Evaluation of film stimuli for the assessment of social-emotional processing: A pilot study

Jenni Leppanen[1], Olivia Patsalos[2], Sophie Surguladze[2], Jess Kerr-Gaffney[2], Steven Williams[1], Kate Tchanturia[2,3,4]

[1] Department of Neuroimaging, Institute of Psychology, Psychiatry and Neuroscience, King's College London, London, United Kingdom
[2] Department of Psychological Medicine, Institute of Psychology, Psychiatry and Neuroscience, King's College London, London, United Kingdom
[3] South London and Maudsley NHS Foundation Trust National Eating Disorder Service, London, United Kingdom
[4] Tbilisi State Medical University, Psychological Set Research and Correction Center

Corresponding Author:
Jenni Leppanen[1]
De Crespigny Park, London, SE5 8AF, United Kingdom
Email address: jenni.leppanen@kcl.ac.uk

Abstract

Background. Difficulties in top-down and bottom-up emotion generation have been proposed to play a key role in the progression of psychiatric disorders. The aim of the current study was to develop more ecologically valid measures of top-down interpretation biases and bottom-up evoked emotional responses.

Methods. 124 healthy female participants aged 18 – 25 took part in the study. We evaluated two sets of 18 brief film clips. The first set of film clips presented ambiguous social situations designed to examine interpretation biases. Participants provided written interpretations of each ambiguous film clip which were subjected to sentiment analysis. We compared the films in terms of the valence of participants interpretations. The second set of film clips presented neutral and emotionally provoking social scenarios designed to elicit subjective and facial emotional responses. While viewing these film clips participants mood ratings and facial affect were recorded and analysed using exploratory factor analyses.

Results. Most of the 18 ambiguous film clips were interpreted in the expected manner while still retaining some ambiguity. However, participants were more attuned to the negative cues in the ambiguous film clips and three film clips were identified as unambiguous. These films clips were deemed unsuitable for assessing interpretation bias. The exploratory factor analyses of participants' mood ratings and evoked facial affect showed that the positive and negative emotionally provoking film clips formed their own factors as expected. However, there was substantial cross-loading of the neutral film clips when participants' facial expression data was analysed.

Discussion. A subset of the film clips from the two tasks could be used to assess top-down interpretation biases and bottom-up evoked emotional responses. Ambiguous negatively valenced film clips should have more subtle negative cues to avoid ceiling effects and to ensure there is enough room for interpretation.

Introduction

48 Emotional responses have been proposed to be generated through two processes: top-down and
49 bottom-up (McRae et al,. 2012; Ochsner et al., 2009). Although everyday emotional experiences
50 are likely to arise from a blend of these two processes, top-down and bottom-up emotion generation
51 have been proposed to utilise distinct psychological pathways (McRae et al,. 2012; Ochsner et al.,
52 2009). Top-down emotion generation involves cognitive appraisal of a situation and is influenced
53 by the person's thinking style and past experiences. For instance, due to excessive focus on
54 negative cues or overgeneralisation of previous unpleasant experiences, a person can end up
55 developing a general tendency to interpret neutral, ambiguous situations as negative resulting in
56 negative emotional responses and low mood (Davis et al., 2018; McRae et al., 2012). Bottom-up
57 emotion generation, on the other hand, refers to reactions arising from exposure to emotionally
58 provoking stimuli rather than from cognitive appraisal (McRae et al,. 2012; Ochsner et al., 2009).
59 As such bottom-up emotion generation is less influenced by cognitive biases, but rather an inherent
60 or habitual reaction with little conscious processing. These include biologically prepared fear
61 responses to seeing a predator and flashbacks following exposure to trauma related places or
62 sounds (Guillery-Girard et al., 2013; McRae et al,. 2012).

63

64 The top-down emotion generation pathway is often linked to interpretation biases, which can
65 interfere with the way a person perceives various situations and influence subjective emotional
66 experiences (Davis et al., 2018; McRae et al., 2012). Over the years, a variety of methods have
67 been used to assess interpretation biases ranging from disambiguating homophones to rating the
68 valence of ambiguous video clips (Schoth & Liossi, 2017). Regardless of the exact method used,
69 stimuli evaluation is essential to reliably assess interpretation bias (Schoth & Liossi, 2017). It is
70 important to ensure that the stimuli used is ambiguous so that different interpretations are possible,
71 but evaluation of any dominance and unbalance in the possible interpretations is also crucial. For
72 instance, in a homophone paradigm, if participants are presented with spoken unbalanced
73 homophones such as "gilt", and then asked to write what they just heard. In this scenario a
74 participant would likely give a negative answer and write "guilt", rather than providing the neutral
75 answer, "gilt". This could create the impression of negative bias, even though the response bias is
76 more likely to be driven by word dominance. Additionally, dominance effects can result in ceiling
77 effects which make it difficult to examine group differences in interpretation bias in a case-control
78 paradigm.

79

80 Although interpretation biases have been extensively studied over the years, recent reviews have
81 reported substantial between study variability which was at least in part explained by methodology
82 (Everaert et al., 2017; Chen et al., 2020). One meta-analysis reported that significant effects were
83 found only when direct measures, such as the sentence completion paradigm and identification of
84 ambiguous facial expressions, were used (Everaert et al., 2017). However, even amongst studies
85 employing direct measures, heterogeneity was substantial. Another meta-analysis found

86   significant differences between studies employing pictures or videos and those that used written
87   or spoken stimuli only (Chen et al., 2020). These findings suggest that, in addition of dominance
88   effects, the type of stimuli can also influence the results. This may be linked to differences in
89   ecological validity. Indeed, over the years questions have been raised about the real-life
90   generalisability of simple tasks, such as emotion recognition tasks, using still images presented
91   without further context (Greenaway et al., 2018; Hogenelst et al., 2015). Performance on these
92   types of tasks has not been found to reflect day-to-day social functioning (Janssens et al., 2012).
93   Similarly, sentence completion paradigms, which involve reading or listening to a sentence that is
94   missing an ending and then providing a short one or two word ending for the sentence, have been
95   criticised for not reflecting complexities of everyday experiences and for lacking context and
96   visual cues (Jurado and Rosselli, 2007; Chan et al., 2008). Taken together, there is need to develop
97   more naturalistic paradigms that reliably assess interpretation bias.
98
99   The use of film stimuli could offer an easy to implement way of overcoming some of above-
100 mentioned limitations. Previous reviews of studies examining interpretation bias have suggested
101 that using film clips rather than still images to depict ambiguous social situations could help
102 increase vividness of the stimuli and thus increase ecological validity (Schoth & Liossi, 2017;
103 Chen et al., 2020). Indeed, films can be used to present naturalistic, dynamic social interactions,
104 which can provide a reasonable approximation of real-life situations (Sonkusare et al., 2019;
105 Schaefer et al., 2010). A recent study examining consensus in appraisal of film stimuli documented
106 low degree of agreement between participants indicating that film stimuli may be an effective way
107 to examine highly personal interpretations (Wallisch & Whritner, 2017). Additionally, tendency
108 to appraise neutral film clips negatively has also been reported to be strongly associated with
109 emotional responses to daily life events among people with depression (Panaite et al., 2017). Taken
110 together, these findings suggest that evaluation of ambiguous film stimuli to effectively and
111 reliably study appraisal tendencies or interpretations biases is of interest.
112
113 Unlike top-down emotion generation, bottom-up emotions tend to be generated through immediate
114 reactions to the stimuli presented and they are believed to be less influenced by cognitive processes
115 (McRae et al,. 2012; Ochsner et al., 2009). Several different methods have been developed to
116 assess bottom-up emotional reactivity ranging from imagination to emotionally provoking pictures
117 and film clips (Siedlecka & Denson, 2019; Quigley et al., 2014). As with interpretation bias,
118 stimuli evaluation is important to reliably examine emotional reactivity. However, unlike
119 interpretation bias, ambiguity should be avoided. If a stimulus is open to interpretation, it is less
120 likely produce in consistent immediate emotional reactions and it could result in blend of bottom-
121 up and top-down emotional responses (Aguado et al., 2019; Sheppes & Gross, 2011).
122
123 Brief film clips are commonly used to study the bottom-up emotion generation pathway. Over the
124 years, film stimuli have been reported to be an effective and powerful tool to reliably provoke
125 emotional responses resulting in several research databases (Deng et al., 2017; Gilman et al., 2017;

126   Samson et al., 2015; Schaefer et al., 2010). However, most of the previous work have only
127   evaluated the film stimuli based on self-reported mood ratings. Incorporating other measures, such
128   as spontaneous facial expressions, could help provide further information about how the stimuli
129   alters a person's emotional experience and how they communicate such experiences. Facial
130   expressions of emotions have a complex role in social interactions, and they are crucial for
131   communication, connection, and building rapport with others (Schmidt and Cohn, 2001;
132   Parkinson, 2005). Indeed, recent work evaluating importance of cultural specificity of emotional
133   film clips reported that stimuli depicting the participants' own culture provoked stronger facial
134   affective responses than non-culturally specific stimuli (Alghowinem et al., 2019). However,
135   similar effects were not observed in the self-reported emotion ratings (Alghowinem et al., 2019).
136   Thus, there is interest in developing paradigms that could be used to assess how evoked facial
137   affect might change as the illness progresses or following an intervention. To achieve these goals,
138   it is important to evaluate emotionally provoking film stimuli not only in terms of self-reported
139   emotional states but also evoked facial expressions.
140
141   Although previous studies have used film clips which have been found to effectively and reliably
142   induce intended self-reported emotional responses, most of the stimuli used were from well known
143   movies and television shows (Deng et al., 2017; Gilman et al., 2017; Schaefer et al., 2010). This
144   can introduce problems as participants can have personal memories attached to the stimuli which
145   can be evoked by watching the film resulting in unpredictable responses (Conway and Loveday,
146   2010; Clark, Mackay and Holmes, 2013; Arnaudova and Hagenaars, 2017; Maksimainen et al.,
147   2018). For instance, personal memories have been found to influence the pleasantness and
148   unpleasantness ratings of emotionally provoking musical and pictorial stimuli more than the
149   features of the stimuli alone (Maksimainen et al., 2018). Additionally, those who report more
150   spontaneous positive autobiographical memories following an experimental task were found to
151   report greater positive mood reactions to the film clip used in the task (Clark, Mackay and Holmes,
152   2013). These findings suggest that personal memories can impact participants' reactions to a given
153   task, indicating that developing paradigms that use stimuli that participants are not familiar with
154   would be of interest to avoid these possible confounding effects.
155
156   Both top-down and bottom-up emotion generation are relevant in the context of psychiatric
157   disorders. Interpretation bias, particularly negative interpretation bias, has been widely researched
158   in psychiatric disorders, including mood, anxiety, and eating disorders, and negative biases have
159   been associated with low mood and worse psychopathology (Rowlands et al., 2020; Everaert et
160   al., 2017; Hirsch et al., 2016). Similarly, alterations in bottom-up emotional reactivity, particularly
161   reduced facial expression of emotions, have been documented in mood and eating disorders (e.g.
162   Panaite et al., 2017; Davies et al., 2016; Leppanen et al., 2017) and these alterations in bottom-up
163   emotion generation are believed to contribute to illness progression (Gross, 2002; Butler et al.,
164   2003; Treasure and Schmidt, 2013). Even though difficulties in these aspects of top-down and
165   bottom-up emotion generation appear to be transdiagnostic, it has been suggested that it is

166   important to tailor stimuli to the specific psychiatric disorder being investigated as cognitive biases
167   and habitual emotional responses may vary substantially (Hirsch et al., 2016). Thus, if stimuli are
168   to be used in a clinical study, it crucial to not only take the above-mentioned methodological points
169   into consideration when selecting the type of stimuli, ensure that the stimuli targets disorder
170   specific processes, but alsi evaluate the stimuli to ensure reliability of any subsequent case-control
171   findings.

