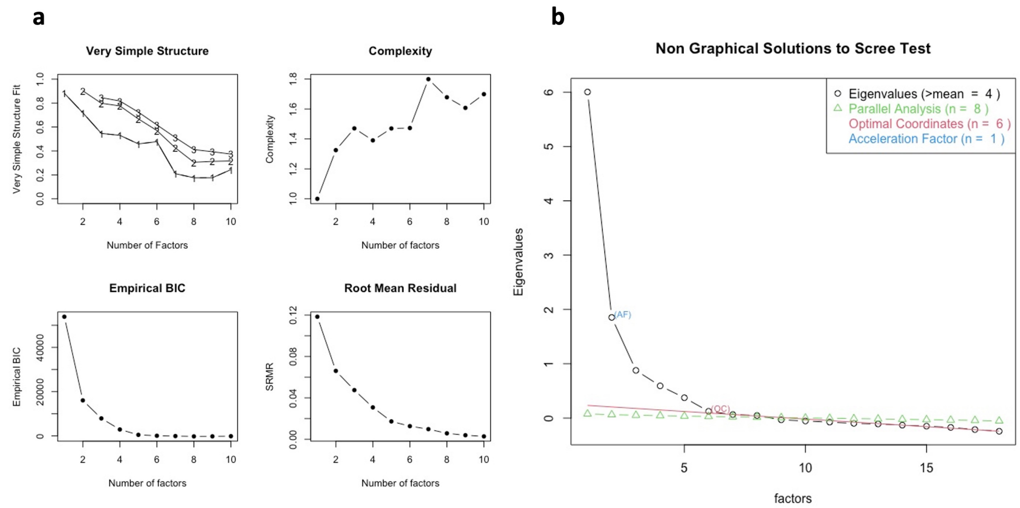

Figure S5.1: Results for the various statistical methods. (a) Very Simple Structure and Empirical BIC (Velicer's MAP is not plotted), and (b) Parallel analysis, the acceleration factor, and the optimal coordinate.

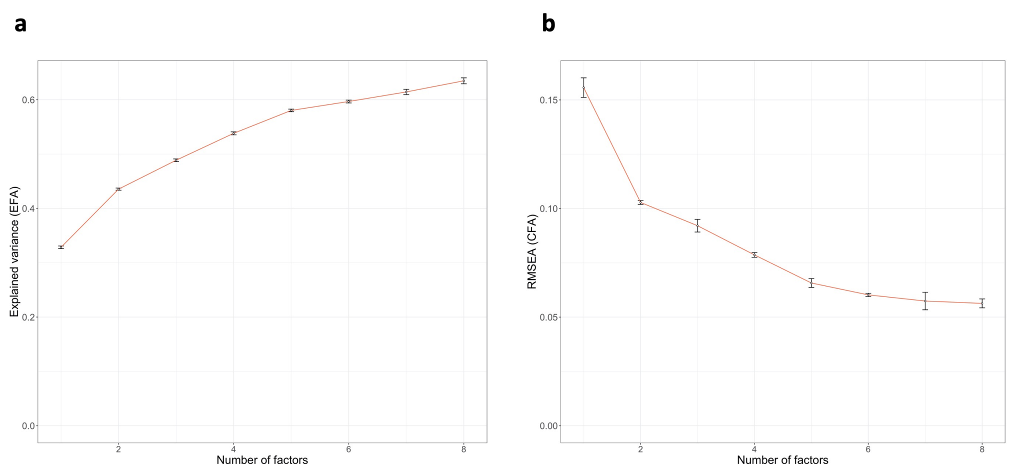

Figure S5.2: Results for the cross validation procedure. The means (points) and standard deviations (error bars, n = 20 iterations) of (a) explained variance from the EFA on training data and (b) root mean square error of approximation (RMSEA) fit index from the CFA on testing data.

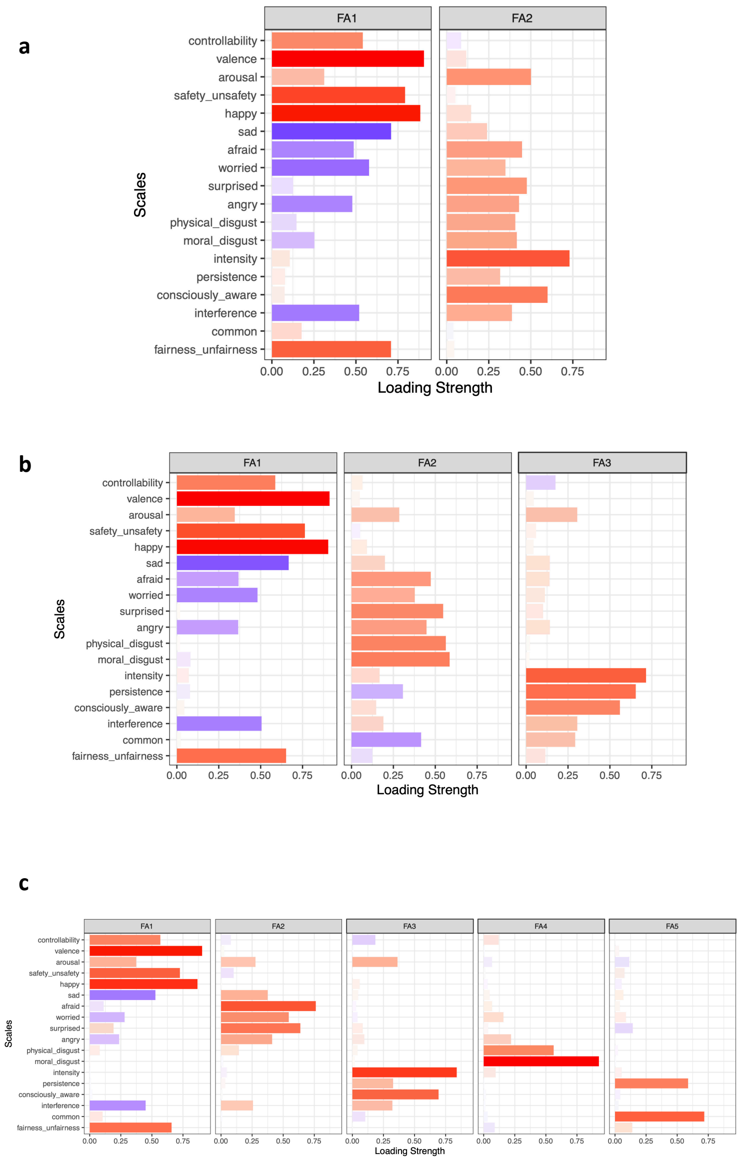

Figure S5.3: Factor loadings of scales on (a) the 2 factors from EFA, (b) the 3 factors from EFA, and (c) the 5 factors from EFA. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

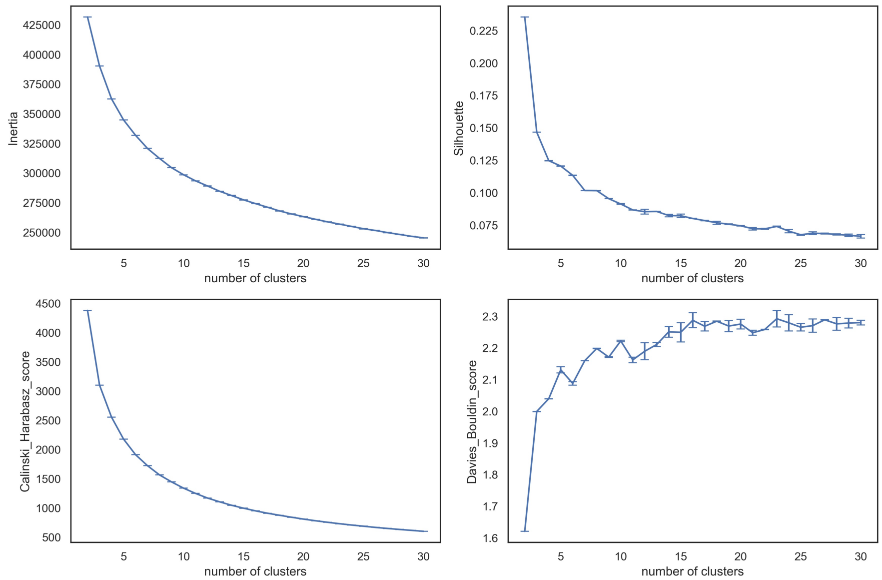

Figure S5.4: Determining the number of clusters for K-means. The means (points) and standard deviations (error bars, n = 20 iterations) of inertia, the Silhouette coefficient, the Calinski-Harabasz index, and the Davies-Bouldin index for K-means results with different number of clusters (2 to 30).

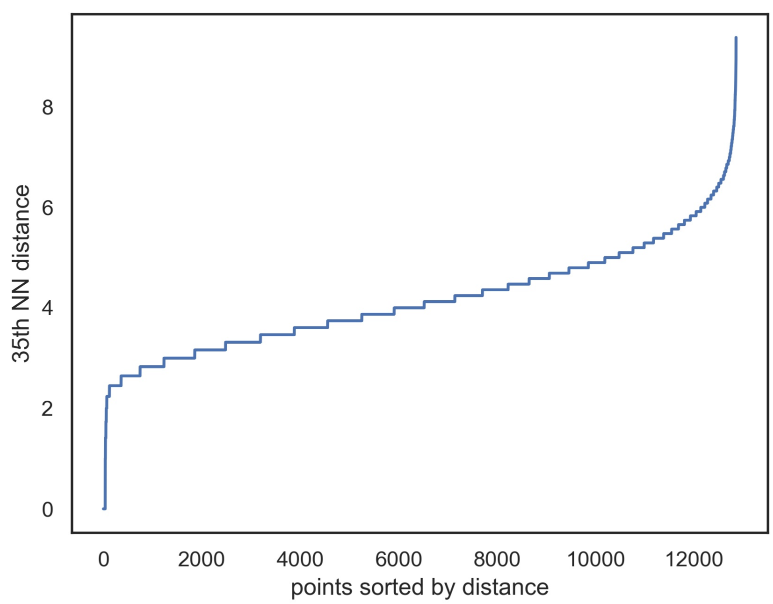

Figure S5.5: The 35th nearest distance plot for DBSCAN.

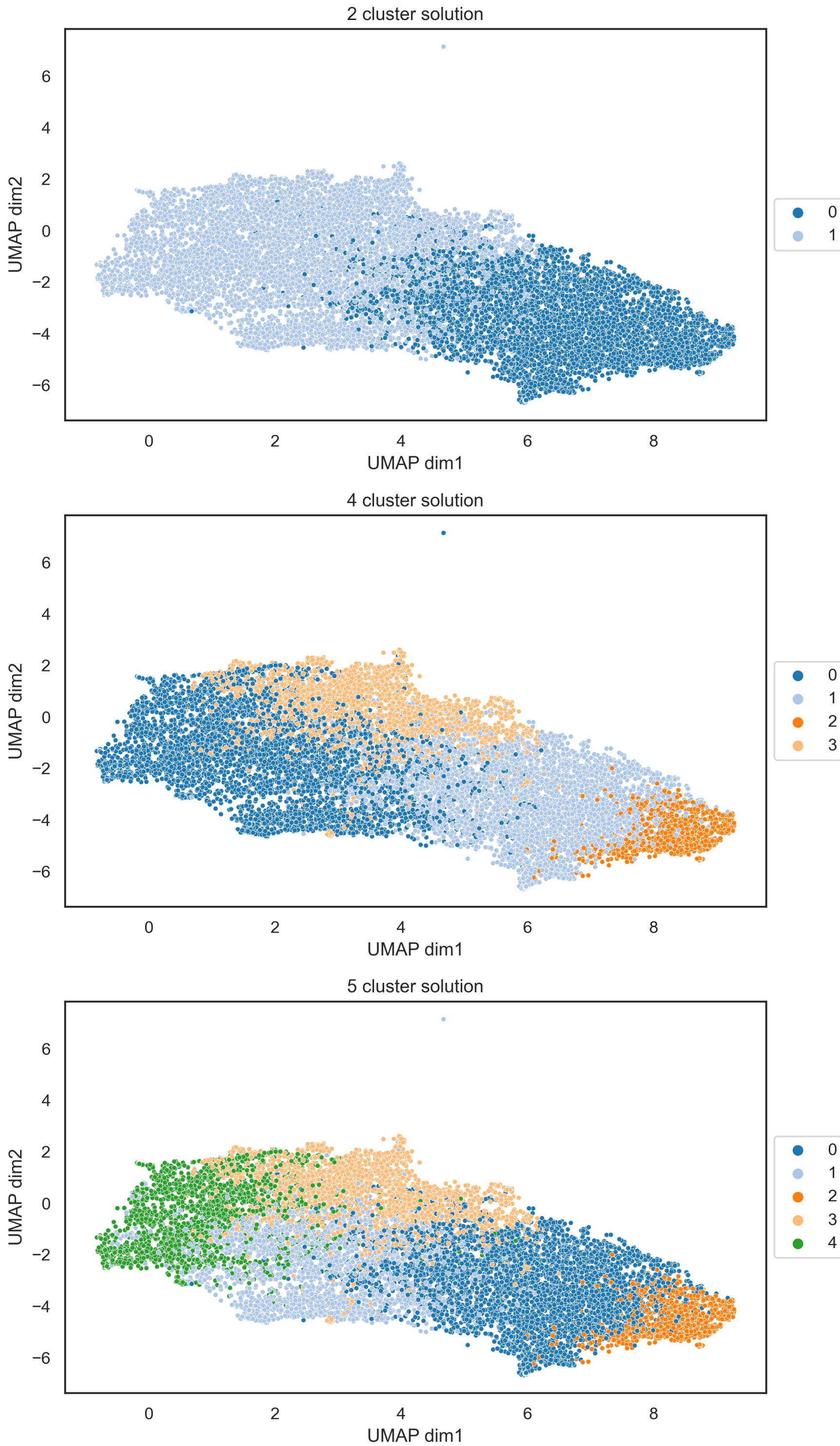

Figure S5.6: Visualization of the different cluster solutions (2, 4, 5 clusters) as determined by HAC. Points were color coded for cluster membership and location was based on UMAP coordinates.

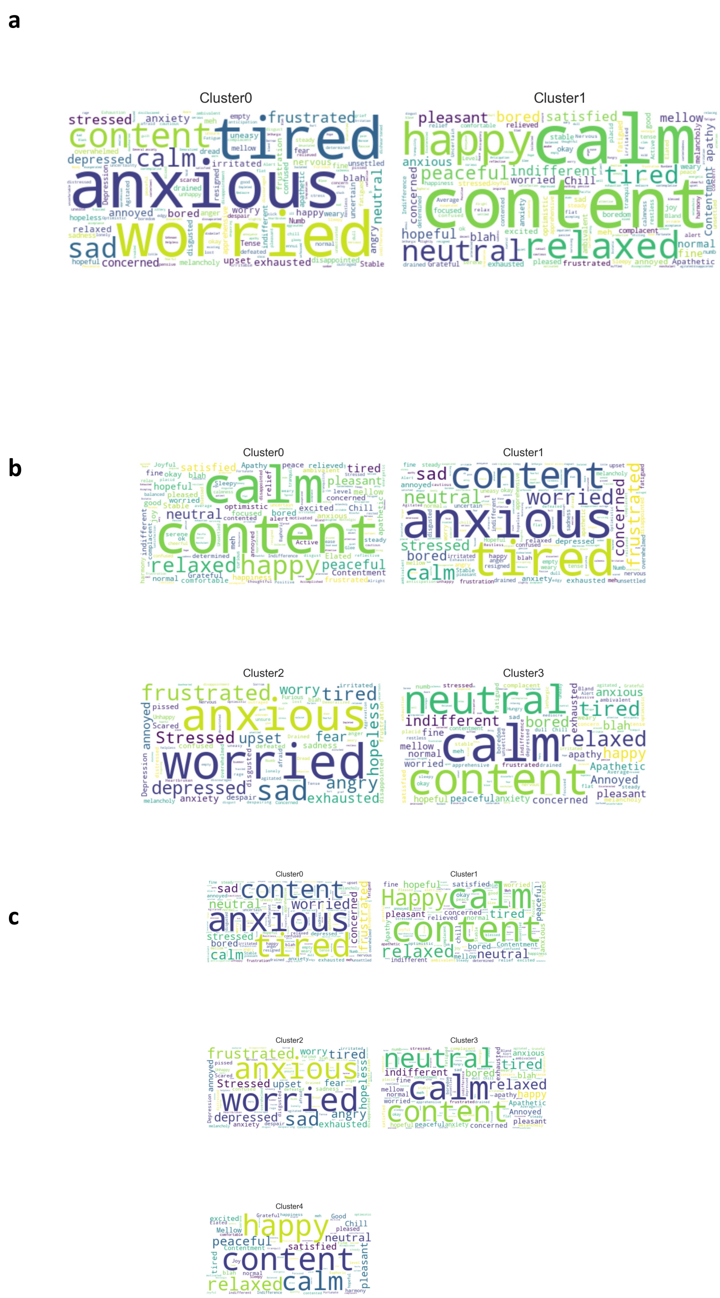

Figure S5.7: Word Clouds of the free descriptions for each cluster of the different cluster solutions as determined by HAC: (a) 2 clusters, (b) 4 clusters, and (c) 5 clusters.

References

[1] John M Zelenski and Randy J Larsen. “The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data”. In: *Journal of Research in Personality* 34.2 (2000), pp. 178–197.

[2] Peter Wilhelm, Dominik Schoebi, and Meinrad Perrez. “Frequency estimates of emotions in everyday life from a diary method’s perspective: a comment on Scherer et al.’s survey-study “Emotions in everyday life””. In: *Social Science Information* 43.4 (2004), pp. 647–665.

[3] Debra Trampe, Jordi Quoidbach, and Maxime Taquet. “Emotions in everyday life”. In: *PloS one* 10.12 (2015), e0145450.

[4] Klaus R Scherer et al. “Emotions in everyday life: Probability of occurrence, risk factors, appraisal and reaction patterns”. In: *Social Science Information* 43.4 (2004), pp. 499–570.

[5] Laura L Carstensen et al. “Emotional experience in everyday life across the adult life span.” In: *Journal of personality and social psychology* 79.4 (2000), p. 644.

[6] Keith Oatley and Elaine Duncan. “The experience of emotions in everyday life”. In: *Cognition & emotion* 8.4 (1994), pp. 369–381.

[7] Katharine E Smidt and Michael K Suvak. “A brief, but nuanced, review of emotional granularity and emotion differentiation research”. In: *Current Opinion in Psychology* 3 (2015), pp. 48–51.

[8] Todd B Kashdan, Lisa Feldman Barrett, and Patrick E McKnight. “Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity”. In: *Current Directions in Psychological Science* 24.1 (2015), pp. 10–16.

[9] Carlos Bailon et al. “CoVidAffect, real-time monitoring of mood variations following the COVID-19 outbreak in Spain”. In: *Scientific Data* 7.1 (2020), pp. 1–10.

Chapter 6

COMPARING EMOTIONS ACROSS STIMULUS TYPES

In the previous chapters, I have covered the structure and distribution of emotion experiences evoked by three different stimulus types on their own. In this chapter, I discuss the similarities and dissimilarities across stimulus types which is a unique strength of my study. The fact that the same participants rated their emotions across these three kinds of stimuli on the same rating scales made it possible to explore whether subjects were reporting on what they think the intended or "correct" emotion is (a reasonable concern for emotions evoked by stories), versus reporting on the actual contents of their conscious experience of the emotion (more likely the case for video-evoked and real-life emotions).

6.1 Comparison of the correlation structure across scales

Similar structure across stimulus types

The first question I asked was whether emotion experiences across stimulus types shared a similar broad correlation structure, a pattern intuitively noted already but not formally quantified yet. For this and other comparisons between all three stimulus types, I used only those 18 rating scales that were used for the real-life emotions (a subset of the 23 used for the stories and videos).

Sorting the correlation matrices across the 18 scales with the same order revealed visually that the representational structure was highly consistent across stimulus types (Fig.6.1). Specifically, the second-order similarity calculated by correlating the correlation matrices was rs = 0.953, 0.923 and 0.909 for stories and videos, stories and real life and video and real life respectively (see details in 2.2.2, all ps < 0.0001, Fig.S6.1). As stories and videos shared more scales, I quantified the more complete correlation matrices between the two and the conclusion did not change (the correlation matrices across 23 scales were significantly correlated at r = 0.944, Fig.S6.1, Fig.S6.2).

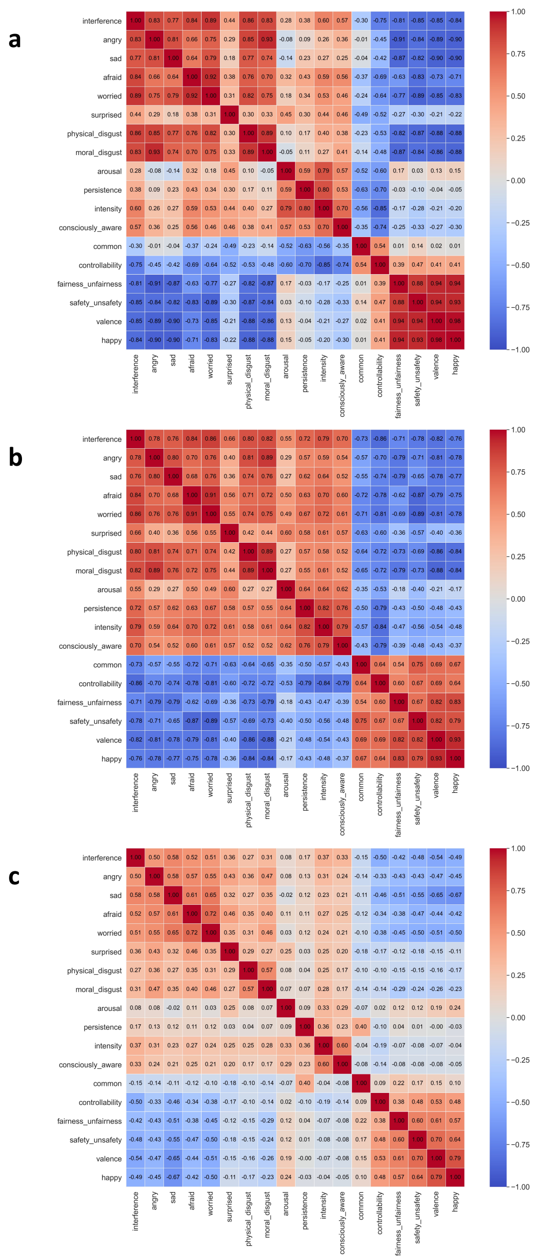

Figure 6.1: Correlation matrices across 18 scales for (a) story-evoked emotions, (b) video-evoked emotions, and (c) real-life emotions, sorted based on real life emotions.

