

Supplementary Materials

1

2

3 **Contents:**

4	1. Pilot Study –Methods and Results	2
5	2. Main Study Additional Results – Manipulation Check	13
6	3. Main Study Additional Results – Difference from Baseline or Zero	13
7	4. Supplement References	14

8 1. Pilot Study - Methods and Results

9 For the pilot study, participants were instructed to view a large selection of emotion-eliciting films and self-
 10 report their emotional experience to these. Due to the variability across different types of positive and negative
 11 emotional responses, two discrete emotions of amusement and sadness were chosen to represent differential
 12 ends along the valence continuum. The pilot study also tested neutral films. Young and older adult participants
 13 were recruited to ensure the final selected film sets evoked the intended emotion for both age groups in
 14 subsequent studies.

15 The primary goal of the pilot study was to develop and validate separate sets of amusing and sad films that
 16 elicit self-reported high and low levels of emotional intensity for young and older adults. The purpose was to
 17 establish a selection of dynamic stimuli with varying intensities to be used in the current study. The intention
 18 was to evoke high and low levels of emotional facial reactivity and arousal to vary the amount of effort
 19 required to regulate emotion.

20 1.1. Method

21 1.1.1 Participants

22 Sixteen young ($M = 27.19$ years, $SD = 2.11$; range 22–32; 10 females) and 15 older adults ($M = 72.53$, $SD =$
 23 6.03 ; range 63–88; 8 females) participated in the pilot study. Participants were recruited through posters and
 24 word of mouth. Exclusion criteria for older adults included scores of 84 or below on Addenbrooke's Cognitive
 25 Examination-Revised (ACE-R), which is a brief cognitive screening measure [1].

26 Age-related scores on background measures are shown in **Table S1**. Compared with young, older adults
 27 performed worse with task switching and processing speed [Trail Making Test; 2]. However, young and older
 28 adults performed the same on a measure of verbal intelligence, as indexed by the National Adult Reading Test
 29 [NART; 3]. Compared to older adults, young adults reported better overall health for the last month on a self-
 30 rated 5-point scale (1 = *poor*, 5 = *excellent*). There were no age differences in self-reported levels of anxiety or
 31 depression over the past week [Hospital Anxiety and Depression Scale, HADS; 4]. At the start of testing there
 32 were no age differences in self-reported positive affect or negative affect on the Positive and Negative Affect
 33 Schedule [PANAS, state version; 5].

34

35 Table S1

36 *Age Performances on Background Measures (N = 31)*

Characteristic	Young Adults <i>n</i> = 16		Older Adults <i>n</i> = 15		<i>t</i> test ^f (<i>df</i> = 29)		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>d</i>
NART ^a : FSIQ ^b	114.02	4.98	113.10	5.87	3.24	.641	0.17
Trail Making	21.03	8.68	40.86	22.77	3.24	.003	1.15
Depression ^c	2.13	2.06	2.33	1.54	0.32	.754	0.11
Anxiety ^c	5.69	4.19	4.47	3.34	0.89	.379	0.32
Positive Affect ^d	34.56	9.33	37.73	5.47	1.14	.262	0.41
Negative Affect ^d	12.69	2.41	12.27	4.33	0.34	.739	0.12
Health last Month ^e	4.31	0.79	3.53	0.74	2.82	.009	1.02

37 ^aNART = National Adult Reading Test; ^bFSIQ = Full Scale Intelligent Quotient; ^cAnxiety and Depression scales from the
 38 Hospital Anxiety and Depression Scale; ^dPositive and Negative Affect scale from the Positive and Negative Affect Scale;
 39 ^eFive-point Likert scale (1=poor, 5=excellent); ^fIndependent *t*-test applied, with Cohen's *d* = effect size.

40

41 1.1.2. Procedure and Materials

42 **Emotion rating task.** This task involved participants viewing a total of 32 short films and rating their emotional
 43 experience to each. Participants viewed a block of 16 amusing and a separate block of 16 sad films, which were
 44 counterbalanced in order between participants. Videos within each block were randomized in order, with

REGULATE FACIAL MUSCLE REACTIVITY IN AGING

45 participants viewing a neutral film following each fourth film to reduce the carryover effect of emotional
 46 experience. Additionally, participants completed background measures and had opportunities for breaks
 47 following each eighth film. Participants were instructed to, “Watch each scene carefully, and following each
 48 video, please rate the emotional experience *you* felt whilst watching it”. Participants manually pressed the
 49 spacebar when they were ready to proceed to the next video. Participants completed an initial practice of one
 50 neutral video to familiarize themselves with the task. Task instructions and stimuli were presented with
 51 PowerPoint (Microsoft Office for Mac, 2011) on an Apple MacBook Pro laptop computer with a 32-inch LCD
 52 screen. Participants wore over-ear Bose audio headphones and were able to adjust volume as required.