172

173   Current pilot study aims to build on previous research by evaluating two sets of film clips with
174   more dynamic expressions of emotion and contextual cues to assess interpretation biases and
175   evoked emotional responses. The film clips were taken from short films that had not received a
176   wide audience in order to avoid impact of personal memories attached to the stimuli and were
177   selected based on themes reported by people with lived experience of eating disorders. First, we
178   aimed to evaluate the reliability of 18 ambiguous film clips. Specifically, we aimed to examine
179   whether these film clips were sufficiently ambiguous so they could be used to assess interpretation
180   bias. We hypothesised that film clips which were initially categorised as neutral, positively
181   valenced, and negatively valenced would mostly elicit interpretations that fell into the
182   corresponding categories, but that other interpretations were also possible, thus demonstrating
183   ambiguity. Second, we aimed to evaluate the reliability of a second set of 18 emotionally
184   provoking film clips. Specifically, we aimed to examine whether the film clips reliably evoked the
185   intended subjective emotional experiences and facial affective responses. We hypothesised that
186   film clips initially categorised as neutral, positive, and negative would evoke the corresponding
187   self-reported mood and facial affect responses. With both stimuli sets, we explored whether any
188   carry-over effects were present or if the present inter-stimulus interval was sufficient. Finally, as
189   top-down and bottom-up emotion generation have been reported to be affected in anxiety,
190   depression, and eating disorders as outlined above, we also explored whether participants
191   interpretations or emotional reactions were associated with relevant psychopathology measures.

Materials & Methods

Participants

Altogether 124 female participants took part in the study. All participants were 18 – 25 years old with no current or past mental health problems or neurological diagnoses. The sample size was based on a power calculation for a X^2 test of independence conducted using the R package *pwr* (Champely, 2020). Participants we recruited through online adverts featured on websites such as www.callforparticipants.com and amongst King's College London staff and students. Participants met inclusion criteria if they had no current or past mental health or neurological problems, sleep disturbances, or alcohol or drug misuse or abuse. All participants were screened for eligibility using the Structured Clinical Interview for DSM-5 (First *et al.*, 2015), which included the non-patient overview and enhanced screening modules. Prior to taking part all participants gave written, informed consent and all study procedures were conducted in accordance with the latest version of the Declaration of Helsinki (2013). Ethical approval was obtained from the King's College London Psychiatry, Nursing and Midwifery Research Ethics sub-committee (ref: HR-19/20-13004).

Film stimuli

Altogether 36 short film clips were selected for the two tasks: one assessing interpretation of ambiguous film clips and another examining evoked facial expressions and mood in response to emotionally provoking film clips. For further information about the films evaluated in the present study see Supplemental Table S1. Prior to inclusion in the evaluation study, all film stimuli were first discussed amongst members of the research team and peer feedback was used to ensure they were appropriate for each task. For the Ambiguous scenarios, we adopted similar methodology used in previous work (Huppert et al., 2007; Cardi et al., 2017) and used the peer feedback to select 18 film clips with equal number of ambiguous film clips that were either somewhat positively valenced, somewhat negatively valenced, or neutral. To ensure ambiguity for the initial evaluation, we selected only film clips that produced a variety of interpretations and predictions regarding how the situation might end. These peer discussions resulted in selection of film clips such as a video depicting a man and a woman on a date in a restaurant. When the couple are seemingly unexpectedly served a dessert, which they did not order, and the woman insists that the man eats it despite him protesting that he is feeling full. The man then finds something hard inside the dessert resulting in a surprised expression. The film clip ends as the woman begins a speech that sounds like a proposal. Similarly, for the Evoked emotions task, the feedback was used to select a set of 18 film clips with equal number of positive and negative emotionally provoking film clips as well as neutral films. Through peer discussion we ensured that the film clips were not ambiguous, producing a variety of reactions, but rather evoked the intended mood states. These discussions

228 resulted in the selection of film clips such as a video depicting two police officers who were newly
229 assigned as partners. In the beginning it is clear that one of the officers is not happy to have a new
230 colleague, but throughout the video the two officers become better acquitted with each other and
231 develop a friendship. The video ends with the two officers driving whilst singing along to the radio.
232

233 The films were chosen based on themes that were identified in a previous qualitative study asking
234 people with lived experience of anorexia nervosa about critical life events (Leppanen et al., 2021)
235 to ensure they will cover themes relevant for people with eating disorders and can thus be used in
236 later case-control studies. The positive themes selected for the two tasks included receiving
237 emotional support, positive messages and reminders from friends and loved ones, and feeling
238 confident. The negative themes selected for the two tasks were feeling unsupported and
239 misunderstood, loneliness, and not feeling good enough. We made sure that these themes touched
240 on or were experienced by the main character in the film clips. Care was taken to ensure that films
241 did not feature themes directly related to eating disorders, such as inpatient treatment, which may
242 alienate healthy participants with no personal experience of eating disorders, thus failing to evoke
243 intended emotions in this group. Additionally, such film clips could trigger unintended emotional
244 responses associated with memories of such situations in future case-control studies.
245

246 The film clips were taken from short films freely available on online sharing platforms such as
247 YouTube and Vimeo. We chose to use short films over feature length films, because scenes in
248 such films tend to be succinct and often do not require extensive prior knowledge of events that
249 preceded the scene in order to understand what was happening. This was important as confusion
250 about what was going on in the scene could lead to unpredictable results. Additionally, many
251 people tend not to be as familiar with short films as with feature length films. Unfamiliarity with
252 the film clips was important to ensure participants were not able to predict what was going to
253 happen next. Unfamiliarity with the film clips also ensured that participants did not have memories
254 attached to the stimuli, which could have influenced their responses in unpredictable ways.
255

Self-report measures

257 Participants were asked to complete self-report questionnaires to obtain information regarding age,
258 body mass index (BMI), ethnicity, employment status, level of education, and marital status.
259 Additionally, participants completed the following standardised questionnaires.
260

261 The Eating Disorder Examination Questionnaire (EDEQ; (Fairburn and Beglin, 1994)) is a 28-
262 item self-report questionnaire, which assesses the level of eating disorder cognitions and
263 behaviours experienced over the past 28 days. In the present study the EDEQ had good internal
264 consistency (Cronbach's alpha = 0.90, 95% CI [0.87, 0.92]).
265

266 The Hospital Anxiety and Depression Scale (HADS; (Zigmond and Snaith, 1983)) is a 14-item
267 self-report questionnaire, which assesses the level of anxiety and depressed mood experienced
268 over the past 2 weeks. In the present study the HADS had good internal consistency (Cronbach's
269 alpha = 0.84, 95% CI [0.80, 0.88]).
270
271 The Beliefs about Emotions Scale (BES; (Rimes and Chalder, 2010)) is a 12-item self-report
272 questionnaire assessing participants' general beliefs about experiencing and expressing emotions.
273 In the present study the BES had good internal consistency (Cronbach's alpha = 0.88, 95% CI
274 [0.84, 0.91]).
275

Procedures
276

277 Prior to the SARS-CoV-2 pandemic participants completed the two tasks at King's College
278 London and those who took part after the pandemic completed the study remotely online. In the
279 office participants first gave written consent, filled in a set of self-report questionnaires, and
280 completed the Ambiguous scenarios task followed by the Evoked emotions task. The two tasks
281 were presented using Psychopy (Peirce et al., 2019). When taking part in the study remotely
282 participants completed the same steps online during a video conference with one of the researchers.
283 The online versions of the tasks were otherwise identical to the those completed in the office but
284 were hosted on http://gorilla.sc/ (Anwyl-Irvine et al., 2020).
285

Ambiguous scenarios task
286

287 As part of the Ambiguous scenarios task participants were asked to watch 18 short film clips
288 presented in random order. The film clips varied in length lasting 1.1 to 1.6 minutes and an inter-
289 stimulus interval of 2.0 seconds was used. After each film clip participants were asked two
290 questions: 1.) "What do you think happened in the video?" and 2.) "What do you think might
291 happen next?". Participants were instructed to type brief answers to each question explaining how
292 they perceived the situation and how they thought the situation might move forward. This was
293 done to ensure participants did not only describe the events in the scenarios, but also provided their
294 own interpretations. After the task had finished, participants were also asked if they had seen any
295 of the film clips before to avoid memory effects.
296

297 The written responses were subjected to sentiment analysis and were coded in terms of their
298 emotional valence into the following categories: positive, neutral, or negative. The answers to the
299 two questions were coded separately. The rating was done by two independent coders (O.P. and
300 S.S.) who were not familiar with the film stimuli to avoid bias. The coders were instructed to focus
301 on how participants described the interactions between the people in the film clips rather than
302 descriptions of the surroundings. Inter-rater agreement was acceptable (Cohen's Kappa = 0.65).

303 All disagreements were resolved by a third independent coder (J.K-G) and the final codes were
304 then taken forward for further statistical analysis. The Ambiguous scenarios task took participants
305 approximately 45 minutes to complete.
306

Evoked emotions task

308 During the Evoked emotions task participants were asked to watch 18 short film clips presented in
309 random order. The film clips varied in length lasting 1.5 to 2.1 minutes and an inter-stimulus
310 interval of 2.0 seconds was used. After each film participants were asked to rate their mood and
311 level of alertness using the affective slider (Betella and Verschure, 2016). The affective slider is a
312 digital visual analogue scale designed to assess mood/pleasure and alertness/arousal. The present
313 study focused on participants' mood/pleasure ratings. This was done to obtain information about
314 how participants rated their own mood in response to the film clips and how well the stimuli
315 captured their attention. During this task participants' faces were also video recorded to obtain
316 information about evoked facial expressions in response to the film stimuli. The video recordings
317 were analysed using a commercially available automated facial affect analysis tool FaceReader
318 version 8.1 (Noldus Information Technologies). The FaceReader uses a 500 key point mesh and
319 deep learning to classify facial expression in each frame of the video into seven emotion categories
320 including neutral. The FaceReader calculated the valence of facial expressions in each frame by
321 subtracting the highest intensity negative emotion from the intensity of 'happy' emotion. Each
322 participant's neutral facial expression recorded at the beginning of the task was used to first
323 calibrate the analysis and sample rate was set to frame rate. Information regarding each
324 participants' mean valence of evoked facial expressions made during each film clip was extracted
325 from the FaceReader output. After the task had finished, participants were also asked if they had
326 seen any of the film clips before to avoid memory effects. The Evoked emotions task took
327 participants approximately 40 minutes to complete.
328

Data analysis

330 All data analyses were conducted using R version 4.0.3 (R Core Team, 2020). In all significance
331 tests, p < 0.05 was considered significant and where appropriate adjustments for multiple
332 comparisons were used. As several participants took part in the study from home due to the SARS-
333 CoV-2 pandemic, we compared participants who completed the tasks in person and those who
334 took part from home. There were no significant differences between the two groups. The results
335 are presented in the Supplemental Tables S2, S3 and S4.
336

Ambiguous scenarios task

First, we assessed whether participants interpreted the ambiguous film clips in an expected manner offering more negative interpretation in response to negative films and positive interpretations in response to positive films. This was done by examining differences between the film categories in the valence of participants' written responses to the two questions after each film clip. We conducted a cumulative link mixed model from the R package *Ordinal* (Christensen, 2019) with film category and question as predictors and valence as the ordinal outcome measure. For ease of interpretation, the *Anova.clmm* command was used to obtain analysis of deviance table. In case of significant main effects or interactions, post-hoc comparisons were conducted using *emmeans* (Russell, 2021) and p-values were adjusted for multiple comparisons using the Tukey method.