Different correlation strengths across stimulus types

I also noted that the magnitudes of the correlation coefficients in the correlation matrices did differ despite the similar overall structure, with the strongest overall correlations for videos and the weakest correlations for real-life emotions. Examining the distributions of the correlation coefficients, I can see that for videos, the correlations across scales were either strongly positive or strongly negative while for stories and real life, many correlations across scales were close to zero (Fig.6.2, means of the absolute correlation coefficients were 0.645, 0.522, and 0.288 for video, story, and real life respectively).

There are at least two possible explanations for why videos have stronger correlations than stories, either because they were more diverse or more potent (or both). It's also possible that the stronger correlations are a result of simply having a larger number of videos. I tested this hypothesis by subsampling the videos to match the number of stories, but still found stronger correlations (Fig.S6.3).

Possible explanations for real-life emotions having the weakest correlations include data being noisier as aggregation across subjects was not possible, and real-life emotions being naturally weaker and less intense.

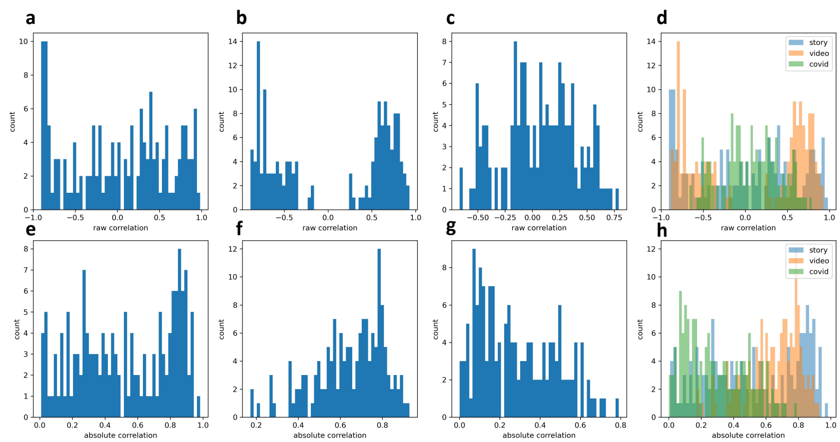

Figure 6.2: Correlation strengths across stimulus types. Top: histograms of the raw correlation coefficients from the correlation matrices across 18 scales for emotions evoked by (a) stories alone, (b) emotions evoked by videos alone, (c) real-life emotions alone, and (d) combined. Bottom: histograms of the absolute correlation coefficients from the correlation matrices across 18 scales for emotions evoked by (e) stories alone, (f) emotions evoked by videos alone, (g) real-life emotions alone, and (h) combined.

Specific differences on certain scales

So far, I have established that correlation structures across stimulus types share a similar broad structure, but with different levels of correlation strengths.

Closer inspection revealed some specific differences, most notably with the "arousal" and "common" scales. Out of the three stimulus types, the "arousal" scale correlated most strongly with other scales for videos, compared with stories and real life. Again, I think it makes sense because the videos were more potent and arousing. For the “common” scale, in particular, I noted that the correlation between “persistence” and “common” was in the opposite direction for stories and videos versus real life. For videos, positive emotions were common but not persistent; instead the uncommon negative emotions were persistent. This made sense as videos evoking negative emotions often involved horrifying scenes that would have a long-lasting impact while videos evoking positive emotions largely involved amusing scenes with more transient impact. In real life, commonly experienced emotions were also the most persistent, which makes logical sense: if those emotions persist and are experienced a lot, then by definition they would become common.

6.2 Comparison of the factors

I had previously reported that I found 3 factors for emotions evoked by stories (interpreted as "valence", "arousal", and "generalizability"), 3 factors for emotions evoked by videos (also interpreted as "valence", "arousal", and "generalizability") and 4 factors for real life emotions (interpreted as "valence", "negative affect", "arousal", and "common").

Noting the consistency of the overall correlation structure and the semantic similarity of the factors across stimulus types, I directly tested the idea of shared latent factors across domains.

Using the correlation matrix across 18 shared scales averaged across three domains for EFA and then applying CFA to each of the three stimulus types allowed me to determine the number of factors that would best explain the shared structure. A 2 factor solution was suggested as increasing from 1 to 2 factors showed the most substantial improvement both in the explained variance for EFA and the model fits for CFA (Fig.6.3). Examining the factor loadings also suggested the clearest interpretations for the 2 factor solution with “valence” and “arousal” as the factors (Fig.S6.4).

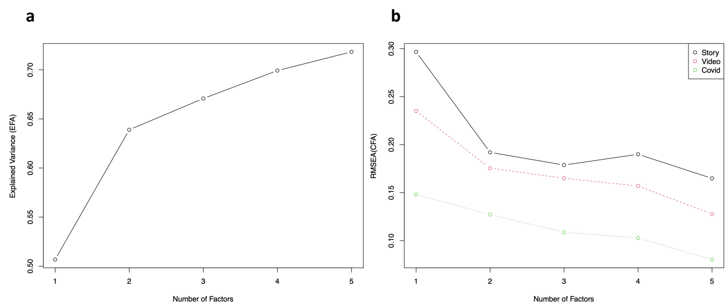

Figure 6.3: Determining the number of factors for the averaged correlation matrix. (a) explained variance from the EFA on the averaged correlation matrix and (b) root mean square error of approximation (RMSEA) fit index from the CFA on each of the three types of data (indicated by different colors).

I therefore extracted 2 factors from each of the three types of data and quantified the relatedness of the factors by calculating factor congruence (Fig.6.4). A two-dimensional structure of emotion experience was indeed consistent across stimulus types as indicated by high levels of factor congruence.

The “generalizability” factor, which was the third factor for both stories and videos did not emerge here. This is at least partially due to the limitation that the specific scales with highest loadings on the generalizability factor were not included for assessing real-life emotions.

I think the generalizability factor (scales with highest loadings: “generalizability over stimuli”, “generalizability over behavior”, “common”, and “surprised”) is related to the interrupting role of emotions that Herbert Simon had conceived [1]. Simon analyzed emotions as the operation of a system that was in parallel with standard goal-oriented cognition, a system that he felt was required by the ecology of survival in the real world, and that consisted of two fundamental components: a continual monitoring (“noticing”) function, in order to detect salient events in the first place, and an “interrupt mechanism” whose function was “setting aside ongoing programs when real-time needs of high priority are encountered.” (p. 34). Presciently, Simon noted that, “. . . the tendency of a particular stimulus to evoke emotional behavior . . . . generally decreases with repetition, “ concluding that, “In general, real-time needs to respond to the environment arise when the environment can change rapidly and unpredictably,” (p.37) going on to remark that emotions would be expected to arise especially in social situations.

Highly generalizable emotions are experienced often and are therefore less surprising and therefore less interruptive. Emotions with low generalizability, on the other hand, are less often experienced, more specific and more surprising, thus warranting a need for interrupting the ongoing goal-attaining system. Simon also hypothesized that learning can change generalizability, that is, if a stimulus is encountered repetitively (as in the above quote), then one can learn to adapt to decrease the interruption.

I also think that it's possible that the evaluation of the factors I identified happens in serial in the sequence of "valence", "arousal", and "generalizability" as they served as the basis for partitioning for the hierarchical clustering at different levels. That is, the most global distinction, the one based on valence, would be implemented first, followed by a distinction of low or high arousal, followed by a distinction of generalizability. This point raises a deep and related question: what exactly is supposed to be implemented by these factor-based distinctions? I would argue that it is, in the first instance, psychological and neural processing; that is, the stimulus, or situation, is classified as of positive or negative valence first, then classified with respect to arousal, and then with respect to generalizability. This is of course a conjecture—one alternative would be that the brain carries out the full classification in a (at least) three-dimensional space in a single step. A second alternative would be that the order in which the classification happens depends on the stimulus. Future studies and analyses based on reaction-time, as well as physiological dependent measures with rapid sampling times, could address these possibilities.

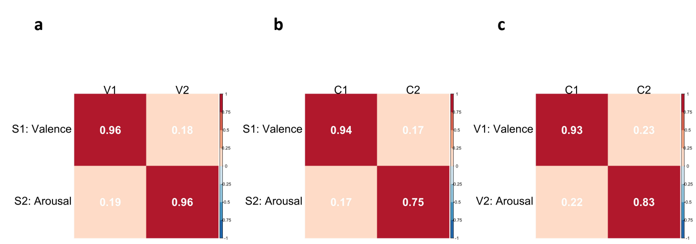

Figure 6.4: Tucker indices of factor congruence (with orthogonal Procrustes rotation) across stimulus types. The first factor (S1/V1/C1) is valence and the second factor (S2/V2/C2) is arousal. (a) rows for stories and columns for videos, (b) rows for stories and columns for real life, and (c) rows for videos and columns for real life.

6.3 Comparison of the emotion experiences

The comparisons across domains so far focused on the structure of the emotion space represented by the correlation structures and latent factors. But it remains unknown how similar or different the individual emotion experiences were evoked by different types of stimuli.

Clustering based on stimulus type?

My first idea was to see if there would be clustering based on stimulus type if I put emotions evoked by different types of stimuli together. To balance the number of emotions across domains, I subsampled 150 video-evoked emotions and 150 real-life emotions to match the number of emotions evoked by stories. A maximum variation sampling procedure (similar to what I did with stimuli selection, see details in 2.1.2) was used to choose the most diverse subset of emotions from videos and real life, instead of random sampling.

Applying hierarchical clustering to the combined set of 450 emotion experiences, I found no prominent clustering based on stimulus type (Fig.6.5). I did notice sequential rows of the same color, especially with real-life emotions indicating similarities of emotion experiences belonging to the same stimulus type.

It's worth noting that even though the number of emotion experiences was matched across domains, the exact distributions were not. I attempted to partially resolve the issue by selecting only emotion experiences of the basic emotion categories. The purpose of matching distributions is to better compare intra- and inter-domain similarities. However, the true distributions of all possible emotions that can be evoked by these three types of stimuli are likely to be different.

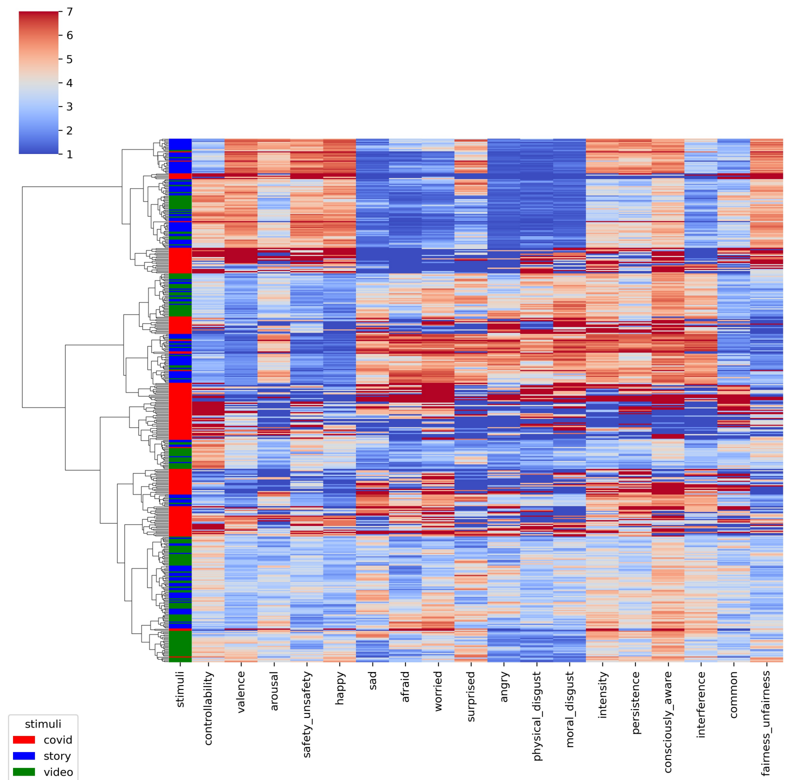

Figure 6.5: Visualization of the hierarchical clustering results where the first column indicates the stimulus type and the subsequent columns indicate ratings on the 18 scales (color indicates rating magnitude: blue for lower ratings and red for higher ratings). Each row represents one emotion, 450 rows in total.

Basic emotions across stimulus types

As mentioned, even with maximum variation sampling done within each stimulus domain, the detailed distributions of ratings of the emotions were still different across domains. One way to alleviate the problem was to select only emotion experiences of the six basic emotion categories.

One immediate question was how to select good examples of basic emotions. As a subset of my scales corresponded to the six basic emotion categories as commonly defined in the literature [2], I color-coded the UMAP plots with these scales to see where the best examples of basic emotions were located (Fig.S6.5,Fig.S6.6,Fig.S6.7). The general observation was that regardless of the stimulus types, even for carefully-engineered stories, it's surprisingly hard to find 'pure' instances of basic emotions, especially for the negative ones which seemed to be experienced frequently at the same time. The finding contradicts the idea of basic emotions being discrete categories.

Nevertheless, I tried to select good examples of basic emotions with the following criteria. The first requirement was that for an emotion to be considered a good example for a given basic emotion category, it needs to have a high rating for the corresponding scale (higher than 6 as my scales ranged from 1 to 7). In addition, it should have as low ratings as possible on the other basic emotion scales (1 being the lowest possible rating). Together, these criteria would produce relative specificity for a basic emotion category. I therefore selected 5 instances (whenever possible) for each basic emotion category for the three stimulus types.

With these basic emotions, I first tried to see if the basic emotions can be well separated into six clusters. However, even for carefully-chosen good examples, the intended basic emotion categories were not perfectly recovered using any of the three types of data alone or when combined (Fig.6.6). The separation was best for emotions evoked by stories, and worst for emotions evoked by videos, probably because of the specificity of the stimuli. Among the six basic emotions, happiness and surprise clustered relatively better than the negative ones, again confirming my observation from the UMAP plots.

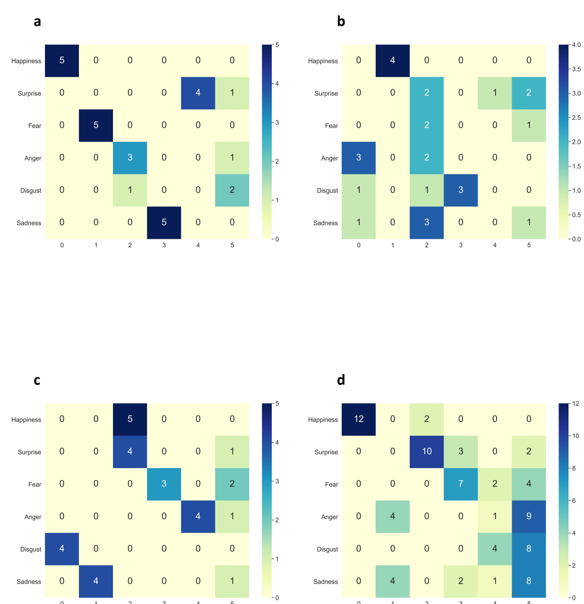

Figure 6.6: Contingency matrices between the discovered categories (columns) and the intended basic emotion categories (rows) for (a) story, (b) video, (c) real life, and (d) combined.

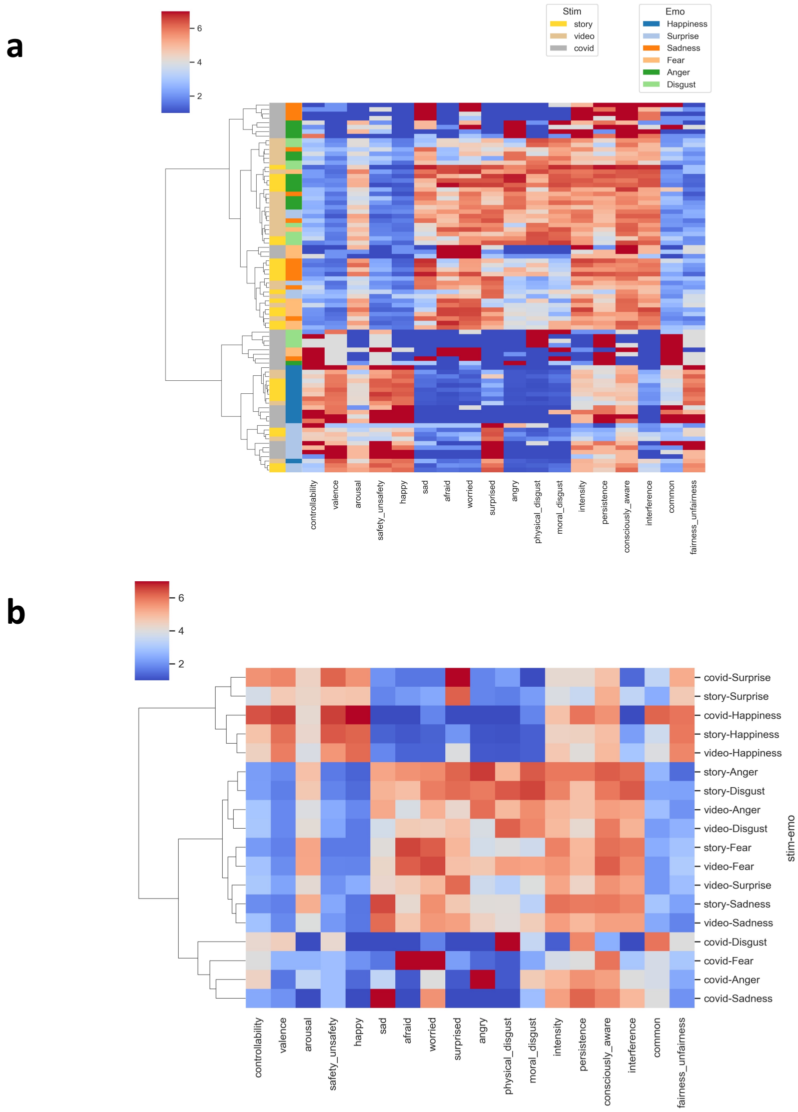

Figure 6.7: Visualization of the hierarchical clustering results for basic emotions. (a) for individual emotion experiences, the first two columns indicate stimulus types and emotion categories and the subsequent columns indicate ratings on the 18 scales (color indicates rating magnitude: blue for lower ratings and red for higher ratings). Each row represents one emotion (b) for emotion experiences averaged within each basic emotion category, each column shows the averaged ratings on the 18 scales and each row represents a combination of emotion category and stimulus type.

Applying hierarchical clustering to the combined set of basic emotions, I again found no prominent clustering based on stimulus type (Fig.6.7 a) though similarities within the same stimulus type were indicated by some blocks of rows of the same color.

I further averaged the different emotion experiences of the same category for each stimulus type and clustering on these averaged basic emotions allowed me to better examine the effect of stimulus domain (Fig.6.7 b). In general, emotions evoked by videos and stories were more similar compared to the ones in real life. The differences between stimuli-evoked emotions and real-life emotions were most prominent for the negative emotions, and less so for happiness and surprise. The separate cluster at the bottom revealed that the negative emotions (sadness, anger, fear, and disgust) in real life were less surprising, less arousing, more controlled, more common, and also purer than the ones evoked by stories or videos. These distinctions between real-life emotions and the stimuli-evoked emotions might be explained by a better understanding of one's own emotions with a more complete context.

6.4 Summary and discussion

In this chapter, I compared emotion experiences across three domains of stimuli. In terms of overall structure, the correlation matrices across scales were highly similar with Spearman rank correlations higher than 0.9. I identified two shared factors ("valence" and "arousal") across domains, in line with literature.

In terms of the actual emotion experiences, I found no domain-based clustering when mixing three types of emotions together. Focusing only on instances belonging to the six basic emotion categories, I reported two main findings. First, basic emotions were not perfectly separated regardless of domains, with happiness and surprise forming more discrete clusters compared to the negative emotions. Second, domain-specific differences were prominent for negative basic emotions, but not for happiness.

I have several limitations that I discuss below.

The first limitation, as already mentioned, was that I didn't use the complete set of scales for real-life emotions. This had direct impacts on the representational similarity analysis and the factor analysis.

Second, I collected emotion experiences of three domains in different experiment sessions, meaning that subjects only rated emotions evoked by stories, or emotions evoked by videos or real-life emotions for any given session. The data collection procedure made it difficult to directly compare emotions of different domains in terms of the magnitudes. For example, a story with a rating of 7 on the intensity scale when rated with other stories would probably have a lower rating when presented in the same session with videos that are more powerful at eliciting intense emotions.

Third, for a fair comparison, ideally I would like to sample all possible emotions that can be evoked by each stimulus type or at least a representative subset of them. In practice, however, it's difficult to control the quantity and diversity of the emotions for each domain. The impact of sampling depends on the research question. While it may be a big issue for characterizing distributions of emotions, it's not much of a concern for factor analysis as I have shown that the factors were remarkably robust with a smaller number of stimuli.