53 Self-report scales. For a subjective measure of emotional intensity, participants recorded their emotional
 54 experience immediately following each film. Participants rated how much of a range of emotions they felt
 55 whilst watching each video on a series of nine-point Likert scales, ranging from 1 (*not at all*) to 9 (*extremely*).
 56 Specified emotions, each with their own nine-point scale, included: amusement, anger, disgust, fear, sadness,
 57 surprise, and confusion. Additionally, participants rated the arousal level of their emotional experience on a
 58 nine-point Likert scale, from 1 (*calm/dull*) to 9 (*intense*). Participants were also asked whether they had seen
 59 the film before, and whether they tried to avoid or suppress their feelings during the scene.

60 Stimuli. Video clips were obtained from films, television shows and YouTube. Criteria for videos selected,
 61 included content requiring no background information or explanation, with minimal cognitive effort to view
 62 and comprehend. Films were also chosen to be appropriate for both young and older adult age groups, based on
 63 the relevance of the content, decade of original film release, and age range of the characters in the film [6].
 64 Guidance on films were also acquired from previously published studies using films to elicit amusement or
 65 sadness [7-9], and also popular online databases, magazines, and forums listing the top movie scenes eliciting
 66 the targeted emotion (e.g., International Movie Database, Rolling Stone website, YouTube). Neutral films were
 67 selected to elicit minimal emotional arousal, and depicted scenes of nature, cooking pasta, watering tomatoes,
 68 and information on a castle, as shown in **Table S2b**. Prior to the selected films being included in the current
 69 study, the videos were initially piloted with several peers, who provided feedback on the effectiveness and
 70 appropriateness of each film. Based on this feedback, several films were replaced, and finalized to the set tested
 71 in the current study. Of the 16 films targeting amusement, eight were selected with the intention of eliciting low
 72 levels of amusement, and another eight for eliciting high levels of amusement, as shown in **Table S2a**. Also of
 73 the 16 films targeting sadness, eight were selected with the intention of eliciting low levels of sadness, and
 74 another eight for eliciting high levels of sadness, as shown in **Table S2a**. All films aimed to primarily elicit the
 75 targeted emotion, with minimal additional emotional states, such as anger, fear, or disgust. Videos ranged from
 76 69 to 141 s in duration, with an average of 109.19 s in duration.

77

78 Table S2a

79 *Film Segments Selected to Elicit Amusement, and Sadness.*

Film – Scene	Date	Source	Length (seconds)
Amusing Films (High-intensity)			
<i>I Love Lucy</i> – Chocolate factory mishap	1952	TV show	122
<i>Just for Laughs</i> – Police in underwear prank	2011	TV show	131
<i>Baby Laughing</i> – Watching paper being torn	2011	YouTube	103
<i>Walk on the Wild Side</i> - Animals talking	2009	TV show	104
<i>Mr Bean</i> – Naked and locked out of hotel room	1993	TV show	87
<i>Japanese Show</i> – Dinosaur scare prank	2013	TV show	85
<i>The Naked Gun</i> – Car chase with learner driver	1988	Movie	95
<i>Just for Laughs</i> – British Guard photo prank	2011	TV show	79
Amusing Films (Low-intensity)			
<i>Some Mothers Do Ave’Em</i> – Rollerskating	1973	TV show	107
<i>Just for Laughs</i> – Horrible makeup prank	2011	TV show	101

REGULATE FACIAL MUSCLE REACTIVITY IN AGING

<i>Caddyshack</i> – Boat causing chaos in harbor	1980	Movie	120
<i>Mad TV</i> – Airplane service skit	2007	TV show	108
<i>Just for Laughs</i> – Waterpark urinate prank	2011	TV show	72
<i>Austin Powers</i> – Three point turn in a hallway	1997	Movie	69
<i>Just for Laughs</i> – News prank, told to duck	2011	TV show	120
<i>Cat Cleans Kitchen</i> – Cat rides vacuum cleaner	2013	YouTube	105
Sad Films (High-intensity)			
<i>Sophie’s Choice</i> – Can’t choose which child	1982	Movie	132
<i>The Champ</i> – Boy crying over boxer’s death	1979	Movie	139
<i>Click</i> – Father dying in the street	2006	Movie	120
<i>My Girl</i> – Best friend’s funeral	1991	Movie	141
<i>The Green Mile</i> – Execution	1999	Movie	141
<i>The NeverEnding Story</i> – Artax sinks in swamp	1984	Movie	120
<i>Remember Me</i> – 9/11 World Trade Centre	2010	Movie	120
<i>The Impossible</i> – Family reunion after disaster	2012	Movie	126
Sad Films (Low-intensity)			
<i>My Dog Skip</i> – Dog with arthritis misses owner	2000	Movie	118
<i>Fresh Prince of Bel-Air</i> – Drug use confession	1993	TV show	69
<i>Rabbit Hole</i> – Crying in car about memory	2010	Movie	99
<i>Dangerous Minds</i> - News of student shooting	1995	Movie	126
<i>E.T.</i> – Alien says goodbye to family	1982	Movie	140
<i>Midnight Cowboy</i> – Man dies during bus ride	1969	Movie	102
<i>Dear John</i> – Relationship breakup	2010	Movie	112
<i>Seven Pounds</i> – Ezra and Emily meet	2008	Movie	78