As the film clips in the Ambiguous scenarios task should be ambiguous to ensure that it could be used as an alternative to other ambiguous scenarios tasks, we conducted an ambiguity check using a X^2 test of independence. This was done by calculating the relative contributions of each film clip to each interpretation category to establish if participants' written responses to any film clip were strongly associated with a particular interpretation category, which would indicate lack of ambiguity.

As the films were presented in randomised order, the presence of possible carry-over effects was examined by comparing the valence of written responses given in response to film clips that were preceded by negative or positive film clips. A cumulative link model was conducted to examine carry-over effects on neutral, positive, and negative film clips, with preceded by negative and preceded by positive and film category as the predictors, and valence of the written responses to the two questions as the ordinal outcome measure. Finally, Kendall's tau correlation matrix was calculated to explore whether the overall mean valence of all interpretations to the two questions participants offered after each ambiguous film clip were correlated with any of the self-report measures. Overall mean valence of all interpretations was used in an attempt to capture participants general tendency to offer positive, neutral or negative interpretations. In the correlation tests the p-values were adjusted for multiple comparisons using the Holm method.

Evoked emotions task

We conducted an exploratory factor analyses to examine the factor structure in participants self-reported mood rating following each emotional and neutral film clip and the valence of participants facial expressions during each film. The factor analyses were conducted using the R package *psych* (Revelle, 2020) with weighted least square factor method to minimise the impact of any potential outliers. First, we determined the number of factors in the data matrices using the parallel method which compares the observed data scree plots with those produced using a same sized random data matrix. The factor exploratory factor analyses were then conducted with the predetermined number of factors and *varimax* matrix rotation utilising Kaiser normalisation. We then examined whether

376 the film clips were split into factors in an expected manner, with neutral film clips forming one
377 factor, positive film clips another, and negative film clips forming a third factor. We also explored
378 if the results from the two factor analyses using self-reported mood ratings and valence of facial
379 expressions produced different factor structures and whether participants mood ratings and facial
380 expressions in response to each video correlated with each other using Kendall's tau correlation
381 tests.
382
383 As the films were presented in randomised order, the presence of carry-over effects was examined
384 by comparing the self-reported mood and valence of facial expressions in response to film clips
385 that were preceded by negative or positive film clips. Two separate analyses were conducted to
386 examine carry-over effects on neutral, positive, and negative film clips by conducting multivariate
387 linear mixed models, with preceded by negative and preceded by positive as the predictors and
388 self-reported mood, and mean valence of facial expression as the outcome measures. Finally,
389 Kendall's tau correlation matrix was calculated to explore whether the overall mean mood ratings
390 and valence of facial expressions across all films were correlated with any of the self-report
391 measures. Overall mean mood ratings and valence of facial expressions across all film clips was
392 used in an attempt to capture participants general tendency to rate and express positive, neutral, or
393 negative emotions.

Results

Sample characteristics

Sample characteristics are presented in Table 1. Overall participants' responses to the self-report questionnaires fell into a non-clinical range, with low incidence of eating disorder cognitions and behaviours, anxiety, and depressed mood. Participants' general beliefs about emotions and quality of life was comparable to previous studies reporting normative data.

Table 1. Participant characteristics (N = 124)
---TABLE 1 HERE ---
[1] (Rø, Reas and Stedal, 2015) recommended using EDEQ total score above 2.5 as an indicator of clinical significance; [2] the mean is based on what was reported in the initial validation of the scale among healthy people (Rimes and Chalder, 2010). BMI = body mass index, HADS = Hospital Anxiety and Depression Scale, EDEQ = Eating Disorder Examination Questionnaire, BES = Beliefs about Emotions Scale, NVQ = National Vocational Qualifications, BTEC = Business and Technology Education Council, SD = standard deviation, N/A = not applicable.

Ambiguous scenarios task

Quality assessment was conducted prior to any data analysis. Data from two participants was excluded due to internet connectivity issues which made it difficult for the participants to follow and understand the story in the film clips. Additionally, due to internet connectivity difficulties, data was missing from three participants and three other participants were unable to answer all questions. One participant had seen one of the film clips before (Film 5). Her written responses to the two questions after this film clip were not included in further analysis. No other participants reported having seen any of the film clips before. Thus, data from 119 participants were included for further analysis, 116 of whom provided complete answers for both of the two questions after each video.

Valence of interpretations per film category

The valence summary statistics per film category are presented in table 2. As expected, there was a significant main effect of film category in the cumulative link mixed model (X^2 (2) = 872.92, p < 0.001). The valence of the written answers to the two questions given in response to film clips categorised as positive were more positive than those given in response to film clips categorised as negative (z = 28.75, p < 0.001) or neutral (z = 11.86, p < 0.001). Additionally, the valence of the answers to the two questions following film clips that were categorised as negative were more negative than those given after film clips that were categorised as neutral (z = -18.27, p < 0.001).

429 The cumulative link mixed model also revealed a significant main effect of question (X^2 (1) =
430 56.59, p < 0.001), such that participants gave more positive responses to the second question asking
431 them to predict what was going to happen next than to the first question asking them what
432 happened in the film clip (z = 7.56, p < 0.001). There was no significant film category by question
433 interaction (X^2 (2) = 3.79, p = 0.151).

434

435 *Table 2. Valence of written interpretations*
436 ---TABLE 2 HERE --
437

438 ## Ambiguity check

439 The X^2 test of independence was significant (X^2 = 1515.9, p < 0.001), suggesting that the film
440 clips and the valence of participants' written response are dependent as would be expected based
441 on the findings above. The relative contributions of each film clip to the model are presented in
442 Fig 1. There were three film clips that appeared to make substantial contributions to one valence
443 category over others when compared to the other film clips, suggesting they lacked ambiguity.
444 Film 1 was strongly associated with negative responses to question 1, Film 4 was strongly
445 associated with negative responses to question 1, and Film 8 was strongly associated with positive
446 responses to question 2. These relative contribution statistics are supported by histograms showing
447 the number of neutral, positively valenced, and negatively valenced answers to each film (Fig 2).
448 Additionally, as shown in the Fig 2, participants primarily interpreted the negatively valenced film
449 clips in a negative way, while such a pattern was not observed in the interpretations of the neutral
450 and positively valenced film clips.

451

452 *Fig 1. X^2 contribution plot.*
453 ---FIG 1 HERE --
454 The dots indicate contribution each film clip made to a specific interpretation category. Larger
455 blue dots indicate higher contribution and lower ambiguity with most participants offering the
456 same type of interpretation.

457

458

459 *Fig 2. Histogram plots of participants interpretations of each film clip.*
460 ---FIG 2 HERE --
461 The histograms show the frequency of positive, neutral and negative answers participants offered
462 to the two questions after each film clip. Films 1 – 6 were categorised as negatively valenced
463 ambiguous film clips, Films 7 – 12 were categorised as neutral ambiguous film clips, and Films
464 13 – 18 were categorised as positively valenced ambiguous film clips.

465

Assessment of carry-over effects

The cumulative link mixed model did not reveal significant effects of preceded by a positive film (X^2 (1) = 0.137, p = 0.711) or a negative film (X^2 (1) = 0.24, p = 0.625), suggesting that a preceding emotional film did not impact the interpretation of a subsequent film clip. There were also no significant interactions between preceded by positive film and film category (X^2 (2) = 0.98, p = 0.612) or preceded by negative film and film category (X^2 (2) = 7.21, p = 0.125) suggesting there were no carry-over effects within film categories.

Correlations between valence of interpretations and clinical variables

There were no significant correlations between the mean valence of the written answers participants gave to question 1 or question 2 and self-reported eating disorder symptomatology, anxiety, depression, or beliefs about emotions (Supplemental Table S5).

Evoked emotions task

Prior to statistical analysis, data quality was evaluated. As with the Ambiguous scenarios task above, data from two participants was excluded due to internet connectivity issues which made it difficult for the participants to follow and understand the story in the film clips. Due to technical difficulties data was missing from one participant and another two participants were unable to watch all 18 videos during the Evoked emotions task leading to partially missing data. Thus, self-reported mood ratings were available from 121 participants. Additionally, video recordings from three participants were excluded because the participants moved outside the camera's range for extended periods of time. Video recordings of another three participants were excluded from further analysis due to low quality. Thus, facial expression data was available from 115 participants. Finally, two participants reported having seem two of the films (Film 5, Film 16) before and their responses to these films were not included in the final analysis.

Exploratory factor analysis with self-reported mood ratings

It was determined that the self-reported mood rating dataset contained three factors (Supplemental Fig S1). Factor 1 comprised of seven film clips and explained 45.1% of the variance with factor loadings ranging from 0.43 to 0.71 (Table 3). Factor 1 included all film clips from the negative category and one neutral film clip. Factor 2 comprised of four film clips, explained 28.9% of the variance and the factor loadings ranged from 0.38 to 0.82. Factor 2 included four of the six neutral film clips with one of the film clips cross-loading with Factor 3. Factor 3 consisted of seven film clips and explained 26.0% of the variance with factor loadings raging from 0.31 to 0.55. Five of

500   the film clips in Factor 3 were from the positive category and two were from the neutral category.
501   Film 14 did not fit in any of the three factors.
502
503   The exploratory factor analysis of mood ratings showed that most of the positive, neutral and
504   negative films tended to form their own unique factors, which supports the initial categorisation
505   of the film clips. This is supported by a visual inspection of participants' mood ratings in response
506   to each film clip (Supplemental Fig S2).
507
508   *Table 3. Self-rated mood data factor loadings*
509   --TABLE 3 HERE --
510   Factor loadings < 0.3 were excluded
511

512   ## Exploratory factor analysis with valence of facial expressions
513   It was determined that three factors were present in the facial expression data (Supplemental Fig
514   S3). Factor 1 consisted of twelve film clips and explained 46.0% of the variance with factor
515   loadings raging from 0.31 to 0.91 (Table 4). Factor 1 included all six film clips from the positive
516   category. The other film clips in Factor 1, five neutral and one negative film clips, cross-loaded
517   with Factors 2 and 3. Factor 2 included eleven film clips and explained 38.2% of the variance with
518   factor loadings ranging from 0.42 to 0.87. Factor 2 included all six film clips from the negative
519   category, four of which cross-loaded with Factors 3 and 1. Factor 2 also included five of the neutral
520   films, four of which cross-loaded with the other two factors. Factor 3 consisted of five film clips,
521   explained 15.7% of the variance, and the factor loadings ranged from 0.34 to 0.76. Factor 3
522   included three negative and two neutral film clips, all of which cross-loaded with the other two
523   factors.
524
525   The exploratory factor analysis showed that most of the positive film clips appeared to form one
526   factor while most of the negative film clips appeared to form another factor based on participants
527   facial affect while watching the films. The neutral film clips on the other hand tended to cross-load
528   across factors based on participants facial expressions while watching these films. This is
529   supported by a visual inspection of the valence of participants facial expressions during each film
530   clip (Supplemental Fig S4).
531
532   *Table 4. Facial expression data factor loadings*
533   --TABLE 4 HERE --
534   Factor loadings < 0.3 were excluded
535

Correlation between self-reported mood ratings and evoked facial affect

The exploratory correlation analysis found significant positive correlations between participants mood ratings and the valence of facial expressions produced during most of the positive film clips ($\tau = 0.24 - 0.32$, all $p < 0.05$) apart from two film clips (Film 13: $\tau = 0.21$, $p = 0.05$, Film 17: $\tau = 0.13$, $p = 0.23$). There were no significant correlations were self-reported mood ratings and valence of facial expressions in response to the neutral (all $\tau < 0.14$, all $p > 0.1$) or negative films clips (all $\tau < 0.13$, all $p > 0.1$).