6.5 Supplementary information

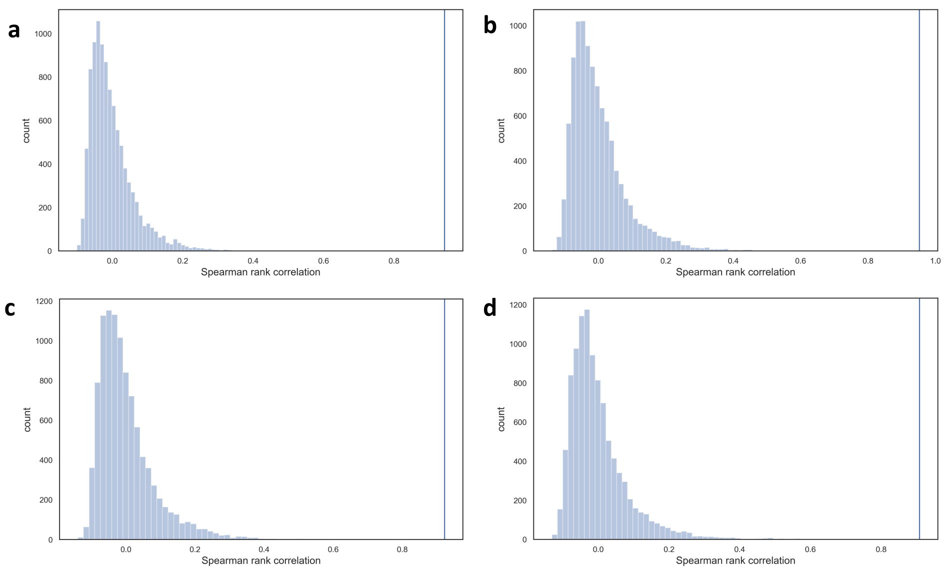

Figure S6.1: Testing relatedness of correlation matrices by randomization. Null distribution of correlations of two unrelated correlation matrices (simulated by randomization of one of the matrices across 10,000 iterations, so the smallest possible estimate is 0.0001) with the vertical line indicating actual correlation for (a) story and video (across 23 scales), (b) story and video (across 18 scales), (c) story and real-life (across 18 scales), and (d) video and real-life (across 18 scales).

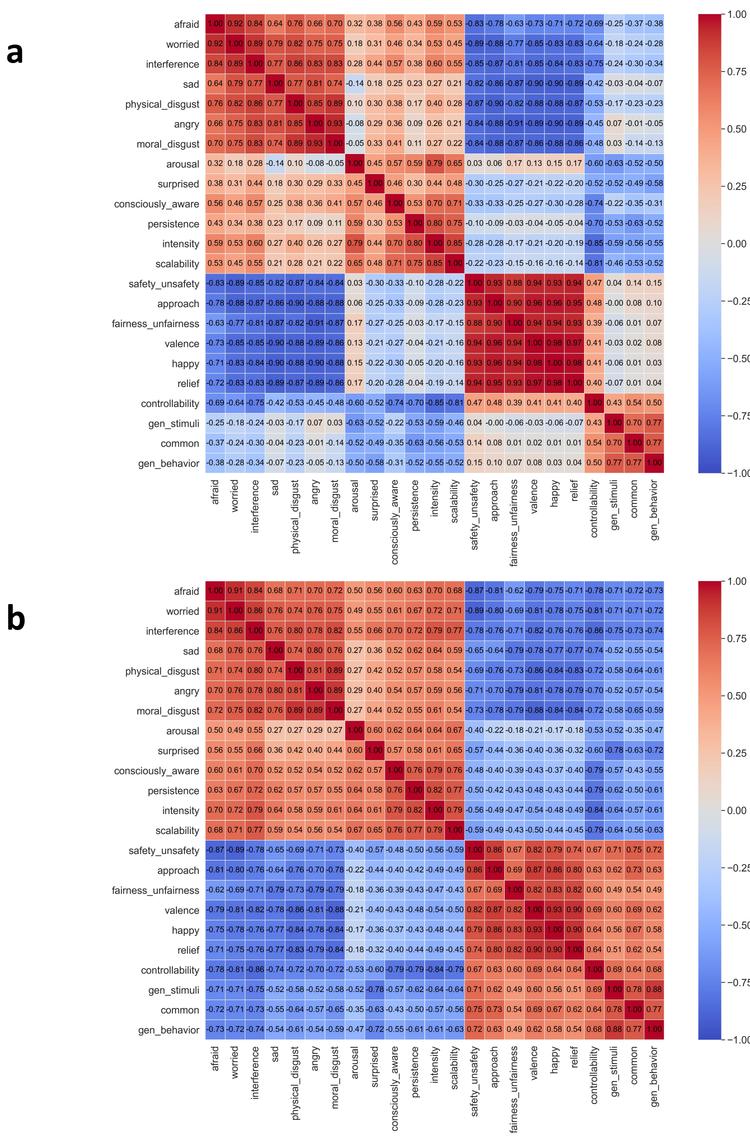

Figure S6.2: Correlation matrices across 23 scales for (a) story-evoked emotions and (b) video-evoked emotions, sorted based on story-evoked emotions

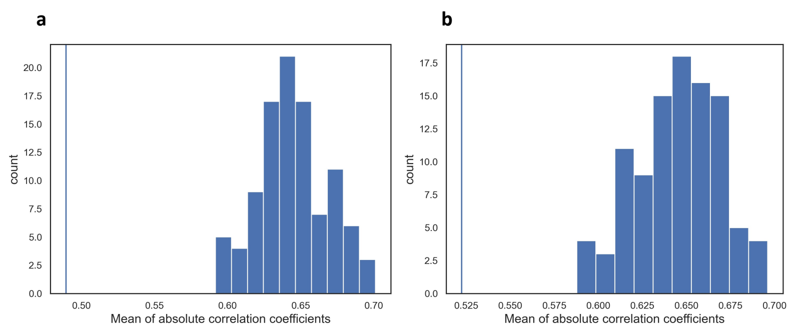

Figure S6.3: Distributions of the means of the absolute correlation coefficients from the correlation matrices for videos (subsampled 100 times to match the number of stories) with the vertical line indicating the mean of the absolute correlation coefficients for stories for (a) using 23 scales and (b) using 18 scales.

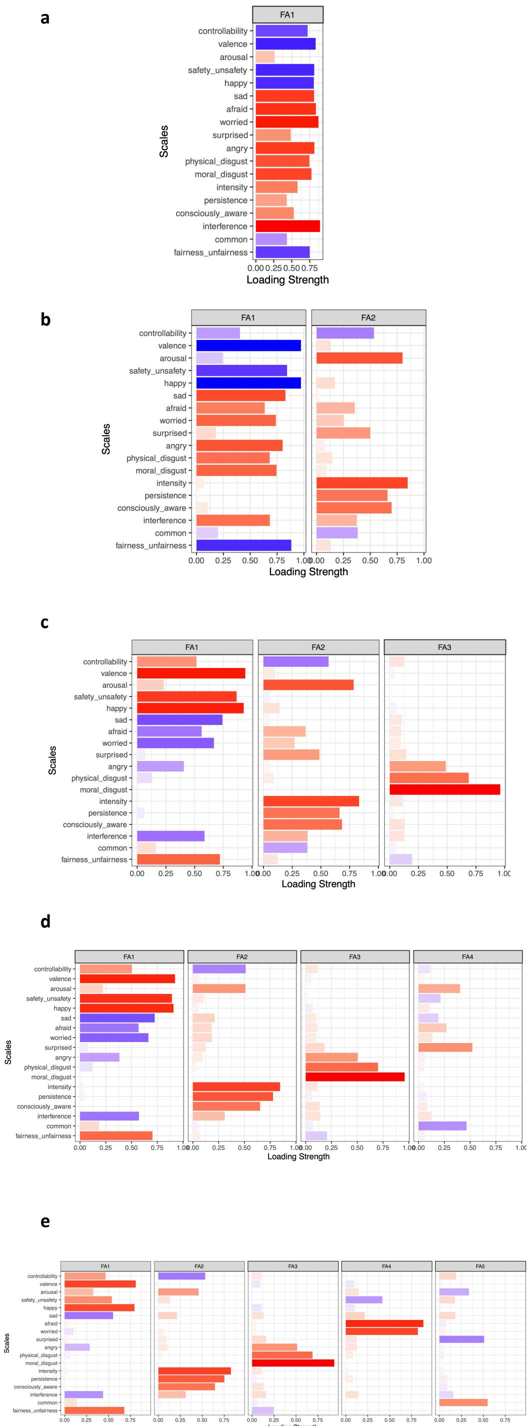

Figure S6.4: Factor loadings of scales on (a) the 1 factor, (b) the 2 factors, (c) the 3 factors, (d) the 4 factors, and (e) the 5 factors from EFA using the averaged correlation matrix. Each column plots the strength of the factor loadings (x-axis, absolute value) across scales (y-axis). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

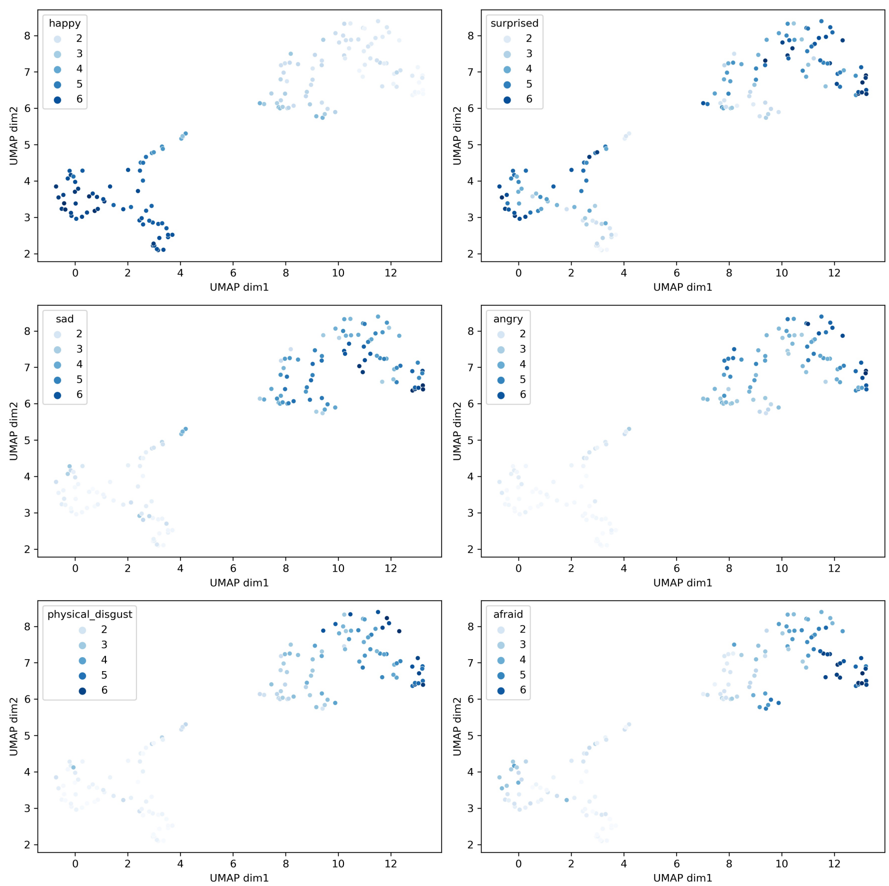

Figure S6.5: UMAP plots, color-coded for ratings on the six basic emotions for story data.

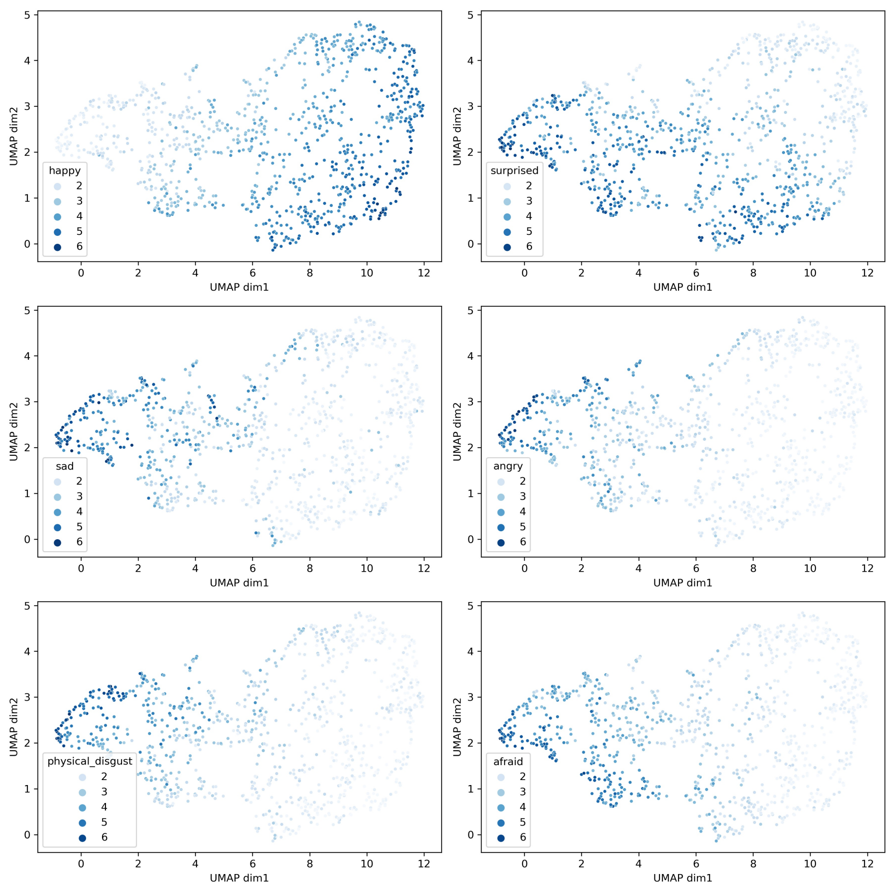

Figure S6.6: UMAP plots, color-coded for ratings on the six basic emotions for video data.

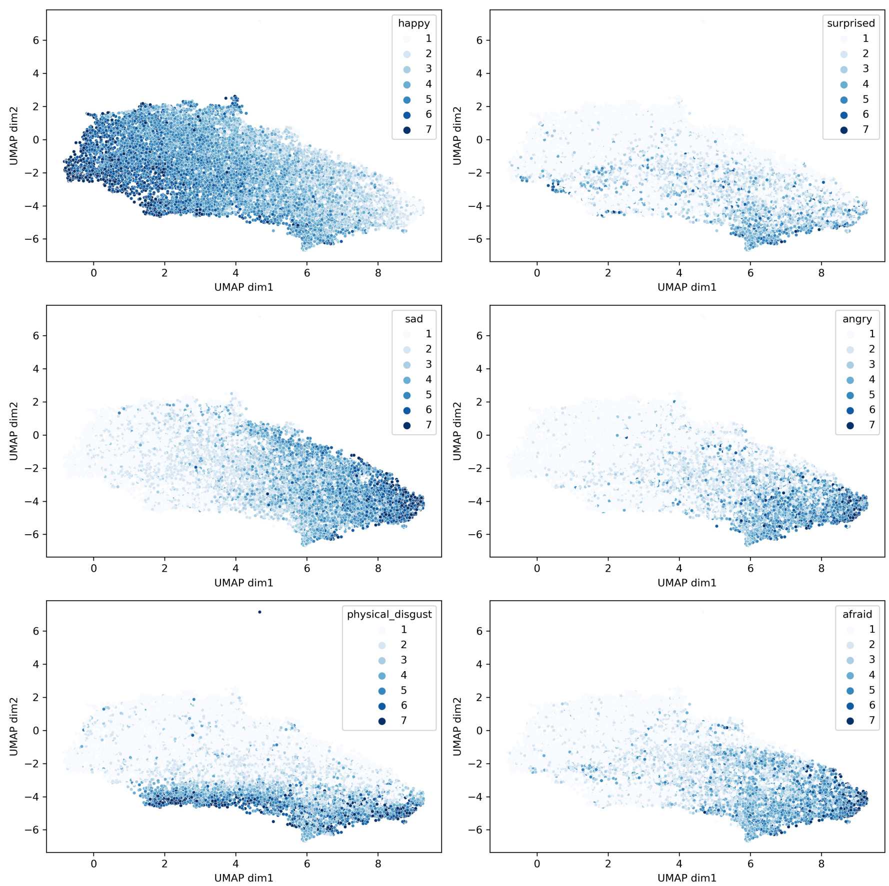

Figure S6.7: UMAP plots, color-coded for ratings on the six basic emotions for real-life data.

References

[1] Herbert A Simon. “Motivational and emotional controls of cognition.” In: *Psychological review* 74.1 (1967), p. 29.

[2] Michelle N Shiota. “Ekman’s theory of basic emotions”. In: *The Sage Encyclopedia of Theory in Psychology* (2016), pp. 248–250.

Chapter 7

INDIVIDUAL DIFFERENCES IN EMOTION EXPERIENCES

7.1 Introduction

In this chapter, I present the results of individual differences in emotion experiences, enabled by the wealth of psychological measures collected as part of the Covid Dynamic study. I begin by reviewing some of the relevant work on individual differences associated with the psychological variables available in my dataset.

Studies have reported age differences in emotion experiences, specifically, improved emotional well-being with age with fewer negative emotions and greater emotional control [1, 2]. Religion is believed to influence both the generation and regulation of emotions [3, 4].

Extensive research has been conducted on sex differences in emotions. Women were reported to have higher levels of emotional awareness [5]. Regarding the frequency of everyday emotions, cultural beliefs often associate women with powerless emotions (such as sadness and fear) and men with powerful emotions (such as anger) [6]. But some studies using self reports have found no sex differences in the averaged momentary ratings of emotions [7, 8]. A study using fMRI to study emotion regulation did find gender differences neurally, but not behaviorally [9]. In another study using film clips to evoke emotions, women were reported to display greater physiological responses for sadness[10].

Personality traits have also been linked with individual differences in emotion experiences, in particular, neuroticism and extraversion have been consistently associated with experiencing negative and positive affect respectively [11, 12, 13]. Extraversion scores robustly predict the frequency and intensity of positive emotions [14, 15]. In contrast, individuals with higher levels of neuroticism tend to experience more frequent and intense negative emotions, associated with a sense of uncontrollability [16, 17, 18]. Besides the big five personality traits, psychological resilience is also associated with positive emotions and lower levels of depression, anxiety, and burnout ([19, 20], see Chapter 8 for a more comprehensive review).

7.2 Individual differences in rating magnitudes

People differ in the emotions they experience in multiple ways. I start by asking whether there are meaningful differences in terms of the overall magnitudes of ratings for different stimuli and if so, how are those differences related to the demographic variables and psychological traits?

For emotions evoked by stories or videos, each experiment session involved rating roughly 75 stories or 100 videos on a single scale. I first centered each stimulus on its average rating, from all ratings of that stimulus on that scale across subjects and then computed the mean rating of a session for each subject. Eliminating the effect of stimuli allowed me to compare each subject's mean ratings across stimuli with different raw means. This also served as a general baseline that represented the bias of each subject.

For real-life emotions, I noticed meaningful temporal patterns across waves on a population level (Fig.7.1). Several interesting patterns emerged that possibly corresponded to real world events. One observation was that in general people felt less negatively over time (most evidently for scales such as "afraid" and "worried"). I also noticed a peak for several scales (such as "moral disgust") around wave 7, possibly explained by the incident of George Floyd. Another interesting peak happened around wave 14 (such as "surprised"), coinciding with the 2020 election (Fig.2.1). Given the temporal variations, I decided to only include subjects with complete data from wave 2 to 16 for the investigation of rating magnitude.

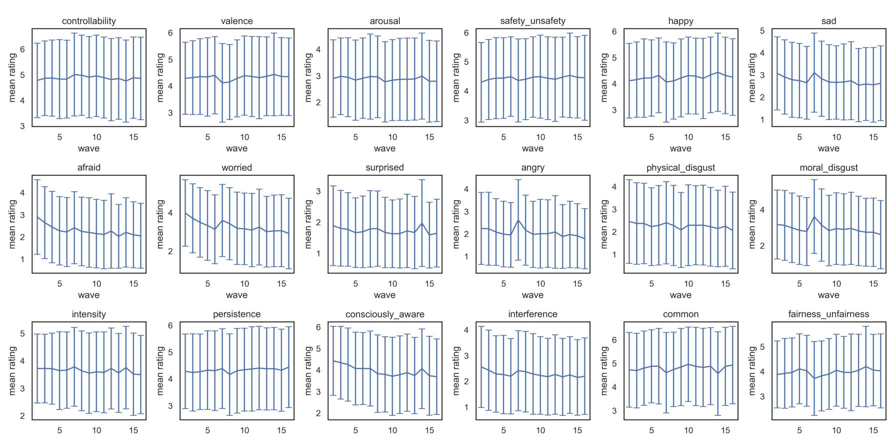

Figure 7.1: Means and standard deviations of ratings across subjects on each of the 18 scales for real-life emotions across administrative waves (wave 2 to wave 16, see figure 2.1 for dates of the waves and associated real-world events).

As part of the COVID-Dynamic dataset, I collected a rich set of psychological assessments at multiple time points (for a complete list, see [21]). I selected a subset of measures most relevant for emotion experiences and since they were generally stable over time, I averaged the multiple assessments across time for each measure.

I first correlated each subject's mean ratings for task-evoked emotions with the psychological traits and found very few significant associations after correcting for multiple testing (Fig.7.2 a, see Fig.S7.1 a for results without Bonferroni correction). This suggested that psychological traits had little effect on the emotions evoked by stories or videos.