80
81
82
83
84
85

Table S2b
Film Segments Selected to Elicit Neutral Feelings

Film – Scene	Date	Source	Length (seconds)
Neutral Films			
<i>Mountain Goats</i> – Information on goats	2010	Documentary	43
<i>Mountain Villagers</i> – Information on villagers	2010	Documentary	35
<i>Lyre Bird Mimicking</i> – Mimics sounds	2013	YouTube	71
<i>Flowers in a Valley</i> – Scenes of flowers	2010	Documentary	53
<i>Castle History</i> – Information on a castle	2010	Documentary	40
<i>Cooking Pasta</i> – Watching pasta boil	2010	TV Show	79
<i>Knitting Demonstration</i> – Viewing hands knit	2012	YouTube	64
<i>Watering Tomatoes</i> – Gardening instructions	2012	YouTube	62

86
87
88

1.2 Results

89 Films were first checked to ensure they elicited the intended target emotion (i.e., greater amusement or sadness
90 mean scores), and minimal comparable levels of non-target discrete emotions [e.g., anger, disgust, fear,
91 surprise; 8]. Second, films were checked for confusion to ensure participants were able to comprehend each
92 film as a standalone scene. For the current study, films receiving confusion scores of $M = 3.00$ or above were
93 deemed unsuitable stimuli and removed from subsequent steps. Third, remaining films were then ranked based

94 on their target emotion mean score (e.g., amusement or sadness), and the four highest and four lowest films
 95 were assigned to respective high- and low-intensity sets. Lastly, obtained high- and low-intensity sets were
 96 analyzed to ensure differences exist *between* intensity sets (based on self-rated target emotion and arousal
 97 levels), and that no differences exist *within* intensity sets (i.e., emotion and arousal ratings don't differ between
 98 films of each independent intensity set). Analyses where assumptions of sphericity were violated, Greenhouse-
 99 Geisser corrected values are reported.

100 *1.2.1 Films Eliciting Amusement*

101 Emotional discreteness. Each amusing film was checked for the intensity of self-reported emotional experience,
 102 and whether amusement differed significantly from non-targeted discrete emotions. Emotion ratings for each
 103 film were analyzed with one-way repeated measures (RM) analysis of variance (ANOVA). The within-subjects
 104 variable was self-reported emotion (amusement, anger, disgust, fear, sadness, surprise, confusion). For each
 105 film, there was a main effect of emotion ($ps < .001$), and according to post hoc analyses, all films were rated
 106 with greater amusement than anger, disgust, fear, sadness, surprise, and confusion. The mean self-rated emotion
 107 scores for all amusing films are shown in **Table S3**.

108 Clarity of film content. Films were then double checked for self-reported confusion. According to the previous
 109 analyses (see **Table S3**), no mean scores of confusion were above $M = 3.00$. Therefore, no films were excluded
 110 for subsequent steps.

111 Assigning films to intensity sets. Based on the previous steps, the films eliciting the highest scores of
 112 amusement were assigned to the high intensity set, and the films eliciting the lowest scores of amusement were
 113 assigned to the low intensity set. *I Love Lucy*, *Police Prank*, *Baby Laughing*, and *Mr Bean* were selected for the
 114 high intensity set, and *Caddyshack*, *Austin Powers*, *News Prank*, and *Cat Cleans Kitchen* were selected for the
 115 low intensity set.

116 Similarities within each intensity set. The selected films were checked for amusement and arousal differences¹
 117 within each high and low intensity sets, as shown in **Figure S1**. Amusement and arousal ratings were analyzed
 118 with separate one-way RM ANOVAs for high and low intensity sets. The within-subjects variables were
 119 respective high intensity films (*I Love Lucy*, *Police Prank*, *Baby Laughing*, *Mr Bean*), and low intensity films
 120 (*Caddyshack*, *Austin Powers*, *News Prank*, *Cat Cleans Kitchen*). For amusement ratings, there was no main
 121 effect of film within the high-intensity set, $F(3, 90) = 1.02, p = .386, \eta_p^2 = .03$, or within the low-intensity set,
 122 $F(2.34, 70.29) = 2.02, p = .133, \eta_p^2 = .06$. For arousal ratings, there was no main effect of film within the high
 123 intensity set, $F(3, 90) = 1.48, p = .224, \eta_p^2 = .05$, or within the low intensity set, $F(3, 90) = 2.38, p = .075, \eta_p^2 =$
 124 $.07$.

¹ Additional *t*-test analyses were conducted to determine whether age differences exist in amusement and arousal ratings for each film. *Mr Bean* received age differences across amusement (young: $M = 5.75, SD = 1.92$; older: $M = 7.60, SD = 1.06$), $t(29) = 3.30, p = .003, d = 1.19$, and arousal ratings (young: $M = 5.00, SD = 1.97$; older: $M = 6.53, SD = 1.19$), $t(29) = 2.61, p = .014, d = 0.94$. *Caddyshack* received age differences in arousal ratings (young: $M = 3.88, SD = 1.93$; older: $M = 5.73, SD = 1.87$), $t(29) = 2.72, p = .011, d = 0.97$. *News Prank* received age differences in amusement ratings (young: $M = 3.50, SD = 1.59$; older: $M = 5.20, SD = 1.82$), $t(29) = 2.77, p = .010, d = 0.99$. There were no further age differences across remaining amusing films ($ps > .052$).