Assessment of carry-over effects

We assessed whether there were any carry-over effects in participants' self-reported mood ratings or valence of facial expressions. The linear mixed effects model of mood ratings showed a significant interaction between preceding a negative film and film category ($F(4,2063.1) = 523.36$, $p < 0.001$). This interaction was driven by a significant impact of film category such that participants reported more negative mood in response to negative films than neutral ($t(2045) = -23.02$, $p < 0.001$) or positive film clips ($t(2047) = -41.27$, $p < 0.001$) when they were not preceded by another negative film clip. Participants also reported more positive mood when watching the positive than neutral film clips when they were not preceded by a negative film clip ($t(2049) = 17.30$, $p < 0.001$). Similarly, when the film clips were preceded by another negative film clip, participants reported more negative mood while watching the negative film clips than when watching the neutral ($t(2076) = -12.64$, $p < 0.001$) or positive film clips ($t(2077) = -21.69$, $p < 0.001$). Participants also reported more positive mood in response to the positive than neutral film clips ($t(2072) = -9.79$, $p < 0.001$) when they were preceded by a negative film. There was no significant effect of being preceded by another negative film within the negative ($t(2071) = 0.10$, $p = 0.923$), neutral ($t(2067) = 0.17$, $p = 0.866$), or positive film categories ($t(2047) = 0.52$, $p = 0.605$). There also was no significant interaction between preceded by a positive film and film category ($F(2,2087.8) = 0.28$, $p = 0.753$) or significant effects of preceded by a positive film ($F(1,2027.2) = 0.001$, $p = 0.972$) or preceded by a negative film ($F(1,2071.3) = 0.01$, $p = 0.923$) across the film categories. These findings indicate that there were no substantial carry-over effects in participants mood ratings.

As above, the linear mixed effects model of the valence of participants' facial expressions showed a significant interaction of preceded by a negative film and film category ($F(4,1556.0) = 53.67$, $p < 0.001$). This interaction was again driven by a significant effect of film category such that participants expressed more negative emotions in response to negative films than neutral ($t(1559) = -4.60$, $p < 0.001$) or positive film clips ($t(1560) = -13.22$, $p < 0.001$) when the films were not preceded by another negative film. Participants also expressed more positive emotions in response to the positive than neutral films when they were not preceded by a negative film clip ($t(1561) = 8.28$, $p < 0.001$). Similarly, participants expressed more negative emotions in response to negative

574 films than neutral (t(1566) = -3.62, p = 0.001) or positive film clips (t(1570) = -6.88, p < 0.001)
575 when the films were preceded by another negative film. Participants also expressed more positive
576 emotions in response to the positive than neutral films when they were preceded by a negative film
577 clip (t(1568) = 3.56, p = 0.001). There was no significant effect of being preceded by another
578 negative film clip on participants' facial affect when watching negative (t(1570) = 1.35, p = 0.179),
579 neutral (t(1565) = 0.06, p = 0.952), or positive films (t(1565) = 1.61, p = 0.108).
580
581 The linear mixed model of facial affect also revealed a similar significant interaction between
582 preceded by a positive film clip and film category (F(2,1573.8) = 4.12, p = 0.016). As with the
583 other interaction, this was driven by a significant impact of film category such that participants
584 expressed more negative emotions in response to negative films than neutral (t(1561) = -4.78, p <
585 0.001) or positive film clips (t(1558) = -14.96, p < 0.001) when the films were not preceded by a
586 positive film. Participants also expressed more positive emotions in response to the positive than
587 neutral films when they were not preceded by another positive film clip (t(1560) = 10.18, p <
588 0.001). Similarly, participants expressed more negative emotions in response to negative films
589 than neutral (t(1570) = -3.50, p = 0.001) or positive film clips (t(1569) = -5.92, p < 0.001) when
590 the films were preceded by a positive film. Participants also expressed more positive emotions in
591 response to the positive than neutral films when they were preceded by another positive film clip
592 (t(1567) = 2.53, p = 0.031). There was no significant effect of being preceded by a positive film
593 clip on participants' facial affect when watching negative (t(1568) = -1.06, p = 0.287), neutral
594 (t(1571) = -1.94, p = 0.052), or positive films (t(1563) = 1.93, p = 0.054). There were no significant
595 effects of being preceded by a negative (F(1,157.8) = 1.81, p = 0.179) or a positive film clip
596 (F(1,1559.7) = 0.43, p = 0.513) across all film categories suggesting no substantial carry-over
597 effects were precent in the facial expression data.
598

599 ## Correlations between self-reported mood, valence of facial expressions and
600 ## self-report questionnaire measures
601 There were no significant correlations between mean mood rating across all films, valence of facial
602 expressions across all films, eating disorder symptomatology as measured by the EDEQ total
603 score, anxiety as measured by HADS, or the BES total score (Supplemental Table S6).
604

Discussion

The present pilot study aimed to conduct initial evaluation to identify film stimuli that could be used in two tasks assessing different aspects of emotion generation, namely top-down interpretation of ambiguous scenarios and bottom-up evoked emotional responses. Most participants were not familiar with the short films used in the present study, which allowed us to avoid memory effects. Most of the film clips in the Ambiguous scenarios task elicited interpretations that fell into the expected neutral, positive, and negative categories. Additionally, with the exception of three film clips, they appeared to also be ambiguous with participants offering different types of interpretations. The exploratory factor analyses revealed that positive and negative film clips similarly formed their own factors when using participants self-reported mood and valence of facial expressions. No significant carry-over effects were observed in either task.

As anticipated participants mostly offered valence-congruent interpretations. Additionally, most film clips appeared to be ambiguous enough to allow participants to offer different interpretations, with only three of the film clips being deemed unambiguous. Although this suggests that most of the film clips would be suitable for use in a more naturalistic version of the sentence completion task to assess interpretation biases, it is important to note that the participants offered primarily negative interpretations for most negatively valenced ambiguous film clips and no or very few positive interpretations. The opposite pattern was not observed in the interpretations of the positively valenced film clips. This suggests that participants were more vigilant to detect negative cues and more subtle negative cues may be needed to preserve ambiguity. Similar issues have been reported in a previous study, where healthy participants' interpretation bias score was negative even after positive training indicating greater vigilance towards negative scenarios (Van Bockstaele et al., 2019). Another factor at paly could be the stimulus duration. A previous study exploring interpretation of negative facial expressions reported that increasing the duration of stimulus presentation appeared to result in increased ambiguity, that is fewer negative and greater number of alternative interpretations (Vassilopoulos, 2011). Having more time with the stimulus could give participants more opportunities to engage with cognitive appraisal and think of alternative explanations for the facial expressions. However, it is important to note that the negative faces were only presented for 200 or 500 milliseconds, which may not be directly comparable to film stimuli. To our knowledge no studies to date have examined the impact of film duration on ambiguity. Future studies may benefit from utilising more subtle negative elements to examine interpretation biases in negatively valenced ambiguous films as well as exploring the impact of stimulus duration.

Interestingly, participants frequently offered answers to question 1 that went beyond neutral description of the scene they were presented with, suggesting that the analysis of participants' responses to the present Ambiguous scenarios task could be expanded. In the future, this task could

644   be used to expand the field of interpretation bias research by separately examining how participants
645   generally view the current situation and how they view the future. Additionally, the present
646   paradigm involves participants writing complete sentences to answer the two questions after each
647   film, a qualitative analysis that goes beyond rating the valence of the answers could also be
648   employed. Using qualitative methods, such as thematic analysis, could enable researchers to
649   examine more subtle differences in the written responses and gain deeper insight into the
650   participants' thinking style. This method has been used to assess participants' responses to the
651   Frith-Happé theory of mind task (Abell, Happé and Frith, 2000). In this study people with AN
652   were found to focus more on the details and showed negative interpretation bias when describing
653   the movements of the triangles than healthy comparison participants even though no significant
654   quantitative differences in task performance were observed (Sedgewick et al., 2019). This suggests
655   that the present version of Ambiguous scenarios task could be a useful tool to assess both
656   quantitative and qualitative differences in top-down cognitive processing of ambiguous, social
657   stimuli.
658
659   In both exploratory factor analyses positive film clips loaded onto the same factor while negative
660   film clip loaded onto another factor, suggesting the positive and negative film clips successfully
661   evoked the intended mood and facial affect. This finding suggests that both measurement
662   modalities produced similar results, but the correlation results suggest that this was only the case
663   with some of the positive film clips. Although strong reactions were recorded and reported in
664   response to the negative film clips, no significant correlations were observed between the two
665   measures. Another recent study also using similar measures also found that while there was a
666   positive correlation between facial affect and self-reported mood when viewing positive pictures,
667   no significant correlation was observed when participants viewed negative pictures (Höfling et al.,
668   2020). Similar lack of association between facial expressions and self-reported mood has been
669   previously reported in response to negative film clips (Wang et al., 2008; Höfling et al., 2020).
670   These findings may reflect the fact that self-reported mood and spontaneous facial expressions can
671   provide different types of information about a person's emotional experience highlighting the need
672   to assess both. This is particularly relevant in the field of mood and eating disorders where people
673   have been documented to report experiencing emotions while suppressing their outward display
674   (Davies et al., 2013; Reed et al., 2007).
675
676   The above findings may suggest that self-reported mood ratings and facial expressions provide
677   different. Similar findings have previously been reported in other studies reporting that facial and
678   physiological reactions, such as changes in skin conductance, in response to emotionally
679   provoking film clips did not significantly correlate with self-reported emotional states (e.g.
680   Fernández et al., 2012; Wilms & Oberfeld, 2017; Gabel et al., 2019). Along the same lines, studies
681   exploring emotion regulation have reported that instructing participants to down regulate their
682   emotions through reappraisal results in reduction in self-reported mood, but not necessarily in
683   facial or physiological reactions to emotionally provoking stimuli (Lalot et al., 2014; Mohammed

684   et al., 2021). Such findings have led suggestions that having an emotional experience is not the
685   same as being aware of it and asking participants to rate or otherwise explain their current feelings
686   would require emotional awareness and possibly some top-down interpretation (Kim &
687   Fesenmaier, 2015; Russell, 2003). This seems to suggest that facial and physiological reactions
688   are distinct from self-reported mood states and utilising both could enable simultaneous
689   examination of both top-down and bottom-up emotion generation pathways. As everyday
690   emotional experiences have been proposed to be a result of a mixture of top-down and bottom-up
691   processes (McRae et al,. 2012; Ochsner et al., 2009), such a paradigm could be even more
692   naturalistic and have greater ecological validity. Further investigation of the mechanisms that
693   underlie differences between self-reported mood ratings and facial expressions may be of interest
694   to explore whether both emotional reactivity and interpretation biases could be examined
695   simultaneously.