I then asked the same question with real-life emotions. Unlike emotions evoked by stories and videos, several psychological traits were associated with people's real-life emotions even after Bonferroni correction (Fig.7.2 b, see Fig.S7.1 b for results without Bonferroni correction). The pattern also didn't change after correcting for baseline from tasks (Fig.7.2 c, see Fig.S7.1 c for results without Bonferroni correction). Positive personality traits (especially resilience) were associated with more positive real-life emotions while negative traits were associated with more negative real-life emotions.

Demographic variables, both the continuous ones (such as age and how religious a person is, Fig.7.2) and the categorical ones (such as sex and education level, analyzed using t tests, see results in Fig.S7.2) had largely no effect on experienced emotions in any of the domains.

It's worth noting that even when correlations between traits and mean ratings were both significant for task-evoked and real-life emotions, the effects were stronger for real-life emotions. For instance, correlations between resilience scores and mean ratings on the happy scale were 0.28 and 0.58 for task-evoked emotions and real-life emotions respectively.

It's also worth mentioning that the stronger effects were not due to larger sample size for real-life emotions as the correlation between resilience scores and mean ratings on the happy scale were 0.38 for real-life emotions when corrected for baseline from tasks which had the same sample size as the task-evoked ones and still a higher correlation.

There are at least two possible explanations for the finding that psychological traits had an effect on people's real-life emotions, but not on the emotions evoked by stories or by videos. One reason might be that the story and video stimuli were all the same for subjects whereas their real-life emotions were idiosyncratic. Relatedly, it could be that real-life emotions were taken more seriously and really interacted with individual differences, compared with artificial stimuli.

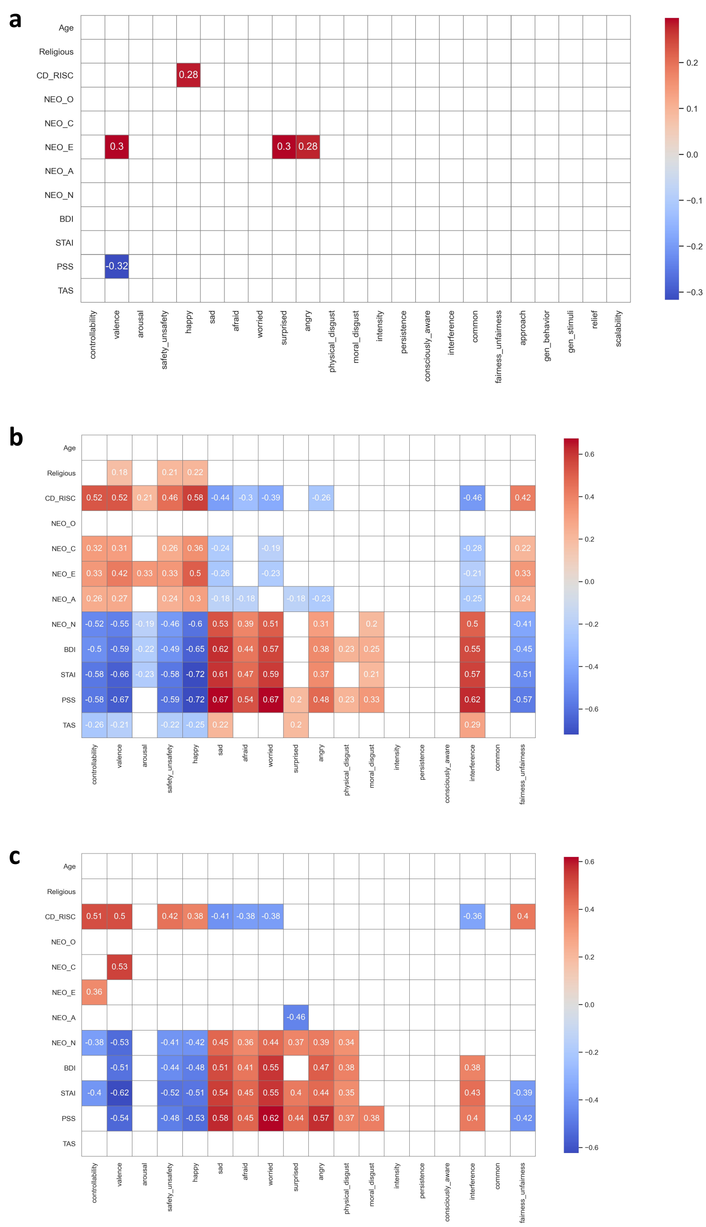

Figure 7.2: Pairwise correlations between ratings and traits (significant Pearson's correlation coefficients were shown, corrs with p >= 0.05 were omitted after Bonferroni correction) for (a) ratings for emotions evoked by tasks (stories and videos), (b) ratings for raw real-life emotions and (c) ratings for real-life emotions corrected for baseline ratings from tasks.

7.3 Individual differences in correlation structures

In addition to investigating whether psychological variables had an effect on people's actual emotion experiences, I also probed individual differences with respect to the psychological space as represented by the correlation structure across scales. This is a separate question because for instance, people may differ in how negative their emotions are, but can still share the same association between valence and intensity.

For most measures, I divided subjects into two groups based on the median values of the measures or a natural way of separation (such as females and males) and thus the sample size was roughly matched, with the exception of Beck Depression Inventory (BDI) and Toronto Alexithymia Scale (TAS) where meaningful cutoffs exist. Specifically, for BDI, I divided people into minimal and mild to severe depression groups using a cutoff score of 13 [22] and the low depression group was roughly twice the size of the high depression group. For TAS, the non-alexithymia group and possible alexithymia groups were divided using a cutoff score of 51 [23] and the non-alexithymia group was roughly three times the size of the alexithymia group.

I then derived correlation matrices across scales for each group and for each stimulus type and then calculated second-order similarity by correlating the correlation matrices. When subjects were separated based on demographic variables (sex, age, education, and religious level), no strong group differences were observed across all three domains (Fig.S7.3).

Negative traits which had an effect on mean ratings for real-life emotions, also affected the correlation structures across scales for real-life emotions only, but not for emotions evoked by stories or videos (Fig.7.3).

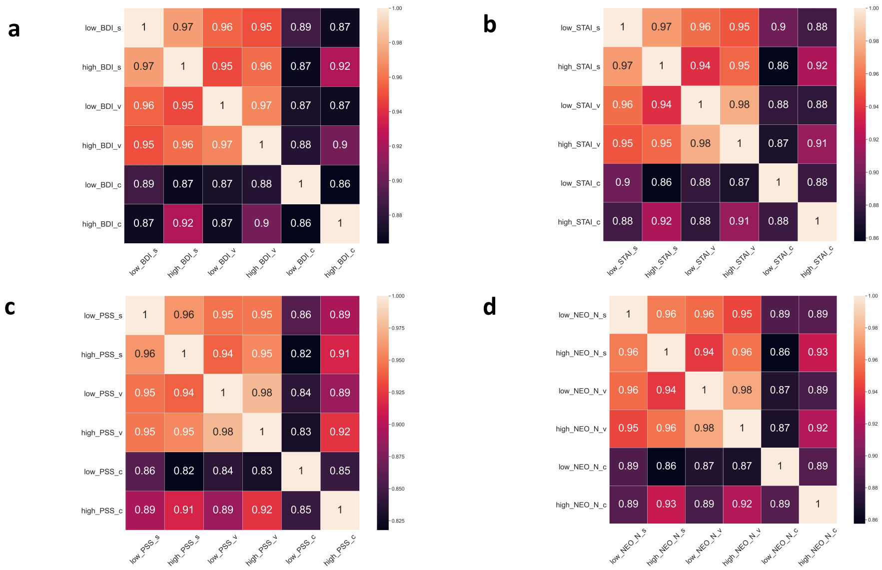

Figure 7.3: Representational similarity across scales and across groups for groups defined by (a) BDI, (b) STAI, (c) PSS, and (d) NEO Neuroticism. The cells in each matrix represent Spearman's rank correlations between two correlation matrices across scales. The order of the cells (from top to bottom): two groups for emotions evoked by stories, two groups for emotions evoked by videos and two groups for real-life emotions.

More specifically, if I compared the correlation matrices for the low and high depression groups, there were notable differences in the following scales: persistence, intensity, consciously aware, and common (Fig.7.4). The correlations revealed that for depressed people, they found the negative emotions in real life to be more persistent, more intense, more consciously aware, and more common.

Among positive personality traits, resilience and extraversion (but not openness, agreeableness and conscientiousness) had an effect on the emotion structure but again the effects were restricted to real life only (Fig.7.5). Closer inspection of the correlation matrices revealed resilience had the exact opposite effect of depression (Fig.S7.4).

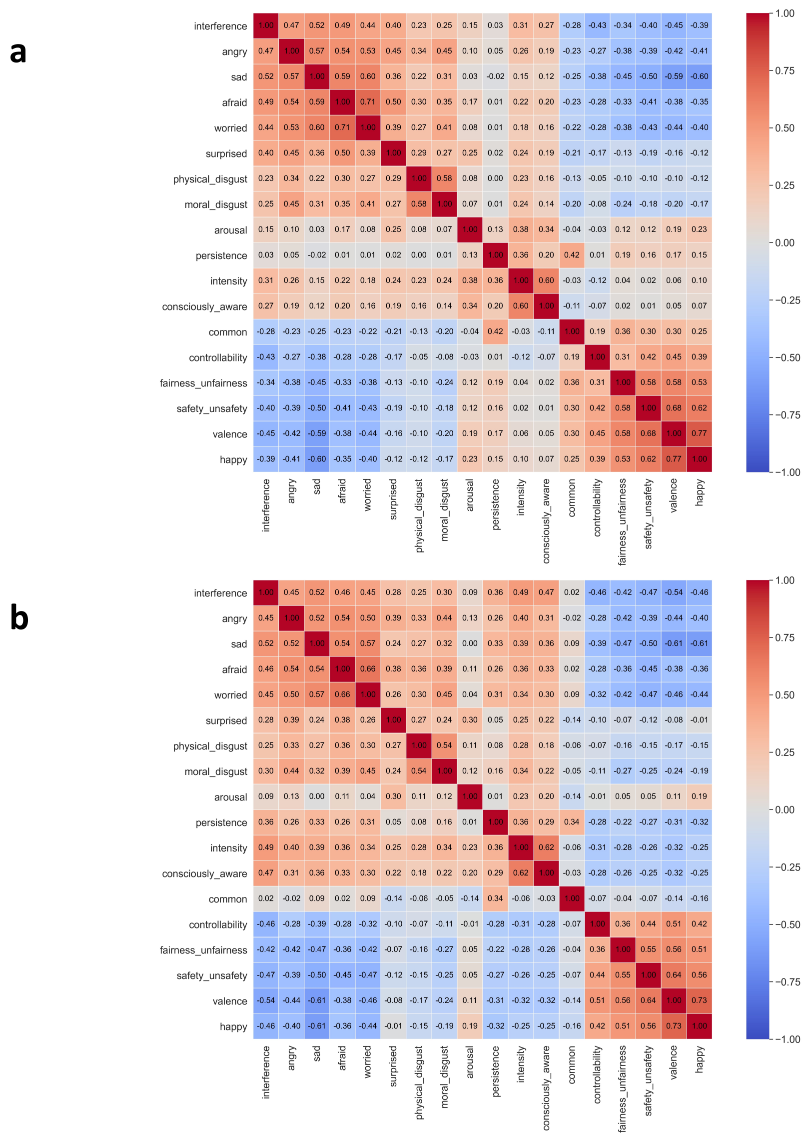

Figure 7.4: Correlation matrices across scales for real-life emotions for (a) people with minimal depression, and (b) people with mild to severe depression.

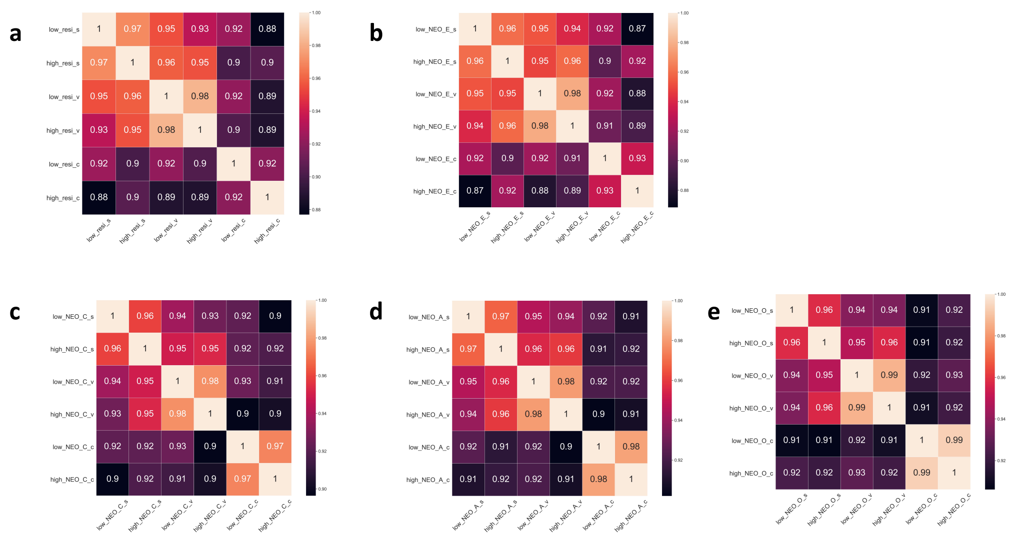

Figure 7.5: Representational similarity across scales and across groups for groups defined by (a) CD-RISC, (b) NEO Extraversion, (c) NEO Conscientiousness, (d) NEO Agreeableness and (e) NEO Openness. The cells in each matrix represent Spearman's rank correlations between two correlation matrices across scales. The order of the cells (from top to bottom): two groups for emotions evoked by stories, two groups for emotions evoked by videos and two groups for real-life emotions.

The Toronto Alexithymia scale (TAS) revealed the biggest group difference for emotions evoked by stories across three domains of stimuli(Fig.7.6). The most notable differences were found for the following scales: arousal, scalability, gen_behavior, gen_stimuli, and common (Fig.7.7). One reasonable concern was that the differences were due to the mismatch of sample size between groups, which I tested by subsampling the non-Alexithymia group and verified against it (Fig.S7.5).

One possible explanation for the group difference that I observed on these specific scales is that they were more difficult to rate, as indicated by the quality metrics (see Fig.2.7: test-retest reliability and Fig.2.8: split half reliability for all the scales). This seems most related to one component of alexithymia, that is, the difficulty in describing feelings [24]. So people with higher levels of alexithymia failed to elaborate their emotions on the scales intended to capture finer differences between emotions, but performed normally on the simpler scales.

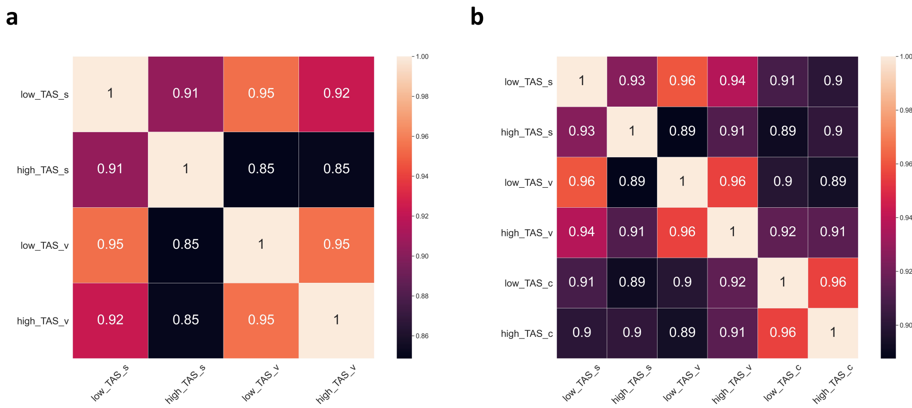

Figure 7.6: Representational similarity across scales and across groups defined by TAS. (a) using correlation matrices across 23 scales for stories and videos, and (b) using correlation matrices across 18 scales for all three domains. The cells in each matrix represent Spearman's rank correlations between two correlation matrices across scales. The order of the cells (from top to bottom): two groups for emotions evoked by stories, two groups for emotions evoked by videos and two groups for real-life emotions (only for the 18 scales on the right).

It's worth noting that the arousal scale had the most significant difference among all scales (Fig.7.7), and the extent varied across stimulus domains. It was most prominent for emotions evoked by stories and least for emotions in real life. Therefore, it's possible that subjects with alexithymia didn't have difficulty judging how arousing emotions are per se, instead, they may have more difficulty experiencing or identifying emotions, especially when induced by less effective stimuli such as stories. In fact, I asked subjects to imagine themselves as the characters in the stories experiencing an emotion-eliciting event, which could be particularly difficult for people with Alexithymia as they have deficits in imagery ability [25].

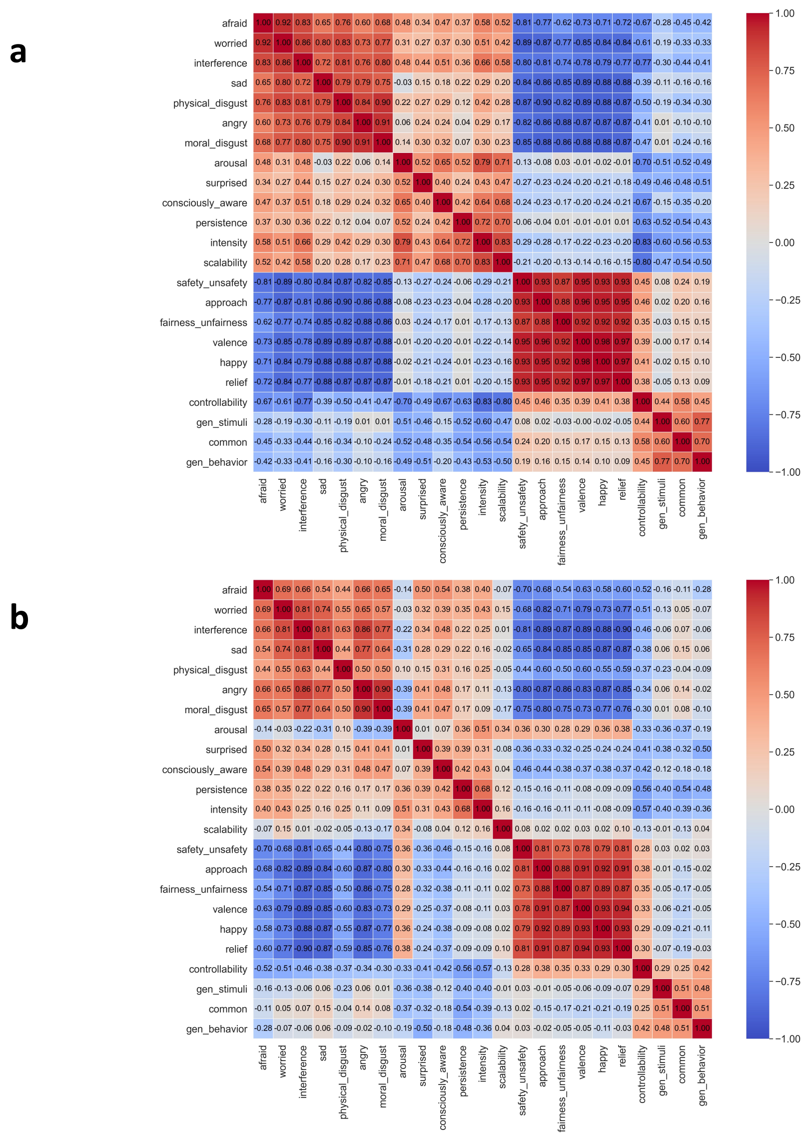

Figure 7.7: Correlation matrices across scales for emotions evoked by stories for (a) non-Alexithymia group, and (b) (possible) Alexithymia group.

7.4 Specific questions

In addition to the two broad questions of individual differences with respect to the mean magnitude of ratings and the correlation structures, I also investigated two specific questions of interest.

Do people with higher education rate the scales better?

To give reliable and accurate ratings on my scales, one needs to have reasonably good verbal intelligence and understanding the scales, especially for the scales with higher levels of semantic difficulty. I wondered if subjects with higher education would rate the scales, especially the difficult ones better than subjects with lower education, as evaluated by test-retest reliability.

Within my sample, subjects differed in their education background. Specifically, on the lower end, I had 7 people with some high school education and 111 with high school education. On the higher end, 140 people have a master's degree and 17 have a PhD.

The distributions of test-retest reliabilities across scales for these two groups revealed that except for the intrinsic_extrinsic scale, the two groups of different education levels didn't seem to differ in their ability to produce reliable ratings on my scales (Fig.7.8). The intrinsic_extrinsic scale was the most difficult scale that I had (see definition in Table 2.1), which required a grade level of 20.48 corresponding to a school level beyond college while most of my scales required a grade level of 12 or lower corresponding to a high school education. Therefore, it's not surprising that the group of subjects with the master's and PhD degrees did better for the intrinsic_extrinsic scale. On the other hand, the fact that the higher education group didn't outperform the lower education group on the other scales was reassuring because it indicated good understanding of my scales for the whole sample.