REGULATE FACIAL MUSCLE REACTIVITY IN AGING

Table S3
 Mean (SD) Emotion Scores for Films Intended to Elicit Feelings of Amusement (N = 31)

Film	F	p	η_p^2	Amusement	Anger	Disgust	Fear	Sadness	Surprise	Confusion
<i>I Love Lucy</i>	102.67	< .001	.77	6.97 (1.89)	1.16 (0.73)	1.16 (0.73)	1.13 (0.72)	1.03 (0.18)	2.35 (2.00)	1.65 (1.58)
<i>Mr Bean</i>	129.40	< .001	.81	6.65 (1.80)	1.06 (0.36)	1.10 (0.40)	1.19 (0.75)	1.29 (1.01)	2.13 (1.63)	1.16 (0.73)
<i>Police Prank</i>	64.42	< .001	.68	6.35 (2.39)	1.10 (0.54)	1.48 (1.09)	1.00 (0.00)	1.03 (0.18)	3.94 (2.66)	1.52 (1.36)
<i>Baby Laughing</i>	117.21	< .001	.75	6.29 (1.66)	1.03 (0.18)	1.06 (0.36)	1.00 (0.00)	1.00 (0.00)	2.23 (2.25)	1.35 (1.08)
<i>Dinosaur Prank</i>	39.37	< .001	.57	6.06 (2.08)	1.13 (0.50)	1.35 (1.02)	2.00 (1.91)	1.03 (0.18)	2.77 (2.20)	1.94 (1.97)
<i>Walk on Wild Side</i>	124.39	< .001	.81	6.06 (1.65)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	2.26 (1.86)	1.16 (0.45)
<i>The Naked Gun</i>	62.34	< .001	.68	6.00 (1.90)	1.00 (0.00)	1.06 (0.36)	1.94 (1.55)	1.00 (0.00)	2.45 (1.95)	1.74 (1.65)
<i>Guard Prank</i>	94.09	< .001	.76	5.77 (2.00)	1.00 (0.00)	1.10 (0.40)	1.00 (0.00)	1.00 (0.00)	2.39 (1.82)	1.10 (0.40)
<i>Some Mothers..</i>	49.48	< .001	.62	5.71 (2.45)	1.13 (0.43)	1.13 (0.43)	1.61 (1.28)	1.00 (0.00)	2.65 (2.24)	1.52 (1.26)
<i>Makeup Prank</i>	72.06	< .001	.71	5.71 (2.07)	1.06 (0.36)	1.26 (0.63)	1.06 (0.36)	1.06 (0.36)	2.77 (2.00)	1.35 (0.91)
<i>Waterpark Prank</i>	28.41	< .001	.49	5.55 (2.68)	1.29 (0.69)	2.23 (1.82)	1.23 (0.80)	1.29 (0.86)	3.26 (2.65)	1.48 (1.29)
<i>Mad TV Flight</i>	68.37	< .001	.70	5.52 (2.03)	1.23 (0.80)	1.10 (0.40)	1.06 (0.36)	1.06 (0.36)	2.10 (1.97)	1.29 (0.86)
<i>Cat Cleans Kitchen</i>	45.80	< .001	.60	5.16 (2.35)	1.03 (0.18)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	2.26 (1.63)	1.87 (1.84)
<i>Austin Powers</i>	48.55	< .001	.62	5.03 (1.96)	1.32 (1.14)	1.13 (0.72)	1.00 (0.00)	1.03 (0.18)	1.52 (1.21)	1.90 (1.76)
<i>Caddyshack</i>	32.07	< .001	.52	4.58 (2.01)	1.29 (0.59)	1.16 (0.45)	1.19 (0.65)	1.03 (0.18)	1.93 (1.50)	2.13 (2.29)
<i>News Prank</i>	41.55	< .001	.58	4.32 (1.89)	1.13 (0.56)	1.06 (0.25)	1.03 (0.18)	1.06 (0.25)	1.48 (1.12)	1.81 (1.62)

Each film was analyzed using separate RM ANOVAs ($df = 6$, $df_{error} = 180$). All post hoc comparisons between the target emotion (amusement) and each discrete emotion differed significantly ($ps < .05$).

Differences between intensity sets. In order to determine whether the intended high- and low-intensity amusing film sets differed significantly, averaged self-reported amusement and arousal scores were first calculated for each participant. Thus, separate amusement and arousal ratings were averaged for the four highest intensity films (*I Love Lucy*, *Police Prank*, *Baby Laughing*, *Mr Bean*) and the four lowest intensity films (*Caddyshack*, *Austin Powers*, *News Prank*, *Cat Cleans Kitchen*). Averaged high- and low-intensity set scores were then compared using paired *t* tests. The high-intensity set ($M = 6.56$, $SD = 1.28$) was rated as eliciting greater feelings of amusement than the low-intensity set ($M = 4.77$, $SD = 1.58$), $t(30) = 6.59$, $p < .001$, $d = 1.24$). The high-intensity set ($M = 6.06$, $SD = 1.48$) was also rated as eliciting greater arousal than the low-intensity set ($M = 4.76$, $SD = 1.53$), $t(30) = 6.26$, $p < .001$, $d = 0.86$).