696
697

698   In the present study we did not find any significant correlations between interpretations of or
699   reactions to emotional film clips and the clinical self-report measures. This is unexpected
700   considering that there is a wealth of previous work indicating that people with depression and
701   eating disorders show reduced facial reactivity to emotionally provoking stimuli (Panaite et al.,
702   2017; Davies et al., 2016; Dapelo et al, 2016; Leppanen et al., 2017). Similarly, several studies
703   have also documented that negative interpretation bias is a key feature in depression and social
704   anxiety (Everaert et al., 2017; Chen et al., 2020) and it has recently also been documented in eating
705   disorders (Rowlands et al., 2020; Dapelo et al., 2016). On one hand, this may be because
706   performance on interpretation of the ambiguous films and reactivity to the emotionally provoking
707   films bear no association with depression, anxiety or eating disorder psychopathology. On the
708   other hand, this finding may also be driven by the fact that the present sample consisted of healthy
709   young women, which naturally led to reduced variability in the clinical self-report measures.
710   Further investigation of whether performance on these tasks is associated with relevant clinical
711   measures among people with anorexia nervosa may be of interest to further shed light on whether
712   these film stimuli target the relevant aspects of social-emotional processing in this illness.

713

Recommendations for use

715   The film clips we recommend for the Ambiguous scenarios and Evoked emotions tasks are listed
716   below in Table 5. Based on our findings most of the negative film clips in the Ambiguous scenarios
717   task appeared to yield primarily negative responses and we would, therefore, not recommend most
718   of them. Instead, we recommend that some of the film clips initially categorised as neutral but
719   were interpreted primarily as neutral or negative. These films would provide enough variance in
720   participants responses to statistically examine differences between clinical and general populations
721   without risking ceiling effects. Additionally, Film 17, which was originally categorised as
722   positively valenced, may also be used to as a negatively valenced ambiguous film, particularly in

723 terms of Question 2, which asks participants to make predictions about how the situation might
724 turn out. We would recommend the use of most of the film clips initially categorised as positively
725 valenced with the exception of Films 17. Finally, based on our findings we cannot recommend any
726 of the films clips evaluated here to be used as truly neutral ambiguous stimuli due to some degree
727 of bias being observed in all films.

728

729 Most of the film clip used in the Evoked emotions task elicited the intended mood ratings and
730 facial affect. Only two films, both of which were initially labelled as neutral (Films 7 and 8), should
731 be excluded based on the present findings for producing unexpectedly positive and negative
732 responses. If these films clips are used to study evoked facial affect, it is of note that there was
733 substantial variability between in the facial affect data. Thus, large sample size may be needed to
734 examine true differences between any clinical and healthy control groups. Additionally, care
735 should be taken during the video recording of participants' faces to ensure the recordings can be
736 analysed accurately. For instance, lighting can impact any automated facial expression analysis
737 software's ability find the face and participants should be instructed to not move too much or cover
738 their faces by resting the heads in their hands.

739

740 *Table 5. Recommended film clips*
741 ---TABLE 5 HERE ---
742 Min = minutes

743

744 ## Limitations

745 The main limitation of this study was the small sample size. Although the sample size was
746 sufficient for the ambiguity check, it may not have been sufficient for the exploratory factor
747 analyses. Even though the two exploratory factor analyses overall produced similar results, the
748 small sample size may go some way to explain the number of cross-loadings observed with the
749 neutral film clips. Furthermore, due to the SARS-CoV-2 pandemic, most participants had to take
750 part in the study remotely from home. This may have impacted data quality introducing variability
751 in the quality of the video recordings of participants' facial expressions, which may have impacted
752 the FaceReader analysis and thus the factor analysis. Furthermore, a recent study found that
753 FaceReader's valence measure may not distinguish well between reactions to unpleasant and
754 neutral stimuli, even when the difference could be established using facial electromyography
755 (Höfling et al., 2020). The authors suggest that FaceReader may lack sensitivity in accurately
756 detecting true negative facial expressions while simultaneously having negative bias while
757 analysing neutral faces. This is relevant as in the present study the neutral film clips loaded across
758 all factors in the factor analysis that used facial expression data, which was not the case in the
759 factor analysis using mood ratings. Thus, greater accuracy of computerised facial affect
760 recognition tools further examination of participants' facial affect under controlled laboratory
761 conditions are needed.

762

763 Additionally, it is possible that the lack of association between facial affect and participants mood
764 ratings was due to the use of FaceReader. We averaged the valence of participants facial
765 expressions over the duration of the film clip. As most participants displayed neutral facial
766 expressions most of the time, this approach led to only small differences in valence of facial affect
767 between the film categories. Other studies have explored the use of dominant basic emotion rather
768 than using averaged emotion intensity or valence scores, which has been found to be comparable
769 to the manual Facial Action Coding System (FACS) (Skiendziel, Rösch and Schultheiss, 2019).
770 However, such an approach would result in categorical data, which does not take other secondary
771 emotions expressed during the film into consideration. Additionally, there is some uncertainty
772 regarding whether FACS-based scoring would provide greater accuracy (Limbrecht-Ecklundt et
773 al., 2016). Therefore, more work is needed to establish suitable methods to summarise the frame-
774 by-frame facial expression data produced by computerised analysis tools, such as FaceReader.

775

776 Another limitation was that we used a convenience sample of young women aged 18 to 25 years.
777 Thus, the results cannot be generalised beyond this population. This was done with the intention
778 of using some of the stimuli in a case-control study to examine social-emotional processing
779 difficulties among people with AN, who are mostly female. Additionally, to avoid effects of severe
780 and enduring form of eating disorder, any future case-control research should focus on young
781 people during the early stage of the illness. To expand beyond this sample, further evaluation of
782 the stimuli is needed with a more diverse and larger sample. Additionally, the sample size is
783 modest, particularly for factor analyses. Thus, further examination of the stimuli in the context of
784 Ambiguous scenarios and Evoked emotions tasks is still needed.

785

786 Finally, although we did not find statistically significant carry-over effects in either task, some
787 trend level effects were observed in the Evoked emotions task. It appeared that participants
788 reactions to neutral and positive films clips may have been somewhat influenced by the films being
789 preceded by other positive films. This suggests that increasing the inter-stimulus interval for a task
790 assessing participants mood and evoked facial expressions beyond 200ms may be required to
791 eliminate any potential carry-over effects. It is possible that similar trend level carry-over effects
792 were not seen in the Ambiguous scenarios task because participants were required to write answers
793 to two questions after each film clip, which took longer than rating their mood and alertness. Thus,
794 the ultimate inter-stimulus interval in the Ambiguous scenarios task extended beyond that on the
795 one in the Evoked emotions task.

796

797

Conclusions

799   This study aimed to evaluate two sets of 18 film clips to be used in computerised tasks assessing
800   interpretation of ambiguous scenarios and evoked emotional responses. Altogether, 124 healthy
801   female participants aged 18 – 25 took part in the evaluation study. The participants completed two
802   tasks in which they watched all film clips and responded to questions after each clip. During the
803   Evoked emotions task participants facial expressions were also recorded. Due to the SARS-CoV-
804   2 pandemic more than half of the participants completed the study remotely. The ambiguous film
805   clips elicited the intended primarily neutral, positive, and negative interpretations while remaining
806   ambiguous. However, three of the film clips were deemed unambiguous. Additionally, participants
807   were generally more attuned to the negative cues suggesting that to preserve ambiguity more subtle
808   negatively valenced film clips should be used to assess interpretation biases. Participants' evoked
809   facial expressions and mood ratings in response to the film clips used in the Evoked emotions task
810   were analysed with two exploratory factor analyses to identify film clips that elicited the intended
811   reactions. The factor analyses revealed that the positive and negative emotionally provoking film
812   clips formed their own factors, while there was substantial cross-loading with the neutral film clips
813   when facial affect data was used. This could reflect reduced data quality due to most of the
814   participants taking part remotely with webcams and internet connections of varying quality. Still,
815   the findings show that a subset of the film clips evaluated in the present study could be used to
816   assess interpretation biases and emotional reactions using a new, more ecologically valid set of
817   stimuli within a clinical population.

818

819   **Acknowledgements**

820

References

Abell, F., Happé, F. and Frith, U. (2000) 'Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development', *Cognitive Development*. Elsevier Ltd, 15(1), pp. 1–16. doi: 10.1016/S0885-2014(00)00014-9.

Alghowinem, S., Goecke, R., Wagner, M., & Alwabil, A. (2019). Evaluating and validating emotion elicitation using English and Arabic movie clips on a Saudi sample. Sensors, 19(10), 2218.

Ambwani, S., Berenson, K. R., Simms, L., Li, A., Corfield, F., & Treasure, J. (2016). Seeing things differently: An experimental investigation of social cognition and interpersonal behavior in anorexia nervosa. International Journal of Eating Disorders, 49(5), 499-506.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. Behavior research methods, 52(1), 388-407.

Aguado, L., Fernández-Cahill, M., Román, F. J., Blanco, I., & de Echegaray, J. (2018). Evaluative and psychophysiological responses to short film clips of different emotional content. Journal of Psychophysiology, 32(1), 1.

Arnaudova, I. and Hagenaars, M. A. (2017) 'Lights … action: Comparison of trauma films for use in the trauma film paradigm', *Behaviour Research and Therapy*. Elsevier Ltd, 93, pp. 67–77. doi: 10.1016/j.brat.2017.02.007.

Betella, A. and Verschure, P. F. M. J. (2016) 'The affective slider: A digital self-assessment scale for the measurement of human emotions', *PLoS ONE*. Public Library of Science, 11(2), p. e0148037. doi: 10.1371/journal.pone.0148037.

Van Bockstaele, B., Notebaert, L., Salemink, E., Clarke, P. J., MacLeod, C., Wiers, R. W., & Bögels, S. M. (2019). Effects of interpretation bias modification on unregulated and regulated emotional reactivity. Journal of Behavior Therapy and Experimental Psychiatry, 64, 123-132.

Burgess, P. W. and Shallice, T. (1996) 'Response suppression, Initiation and strategy use following frontal lobe lesions', *Neuropsychologia*. Elsevier Ltd, 34(4), pp. 263–272. doi: 10.1016/0028-3932(95)00104-2.

Butler, E. A., Egloff, B., Wlhelm, F. H., Smith, N. C., Erickson, E. A., & Gross, J. J. (2003). The social consequences of expressive suppression. Emotion, 3(1), 48.

Caglar-Nazali, H.P., Corfield, F., Cardi, V., Ambwani, S., Leppanen, J., Olabintan, O., Deriziotis, S., Hadjimichalis, A., Scognamiglio, P., Eshkevari, E. & Micali, N. (2014) 'A systematic review and meta-analysis of "Systems for Social Processes" in eating disorders', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, pp. 55–92. doi: 10.1016/j.neubiorev.2013.12.002.

Cardi, V., Turton, R., Schifano, S., Leppanen, J., Hirsch, C. R., & Treasure, J. (2017). Biased interpretation of ambiguous social scenarios in anorexia nervosa. European Eating Disorders Review, 25(1), 60-64.