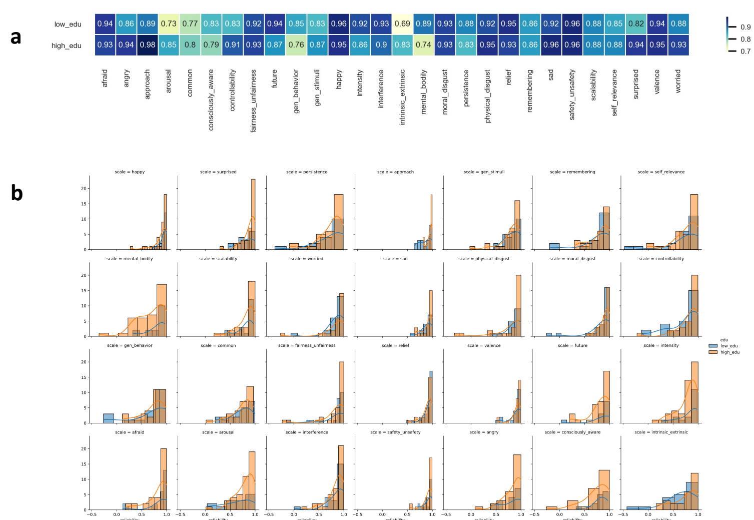

Figure 7.8: Test-retest reliability for each scale. (a) median value for each scale for each group, and (b) histograms for each scale, color coded for different groups.

Did people who tested positive for COVID have different real-life emotions?

Another specific question of interest was to see if subjects tested positive for COVID-19 had significantly different emotion experiences in real life.

Within my sample, 35 people out of 1000 self-reported to have tested positive for COVID-19 after exclusion. The positive percentage in my sample is considerably lower than what's been estimated for the US population (about one third of the US population had been infected by the end of 2020 [26]). Possible explanations for the discrepancy could be that subjects were not aware of being positive (for example, being asymptomatic), or lack of testing despite being positive or they were unwilling to disclose the information and more. Since there's no way to verify those possibilities, I assumed the subjects to have never been positive with covid unless they reported otherwise.

Given the covid negative group outnumbered the covid positive group significantly, I subsampled the covid negative population to match that of the positive group (1000 times) and computed mean ratings on the 18 scales for different groups, across waves and across subjects.

The results revealed that subjects who have tested positive for covid had significantly different emotion experiences in some aspects (Fig.7.9). Specifically, their emotions were more persistent, more aware of and lasted longer. Also, they were more morally disgusted towards acts of violating social norms than people who haven't got covid. The moral disgust scale was the one with the highest loading for the "negative affect" factor that I identified, supporting the idea of the "negative affect" factor being COVID-specific.

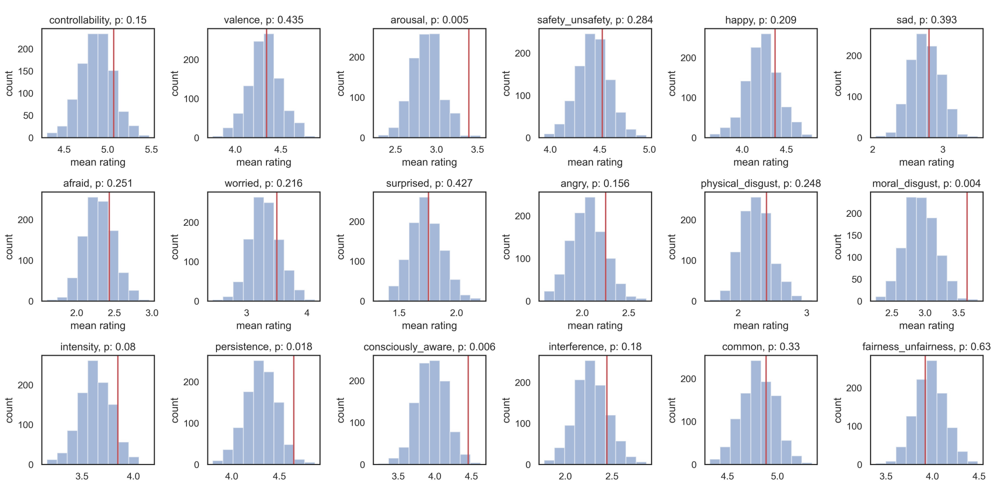

Figure 7.9: Distributions of mean rating across waves and across subjects on each of the 18 scales for the covid negative group (sampled 1000 times to match the number of subjects of the covid positive group), the red line indicates the mean rating of the covid positive group.

7.5 Summary and discussion

In this chapter, I investigated how individuals differ both in terms of the overall magnitudes of ratings and the psychological space as represented by the correlation structure across scales.

I didn't find support for individual differences related to demographics variables (such as sex and age). Negative traits, for instance neuroticism, were associated with more negative emotions in real-life (but not ones evoked by stories or videos). Neurotic individuals frequently experienced intense and persistent negative emotions. The opposite pattern was found for positive traits, such as extraversion and resilience. I observed interesting differences between the low and high alexithymia groups, in particular, with the correlation structure for emotions evoked by stories.

My investigation of individual differences with respect to the correlation structure across scales has been carried out in a dichotomous manner so far. This was mainly limited by the data sparseness issue that I touched on before. I was unable to construct a correlation matrix across scales for each subject for emotions evoked by stories and videos and therefore can't associate the conceptual space with psychological traits continuously.

However, for real-life emotions, as subjects rated their emotions on all scales across multiple waves, it's possible to construct the correlation structure at the individual level and is a future direction of mine. It would be interesting to see whether the four factors that I identified using aggregate data are preserved at the individual level. For the negative affect factor, for instance, my expectation would be that it might be absent for those less impacted by the pandemic.

7.6 Supplementary information

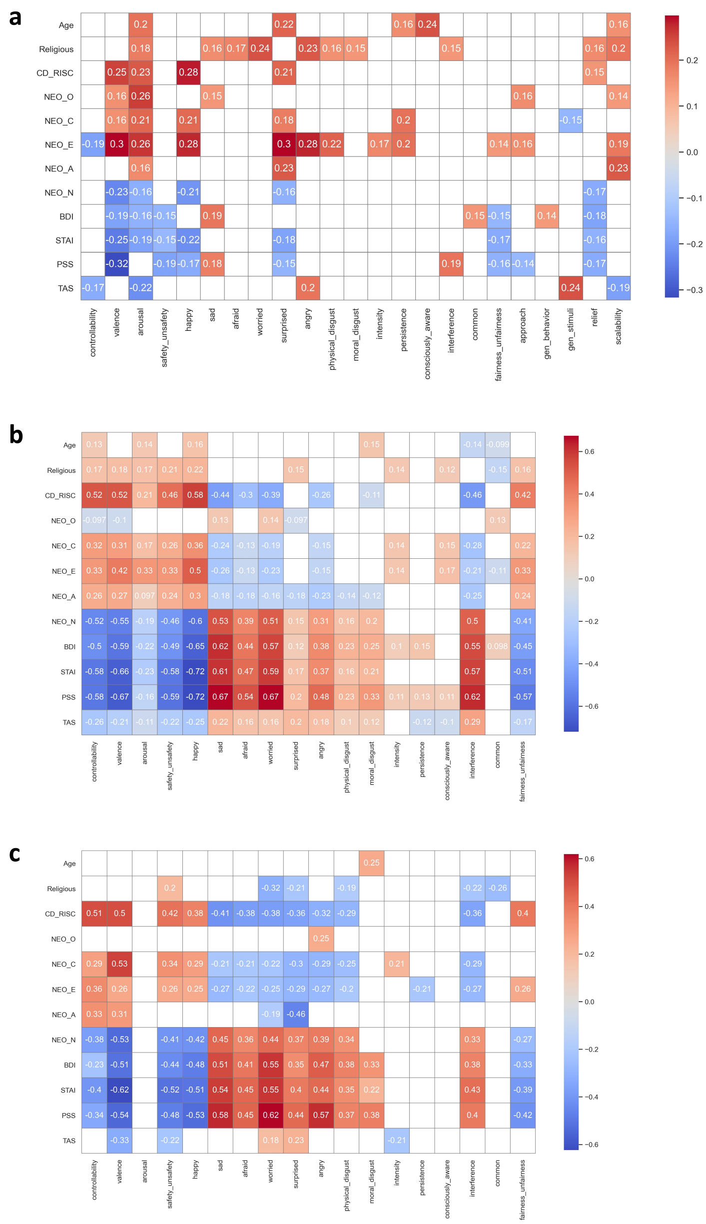

Figure S7.1: Pairwise correlations between ratings and traits (significant Pearson's correlation coefficients were shown, corrs with p >= 0.05 were omitted without Bonferroni correction) for (a) ratings for emotions evoked by tasks (stories and videos), (b) ratings for raw real-life emotions, and (c) ratings for real-life emotions corrected for baseline ratings from tasks.

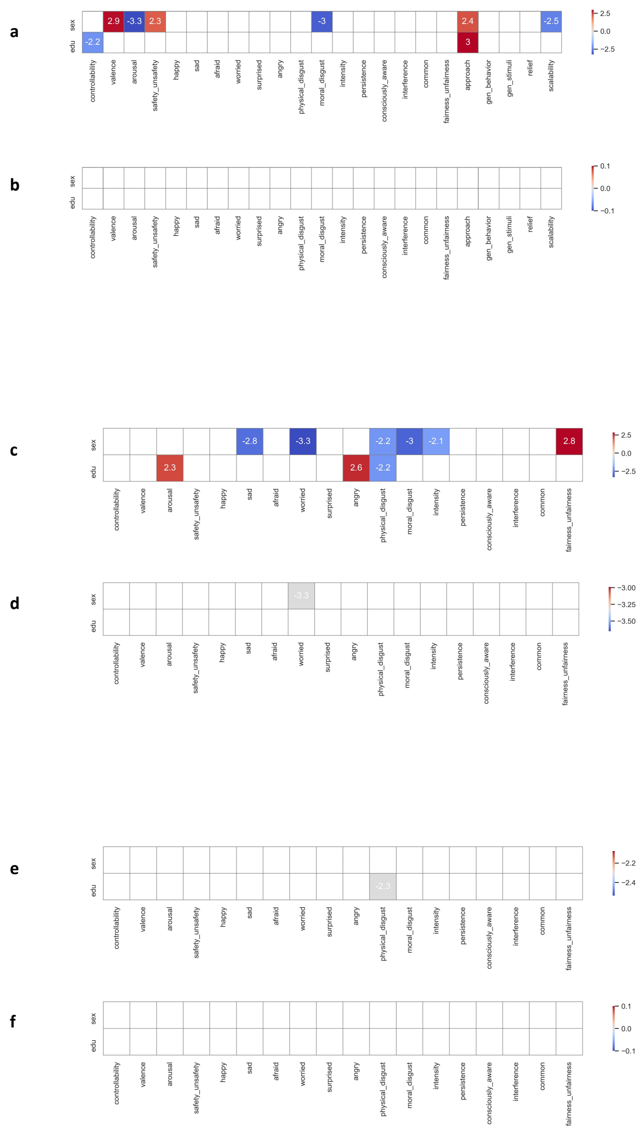

Figure S7.2: Welch's t-test for means of ratings of different groups (divided based on sex and education). T-statistics (male - female and high - low education for the two groups respectively) with significant significant p < 0.05 were shown, insignificant results were omitted for (a) ratings for emotions evoked by tasks (stories and videos) without Bonferroni correction, (b) ratings for emotions evoked by tasks (stories and videos) after Bonferroni correction, (c) ratings for raw real-life emotions without Bonferroni correction, (d) ratings for raw real-life emotions after Bonferroni correction, (e) ratings for real-life emotions corrected for baseline ratings from tasks without Bonferroni correction, and (f) ratings for real-life emotions corrected for baseline ratings from tasks after Bonferroni correction.

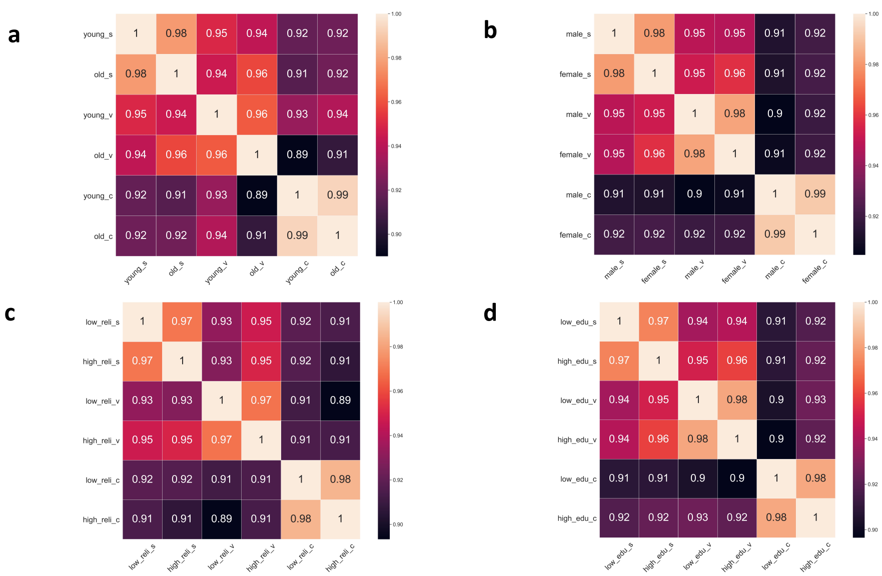

Figure S7.3: Representational similarity across scales and across groups for groups defined by (a) age, (b) sex, (c) religious level, and (d) education. The cells in each matrix represent Spearman's rank correlations between two correlation matrices across scales. The order of the cells (from top to bottom): two groups for emotions evoked by stories, two groups for emotions evoked by videos, and two groups for real-life emotions.

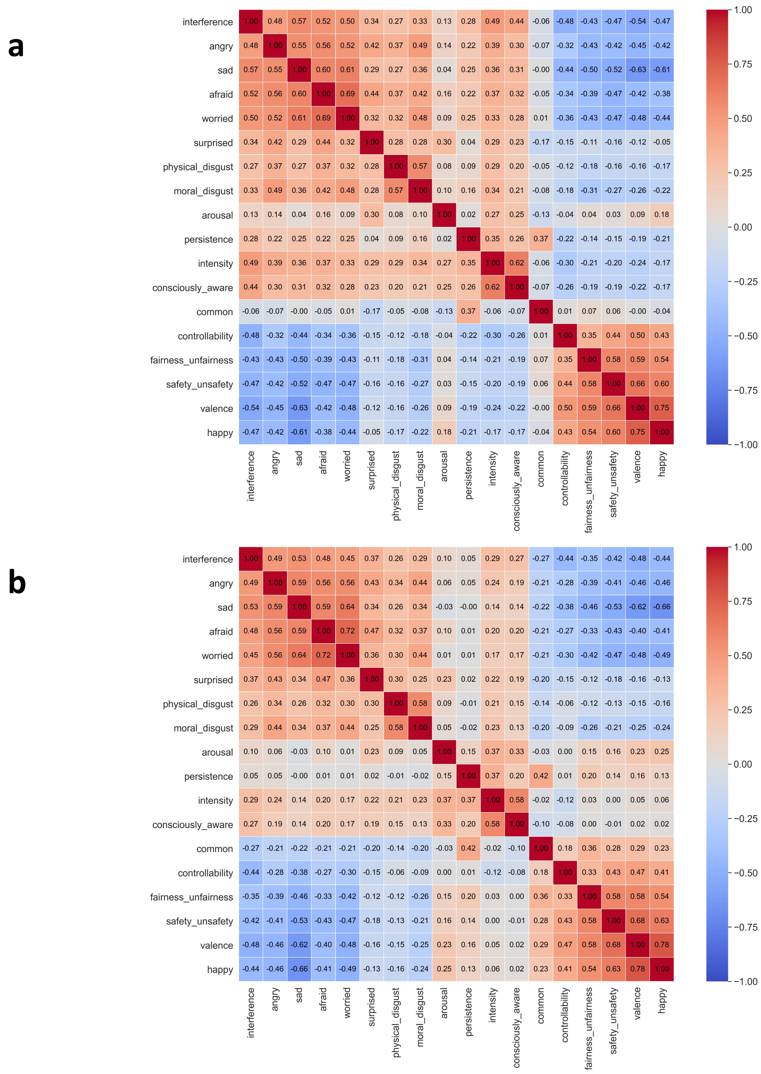

Figure S7.4: correlation matrices across scales for real-life emotions for (a) low resilience group and (b) high resilience group.

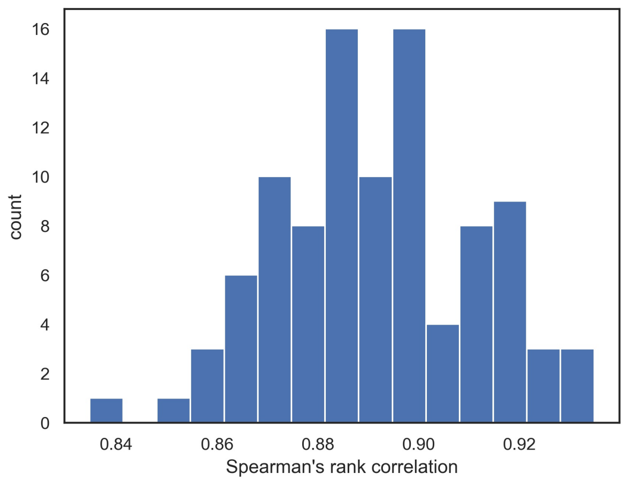

Figure S7.5: Distributions of the spearman correlations between the correlation matrices across 23 scales for emotions evoked by stories for the alexithymia group and the non-alexithymia group (sub-sampled to match the alexithymia group, 100 times).

References

[1] James J Gross et al. “Emotion and aging: experience, expression, and control.” In: *Psychology and aging* 12.4 (1997), p. 590.

[2] Laura L Carstensen et al. “Emotional experience improves with age: evidence based on over 10 years of experience sampling.” In: *Psychology and aging* 26.1 (2011), p. 21.

[3] Allon Vishkin, Yochanan Bigman, and Maya Tamir. “Religion, emotion regulation, and well-being”. In: *Religion and spirituality across cultures*. Springer, 2014, pp. 247–269.

[4] Robert A Emmons. “Emotion and religion”. In: *Handbook of the psychology of religion and spirituality* 24 (2005), pp. 235–52.

[5] Lisa Feldman Barrett et al. “Sex differences in emotional awareness”. In: *Personality and Social Psychology Bulletin* 26.9 (2000), pp. 1027–1035.

[6] Agneta H Fischer et al. “Gender and culture differences in emotion.” In: *Emotion* 4.1 (2004), p. 87.

[7] Robin W Simon and Leda E Nath. “Gender and emotion in the United States: Do men and women differ in self-reports of feelings and expressive behavior?” In: *American journal of sociology* 109.5 (2004), pp. 1137–1176.

[8] Lisa Feldman Barrett et al. “Are women the “more emotional” sex? Evidence from emotional experiences in social context”. In: *Cognition & Emotion* 12.4 (1998), pp. 555–578.

[9] Kateri McRae et al. “Gender differences in emotion regulation: An fMRI study of cognitive reappraisal”. In: *Group processes & intergroup relations* 11.2 (2008), pp. 143–162.

[10] Cristina Fernández et al. “Physiological responses induced by emotion-eliciting films”. In: *Applied psychophysiology and biofeedback* 37.2 (2012), pp. 73–79.

[11] William Revelle and Klaus R Scherer. “Personality and emotion”. In: *Oxford companion to emotion and the affective sciences* 1 (2009), pp. 304–306.

[12] Carroll E Izard et al. “Stability of emotion experiences and their relations to traits of personality.” In: *Journal of personality and social psychology* 64.5 (1993), p. 847.

[13] Rainer Reisenzein, Andrea Hildebrandt, and Hannelore Weber. “Personality and emotion.” In: (2020).

[14] Leslie G Eaton and David C Funder. “Emotional experience in daily life: valence, variability, and rate of change.” In: *Emotion* 1.4 (2001), p. 413.

[15] Michelle N Shiota, Dacher Keltner, and Oliver P John. "Positive emotion dispositions differentially associated with Big Five personality and attachment style". In: *The journal of positive psychology* 1.2 (2006), pp. 61–71.

[16] Thomas A Widiger. "Neuroticism." In: (2009).

[17] Elise K Kalokerinos et al. "Neuroticism may not reflect emotional variability". In: *Proceedings of the National Academy of Sciences* 117.17 (2020), pp. 9270–9276.

[18] David H Barlow et al. "The origins of neuroticism". In: *Perspectives on Psychological Science* 9.5 (2014), pp. 481–496.

[19] Christian T Gloria and Mary A Steinhardt. "Relationships among positive emotions, coping, resilience and mental health". In: *Stress and Health* 32.2 (2016), pp. 145–156.

[20] Georgia Libera Finstad et al. "Resilience, Coping Strategies and Posttraumatic Growth in the Workplace Following COVID-19: A Narrative Review on the Positive Aspects of Trauma". In: *International Journal of Environmental Research and Public Health* 18.18 (2021), p. 9453.

[21] Tessa Rusch et al. "COVID-Dynamic: A large-scale multifaceted longitudinal study of socioemotional and behavioral change across the pandemic". In: (2021).

[22] Aaron T Beck, Robert A Steer, Gregory K Brown, et al. *Manual for the beck depression inventory-II*. 1996.

[23] R Michael Bagby, James DA Parker, and Graeme J Taylor. "The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure". In: *Journal of psychosomatic research* 38.1 (1994), pp. 23–32.

[24] Graeme J Taylor, R Michael Bagby, and James DA Parker. *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge University Press, 1999.

[25] Catherine E Peasley-Miklus, Georgia Panayiotou, and Scott R Vrana. "Alexithymia predicts arousal-based processing deficits and discordance between emotion response systems during emotional imagery." In: *Emotion* 16.2 (2016), p. 164.

[26] Sen Pei et al. "Burden and characteristics of COVID-19 in the United States during 2020". In: *Nature* 598.7880 (2021), pp. 338–341.