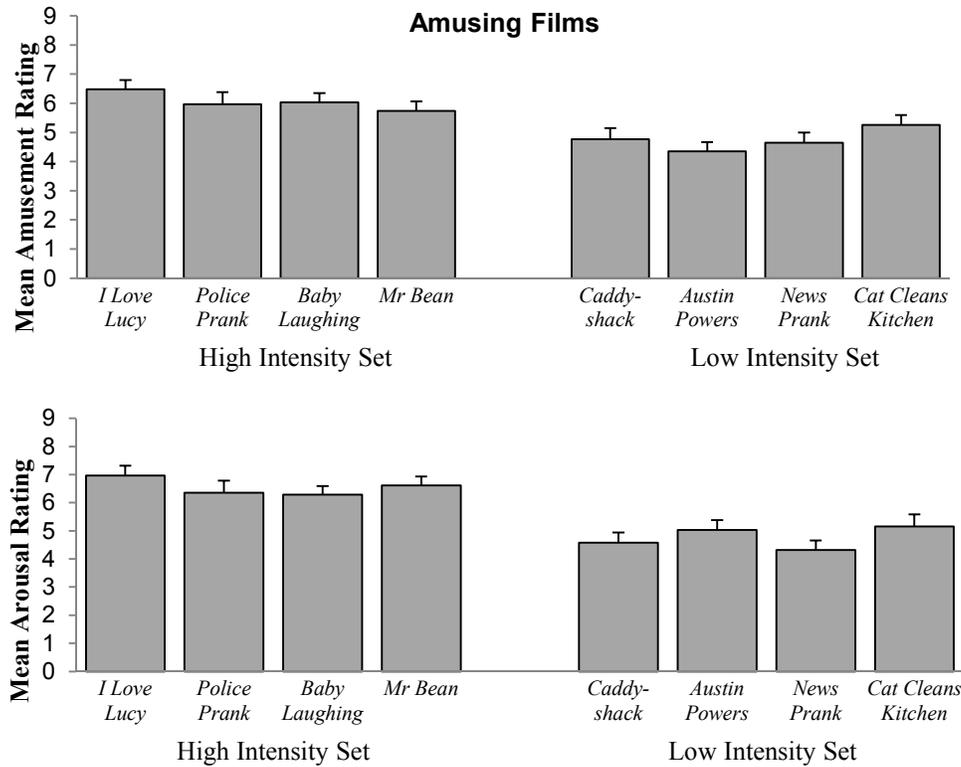


Figure S1. Mean amusement and arousing ratings for selected high- and low-intensity amusing films. Error bars indicate standard errors.

12.2. Films Eliciting Sadness

Emotional discreteness. Each sad film was checked for the intensity of self-reported emotional experience, and whether sadness differed significantly from non-targeted discrete emotions. Emotion ratings for each film were analyzed with one-way RM ANOVAs. The within-subjects variable was self-reported emotion (amusement, anger, disgust, fear, sadness, surprise, confusion). For each film, there was a main effect of emotion ($ps < .001$), and according to post hoc analyses, majority of films were rated with greater sadness than amusement, anger, disgust, fear, surprise, and confusion. For two films, sadness did not differ significantly from surprise (*The Impossible*, $p = .167$; *Seven Pounds*, $p = 1.00$), and were thus removed from subsequent steps. Confusion was also found to not differ significantly from sadness in five films (*Rabbit Hole*, $p = 1.00$; *Midnight Cowboy*, $p = .577$; *The Impossible*, $p = .891$; *Dear John*, $p = 1.00$; *Seven Pounds*, $p = .324$), and thus examined further in the next step. The mean self-rated emotion scores for all sad films are shown in **Table S4**.

Clarity of film content. Films were then double checked for self-reported confusion. According to the previous analyses (see **Table S4**), mean scores of confusion were above $M = 3.00$ for *Rabbit Hole*, *Dear John*, and *Seven Pounds*, and were thus excluded for subsequent steps. All other remaining films did not meet the cut off criteria on confusion (including *Midnight Cowboy*), and remained for subsequent steps.

Assigning films to intensity sets. Of the remaining stimuli, the films eliciting the highest scores of sadness were assigned to the high-intensity set, and the films eliciting the lowest scores of sadness were assigned to the low-intensity set. *Sophie's Choice*, *NeverEnding Story*, *My Girl*, and *The Champ* were selected for the

high-intensity set, and *My Dog Skip*, *Fresh Prince of Bel Air*, *Dangerous Minds*, and *Midnight Cowboy* were selected for the low-intensity set.