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

861   Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive
862   functions: Review of instruments and identification of critical issues. Archives of clinical
863   neuropsychology, 23(2), 201-216.
864   Chen, J., Short, M., & Kemps, E. (2020). Interpretation bias in social anxiety: A systematic
865   review and meta-analysis. Journal of Affective Disorders, 276, 1119-1130.
866   Christensen, R. H. B. (2019) 'ordinal - Regression Models for Ordinal Data'.
867   Cisler, J. M., Olatunji, B. O., Feldner, M. T., & Forsyth, J. P. (2010). Emotion regulation and the
868   anxiety disorders: An integrative review. Journal of psychopathology and behavioral assessment,
869   32(1), 68-82.
870   Clark, I. A., Mackay, C. E. and Holmes, E. A. (2013) 'Positive involuntary autobiographical
871   memories: You first have to live them', *Consciousness and Cognition*. Academic Press Inc.,
872   22(2), pp. 402–406. doi: 10.1016/j.concog.2013.01.008.
873   Conway, M. A. and Loveday, C. (2010) 'Accessing Autobiographical Memories', in *The Act of*
874   *Remembering*. Oxford, UK: Wiley-Blackwell, pp. 56–70. doi: 10.1002/9781444328202.ch4.
875   Dapelo, M. M., Surguladze, S., Morris, R., & Tchanturia, K. (2016). Emotion Recognition in
876   Blended Facial Expressions in Women with Anorexia Nervosa. European eating disorders
877   review : the journal of the Eating Disorders Association, 24(1), 34–42.
878   https://doi.org/10.1002/erv.2403
879   Dapelo, M. M., Hart, S., Hale, C., Morris, R., & Tchanturia, K. (2016). Expression of positive
880   emotions differs in illness and recovery in anorexia nervosa. Psychiatry research, 246, 48–51.
881   https://doi.org/10.1016/j.psychres.2016.09.014
882   Davies, H., Wolz, I., Leppanen, J., Fernandez-Aranda, F., Schmidt, U., & Tchanturia, K. (2016).
883   Facial expression to emotional stimuli in non-psychotic disorders: A systematic review and
884   meta-analysis. Neuroscience & Biobehavioral Reviews, 64, 252-271.
885   Davies, H., Schmidt, U., & Tchanturia, K. (2013). Emotional facial expression in women
886   recovered from anorexia nervosa. BMC psychiatry, 13(1), 1-9.
887   Davis, E. G., Foland-Ross, L. C., & Gotlib, I. H. (2018). Neural correlates of top-down
888   regulation and generation of negative affect in major depressive disorder. Psychiatry Research:
889   Neuroimaging, 276, 1-8.
890   Deng, Y., Yang, M., & Zhou, R. (2017). A new standardized emotional film database for Asian
891   culture. Frontiers in psychology, 8, 1941.
892   Everaert, J., Podina, I. R., & Koster, E. H. (2017). A comprehensive meta-analysis of
893   interpretation biases in depression. Clinical Psychology Review, 58, 33-48.
894   Fairburn, C. G. and Beglin, S. J. (1994) 'Assessment of eating disorders: interview or self-report
895   questionnaire?', *The International journal of eating disorders*. United States, 16(4), pp. 363–370.
896   Fernández, C., Pascual, J. C., Soler, J., Elices, M., Portella, M. J., & Fernández-Abascal, E.
897   (2012). Physiological responses induced by emotion-eliciting films. Applied psychophysiology
898   and biofeedback, 37(2), 73-79.

899 First MB, Williams JBW, Karg RS, Spitzer RL: Structured Clinical Interview for DSM-5—
900 Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV). Arlington, VA,
901 American Psychiatric Association, 2015
902 Gabel, L. N., Daoust, A. R., Salisbury, M. R., Grahn, J. A., Durbin, C. E., & Hayden, E. P.
903 (2019). Development and validation of a battery of emotionally evocative film clips for use with
904 young children. Psychological Assessment, 31(8), 1040.
905 Gilman, T. L., Shaheen, R., Nylocks, K. M., Halachoff, D., Chapman, J., Flynn, J. J., Matt, L.
906 M., & Coifman, K. G. (2017). A film set for the elicitation of emotion in research: A
907 comprehensive catalog derived from four decades of investigation. Behavior research methods,
908 49(6), 2061–2082. https://doi.org/10.3758/s13428-016-0842-x
909 Greenaway, K. H., Kalokerinos, E. K., & Williams, L. A. (2018). Context is everything (in
910 emotion research). Social and Personality Psychology Compass, 12(6), e12393.
911 Gross, J. J. (2002) 'Emotion regulation: Affective, cognitive, and social consequences',
912 *Psychophysiology*. Cambridge University Press, pp. 281–291. doi:
913 10.1017/S0048577201393198.
914 Guala, F. (2012) 'Reciprocity: Weak or strong? What punishment experiments do (and do not)
915 demonstrate', *Behavioral and Brain Sciences*. Cambridge University Press, 35(1), pp. 1–15. doi:
916 10.1017/S0140525X11000069.
917 Guillery-Girard, B., Clochon, P., Giffard, B., Viard, A., Egler, P. J., Baleyte, J. M., Eustache, F.,
918 & Dayan, J. (2013). "Disorganized in time": impact of bottom-up and top-down negative
919 emotion generation on memory formation among healthy and traumatized adolescents. Journal of
920 physiology, Paris, 107(4), 247–254. https://doi.org/10.1016/j.jphysparis.2013.03.004
921 Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional
922 disorders: The nature and role of interpretation biases. Annual review of clinical psychology, 12,
923 281-305.
924 Höfling, T. T. A., Gerdes, A. B., Föhl, U., & Alpers, G. W. (2020). Read my face: automatic
925 facial coding versus psychophysiological indicators of emotional valence and arousal. Frontiers
926 in psychology, 11, 1388.
927 Hogenelst, K., Schoevers, R. A. and aan het Rot, M. (2015) 'Studying the neurobiology of
928 human social interaction: Making the case for ecological validity', *Social Neuroscience*.
929 Routledge, 10(3), pp. 219–229. doi: 10.1080/17470919.2014.994786.
930 Huppert, J. D., Pasupuleti, R. V., Foa, E. B., & Mathews, A. (2007). Interpretation biases in
931 social anxiety: Response generation, response selection, and self-appraisals. Behaviour research
932 and therapy, 45(7), 1505-1515.
933 Jansen, M., Overgaauw, S. and De Bruijn, E. R. A. (2020) 'Social Cognition and Obsessive-
934 Compulsive Disorder: A Review of Subdomains of Social Functioning', *Frontiers in Psychiatry*.
935 Frontiers Media S.A. doi: 10.3389/fpsyt.2020.00118.
936 Janssens, M., Lataster, T., Simons, C. J. P., Oorschot, M., Lardinois, M., Van Os, J., & Myin-
937 Germeys, I. (2012). Emotion recognition in psychosis: no evidence for an association with real
938 world social functioning. Schizophrenia research, 142(1-3), 116-121.

939 Jurado, M. B. and Rosselli, M. (2007) 'The elusive nature of executive functions: A review of
940 our current understanding', *Neuropsychology Review*. Neuropsychol Rev, pp. 213–233. doi:
941 10.1007/s11065-007-9040-z.
942 Kerr-Gaffney, J., Mason, L., Jones, E., Hayward, H., Ahmad, J., Harrison, A., Loth, E., Murphy,
943 D. & Tchanturia, K. (2020) 'Emotion Recognition Abilities in Adults with Anorexia Nervosa are
944 Associated with Autistic Traits', *Journal of Clinical Medicine 2020, Vol. 9, Page 1057*.
945 Multidisciplinary Digital Publishing Institute, 9(4), p. 1057. doi: 10.3390/JCM9041057.
946 Kim, J., & Fesenmaier, D. R. (2015). Measuring emotions in real time: Implications for tourism
947 experience design. Journal of Travel Research, 54(4), 419-429.
948 Kupferberg, A., Bicks, L. and Hasler, G. (2016) 'Social functioning in major depressive
949 disorder', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, pp. 313–332. doi:
950 10.1016/j.neubiorev.2016.07.002.
951 Lalot, F., Delplanque, S., & Sander, D. (2014). Mindful regulation of positive emotions: a
952 comparison with reappraisal and expressive suppression. Frontiers in psychology, 5, 243.
953 Leppanen, J., Dapelo, M. M., Davies, H., Lang, K., Treasure, J., & Tchanturia, K. (2017).
954 Computerised analysis of facial emotion expression in eating disorders. PLoS One, 12(6),
955 e0178972.
956 Leppanen, J., Tosunlar, L., Blackburn, R., Williams, S., Tchanturia, K., & Sedgewick, F. (2021).
957 Critical incidents in anorexia nervosa: perspectives of those with a lived experience. Journal of
958 eating disorders, 9(1), 1-14.
959 Leslie, M., Halls, D., Leppanen, J., Sedgewick, F., Smith, K., Hayward, H., Lang, K., Fonville,
960 L., Simic, M., Mandy, W. & Nicholls, D. (2020) 'Neural Correlates of Theory of Mind Are
961 Preserved in Young Women With Anorexia Nervosa', *Frontiers in psychology*. Front Psychol,
962 11. doi: 10.3389/FPSYG.2020.568073.
963 Levitt, S. D. and List, J. A. (2007) 'What Do Laboratory Experiments Measuring Social
964 Preferences Reveal About the Real World?', *Journal of Economic Perspectives*, 21(2), pp. 153–
965 174. doi: 10.1257/jep.21.2.153.
966 Limbrecht-Ecklundt, K., Werner, P., Traue, H. C., Al-Hamadi, A., & Walter, S. (2016). Mimic
967 activity of differentiated pain intensities: Correlation of characteristics of Facial Action Coding
968 System and electromyography. Schmerz (Berlin, Germany), 30(3), 248-256.
969 McRae, K., Misra, S., Prasad, A. K., Pereira, S. C., & Gross, J. J. (2012). Bottom-up and top-
970 down emotion generation: implications for emotion regulation. Social cognitive and affective
971 neuroscience, 7(3), 253-262.
972 Maksimainen, J., Wikgren, J., Eerola, T., & Saarikallio, S. (2018). The effect of memory in
973 inducing pleasant emotions with musical and pictorial stimuli. Scientific reports, 8(1), 1-12.
974 Mohammed, A. R., Kosonogov, V., & Lyusin, D. (2021). Expressive suppression versus
975 cognitive reappraisal: effects on self-report and peripheral psychophysiology. International
976 Journal of Psychophysiology, 167, 30-37.
977 Ochsner, K. N., Ray, R. R., Hughes, B., McRae, K., Cooper, J. C., Weber, J., Gabrieli, J. D. E.,
978 & Gross, J. J. (2009). Bottom-Up and Top-Down Processes in Emotion Generation: Common

979 and Distinct Neural Mechanisms. Psychological Science, 20(11), 1322–1331.

980 https://doi.org/10.1111/j.1467-9280.2009.02459.x

981 Panaite, V., Whittington, A., & Cowden Hindash, A. (2018). The role of appraisal in dysphoric

982 affect reactivity to positive laboratory films and daily life events in depression. Cognition and

983 Emotion, 32(6), 1362-1373.

984 Parkinson, B. (2005). Do facial movements express emotions or communicate motives?.

985 Personality and Social Psychology Review, 9(4), 278-311.

986 Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E. and

987 Lindeløv, J.K. (2019) 'PsychoPy2: Experiments in behavior made easy', *Behavior Research*

988 *Methods*. Springer New York LLC, 51(1), pp. 195–203. doi: 10.3758/s13428-018-01193-y.

989 Quigley, K. S., Lindquist, K. A., & Barrett, L. F. (2014). Inducing and measuring emotion and

990 affect: Tips, tricks, and secrets.