Chapter 8

TRAIT RESILIENCE PROTECTS AGAINST DEPRESSION CAUSED BY LONELINESS DURING THE COVID PANDEMIC

The following chapter is adapted from Han, Y., & Adolphs, R. (2022). "Trait resilience protects against depression caused by loneliness during the COVID pandemic." (https://doi.org/10.31234/osf.io/9dac6) *[Under review at Affective Science]*. The content is modified according to the format of a Caltech Thesis.

Abstract

We hypothesized that resilience would buffer people from depression caused by loneliness and social isolation during the COVID pandemic. Capitalizing on a unique longitudinal dataset of 447 American adults, we used well established self-report instruments to find that resilience at time 1 buffered individuals against the effects of loneliness at time 2 causing depression at time 3. Effects were robust across age, sex, and education level, and generalized to trait variables we believe are partly constitutive of resilience: conscientiousness, extraversion, and (negatively) neuroticism. However, our results were relatively specific to depression as the outcome, and did not generalize to other adverse outcomes, such as stress and anxiety. Future studies can use the open dataset on which our study is based together with new resilience factors that we propose in order to further test the interventional potential of our findings.

8.1 Introduction

Resilience (Latin: *resilire*, to spring back) refers to an individual's or community's ability to maintain or recover mental health despite challenges from adverse events, a feature most clearly highlighted in theories of resilience that focus on stress and trauma [1]. Resilience has been construed as a trait (e.g., possibly corresponding to a particular polygenic profile), a dynamic process (e.g., the active coping process itself), or even an outcome (e.g., a better outcome in the face of a stressor is sometimes simply defined as resilience) [2]. These three construals are typically closely related (indeed, on some treatments, resilience subsumes all three of the above definitions [3]). Some studies incorporate resilience into the dynamics of an effect—that is, resilience is conceptualized as a state variable, for instance causally influenced by loneliness [4].

Here we specifically treat resilience as a latent trait variable: it predisposes individuals to cope better and to have better mental health outcomes, but it is causally distinct from either of these. This conceptualization allowed us to test a specific causal model, leveraging a unique longitudinal dataset in which resilience, loneliness, and depression could be temporally separated in epochs corresponding to the causal model. The COVID-19 pandemic has been a unique and global stressor, offering a natural test-bed for the protective effects of resilience [5]. We capitalized on a longitudinal dataset acquired during the COVID pandemic [6] in order to test a specific causal model: that resilience protects against the deleterious effects of loneliness and social isolation in causing depression (Fig. 8.1a).

Loneliness is usually conceptualized as a discrepancy between a person's actual social interactions and their desired ideal, resulting in a negatively valenced emotional experience: the number and/or quality of social interactions are insufficient, for that person [7]. While there are enormous individual differences here, with some people needing intense social interaction while others prefer solitude, the need for some social interaction is thought to be a fundamental aspect of human nature (shared with other social animals) [8]. Loneliness appears to be surprisingly impervious to interventions, with generally small-moderate effect sizes across studies [9]. Importantly, loneliness and depression are distinct, with loneliness as a separable risk factor for depression (the core causal model whose moderation by resilience we tested in the present study) [10, 11].

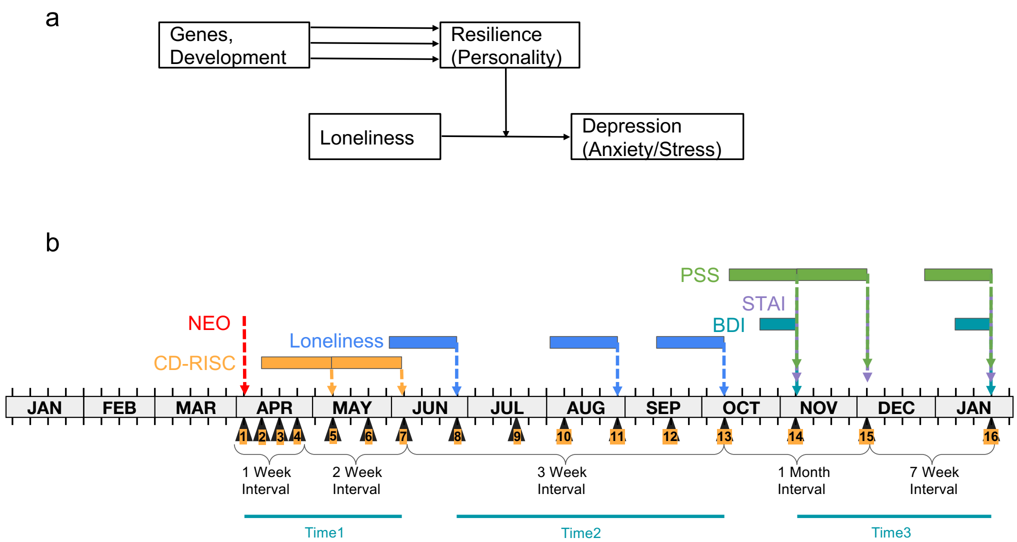

Figure 8.1: Causal model and selection of time windows. **a** The model that we tested. We did not investigate the causal antecedents of resilience but capitalized on the effect of social isolation during the COVID pandemic in testing how loneliness (measured at Time 2) could have an effect on depression (measured at Time 3), possibly moderated by resilience (measured at Time 1, but empirically stable across all time windows). By sequencing our measures in time, within the same subject sample, and by ensuring relative temporal stability (and/or low measurement error) for our variables, we were able to provide a stronger causal inference. **b** Timeline of the wave administrations and our selection of time windows. Black triangles denote each wave administration with varying time intervals. Vertical arrows indicate the selected waves of data collection for different measures, while horizontal bars indicate the temporal range covered by each assessment (NEO: subjects were instructed to answer generally with no timeframe, STAI: subjects were instructed to answer their feelings at the moment). Note that we only indicated the subset of waves selected for this study, total number of waves available for each measure are CD-RISC: n = 6; NEO: n = 6; NIH-Loneliness: n = 8; BDI: n = 7; STAI: n = 16; and PSS: n = 15.

The COVID pandemic clearly provided an acute change in social isolation, with loneliness prominently increased [12, 13]. Of studies specifically during the COVID pandemic, one examined loneliness and resilience using separate instruments [14]. However, a single cross-sectional data collection was undertaken and no causal model was tested. Another study found an association between resilience (measured with the Chinese version of the Connor-Davidson resilience scale) and depression—but was again cross-sectional and did not test a causal model [15]. A few other studies have also probed resilience during the COVID pandemic [16, 17](see [18] for a review), but all remain cross-sectional. Similarly cross-sectional are studies outside the scope of the COVID pandemic, showing that resilience protects against the effects of loneliness on mental health in the elderly [19], in homeless people [20], and in students [21]: all demonstrate the expected associations, but none permit strong causal inferences.

Resilience can result in substantial individual differences in mental health outcomes to stressors, but is a complex construct often viewed more dynamically than we do here [22]. We here focus on an individual's ability to endure an adverse event of some duration while buffering against mental illness [23]. A protective role for resilience in the effects of a stressor on depression has been studied in a number of studies; for instance, the effect childhood adversity on later depression in adulthood is influenced by resilience [24]. Polygenic risk scores for depression are mediated by both resilience and neuroticism (as separately measured variables)[25], perhaps one of the clearest role for resilience in protecting against depression within a causal model. We take resilience to be a latent psychological variable that shows substantial individual differences, but that is relatively trait-like (i.e. temporally stable) within an individual, at least under the conditions of our study.

A widely used, psychometrically very extensively validated instrument to assess such a resilience construct is the Connor-Davidson resilience scale (CD-RISC), whose original 25-item self-report instrument [26] is well correlated with the shorter 10-item version that we used here [27]. The CD-RISC provides high reliability and validity across a very large number of studies and across cultures [28]. Of particular importance, given that it is a self-report instrument, validity is borne out by relations to many other measures. For instance, CD-RISC scores moderate depression in people exposed to childhood trauma [29], and lower CD-RISC scores are associated with increased risk of PTSD [30], postpartum depression [31], and suicide [32, 33]. Similarly, higher CD-RISC scores have been associated with better outcomes following serious illness, such as spinal cord injury [34] or traumatic brain injury [35]. In relation to positive outcomes, higher CD-RISC scores have also been associated with increased response readiness in paramedics [36], positive affect [37] and superior communication ability [38] in nurses, better relationships in couples who had combat exposure [39], and more successful aging [40]. Of particular interest, CD-RISC scores have been shown to protect against depression and PTSD following specific disasters such as earthquakes [41, 42], tsunamis [43], and oil spills [44]. As one might expect, resilience is correlated with personality factors: neuroticism (r=-0.47), extraversion (r=0.43), openness (r=0.27), agreeableness (r=0.36), and conscientiousness (r=0.64) from the NEO-FFI were found to correlate with CD-RISC scores in a Chinese study [45], as well as in one of the original validations of the CD-RISC (neuroticism: r=-0.65; extraversion: r=0.61; openness: r=0.20; agreeableness: r=0.15; conscientiousness: r=0.46) [46]. Our use of the CD-RISC here as measuring a trait-like variable is further supported by the high test-retest reliability of this instrument, which ranges from 0.7-0.9 [47, 26, 48, 49, 50], showing remarkably stable scores over epochs from months [51] to years [52, 53].

Premises built into our study are schematized in Fig. 8.1. In each case, we aimed to have the variable in question be temporally stable over the time window of interest; this also permitted a more precise estimation of the latent variable from the average of its sampled measures.

8.2 Methods

8.2.1 Data

Data used in this study were acquired in the COVID-Dynamic longitudinal dataset, which was pre-registered before data collection began (`https://osf.io/sb6qx`). Details about the entire dataset can be found in the data release paper [6], and a pre-registered data-request provides the broad aims and variables requested for the present study (`https://osf.io/3tfnh/`). Here, we provide brief descriptions of the subject recruitment and psychological measures used in this study.

The recruitment was done through Prolific (www.prolific.co) and subjects were required to be adults 18 or older, fluent in English and reside in the United States. In addition, they had to have a Prolific approval rating of 98% or higher, and a minimum of 5 Prolific studies completed. In total, 1797 subjects completed Wave 1 of the COVID-Dynamic study (see Fig.8.1b for a timeline of the wave administrations).

The Connor-Davidson Resilience Scale - 10 Item (CD-RISC) [26] is a self-report questionnaire of coping responses in the past month that is the most common measure of psychological resilience. The NEO Five-Factor Personality Inventory (NEO) [54] is a 60-item self-report questionnaire that assesses an individual on five dimensions of personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The NIH toolbox: Loneliness scale (NIH-Loneliness) [55] is a 5-item self-report questionnaire of how often an individual felt lonely or alone in the past month. Beck Depression Inventory – II (BDI) [56] is a 21-item self-report questionnaire that examines depressive symptomatology over the past two weeks. The

State Trait Anxiety Inventory (STAI) [57] is a 20-item self-report questionnaire that we used to measure current (state) anxiety. The Perceived Stress Scale (PSS) [58] is a 10-item self-report questionnaire that measures the extent to which a participant perceives personal life events in the past month as stressful.

8.2.2 Exclusion criteria

We applied the following exclusion criteria to ensure data quality. Subjects were excluded if they failed one or more attention checks per data collection wave (on average, three attention questions were included for each wave) across a total of three or more waves. Subjects were also excluded if they self-reported that they were diagnosed with any of the following mental health conditions not related to our hypothesis: schizophrenia, bipolar disorder, or posttraumatic stress disorder; or multiple comorbid psychiatric conditions other than depression and anxiety. After exclusion, the total number of unique subjects was reduced from 1797 to 1580.

8.2.3 Further subject selection

The longitudinal nature of the COVID Dynamic study allowed us to select specific time windows (Fig.8.1b) to establish baseline resilience and personality trait measures at time 1 (CD-RISC: waves 5, 7; NEO: wave 1), loneliness at time 2 (waves 8,11,13), and subsequent depression, anxiety, and stress at time 3 (BDI: waves 14,16; STAI: waves 14,15,16; PSS: waves 14,15,16) out of all available waves. Since we consider resilience and NEO personality to be temporally stable traits, all six available waves of data for these two measures were used in this further selection step (however, only the initial waves were used in regression analyses to ensure clean temporal separation).

We first selected only subjects who had at least two waves of data for each measure to ensure relatively complete data from each subject (there were additional requirements specifically for CD-RISC: at least one wave of valid data was available for waves 5,7 and for NEO: wave 1 data must be complete, to avoid missing values for regression analyses). This step resulted in a total number of 634 subjects with relatively complete data across all waves of interest (note that because of attrition, the total number of subjects completing wave 16 was 876; subject attrition across the longitudinal waves was the main factor in reduction of our final sample size).

We restricted the scope of our study to resilience and personality as temporally stable traits. While we conceive of loneliness, depression, anxiety, and stress as state variables that vary within an individual over time, we nonetheless also aimed to have these variables be relatively stable within our specific time windows (since we would not be able to distinguish rapid variability from one wave to the next from measurement error–that is, within our time window, we required good test-retest reliability even for these state variables). We therefore assessed the within-subject temporal stability in the relevant time windows for each set of variables. Specifically, we calculated the difference between the maximum and minimum values of a measure across the selected windows of waves for each subject and normalized the difference by the total points range possible for each measure (for example, STAI raw scores range from 20 to 80, so we divided the temporal variation by 60) for easier comparison across measures (see Fig.S8.1 in the Supplementary Material for distributions of the normalized differences for all measures). We consequently selected subjects whose normalized differences were no more than 0.3 for all measures; i.e., for any of the measures, the within-subject temporal variation was no more than 30% of the range of the measure. These further selection restrictions resulted in a final sample of 447 subjects (see Table S8.1 in the Supplementary Material for characterization of the sample).

Final variables: We used the raw (untransformed) scores produced by the CD-RISC, NEO, NIH-Loneliness, BDI, STAI and PSS. To obtain the smallest measurement error, we averaged the scores for each individual across the multiple valid longitudinal measure collections (CD-RISC: waves 5, 7; NEO: wave 1; Loneliness: waves 8,11,13; BDI: waves 14,16; STAI: waves 14,15,16; PSS: waves 14,15,16). Note that from the distributions of normalized differences (Fig.S8.1 in the Supplementary Material), the scores for CD-RISC and NEO were indeed empirically stable across six waves (the majority of normalized differences fell within 0.3), justifying our operationalization of resilience as a trait variable for the purposes of the present study.

8.2.4 Multiple Regression analysis

We carried out several linear regression models (lm function in R) to test the associations between the selected variables. We first quantified the associations of loneliness and resilience (as independent variable) with depression (as outcome variable) in separate models. We then further tested the moderating effect of resilience by incorporating loneliness, resilience, and their interaction term in the same model. Loneliness and resilience scores were centered to help alleviate multicollinearity [59]. We repeated the same analysis scheme for different outcome variables (anxiety and stress), and for different moderators (five original NEO scores, and the revised factors obtained from our exploratory factor analysis, see below).

8.2.5 Exploratory factor analysis

We conducted a factor analysis (Psych package in R) to explore the possible constitutive components of resilience, and to obtain initial data for revising trait-resilience measures. Since conscientiousness (positively), extraversion (positively), and neuroticism (negatively) were strongly correlated with resilience, we applied EFA across the itemwise Pearson correlation matrix of all our final subjects to extract four factors (with oblimin rotation), based on the prior decision to consider resilience, conscientiousness, extraversion, and neuroticism as four distinct factors. We interpret the results of this factor analysis in the Results.

8.2.6 Statistical treatment

We report effect sizes and confidence intervals, as well as exact p-values for the main hypotheses. We interpret p-values <0.05 and large effect sizes with the adjective "substantial" and avoid use of the word "significant," so as to avoid dichotomous interpretation.

8.3 Results

8.3.1 Resilience protects against depression caused by loneliness

To test our primary model (Fig.8.1a), we first examined the associations between each pair of our main variables of interest: resilience, loneliness, and depression (Fig.8.2). As expected, resilience was negatively correlated with both loneliness ($r = -0.52$, $p < 0.001$) and depression ($r = -0.59$, $p < 0.001$), while loneliness and depression were positively correlated ($r = 0.67$, $p < 0.001$). Linear regression models (Table 8.1), confirmed that while both loneliness ($b = 0.97$, 95% CI = [0.82, 1.11], $p < .001$) and resilience ($b = -0.38$, 95% CI = [-0.48, -0.29], $p < .001$) predicted depression (positively and negatively, respectively), these two variables also interacted ($b = -0.03$, 95% CI = [-0.05, -0.02], $p < .001$): more resilient individuals were less depressed following loneliness compared to less resilient individuals (Fig.8.3).

Table 8.1: Results of multiple regression models testing whether the association between loneliness and an outcome variable depends on a moderator variable. Results were divided into three sections of different outcome variables: depression, anxiety, and stress. For a given outcome variable, six moderators were tested: resilience, openness, conscientiousness, extraversion, agreeableness, and neuroticism.

Outcome (Y) Moderator (M)	**Loneliness (X)**			**Moderator (M)**			**Interaction term (X*M)**		
	estimate	95% CI	p	estimate	95% CI	p	estimate	95% CI	p
Depression, Resilience	0.97	[0.82,1.11]	< 0.001 ***	-0.38	[-0.48,-0.29]	< 0.001 ***	-0.03	[-0.05,-0.02]	< 0.001 ***
Depression, Openness	1.35	[1.21,1.49]	< 0.001 ***	0	[-0.10,0.10]	0.99	-0.02	[-0.04,-0.00]	0.023 *
Depression, Conscientiousness	1.15	[1.00,1.29]	< 0.001 ***	-0.23	[-0.33,-0.14]	< 0.001 ***	-0.03	[-0.04,-0.01]	0.002 **
Depression, Extraversion	1.05	[0.90,1.20]	< 0.001 ***	-0.28	[-0.36,-0.19]	< 0.001 ***	-0.03	[-0.05,-0.02]	< 0.001 ***
Depression, Agreeableness	1.27	[1.12,1.42]	< 0.001 ***	-0.12	[-0.23,-0.01]	0.039 *	-0.02	[-0.04,0.00]	0.121
Depression, Neuroticism	0.6	[0.44,0.76]	< 0.001 ***	0.43	[0.36,0.51]	< 0.001 ***	0.03	[0.02,0.04]	< 0.001 ***
Anxiety, Resilience	0.68	[0.50,0.87]	< 0.001 ***	-0.78	[-0.89,-0.66]	< 0.001 ***	0	[-0.02,0.02]	0.925
Anxiety, Openness	1.33	[1.14,1.51]	< 0.001 ***	-0.03	[-0.17,0.10]	0.656	-0.01	[-0.04,0.01]	0.321
Anxiety, Conscientiousness	1.04	[0.84,1.23]	< 0.001 ***	-0.41	[-0.54,-0.29]	< 0.001 ***	-0.01	[-0.03,0.02]	0.589
Anxiety, Extraversion	1.01	[0.81,1.21]	< 0.001 ***	-0.42	[-0.53,-0.31]	< 0.001 ***	0.01	[-0.01,0.03]	0.234
Anxiety, Agreeableness	1.2	[1.01,1.40]	< 0.001 ***	-0.27	[-0.42,-0.12]	< 0.001 ***	0.01	[-0.02,0.04]	0.399
Anxiety, Neuroticism	0.47	[0.25,0.69]	< 0.001 ***	0.62	[0.52,0.72]	< 0.001 ***	-0.01	[-0.02,0.01]	0.449
Stress, Resilience	0.54	[0.44,0.64]	< 0.001 ***	-0.53	[-0.59,-0.46]	< 0.001 ***	0.01	[-0.00,0.02]	0.189
Stress, Openness	0.97	[0.86,1.07]	< 0.001 ***	0.02	[-0.06,0.10]	0.676	-0.01	[-0.03,0.00]	0.055
Stress, Conscientiousness	0.79	[0.68,0.91]	< 0.001 ***	-0.27	[-0.35,-0.20]	< 0.001 ***	0.01	[-0.01,0.02]	0.27
Stress, Extraversion	0.78	[0.66,0.90]	< 0.001 ***	-0.24	[-0.31,-0.18]	< 0.001 ***	0.01	[-0.01,0.02]	0.366
Stress, Agreeableness	0.88	[0.77,1.00]	< 0.001 ***	-0.2	[-0.29,-0.12]	< 0.001 ***	0.02	[-0.00,0.03]	0.053
Stress, Neuroticism	0.37	[0.25,0.48]	< 0.001 ***	0.44	[0.38,0.49]	< 0.001 ***	-0.01	[-0.01,0.00]	0.178

Note: *p < .05. **p < .01. ***p < .001.

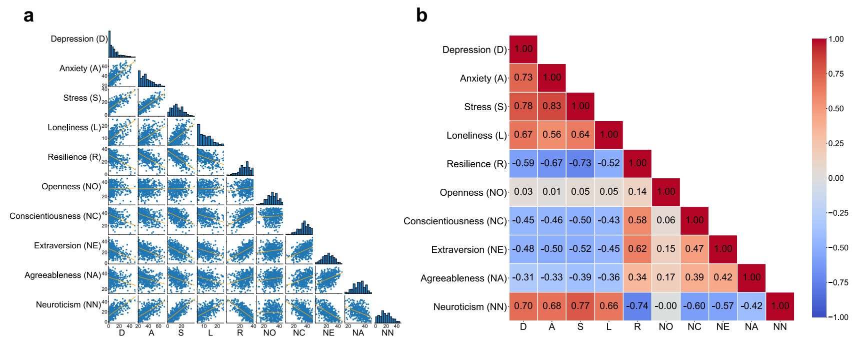

Figure 8.2: Associations between measures. **a** Histograms on the diagonal show the distributions of each measure. Off-diagonal scatterplots show associations between each pair of measures (with fitted regression lines). **b** Pearson correlation coefficients between each pair of measures.