Similarities within each intensity set. Selected films were checked for sadness and arousal differences² within each high- and low-intensity sets, as shown in **Figure S2**. Sadness and arousal ratings were analyzed with separate one-way RM ANOVAs for high- and low-intensity sets. The within-subjects variables were respective high-intensity films (*Sophie's Choice*, *The Champ*, *My Girl*, *NeverEnding Story*), and low-intensity films (*Dangerous Minds*, *Fresh Prince of Bel Air*, *My Dog Skip*, *Midnight Cowboy*). For sadness ratings, there was no main effect of film within the high-intensity set, $F(3, 90) = 1.31, p = .275, \eta_p^2 = .04$, or within the low-intensity set, $F(3, 90) = 1.26, p = .293, \eta_p^2 = .04$. For arousal ratings, there was no main effect of film within the high-intensity set, $F(2.36, 70.67) = 2.00, p = .136, \eta_p^2 = .06$, or within the low-intensity set, $F(3, 90) = 1.60, p = .195, \eta_p^2 = .05$.

Differences between intensity sets. In order to determine whether the intended high and low intensity sad film sets differed significantly, averaged self-reported sadness and arousal scores were first calculated for each participant. Thus, separate sadness and arousal ratings were averaged for the four highest intensity films (*Sophie's Choice*, *The Champ*, *My Girl*, *NeverEnding Story*) and the four lowest intensity films (*My Dog Skip*, *Fresh Prince of Bel Air*, *Dangerous Minds*, *Midnight Cowboy*). Averaged high- and low-intensity set scores were then compared using paired *t*-tests. The high-intensity set ($M = 6.85, SD = 1.38$) was rated as eliciting greater feelings of sadness than the low-intensity set ($M = 4.39, SD = 1.78$), $t(30) = 10.76, p < .001, d = 1.54$. The high-intensity set ($M = 6.99, SD = 1.26$) was also rated as eliciting greater arousal than the low-intensity set ($M = 5.02, SD = 1.33$), $t(30) = 9.28, p < .001, d = 1.52$.

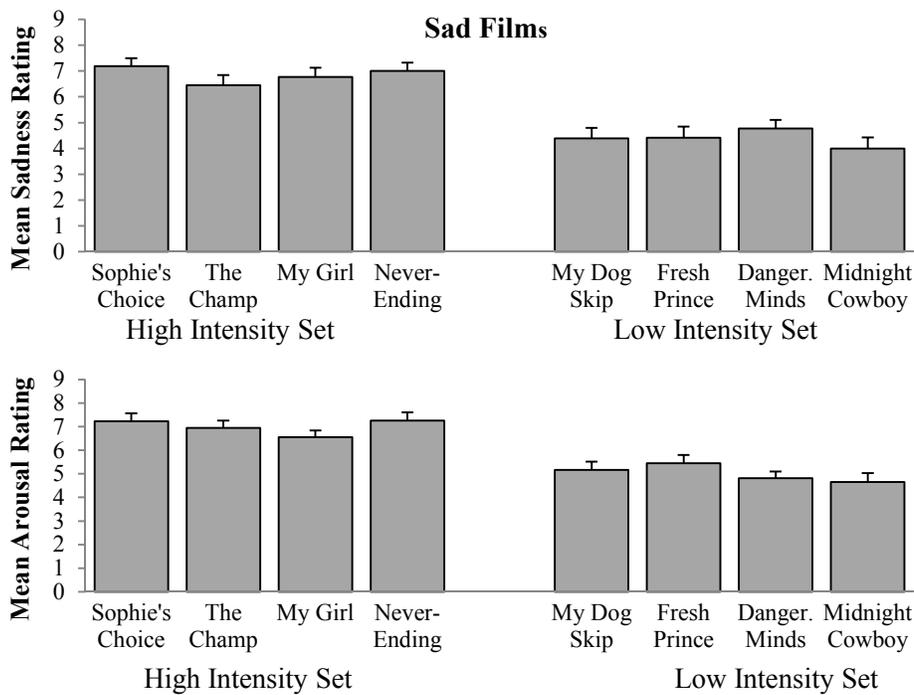


Figure S2. Mean sadness and arousal ratings for selected high- and low-intensity sad films. Error bars indicate standard errors.

² Additional *t*-test analyses were conducted to determine whether age differences exist in sadness and arousal ratings for each film. *NeverEnding Story* received age differences in sadness ratings (young: $M = 6.38, SD = 1.89$; older: $M = 7.67, SD = 1.40$), $t(29) = 2.15, p = .040, d = 0.78$. *The Champ* received age differences in arousal ratings (young: $M = 6.31, SD = 1.82$; older: $M = 7.60, SD = 1.40$), $t(29) = 2.20, p = .036, d = 0.79$. *Midnight Cowboy* received age differences in sadness ratings (young: $M = 3.00, SD = 1.46$; older: $M = 5.07, SD = 2.79$), $t(29) = 2.61, p = .014, d = 0.93$. *Fresh Prince* received age differences in sadness ratings (young: $M = 3.44, SD = 1.97$; older: $M = 5.47, SD = 2.30$), $t(29) = 2.65, p = .013, d = 0.95$. There were no further age differences across remaining sad films ($ps > .050$).