991 R Core Team (2020) 'R: A Language and Environment for Statistical Computing'. Vienna,

992 Austria. Available at: https://www.r-project.org/.

993 Reed, L. I., Sayette, M. A., & Cohn, J. F. (2007). Impact of depression on response to comedy: A

994 dynamic facial coding analysis. Journal of abnormal psychology, 116(4), 804.

995 Revelle, W. (2020) 'psych: Procedures for Personality and Psychological Research'.

996 Northwestern University, Evanston, Illinois, USA.

997 Rimes, K. A. and Chalder, T. (2010) 'The Beliefs about Emotions Scale: Validity, reliability and

998 sensitivity to change', *Journal of Psychosomatic Research*, 68(3), pp. 285–292. doi:

999 10.1016/j.jpsychores.2009.09.014.

1000 Rø, Ø., Reas, D. L. and Stedal, K. (2015) 'Eating Disorder Examination Questionnaire (EDE-Q)

1001 in Norwegian Adults: Discrimination between Female Controls and Eating Disorder Patients',

1002 *European Eating Disorders Review*. John Wiley and Sons Ltd, 23(5), pp. 408–412. doi:

1003 10.1002/erv.2372.

1004 Rowlands, K., Wilson, E., Simic, M., Harrison, A., & Cardi, V. (2020). A Critical Review of

1005 Studies Assessing Interpretation Bias Towards Social Stimuli in People With Eating Disorders

1006 and the Development and Pilot Testing of Novel Stimuli for a Cognitive Bias Modification

1007 Training. Frontiers in Psychology, 11, 2374.

1008 Russell, V. L. (2021) 'emmeans: Estimated Marginal Means, aka Least-Squares Means'.

1009 Russell, J. A. (2003). Core affect and the psychological construction of emotion. Psychological

1010 review, 110(1), 145.

1011 Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A., & Gross, J. J. (2016). Eliciting

1012 positive, negative and mixed emotional states: A film library for affective scientists. Cognition

1013 and emotion, 30(5), 827-856.

1014 Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large

1015 database of emotion-eliciting films: A new tool for emotion researchers. Cognition and emotion,

1016 24(7), 1153-1172.

1017   Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary
1018   questions in facial expression research. American Journal of Physical Anthropology: The
1019   Official Publication of the American Association of Physical Anthropologists, 116(S33), 3-24.
1020   Schoth, D. E., & Liossi, C. (2017). A systematic review of experimental paradigms for exploring
1021   biased interpretation of ambiguous information with emotional and neutral associations.
1022   Frontiers in psychology, 8, 171.
1023   Sedgewick, F., Leppanen, J., Goh, F., Hayward, H., Happé, F., & Tchanturia, K. (2019).
1024   Similarities and differences in theory of mind responses of patients with anorexia nervosa with
1025   and without autistic features. Frontiers in Psychiatry, 10, 318.
1026   Sheppes, G., & Gross, J. J. (2011). Is timing everything? Temporal considerations in emotion
1027   regulation. Personality and Social Psychology Review, 15(4), 319-331.
1028   Siedlecka, E., & Denson, T. F. (2019). Experimental methods for inducing basic emotions: A
1029   qualitative review. Emotion Review, 11(1), 87-97.
1030   Skiendziel, T., Rösch, A. G. and Schultheiss, O. C. (2019) 'Assessing the convergent validity
1031   between the automated emotion recognition software Noldus FaceReader 7 and Facial Action
1032   Coding System Scoring', *PLoS ONE*. Public Library of Science, 14(10). doi:
1033   10.1371/journal.pone.0223905.
1034   Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: critically
1035   acclaimed. Trends in cognitive sciences, 23(8), 699-714.
1036   Treasure, J. and Schmidt, U. (2013) 'The cognitive-interpersonal maintenance model of anorexia
1037   nervosa revisited: a summary of the evidence for cognitive, socio-emotional and interpersonal
1038   predisposing and perpetuating factors', *Journal of Eating Disorders*, 1(1), p. 13. doi:
1039   10.1186/2050-2974-1-13.
1040   Treasure, J., Stein, D. and Maguire, S. (2015) 'Has the time come for a staging model to map the
1041   course of eating disorders from high risk to severe enduring illness? An examination of the
1042   evidence', *Early Intervention in Psychiatry*. Wiley Online Library, 9(3), pp. 173–184. doi:
1043   10.1111/eip.12170.
1044   Vassilopoulos, S. P. (2011). Interpretation bias for facial expressions in high and low socially
1045   anxious individuals: Effects of stimulus duration. The Hellenic Journal of Psychology, 8, 44-65.
1046   Wallisch, P., & Whritner, J. A. (2017). Strikingly low agreement in the appraisal of motion
1047   pictures. Projections, 11(1), 102-120.
1048   Wang, N., Marsella, S., & Hawkins, T. (2008, May). Individual differences in expressive
1049   response: a challenge for ECA design. In AAMAS (3) (pp. 1289-1292).
1050   Wilms, L., & Oberfeld, D. (2018). Color and emotion: effects of hue, saturation, and brightness.
1051   Psychological research, 82(5), 896-914.
1052   Zigmond, A. S. and Snaith, R. P. (1983) 'The hospital anxiety and depression scale.', *Acta
1053   psychiatrica Scandinavica*, 67(6), pp. 361–70. Available at:
1054   http://www.ncbi.nlm.nih.gov/pubmed/6880820 (Accessed: 16 October 2018).

Figure 1

Fig 1 . X^2 contribution plot.

The dots indicate contribution each film clip made to a specific interpretation category. Larger blue dots indicate higher contribution and lower ambiguity with most participants offering the same type of interpretation

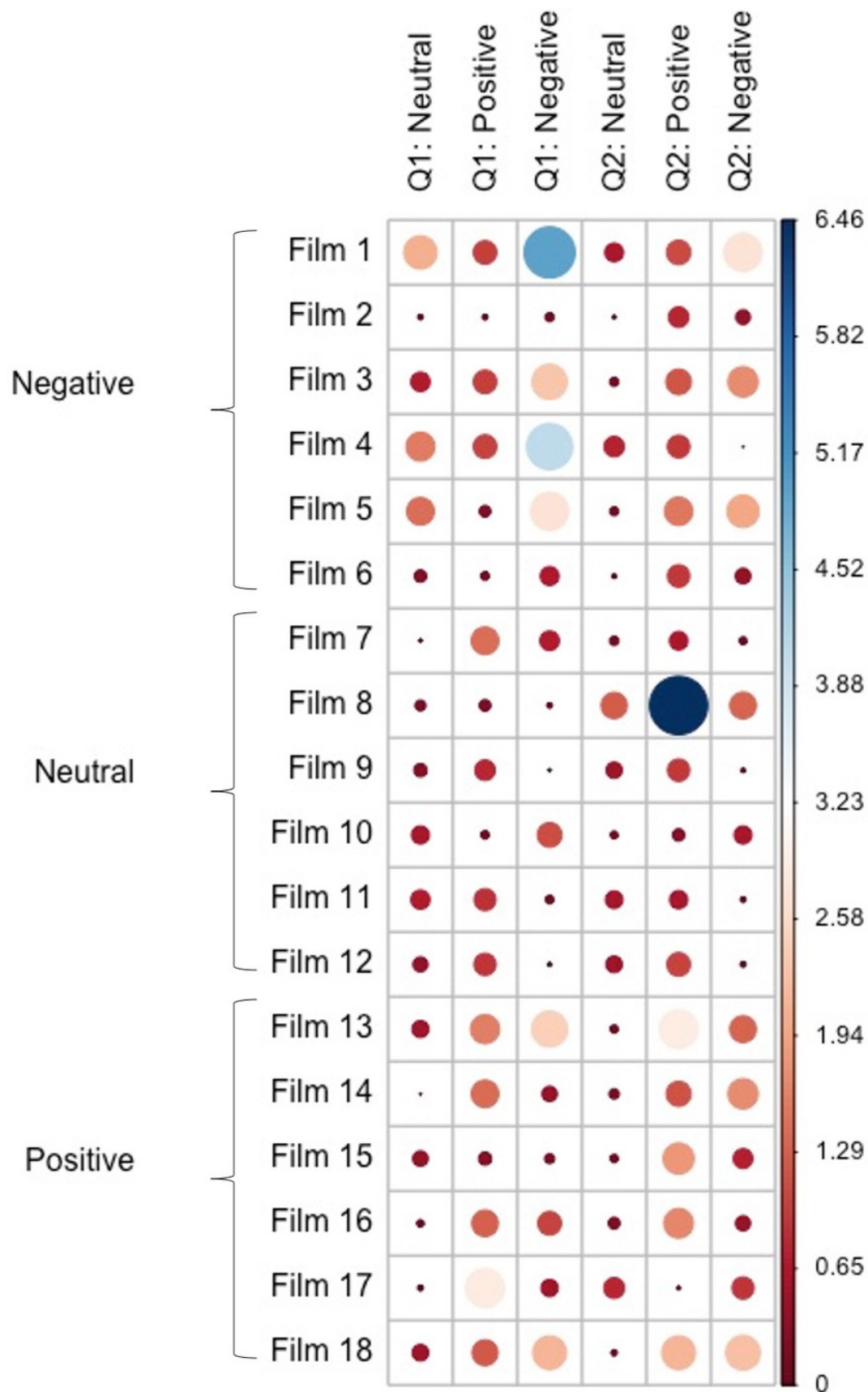
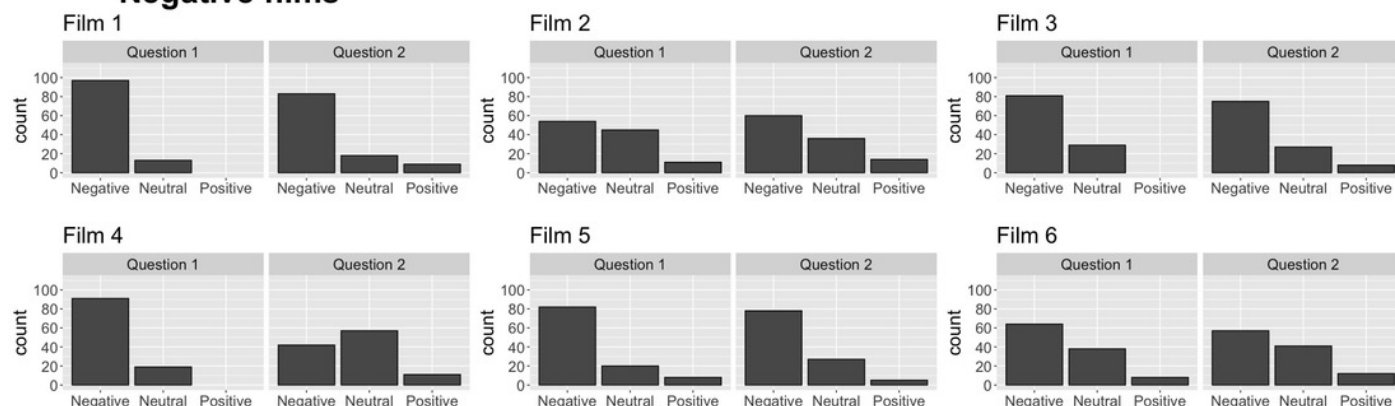



Figure 2

Evaluation of film stimuli for the assessment of social-emotional processing: A pilot study Fig 2. Histogram plots of participants interpretations of each film clip

Evaluation of film stimuli for the assessment of social-emotional processing: A pilot study The histograms show the frequency of positive, neutral and negative answers participants offered to the two questions after each film clip. Films 1 - 6 were categorised as negatively valenced ambiguous film clips, Films 7 - 12 were categorised as neutral ambiguous film clips, and Films 13 - 18 were categorised as positively valenced ambiguous film clips.