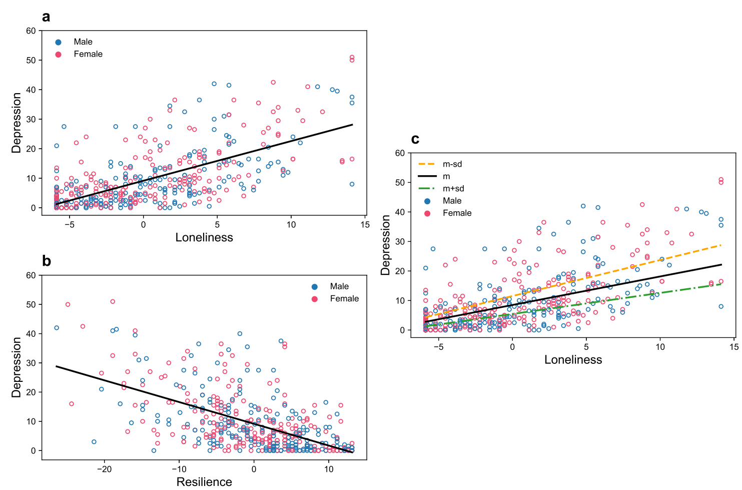

Figure 8.3: Regression results testing the effect of loneliness and/or resilience on depression. **a** Regression result showing the influence of loneliness on depression: higher loneliness predicted higher depression. **b** Regression result showing the influence of resilience on depression: higher resilience predicted lower depression. **c** Simple slopes for the association between loneliness and depression were tested for low (-1 SD below the mean), moderate (mean), and high (+1 SD above the mean) levels of resilience, showing the effect of loneliness on depression was moderated by resilience. Loneliness and resilience were both centered to avoid multicollinearity. Males and females are indicated using blue and pink circles respectively; equivalent effects were found for either sex.

To ensure generalizability of our findings across demographics, we tested whether sex, age, or education level changed these findings: testing female/male, old/young, or low/high education groups separately produced qualitatively similar results (see Table S8.2 in the Supplement Material). Entering these demographic variables into a full regression model did not change the previously observed effects and did not yield any additional effects for sex, age, or education (sex: b = 0.55, 95% CI = [-0.73, 1.83], p = 0.396; age: b = 0.01, 95% CI = [-0.04, 0.06], p = 0.707; education: b = 0.59, 95% CI = [-0.72, 1.90], p = 0.379).

8.3.2 Generalizability to other mental health variables

In addition to depression, we queried anxiety and stress. Anxiety was correlated with both loneliness (r = 0.56, p < 0.001) and resilience (r = -0.67, p < 0.001). Similarly, stress was correlated with both loneliness (r = 0.64, p < 0.001) and resilience (r = -0.73, p < 0.001). To test whether our findings extend beyond depression to other mental health variables, we carried out identical regression analyses as above, using anxiety or stress as outcome variables (Table 8.1).

While both loneliness (b = 0.68, 95% CI = [0.50, 0.87], p < .001) and resilience (b = -0.78, 95 % CI = [-0.89, -0.66], p < .001) predicted anxiety (positively and negatively, respectively), they didn't interact substantially (b = 0.00, 95% CI = [-0.02, 0.02], p = 0.925). The same finding also generalized across demographics (see Table S8.3 in the Supplement Material).

Similar results were found for stress (Table 8.1), where both loneliness (b = 0.54, 95% CI = [0.44, 0.64], p < .001) and resilience (b = -0.53, 95% CI = [-0.59, -0.46], p < .001) predicted stress, but no moderating effect was found (b =0.01, 95% CI = [-0.00, 0.02], p = 0.189). The same finding also generalized across demographics (see Table S8.4 in the Supplement Material).

8.3.3 Exploring trait-resilience as a construct

Our theoretical focus was on individual resilience as a trait, ensured by the way in which we selected subjects with temporally stable CD-RISC scores. Since we also collected personality measures (from the NEO-FFI), we expected empirically that CD-RISC scores would be correlated with at least some personality trait scores, and we expected theoretically that resilience is comprised, at least in part, of certain personality traits. We tested correlations with all five personality traits (Fig.8.2) and found three that correlated substantially with resilience scores: conscientiousness (r =0.58, p < 0.001), extraversion (r = 0.62, p < 0.001), and neuroticism (r = -0.74, p < 0.001). These three personality traits were also entered in the same set of regression models as carried out with resilience above (Table 8.1). Being more conscientious, more extraverted, or less neurotic protected against depression but not anxiety or stress caused by loneliness, a pattern of results for these personality factors that was qualitatively the same as what we found with resilience.

We assume that personality is (partly) constitutive of resilience, and therefore sought further insight into this complex construct using itemwise data from our four inter-correlated variables (CD-RISC, NEO-Conscientiousness, NEO-Extraversion, NEO-Neuroticism; ten items for CD-RISC, twelve for each NEO trait). Since we had six longitudinal data points for both CD-RISC, and NEO-FFI, we decided to use the average scores, for each individual, across all available waves to obtain the most precise estimate of the relevant variables. These data were then submitted to an exploratory factor analysis with oblimin rotation. We set the number of factors at four a priori, assuming that resilience, conscientiousness, extraversion, and neuroticism were at least partly independent psychological variables. This analysis produced four factors (referred to as revised resilience, conscientiousness, neuroticism, and extraversion) that accounted for 19%, 17%, 16%, and 13% of the total variance in the data, respectively (see Table S8.5 in the Supplementary Material for the factor loadings).

We then repeated the same regression analyses (see Table 8.2) and found qualitatively identical moderating effects of our revised resilience, conscientiousness, neuroticism, and extraversion on depression caused by loneliness, as with the original scores. The revised extraversion factor in fact was found to have substantial main and moderating effects across the three mental health outcomes (depression, anxiety, stress), but the direction of the moderation differed for depression compared to anxiety or stress. The revised extraversion factor revealed that more extraverted individuals were less depressed in general ($b = -2.28$, 95% CI = [-3.04, -1.52], $p < .001$) and more protected against depression caused by loneliness (the interaction between loneliness and revised extraversion: $b = -0.27$, 95% CI = [-0.41, -0.12], $p < .001$). However, more extraverted individuals were affected more strongly by loneliness (positive moderating effect of revised extraversion on anxiety: $b = 0.23$, 95% CI = [0.04, 0.43], $p = 0.017$, positive moderating effect of revised extraversion on stress: $b = 0.12$, 95% CI = [0.00, 0.23], $p = 0.041$) despite being less anxious or stressed in general (main effect of revised extraversion on anxiety: $b = -2.67$, 95% CI = [-3.68, -1.65], $p < .001$, main effect of revised extraversion on stress: $b = -1.58$, 95% CI = [-2.18, -0.98], $p < .001$). The original extraversion score showed similar trends, but with smaller effect sizes that were not substantial. We refrain from drawing strong conclusions about the moderating effects of extraversion on anxiety or stress, for which larger-scale future studies will be needed.

Table 8.2: Results of multiple regression models testing whether the association between loneliness and an outcome variable depends on a moderator variable. Results were divided into three sections of different outcome variables: depression, anxiety, and stress. For a given outcome variable, four moderators were tested: revised resilience, conscientiousness, extraversion, and neuroticism as obtained from the exploratory factor analysis.

Outcome (Y) Moderator (M)	**Loneliness (X)**			**Moderator (M)**			**Interaction term (X*M)**		
	estimate	95% CI	p	estimate	95% CI	p	estimate	95% CI	p
Depression, Resilience	0.98	[0.83,1.13]	< 0.001 ***	-2.83	[-3.60,-2.06]	< 0.001 ***	-0.26	[-0.37,-0.14]	< 0.001 ***
Depression, Conscientiousness	1.15	[1.00,1.30]	< 0.001 ***	-1.87	[-2.63,-1.10]	< 0.001 ***	-0.21	[-0.34,-0.08]	0.002 **
Depression, Extraversion	1.1	[0.95,1.25]	< 0.001 ***	-2.28	[-3.04,-1.52]	< 0.001 ***	-0.27	[-0.41,-0.12]	< 0.001 ***
Depression, Neuroticism	0.53	[0.36,0.69]	< 0.001 ***	5.31	[4.51,6.11]	< 0.001 ***	0.39	[0.26,0.52]	< 0.001 ***
Anxiety, Resilience	0.7	[0.51,0.88]	< 0.001 ***	-6.14	[-7.09,-5.19]	< 0.001 ***	0.04	[-0.11,0.18]	0.616
Anxiety, Conscientiousness	1.06	[0.87,1.26]	< 0.001 ***	-2.97	[-3.98,-1.96]	< 0.001 ***	-0.07	[-0.24,0.11]	0.461
Anxiety, Extraversion	1.16	[0.96,1.36]	< 0.001 ***	-2.67	[-3.68,-1.65]	< 0.001 ***	0.23	[0.04,0.43]	0.017 *
Anxiety, Neuroticism	0.36	[0.15,0.57]	< 0.001 ***	7.66	[6.61,8.70]	< 0.001 ***	-0.09	[-0.25,0.07]	0.282
Stress, Resilience	0.56	[0.46,0.67]	< 0.001 ***	-4.06	[-4.60,-3.53]	< 0.001 ***	0.07	[-0.02,0.15]	0.111
Stress, Conscientiousness	0.79	[0.68,0.90]	< 0.001 ***	-2.21	[-2.79,-1.62]	< 0.001 ***	0.05	[-0.05,0.15]	0.309
Stress, Extraversion	0.86	[0.75,0.98]	< 0.001 ***	-1.58	[-2.18,-0.98]	< 0.001 ***	0.12	[0.00,0.23]	0.041 *
Stress, Neuroticism	0.28	[0.16,0.39]	< 0.001 ***	5.44	[4.89,5.98]	< 0.001 ***	-0.05	[-0.13,0.04]	0.294

Note: *p < .05. **p < .01. ***p < .001.

8.4 Discussion

A unique longitudinal dataset in 447 American subjects exceptionally well assessed during the course of the COVID pandemic provided us with a strong test of a causal model, according to which resilience protects against the effects of loneliness on depression. The strong shared environmental intervention of social isolation, experienced across the country during lockdowns, gathering restrictions, and quarantine measures led to an increase in both loneliness and depression across our sample, and, as expected, these two variables were correlated. Regression models provided support for our hypothesis, restricted it to depression (but not anxiety or stress), and extended it to encompass personality traits as protective factors as well.

A host of other variables have been associated with resilience, often without specifying whether those should be thought of as constitutively or causally related. Our scheme in Fig.8.1 makes clear that loneliness and negative affect are distinct in our model, and thus not constitutive of resilience, and that either or both could be causally affected by resilience. We would expect all environmental variables, if they are associated with resilience at all, to be causally related: but they could be either antecedent or consequent (certain environmental events could make a person more or less resilient, and resilience could cause people to seek out and structure their environment). While we interpret personality factors as constitutively related to resilience (they are part of what it means to be a resilient person), it remains possible that the relation is causal as well (extraverted people might tend to become more resilient). As we have noted previously [60] future studies will have to test clearly specified causal models to address these further questions and explore the full richness of what it is that constitutes resilience, and how it interacts with environment and mental health.

While the supposition of temporal stability in resilience (i.e., treating this variable as a trait) was our theoretical interest, and while empirically valid in the context of our study and over our time window, this in no way argues against resilience as a dynamic process. It is well established that resilience emerges throughout development and adolescence [61], and that it can change even during adulthood [62, 63]. The dynamic nature of resilience plays out in a complex psychosocial context, through relationships, attachments, and social support networks [64]. There is good evidence that, at least for some people, exposure to adverse events can build resilience over time [65]. All interventions targeted at resilience itself require such a dynamic view—but it is not within the scope of the present study. Instead, we wanted to ensure temporally stable (and more precisely measured) variables within specific time windows.

An important aspect of our study was the intervening variable of loneliness, a particularly salient feature during the COVID pandemic. This particular stressor differs from others in that social relations are widely acknowledged to be a key part of the mechanism whereby resilience acts: building relationships, seeking advice, and relying on others for emotional support are critically important for resilience [66]. Yet these may have been precisely the mechanisms intervened upon during social isolation during the COVID pandemic. This would suggest a somewhat different interpretation of our findings: rather than thinking of social isolation as one among many possible stressors that can cause negative affect, it may interfere with one of the main mechanisms of resilience itself. Plausible future tests of this hypothesis could test the prediction that social isolation predisposes people to become depressed by other adverse events even when they are otherwise resilient.

A constraint of scope in our study were the measures we used, and hence the extent of the domain of latent variables that could be studied: for our main analysis, we used only a single resilience measure (the Connor-Davidson resilience scale), a single measure of loneliness (the NIH toolbox: Loneliness scale), and a single measure of depressive symptomatology (the Beck Depression Inventory – II). The generalizability of our findings was addressed to some extent by testing other measures of negative affect (the State Trait Anxiety Inventory and the Perceived Stress Scale, which test anxiety/stress rather than depression), and by testing traits other than resilience (the NEO personality factors). Whereas the latter did show an interesting generalization that allowed us to extract putative novel trait-resilience factors, the former showed that our findings are specific to depression and do not generalize as well to anxiety or stress–again, at least as available from our measures. We want to emphasize the fact that all of the measures used in our study are self-report questionnaires, and one clearly important future direction would be to reproduce our findings with measures that do not depend on self report.

A factor analysis provided exploratory factors that could be further tested for their moderation of specific effects of resilience. Of interest, we found that the revised extraversion factor substantially moderated the effects of loneliness on all mental health outcomes (depression, anxiety, and stress; Table 8.2). On closer inspection, the revised extraversion factor highlights the items most related to outgoing behaviors (see Table S8.5 in the Supplement Material for factor loadings; for the original extraversion score, all 12 items would contribute equally). More specifically, the three NEO items most strongly related to the revised extraversion factor were "I like to have a lot of people around me" (NEO2, positively), "I usually prefer to do things alone" (NEO27, negatively), and "I like to be where the action is" (NEO22, positively). While speculative, high loading on these items suggest to us that this factor may tap into the social support aspects of resilience, perhaps explaining the factor's pervasive moderation effect in the case of loneliness. Future studies, ideally with a bigger sample size, could use the exploratory factors that we have identified to further uncover the multiple facets of psychological resilience.

8.5 Supplementary information

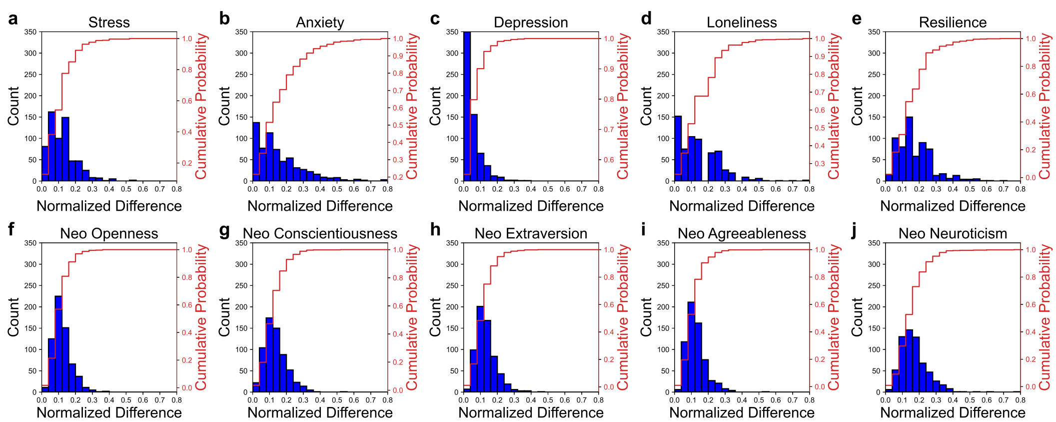

Figure S8.1: Distribution of normalized within-subject difference (maximum – minimum, divided by the total points range possible for each measure) across waves for all selected measure: (**a**) Stress, (**b**) Anxiety, (**c**) Depression, (**d**) Loneliness, (**e**) Resilience, (**f**) Openness, (**g**) Conscientiousness, (**h**) Extraversion, (**i**) Agreeableness, and (**j**) Neuroticism. Red traces indicate the cumulative probability.

Table S8.1: Demographic characteristics, means, and standard deviations of all measures used for the final sample.

	All (N = 447)	Female (N=227)	Male (N=220)
Age (in years) (mean; sd)	43.21 (14.03)	44.49 (13.83)	41.89 (14.15)
Education: below Bachelor (n; %)	169 (37.81%)	78 (34.36%)	91 (41.36%)
Education: Bachelor and above (n; %)	278 (62.19%)	149 (65.64%)	129 (58.64%)
CD-RISC 10 (mean; sd)	26.89 (7.9)	26.34 (8.02)	27.46 (7.75)
NEO Openness (mean; sd)	30.8 (6.81)	31.9 (6.81)	29.66 (6.64)
NEO Conscientiousness (mean; sd)	34.88 (7.74)	35.06 (7.34)	34.69 (8.15)
NEO Extraversion (mean; sd)	23.11 (8.85)	22.78 (8.36)	23.45 (9.34)
NEO Agreeableness (mean; sd)	33.33 (6.54)	34.14 (6.45)	32.5 (6.55)
NEO Neuroticism (mean; sd)	18.62 (10.72)	19.51 (10.86)	17.7 (10.52)
NIH Loneliness (mean; sd)	10.88 (5.01)	10.88 (5.08)	10.87 (4.95)
BDI (mean; sd)	9.15 (10.02)	9.74 (10.25)	8.55 (9.76)
STAI (mean; sd)	35.46 (11.86)	36.1 (12.64)	34.8 (10.99)
PSS (mean; sd)	13.83 (7.55)	14.73 (7.75)	12.89 (7.23)

Table S8.2: Results of multiple regression models testing whether the association between loneliness and depression depends on resilience for different groups divided by demographic variables.

Group	**Loneliness (X)**			**Moderator (M)**			**Interaction term (X*M)**		
	estimate	95% CI	p	estimate	95% CI	p	estimate	95% CI	p
Female	0.99	[0.76,1.21]	< 0.001 ***	-0.33	[-0.47,-0.19]	< 0.001 ***	-0.03	[-0.05,-0.01]	< 0.001 ***
Male	0.96	[0.76,1.16]	< 0.001 ***	-0.43	[-0.56,-0.30]	< 0.001 ***	-0.03	[-0.05,-0.01]	0.006 **
Old	0.97	[0.77,1.17]	< 0.001 ***	-0.36	[-0.49,-0.23]	< 0.001 ***	-0.02	[-0.04,-0.00]	0.024 *
Young	0.96	[0.72,1.19]	< 0.001 ***	-0.4	[-0.55,-0.26]	< 0.001 ***	-0.04	[-0.06,-0.02]	< 0.001 ***
Edu Low	1.02	[0.76,1.28]	< 0.001 ***	-0.36	[-0.51,-0.21]	< 0.001 ***	-0.03	[-0.05,-0.00]	0.047 *
Edu High	0.94	[0.75,1.12]	< 0.001 ***	-0.4	[-0.52,-0.27]	< 0.001 ***	-0.04	[-0.05,-0.02]	< 0.001 ***

Note: *p < .05. **p < .01. ***p < .001.

Table S8.3: Results of multiple regression models testing whether the association between loneliness and anxiety depends on resilience for different groups divided by demographic variables.

Group	Loneliness (X)			Moderator (M)			Interaction term (X*M)		
	estimate	95% CI	p	estimate	95% CI	p	estimate	95% CI	p
Female	0.7	[0.41,1.00]	< 0.001 ***	-0.76	[-0.95,-0.58]	< 0.001 ***	-0.01	[-0.04,0.02]	0.49
Male	0.65	[0.42,0.88]	< 0.001 ***	-0.77	[-0.92,-0.63]	< 0.001 ***	0.02	[-0.01,0.04]	0.218
Old	0.69	[0.43,0.96]	< 0.001 ***	-0.78	[-0.96,-0.61]	< 0.001 ***	0	[-0.02,0.03]	0.701
Young	0.64	[0.38,0.90]	< 0.001 ***	-0.74	[-0.91,-0.58]	< 0.001 ***	-0.01	[-0.03,0.02]	0.654
Edu Low	0.65	[0.39,0.92]	< 0.001 ***	-0.81	[-0.96,-0.65]	< 0.001 ***	0	[-0.03,0.02]	0.801
Edu High	0.71	[0.46,0.96]	< 0.001 ***	-0.75	[-0.92,-0.58]	< 0.001 ***	0	[-0.02,0.03]	0.795

Note: *p < .05. **p < .01. ***p < .001.

Table S8.4: Results of multiple regression models testing whether the association between loneliness and stress depends on resilience for different groups divided by demographic variables.

Group	**Loneliness (X)**			**Moderator (M)**			**Interaction term (X*M)**		
	estimate	95% CI	p	estimate	95% CI	p	estimate	95% CI	p
Female	0.52	[0.36,0.67]	< 0.001 ***	-0.54	[-0.63,-0.44]	< 0.001 ***	0.01	[-0.01,0.02]	0.271
Male	0.57	[0.44,0.71]	< 0.001 ***	-0.51	[-0.59,-0.43]	< 0.001 ***	0.01	[-0.01,0.02]	0.338
Old	0.57	[0.43,0.71]	< 0.001 ***	-0.56	[-0.65,-0.47]	< 0.001 ***	0.01	[-0.01,0.02]	0.262
Young	0.51	[0.36,0.66]	< 0.001 ***	-0.51	[-0.60,-0.41]	< 0.001 ***	0.00	[-0.01,0.02]	0.679
Edu Low	0.58	[0.41,0.75]	< 0.001 ***	-0.49	[-0.59,-0.39]	< 0.001 ***	0.00	[-0.01,0.02]	0.560
Edu High	0.51	[0.38,0.64]	< 0.001 ***	-0.56	[-0.65,-0.47]	< 0.001 ***	0.01	[-0.00,0.02]	0.193

Note: *p < .05. **p < .01. ***p < .001.