REGULATE FACIAL MUSCLE REACTIVITY IN AGING

Table S4

Mean (SD) Emotion Scores for Films Intended to Elicit Feelings of Sadness (N = 31)

Film	F	p	η_p^2	Sadness	Amusement	Anger	Disgust	Fear	Surprise	Confusion
<i>Sophie's Choice</i>	48.94	<.001	.62	7.19 (1.72)	1.00 (0.00)	4.71 (2.44)	5.03 (2.56)	3.03 (2.33)	2.13 (1.88)	1.84 (1.57)
<i>NeverEnding..</i>	50.75	<.001	.63	7.00 (1.77)	1.13 (0.43)	2.32 (2.20)	1.42 (1.12)	3.55 (2.69)	2.00 (1.67)	2.06 (1.77)
<i>My Girl</i>	87.75	<.001	.75	6.77 (1.98)	1.13 (0.43)	1.26 (1.09)	1.39 (1.09)	1.55 (1.69)	1.77 (1.41)	1.77 (1.36)
<i>The Champ</i>	42.23	<.001	.59	6.45 (2.20)	1.13 (0.43)	1.97 (1.66)	1.74 (1.55)	2.19 (1.85)	1.97 (1.70)	2.45 (2.11)
<i>The Green Mile</i>	32.28	<.001	.52	6.42 (2.01)	1.00 (0.00)	3.06 (2.59)	3.74 (2.90)	2.19 (2.18)	2.10 (1.97)	2.29 (2.08)
<i>Remember Me</i>	18.14	<.001	.38	5.29 (2.52)	1.06 (0.25)	2.03 (1.78)	1.90 (1.81)	2.71 (1.90)	2.29 (2.19)	2.90 (2.48)
<i>Click</i>	24.75	<.001	.45	5.10 (2.18)	2.19 (1.49)	1.26 (0.93)	1.23 (0.92)	1.84 (1.68)	1.84 (1.57)	2.84 (2.24)
<i>Danger. Minds</i>	35.90	<.001	.55	4.77 (1.86)	1.42 (0.99)	1.32 (1.01)	1.42 (1.03)	1.48 (1.31)	1.84 (1.46)	2.81 (1.76)
<i>E.T.</i>	17.61	<.001	.37	4.45 (2.74)	1.61 (1.20)	1.16 (0.73)	1.26 (0.82)	1.77 (1.65)	1.87 (1.43)	1.94 (1.53)
<i>Fresh Prince</i>	21.48	<.001	.42	4.42 (2.33)	1.10 (0.40)	1.48 (1.29)	1.68 (1.35)	1.84 (1.90)	2.26 (2.03)	1.81 (1.66)
<i>My Dog Skip</i>	30.29	<.001	.50	4.39 (2.30)	1.26 (0.96)	1.00 (0.00)	1.06 (0.56)	1.13 (0.56)	1.52 (1.12)	2.10 (1.87)
<i>Rabbit Hole</i>	38.08	<.001	.56	4.03 (1.78)	1.29 (1.44)	1.13 (0.43)	1.10 (0.30)	1.19 (0.54)	1.68 (1.25)	4.39 (2.26)*
<i>Mid. Cowboy</i>	10.46	<.001	.26	4.00 (2.41)	1.45 (1.03)	1.58 (1.48)	1.84 (1.51)	1.77 (1.86)	2.42 (1.95)	2.68 (2.09)*
<i>The Impossible</i>	15.48	<.001	.34	3.97 (2.26)	1.26 (0.58)	1.23 (0.76)	1.39 (1.02)	2.29 (1.85)	2.61 (2.14)*	2.77 (1.96)*
<i>Dear John</i>	22.40	<.001	.43	3.77 (1.82)	1.23 (0.80)	1.29 (0.82)	1.29 (0.90)	1.23 (0.67)	1.71 (1.47)	3.77 (2.74)*
<i>Seven Pounds</i>	26.02	<.001	.47	2.74 (1.69)	1.16 (0.45)	1.13 (0.50)	1.00 (0.00)	1.00 (0.00)	2.16 (1.66)*	3.94 (2.41)*

Each film was analyzed using separate RM ANOVAs ($df = 6$, $df_{error} = 180$).

*Represents non-significant post hoc comparisons between sadness and discrete emotion. All other comparisons with sadness were significant ($ps < .05$).

1.2.3 Neutral Films

Ratings of emotion. To ensure neutral films did not elicit unwanted emotions, each film was checked for the intensity of self-reported emotional experience. Descriptive statistics were used to analyze the emotion ratings (amusement, anger, disgust, fear, sadness, surprise) for each film. For each film, any mean emotion rating above $M = 2.00$ was used as a cut off, with only one film (*Lyre Bird Mimicking*) meeting this exclusion criteria (amusement, $M = 2.68$; surprise, $M = 2.52$) and excluded from subsequent analysis. **Table S5** presents the mean self-rated emotion scores.

Clarity of film content. Films were then double checked for self-reported confusion. According to descriptive statistics (and also shown in **Table S5**), no mean scores of confusion were above $M = 3.00$. Therefore, no films were excluded for subsequent steps.

Assigning films to the neutral set. Remaining films were checked for levels of self-reported arousal ratings, as displayed in **Figure S3**. The four films eliciting the lowest arousal scores were assigned to the neutral film set, resulting in the inclusion of *Mountain Goats*, *Cooking Pasta*, *Knitting Demonstration*, and *Watering Tomatoes*.