Negative films

Film 1

Film 2

Film 3

Film 4

Film 5

Film 6

Neutral films

Film 7

Film 8

Film 9

Film 10

Film 11

Film 12

Positive films

Film 13

Film 14

Film 15

Film 16

Film 17

Film 18

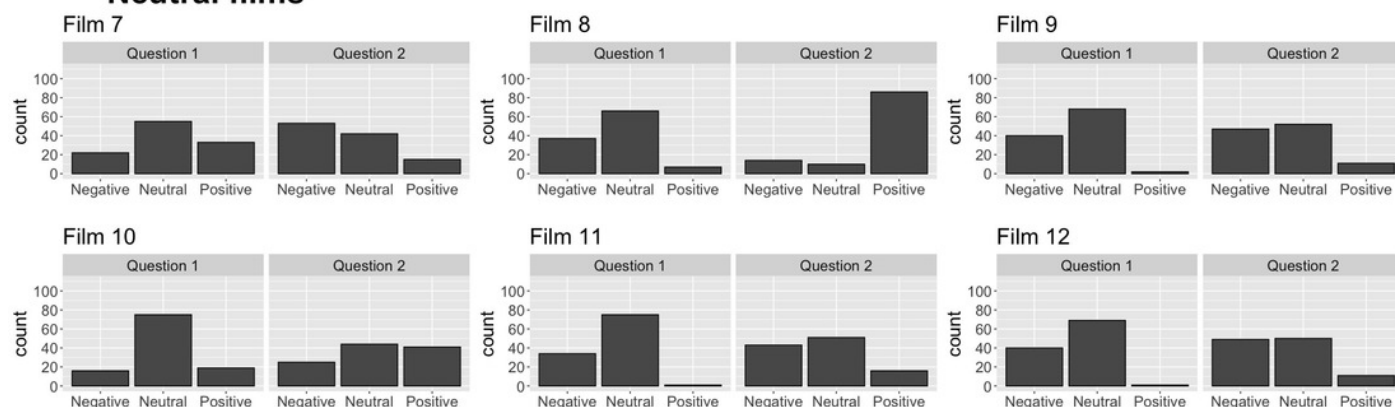
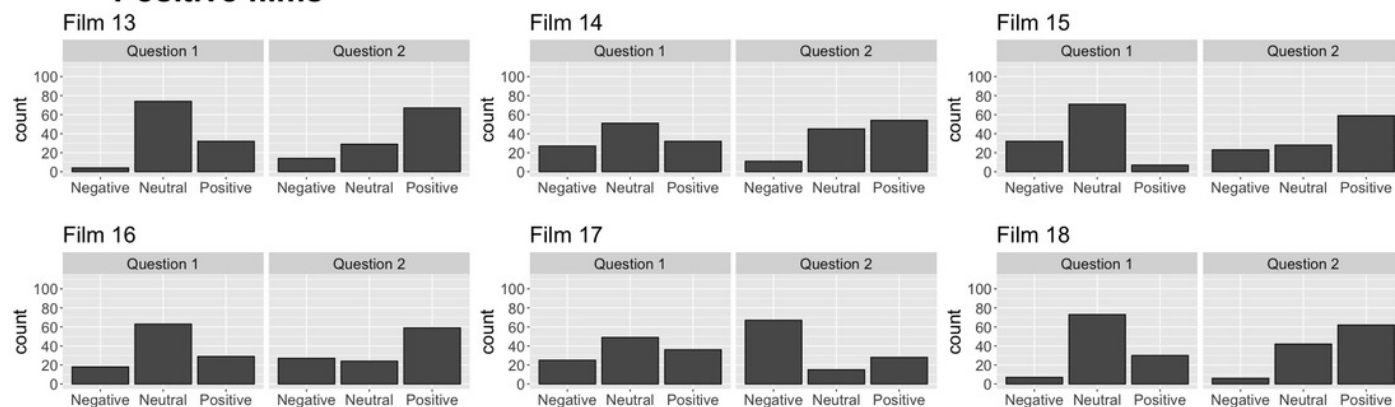



Table 1. Participant characteristics (N = 124)

[1] (Rø, Reas and Stedal, 2015) recommended using EDEQ total score above 2.5 as an indicator of clinical significance; [2] the mean is based on what was reported in the initial validation of the scale among healthy people (Rimes and Chalder, 2010). BMI = body mass index, HADS = Hospital Anxiety and Depression Scale, EDEQ = Eating Disorder Examination Questionnaire, BES = Beliefs about Emotions Scale, NVQ = National Vocational Qualifications, BTEC = Business and Technology Education Council, SD = standard deviation, N/A = not applicable.

1   *Table 1. Participant characteristics (N = 124)*

	Current sample	Normative data
Age M (SD); (range)	22.0 (2.3); (18 – 25)	N/A
BMI M (SD); (range)	22.2 (5.4); (17.8 – 39.7)	18.5 – 24.9
EDEQ Total M (SD); (range)	0.6 (0.6); (0.0 – 2.46)	$< 2.5_1$
HADS Anxiety M (SD); (range)	3.3 (2.8); (0.0 – 13.0)	0 – 7
HADS Depression M (SD); (range)	2.2 (2.5); (0.0 – 14.0)	0 – 7
BES M (SD); (range)	26.1 (11.6); (3.0 – 62.0)	Mean=27.9 (SD=11.3)$_2$
Employment status N (%)	Full-time student: 70 (56.5%) Full-time employment: 18 (14.5%) Student in part-time employment: 16 (12.9%) Part-time employment: 8 (6.5%) Unemployment: 6 (4.8%) Student in full-time employment: 1 (0.8%) Did not disclose: 5 (4.0%)	N/A
Level of education N (%)	A Level, NVQ or equivalent qualification: 47 (37.9%) Undergraduate degree: 40 (32.3%) Postgraduate degree: 28 (22.6%) BTEC or equivalent Diploma: 3 (2.4%) O Level or General Certificate of Secondary education: 1 (0.8%) No qualifications: 1 (0.8%) Did not disclose: 3 (2.4%)	N/A
Marital status N (%)	Single: 105 (84.7%) Domestic partnership: 14 (11.3%) Married: 1 (0.8%) In a relationship: 1 (0.8%) Did not disclose: 3 (2.4%)	N/A

Ethnicity N (%)	White: 71 (57.3%) Asian: 29 (23.4%) Mixed: 9 (7.3%) Black: 8 (6.5%) Middle Eastern: 4 (3.2%) Did not disclose: 3 (2.4%)	N/A

2 [1] (Rø, Reas and Stedal, 2015) recommended using EDEQ total score above 2.5 as an indicator of
3 clinical significance; [2] the mean is based on what was reported in the initial validation of the scale
4 among healthy people (Rimes and Chalder, 2010) . BMI = body mass index, HADS = Hospital
5 Anxiety and Depression Scale, EDEQ = Eating Disorder Examination Questionnaire, BES =
6 Beliefs about Emotions Scale, NVQ = National Vocational Qualifications, BTEC = Business and
7 Technology Education Council, SD = standard deviation, N/A = not applicable.

Table 2. Valence of and number of words in written interpretations

1   *Table 2. Valence of and number of words in written interpretations*

Film category	Question 1 Valence M (SD)	Question 2 Valence M (SD)
Neutral	-0.18 (0.58)	-0.07 (0.78)
Positive	0.09 (0.65)	0.29 (0.80)
Negative	-0.66 (0.56)	-0.51 (0.65)
Total	-0.25 (0.67)	-0.10 (0.82)

Table 3 . Self-rated mood data factor loadings

Factor loadings < 0.3 were excluded

1   *Table 3. Self-rated mood data factor loadings*

Film	Film category	Factor 1	Factor 2	Factor 3
Film 6	Negative	0.709		
Film 2	Negative	0.683		
Film 5	Negative	0.670		
Film 4	Negative	0.614		
Film 3	Negative	0.449		
Film 8	Neutral	0.531		
Film 1	Negative	0.430		
Film 12	Neutral		0.819	
Film 11	Neutral		0.498	
Film 10	Neutral		0.428	0.406
Film 9	Neutral		0.375	
Film 16	Positive			0.547
Film 13	Positive			0.518
Film 18	Positive			0.511
Film 7	Neutral			0.402
Film 17	Positive			0.344
Film 15	Positive			0.313
Film 14	Positive			

2   Factor loadings < 0.3 were excluded

Table 4 . Facial expression data factor loadings

Factor loadings < 0.3 were excluded

1   *Table 4. Facial expression data factor loadings*

Film	Film category	Factor 1	Factor 2	Factor 3
Film 14	Positive	0.912		
Film 16	Positive	0.809		
Film 15	Positive	0.779		
Film 18	Positive	0.761		
Film 13	Positive	0.721		
Film 10	Neutral	0.608	0.511	
Film 8	Neutral	0.598	0.415	
Film 17	Positive	0.463		
Film 1	Negative		0.865	
Film 2	Negative		0.806	
Film 6	Negative	0.314	0.690	
Film 9	Neutral	0.604	0.604	0.350
Film 3	Negative		0.599	0.594
Film 5	Negative		0.589	0.337
Film 11	Neutral	0.495	0.514	
Film 7	Neutral		0.473	
Film 12	Neutral	0.397		0.755
Film 4	Negative		0.473	0.611

2   Factor loadings < 0.3 were excluded

Table 5. Recommended film clips

Min = minutes

1 *Table 5. Recommended film clips*

Ambiguous scenarios task			Evoked emotions task		
Category	Film code	Film name (min)	Category	Film code	Film name
Negatively valenced	Film 2	Fast and Loose (0:00 – 1:42)	Negative	Film 1	In a heartbeat (0:12 – 2:13)
	Film 6	Pillars (6:12 – 7:39)		Film 2	Invisible Strings (1:57 – 2:30, 4:02 – 5:37)
	Film 9	Pregnant Pause (0:41 – 2:14)		Film 3	Lifeboat (5:32 – 8:33)
	Film 11	Whenever You're Ready (1:47 – 3:30)		Film 4	Presentation (4:47 – 7:10)
	Film 12	Youth, full (0:35 – 2:36)		Film 5	Pride and Pack – Pride of Lions (17:58 – 20:09)
	Film 17	The Proposal (0:10 – 1:49)		Film 6	Work (0:18 – 2:22)
Positively valenced	Film 7	Bare (3:58 – 5:52)	Neutral	Film 9	Dreaming whilst Black (1:16 – 3:32)
	Film 10	The Liberty (5:45 – 7:39)		Film 10	Ohio (5:36 – 7:47)

Film 13	City Lights (3:57 – 5:12)		Film 11	Reception (0:07 – 2:24)
Film 14	Don't Be a Hero (7:46 – 9:31)		Film 12	RPG (0:00 – 2:10)
Film 15	Hello, Again (3:37 – 5:16)	Positive	Film 13	Blessing in Disguise (3:14 – 5:32)
Film 16	Palm Trees and Power Lines (3:01 – 4:45)		Film 14	Reality 2.0: Catcalling (0:07 – 2:50)
Film 18	So It Goes (3:28 – 5:24)		Film 15	Chinese Hi-Five (0:11 – 2:17)
			Film 16	Hot mess (0:24 – 2:49)
			Film 17	Russian Roulette (1:12 – 3:46)
			Film 18	Standby (0:21 – 2:26)

2 Min = minutes