Table S8.5: Factor loadings of individual resilience and NEO items on the four factors identified in an exploratory factor analysis with oblimin rotation.

	MR1 (Resilience)	MR2 (Conscientiousness)	MR4 (Neuroticism)	MR3 (Extraversion)
RISC1_1	0.76	0.04	-0.17	0.00
RISC1_2	0.76	0.06	-0.19	0.00
RISC1_3	0.70	-0.16	-0.01	0.15
RISC1_4	0.80	0.02	0.07	0.12
RISC1_5	0.81	0.04	-0.11	0.02
RISC1_6	0.74	0.16	-0.04	0.09
RISC1_7	0.69	0.14	-0.18	-0.02
RISC1_8	0.61	0.08	-0.28	0.10
RISC1_9	0.78	0.10	-0.11	0.07
RISC1_10	0.76	0.03	-0.17	-0.03
NEO1.n_neg.	0.07	-0.12	-0.71	0.10
NEO2.e_pos.	-0.16	0.00	-0.03	0.92
NEO5.c_pos.	-0.09	0.65	-0.05	0.08
NEO6.n_pos.	-0.14	-0.14	0.59	-0.11
NEO7.e_pos.	0.28	-0.13	-0.05	0.42
NEO10.c_pos.	0.04	0.74	-0.11	0.02
NEO11.n_pos.	-0.24	-0.08	0.67	0.09
NEO12.e_neg.	-0.27	0.10	0.10	-0.42
NEO15.c_neg.	-0.04	-0.57	-0.05	0.18
NEO16.n_neg.	0.16	0.04	-0.62	0.07
NEO17.e_pos.	0.24	-0.02	-0.01	0.65
NEO20.c_pos.	0.26	0.53	0.04	-0.16
NEO21.n_pos.	-0.09	-0.09	0.73	0.03
NEO22.e_pos.	0.08	-0.04	0.09	0.72
NEO25.c_pos.	0.12	0.77	0.05	0.09
NEO26.n_pos.	-0.07	-0.17	0.64	-0.07
NEO27.e_neg.	0.16	-0.02	0.16	-0.79
NEO30.c_neg.	0.11	-0.66	0.31	-0.06
NEO31.n_neg.	0.08	-0.07	-0.84	0.04
NEO32.e_pos.	0.13	0.09	-0.05	0.62
NEO35.c_pos.	0.32	0.68	0.14	0.11
NEO36.n_pos.	-0.02	-0.18	0.45	-0.07
NEO37.e_pos.	0.30	-0.05	-0.12	0.64
NEO40.c_pos.	0.11	0.72	-0.01	-0.02
NEO41.n_pos.	-0.31	-0.25	0.45	-0.09
NEO42.e_neg.	-0.33	0.02	0.22	-0.46
NEO45.c_neg.	0.12	-0.74	0.18	-0.06
NEO46.n_neg.	0.10	-0.01	-0.72	0.11
NEO47.e_pos.	0.06	0.28	0.13	0.57
NEO50.c_pos.	0.13	0.87	0.05	0.05
NEO51.n_pos.	-0.30	-0.34	0.39	0.08
NEO52.e_pos.	0.14	0.33	0.04	0.58
NEO55.c_neg.	0.20	-0.79	0.20	-0.02
NEO56.n_pos.	-0.01	-0.17	0.66	-0.12
NEO57.e_neg.	-0.12	-0.10	0.06	-0.48
NEO60.c_pos.	0.41	0.49	0.19	0.07

Note: NEO items are labeled with additional information on how they are related to the original summary scores.

References

[1] Tuppett M Yates, Fanita A Tyrell, and Ann S Masten. "Resilience theory and the practice of positive psychology from individuals to societies". In: *Positive psychology in practice: Promoting human flourishing in work, health, education, and everyday life* (2015), pp. 773–788.

[2] Gregory C Smith and Bert Hayslip Jr. "Resilience in adulthood and later life: What does it mean and where are we heading?" In: *Annual review of gerontology & geriatrics* 32.1 (2012), p. 3.

[3] Kelsey T Laird et al. "Psychobiological factors of resilience and depression in late life". In: *Translational Psychiatry* 9.1 (2019), pp. 1–18.

[4] Eva Gerino et al. "Loneliness, resilience, mental health, and quality of life in old age: A structural equation model". In: *Frontiers in psychology* 8 (2017), p. 2003.

[5] Christiaan H Vinkers et al. "Stress resilience during the coronavirus pandemic". In: *European Neuropsychopharmacology* 35 (2020), pp. 12–16.

[6] Tessa Rusch et al. "COVID-Dynamic: A large-scale multifaceted longitudinal study of socioemotional and behavioral change across the pandemic". In: (2021).

[7] L. A. Peplau and D." Perlman. "Perspective on loneliness". In: *In L. A. Peplau and D. Perlman (Eds.), Loneliness: A sourcebook of current theory, research and therapy (pp. 1-18)*. New York: John Wiley and Sons, 1982.

[8] John T Cacioppo and William Patrick. *Loneliness: Human nature and the need for social connection*. WW Norton & Company, 2008.

[9] Christopher M Masi et al. "A meta-analysis of interventions to reduce loneliness". In: *Personality and social psychology review* 15.3 (2011), pp. 219–266.

[10] John T Cacioppo et al. "Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses." In: *Psychology and aging* 21.1 (2006), p. 140.

[11] Liesl M Heinrich and Eleonora Gullone. "The clinical significance of loneliness: A literature review". In: *Clinical psychology review* 26.6 (2006), pp. 695–718.

[12] William DS Killgore et al. "Loneliness: A signature mental health concern in the era of COVID-19". In: *Psychiatry research* 290 (2020), p. 113117.

[13] Christine M Wickens et al. "Loneliness in the COVID-19 pandemic: Associations with age, gender and their interaction". In: *Journal of psychiatric research* 136 (2021), pp. 103–108.

[14] Simone Marchini et al. “Study of resilience and loneliness in youth (18–25 years old) during the COVID-19 pandemic lockdown measures”. In: *Journal of community psychology* 49.2 (2021), pp. 468–480.

[15] Liuyi Ran et al. “Psychological resilience, depression, anxiety, and somatization symptoms in response to COVID-19: A study of the general population in China at the peak of its epidemic”. In: *Social Science & Medicine* 262 (2020), p. 113261.

[16] Simon Coulombe et al. “Risk and resilience factors during the COVID-19 pandemic: a snapshot of the experiences of Canadian workers early on in the crisis”. In: *Frontiers in psychology* (2020), p. 3225.

[17] Shaul Kimhi et al. “Distress and Resilience in Days of COVID-19: International Study of Samples from Israel, Brazil, and the Philippines”. In: *Cross-Cultural Research* 55.5 (2021), pp. 415–437.

[18] Georgia Libera Finstad et al. “Resilience, Coping Strategies and Posttraumatic Growth in the Workplace Following COVID-19: A Narrative Review on the Positive Aspects of Trauma”. In: *International Journal of Environmental Research and Public Health* 18.18 (2021), p. 9453.

[19] Xia Zhao et al. “Loneliness and depression symptoms among the elderly in nursing homes: A moderated mediation model of resilience and social support”. In: *Psychiatry research* 268 (2018), pp. 143–151.

[20] Jeff L Perron, Kristin Cleverley, and Sean A Kidd. “Resilience, loneliness, and psychological distress among homeless youth”. In: *Archives of psychiatric nursing* 28.4 (2014), pp. 226–229.

[21] Ida Skytte Jakobsen et al. “The relationship between resilience and loneliness elucidated by a Danish version of the resilience scale for adults”. In: *BMC psychology* 8.1 (2020), pp. 1–10.

[22] Michael Rutter. “Resilience as a dynamic concept”. In: *Development and psychopathology* 24.2 (2012), pp. 335–344.

[23] Shae-Leigh Cynthia Vella, Nagesh B Pai, et al. “A theoretical review of psychological resilience: Defining resilience and resilience research over the decades”. In: *Archives of Medicine and Health Sciences* 7.2 (2019), p. 233.

[24] Julia C Poole, Keith S Dobson, and Dennis Pusch. “Childhood adversity and adult depression: The protective role of psychological resilience”. In: *Child abuse & neglect* 64 (2017), pp. 89–100.

[25] LB Navrady et al. “Genetic risk of major depressive disorder: the moderating and mediating effects of neuroticism and psychological resilience on clinical and self-reported depression”. In: *Psychological medicine* 48.11 (2018), pp. 1890–1899.

[26] Kathryn M Connor and Jonathan RT Davidson. “Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC)”. In: *Depression and anxiety* 18.2 (2003), pp. 76–82.

[27] Laura Campbell-Sills, David R Forde, and Murray B Stein. “Demographic and childhood environmental predictors of resilience in a community sample”. In: *Journal of psychiatric research* 43.12 (2009), pp. 1007–1012.

[28] K. M. Connor and J. R. T. Davidson. *The Connor-Davidson Resilience Scale User Guide*. 2014. URL: `http://www.connordavidson-resiliencescale.com/user-guide.php`.

[29] Aliza P Wingo et al. “Moderating effects of resilience on depression in individuals with a history of childhood abuse or trauma exposure”. In: *Journal of affective disorders* 126.3 (2010), pp. 411–414.

[30] Kimberly T Green et al. “Exploration of the resilience construct in posttraumatic stress disorder severity and functional correlates in military combat veterans who have served since September 11, 2001”. In: *The Journal of clinical psychiatry* 71.7 (2010), p. 6966.

[31] Minden B Sexton et al. “The roles of resilience and childhood trauma history: main and moderating effects on postpartum maternal mental health and functioning”. In: *Journal of affective disorders* 174 (2015), pp. 562–568.

[32] Alec Roy, Marco Sarchiapone, and Vladimir Carli. “Low resilience in suicide attempters: relationship to depressive symptoms”. In: *Depression and Anxiety* 24.4 (2007), pp. 273–274.

[33] Alec Roy, Vladimir Carli, and Marco Sarchiapone. “Resilience mitigates the suicide risk associated with childhood trauma”. In: *Journal of affective disorders* 133.3 (2011), pp. 591–594.

[34] Denise Catalano et al. “The buffering effect of resilience on depression among individuals with spinal cord injury: a structural equation model.” In: *Rehabilitation psychology* 56.3 (2011), p. 200.

[35] Stephen R McCauley et al. “Preinjury resilience and mood as predictors of early outcome following mild traumatic brain injury”. In: *Journal of neurotrauma* 30.8 (2013), pp. 642–652.

[36] Garry Stevens et al. “Determinants of paramedic response readiness for CBRNE threats”. In: *Biosecurity and bioterrorism: biodefense strategy, practice, and science* 8.2 (2010), pp. 193–202.

[37] Allison S Gabriel, James M Diefendorff, and Rebecca J Erickson. “The relations of daily task accomplishment satisfaction with changes in affect: a multilevel study in nurses.” In: *Journal of Applied Psychology* 96.5 (2011), p. 1095.

[38] Linghua Kong et al. “Resilience moderates the relationship between emotional intelligence and clinical communication ability among Chinese practice nursing students: A structural equation model analysis”. In: *Nurse Education Today* 46 (2016), pp. 64–68.

[39] Kristal C Melvin et al. “Couple functioning and post-traumatic stress symptoms in US army couples: The role of resilience”. In: *Research in Nursing & Health* 35.2 (2012), pp. 164–177.

[40] Dilip V Jeste et al. “Association between older age and more successful aging: critical role of resilience and depression”. In: *American Journal of Psychiatry* 170.2 (2013), pp. 188–196.

[41] Samoon Ahmad et al. “Earthquake impact in a remote South Asian population: Psychosocial factors and posttraumatic symptoms”. In: *Journal of traumatic stress* 23.3 (2010), pp. 408–412.

[42] Liuhua Ying et al. “Traumatic severity and trait resilience as predictors of posttraumatic stress disorder and depressive symptoms among adolescent survivors of the Wenchuan earthquake”. In: *PloS one* 9.2 (2014), e89401.

[43] I Irmansyah et al. “Determinants of psychological morbidity in survivors of the earthquake and tsunami in Aceh and Nias”. In: *International Journal of Mental Health Systems* 4.1 (2010), pp. 1–10.

[44] Lynn M Grattan et al. “The early psychological impacts of the Deepwater Horizon oil spill on Florida and Alabama communities”. In: *Environmental health perspectives* 119.6 (2011), pp. 838–843.

[45] Xiaonan Yu and Jianxin Zhang. “Factor analysis and psychometric evaluation of the Connor-Davidson Resilience Scale (CD-RISC) with Chinese people”. In: *Social Behavior and Personality: an international journal* 35.1 (2007), pp. 19–30.

[46] Laura Campbell-Sills, Sharon L Cohan, and Murray B Stein. “Relationship of resilience to personality, coping, and psychiatric symptoms in young adults”. In: *Behaviour research and therapy* 44.4 (2006), pp. 585–599.

[47] Hyun-Sook Baek et al. “Reliability and validity of the Korean version of the Connor-Davidson Resilience Scale”. In: *Psychiatry investigation* 7.2 (2010), p. 109.

[48] Timo Giesbrecht et al. “Adversity does not always lead to psychopathology: cognitive reactivity is related to longitudinal changes in resilience”. In: *Netherlands journal of Psychology* 65.2 (2009), pp. 62–68.

[49] Li Wang et al. “Psychometric properties of the 10-item Connor–Davidson Resilience Scale in Chinese earthquake victims”. In: *Psychiatry and clinical neurosciences* 64.5 (2010), pp. 499–504.

[50] Zeng Jie Ye et al. “Validation and application of the Chinese version of the 10-item Connor-Davidson Resilience Scale (CD-RISC-10) among parents of children with cancer diagnosis”. In: *European Journal of Oncology Nursing* 27 (2017), pp. 36–44.

[51] Brian White, Simon Driver, and Ann Marie Warren. “Resilience and indicators of adjustment during rehabilitation from a spinal cord injury.” In: *Rehabilitation psychology* 55.1 (2010), p. 23.

[52] Kimberly T Green et al. “Alcohol misuse and psychological resilience among US Iraq and Afghanistan era veterans”. In: *Addictive behaviors* 39.2 (2014), pp. 406–413.

[53] Evan E Rainey et al. “Psychological factors predicting outcome after traumatic injury: the role of resilience”. In: *The American Journal of Surgery* 208.4 (2014), pp. 517–523.

[54] PT Costa and RR McCrae. “NEO Five-factor inventory professional manual”. In: *Odessa, FL: Psychological Assessment Resources* (1992).

[55] Richard C Gershon et al. “NIH toolbox for assessment of neurological and behavioral function”. In: *Neurology* 80.11 Supplement 3 (2013), S2–S6.

[56] Aaron T Beck, Robert A Steer, Gregory K Brown, et al. *Manual for the beck depression inventory-II*. 1996.

[57] CD Spielberger et al. “Manual for the State-Trait Anxiety Inventory (STAI).” In: *Palo Alto: Consulting Psychologists Press* (1983).

[58] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. “A global measure of perceived stress”. In: *Journal of health and social behavior* (1983), pp. 385–396.

[59] Dawn Iacobucci et al. “Mean centering helps alleviate “micro” but not “macro” multicollinearity”. In: *Behavior research methods* 48.4 (2016), pp. 1308–1317.

[60] Julien Dubois et al. “Personality beyond taxonomy”. In: *Nature human behaviour* 4.11 (2020), pp. 1110–1117.

[61] Angie Hart, Derek Blincow, and Helen Thomas. *Resilient therapy: Working with children and families*. Routledge, 2007.

[62] Anthony D Ong et al. “Psychological resilience, positive emotions, and successful adaptation to stress in later life.” In: *Journal of personality and social psychology* 91.4 (2006), p. 730.

[63] Anthony D Ong, Cindy S Bergeman, and Steven M Boker. “Resilience comes of age: Defining features in later adulthood”. In: *Journal of personality* 77.6 (2009), pp. 1777–1804.

[64] Helen Herrman et al. “What is resilience?” In: *The Canadian Journal of Psychiatry* 56.5 (2011), pp. 258–265.

[65] Mark D Seery. "Resilience: A silver lining to experiencing adverse life events?" In: *Current Directions in Psychological Science* 20.6 (2011), pp. 390–394.

[66] George A Bonanno et al. "What predicts psychological resilience after disaster? The role of demographics, resources, and life stress." In: *Journal of consulting and clinical psychology* 75.5 (2007), p. 671.

Chapter 9

DISCUSSION

9.1 Summary of findings

How many dimensions do we need to characterize the structure of emotion experiences? Using three types of stimuli: a validated set of short stories (Chapter 3), a validated set of short videos (Chapter 4), and actual experiences in real life (Chapter 5), I have found that the psychological spaces for all three types of emotion experiences were remarkably low dimensional.

Given the diversity of ratings that I used, this is somewhat surprising (or perhaps reassuringly). I included not only the most common scales for basic emotions, but also several other complex ones relating to how people interpret the emotion that they experienced (including some that were appraisal-motivated). In addition to these diverse sources, I added scales that, to my knowledge, nobody had ever used (the ones motivated by the biological emotion features discussed by Anderson and Adolphs). That all these diverse ratings scales should yield such a low-dimensional space argues for the robustness of the dimensions that I found. I further quantified the robustness of the factors by decimating the number of ratings scales and stimuli, and still finding the factors (up to a point).

More specifically, three or four dimensions captured most of the variance, with "valence" and "arousal" shared across domains (Chapter 6). These two factors capture "core affect", a component of emotion experience that most psychological work considers essential. It is thus reassuring to find these two dimensions also in my study, aligning with the idea that core affect is indeed a necessary part of all emotion experience. However, the third factor ("generalizability") that I discovered is, to my knowledge, novel and suggests a new and important aspect of emotion experience: the extent to which the experience can be applied to a broad range of situations, or is modular and specific with respect to only certain of them (discussed extensively in Chapter 6).

I also characterized the distributions of the three types of emotion experiences and found that emotions were distributed along continuous gradients (most notably, the valence dimension), with no well-separated clusters even for emotions belonging to the six basic emotion categories (Chapter 6). Among the six basic emotions, happiness and surprise clustered relatively better than the negative ones. Negative emotions (sadness, anger, fear, and disgust) in real life differed from stimuli-evoked ones: they were less surprising, less arousing, more controlled, more common, and also purer.

My thesis also addresses how people differ in their emotions as evoked by stimuli and experienced in real life. Some findings were specific to emotions experienced during the COVID pandemic, for example, Chapter 8 investigated how resilience buffered individuals against the effect of loneliness on depression. In addition, I explored the association between psychological traits and differences in emotion experiences both in terms of the magnitudes of the ratings and the overall correlation structure across scales (Chapter 7). Positive personality traits (especially resilience) were associated with more positive real-life emotions while negative traits (for instance, neuroticism) were associated with more negative real-life emotions. Interesting differences between the low and high alexithymia groups were observed with respect to the correlation structure for emotions evoked by stories.

The core parts of many of the investigations were pre-registered. I will also make all data, experiment codes, and analysis codes publicly available. It is our hope that this unique and rich dataset will be a valuable resource for all researchers. It's worth noting that all the psychological assessments collected in the COVID Dynamic study will also be publicly available and it's possible in principle to contact the participants recruited in our studies for additional data collection if desired.

9.2 Limitations and discussion

As already mentioned, a general caveat is that it's unclear what exactly it is that subjects are reporting when they produce ratings.

In Chapter 6, I made comprehensive comparisons across different types of stimuli and found overall coherence in terms of the correlation structure across scales (section 6.1), the low dimensional factors (section 6.2), and the actual experiences (section 6.3). It's certainly possible that the coherent findings across domains suggest that subjects were rating the emotions that they actually experienced. But that's not the only explanation. It could also be that they were simply good at guessing the "the right answer" (a reasonable concern especially for stories).

I think one specific finding when making comparisons across participants (in Chapter 7) helps address this question to some extent. Individuals with high levels of alexithymia differed from non-alexithymia individuals the most in the correlational structure for emotions evoked by stories, and less for videos and real life. If ratings were all based on actual experiences, then one might expect a low level of agreement across all domains. Instead, the disagreement was most evident for stories possibly because people with high levels of alexithymia failed at guessing the "the right answer".

9.3 Future directions

One future direction is to further analyze the two factors unique to real-life emotions (Chapter 5). For the "negative affect" factor, I will test whether it is more prominent using data from subjects who were more affected by Covid (for example, tested positive) than ones who were less affected. If possible, I'd also like to test whether this factor would disappear by sampling emotions in post-covid times. For the "common" factor describing how common/generalizable emotions are, I plan to analyze the free descriptions (for example, using topic modeling) that subjects gave for the cause of their emotions to further verify my interpretation.

More open questions about individual differences can be investigated using this dataset. So far, I have focused on the means of ratings on the scales, but this is just one metric to characterize. In the context of evoked emotions, the variance of ratings, especially compared with subjects who viewed the same stimuli and rated the same scale, may be a good indication of subjects' ability to differentiate their emotion experiences. Variability of real-life emotions, on the other hand, signals the richness and stability of emotion experiences.

Indeed, some personality traits have been linked with emotional variability, for example, neuroticism with greater variability between high and low levels of negative emotion [1] and openness with experiencing a wider range of emotions in general [2]. It would also be interesting to investigate the longitudinal trajectory of emotions in real life and how that is impacted by real world events.

References

[1] Elise K Kalokerinos et al. "Neuroticism may not reflect emotional variability". In: *Proceedings of the National Academy of Sciences* 117.17 (2020), pp. 9270–9276.

[2] Philip J Corr and Gerald Ed Matthews. *The Cambridge handbook of personality psychology*. Cambridge University Press, 2020.