Differences within neutral set. The selected films were checked for any arousal differences³ within the neutral set. Arousal ratings were analysed with a one-way RM ANOVA. The within-subjects variable was the neutral films (*Mountain Goats*, *Cooking Pasta*, *Knitting Demonstration*, *Watering Tomatoes*). For arousal ratings, there was no main effect of film, $F(3, 90) = 0.42, p = .742, \eta_p^2 = .01$.

Differences between neutral and emotion sets. In order to determine whether the neutral film set differed significantly from the amusing and sad film sets (of high and low intensity), averaged self-reported emotion and arousal scores were first calculated for each participant. Thus, separate amusement, sadness, and arousal ratings were averaged for the four selected neutral films (*Mountain Goats*, *Cooking Pasta*, *Knitting Demonstration*, *Watering Tomatoes*) and compared with the averaged amusing and sad film set scores using paired *t*-tests.

For comparisons with the highly amusing film set, the neutral film set was rated as eliciting less feelings of amusement ($M = 1.39, SD = 0.94, t(30) = 17.98, p < .001, d = 4.61$), and less arousal ($M = 2.89, SD = 1.60, t(30) = 13.61, p < .001, d = 2.06$). For comparisons with the low amusing film set, the neutral film set was rated as eliciting less feelings of amusement, $t(30) = 10.22, p < .001, d = 2.61$, and less arousal, $t(30) = 8.19, p < .001, d = 1.19$. For comparisons with the highly sad film set, the neutral film set was rated as eliciting less feelings of sadness ($M = 1.02, SD = 0.13, t(30) = 22.46, p < .001, d = 5.95$), and less arousal, $t(30) = 14.11, p < .001, d = 2.85$. For comparisons with the low sad film set, the neutral film set was rated as eliciting less feelings of sadness, $t(30) = 10.41, p < .001, d = 2.68$, and less arousal, $t(30) = 7.31, p < .001, d = 1.45$.

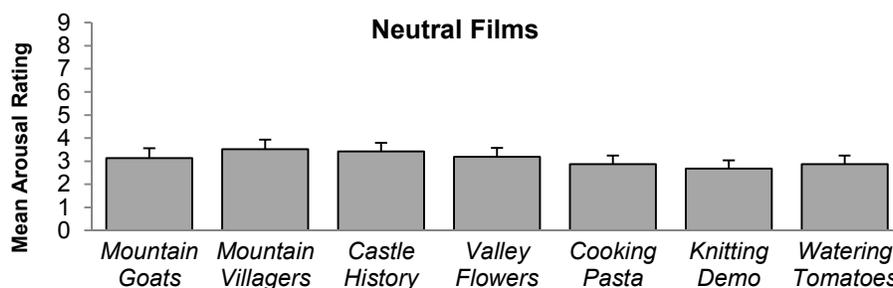


Figure S3. Mean arousal ratings for neutral films. Error bars indicate standard errors.

³ Additional *t*-test analyses were conducted to determine whether age differences exist in amusement, sadness, and arousal ratings for each neutral film. *Knitting Demonstration* received age differences in arousal ratings (young: $M = 1.94, SD = 1.29$; older: $M = 3.47, SD = 2.26$), $t(29) = 2.33, p = .027, d = 0.83$. *Watering Tomatoes* received age differences across amusement (young: $M = 2.56, SD = 2.63$; older: $M = 1.00, SD < 0.01$), $t(29) = 2.30, p = .029, d = 0.84$, and arousal ratings (young: $M = 1.88, SD = 1.36$; older: $M = 3.93, SD = 2.22$), $t(29) = 3.14, p = .004, d = 1.11$. There were no further age differences across remaining neutral films ($ps > .067$).

REGULATE FACIAL MUSCLE REACTIVITY IN AGING

Table S5

Mean (SD) Emotion Scores for Films Intended to Elicit Neutral Feeling (N = 31)

Film – Scene	Amusement	Anger	Disgust	Fear	Sadness	Surprise	Confusion
<i>Mountain Goats</i>	1.13 (0.72)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.29 (0.69)	1.22 (0.62)
<i>Maintain Villagers</i>	1.03 (0.18)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.03 (0.18)	1.23 (0.80)	1.13 (0.56)
<i>Lyre Bird Mimicking</i>	2.68 (2.47)	1.00 (0.00)	1.00 (0.00)	1.13 (0.72)	1.23 (0.92)	2.52 (2.10)	1.35 (0.95)
<i>Valley Flowers</i>	1.10 (0.54)	1.06 (0.36)	1.00 (0.00)	1.00 (0.00)	1.06 (0.36)	1.87 (1.63)	1.06 (0.36)
<i>Castle History</i>	1.19 (0.75)	1.03 (0.18)	1.00 (0.00)	1.00 (0.00)	1.06 (0.36)	1.06 (0.25)	1.13 (0.50)
<i>Cooking Pasta</i>	1.45 (1.52)	1.13 (0.72)	1.00 (0.00)	1.00 (0.00)	1.03 (0.18)	1.16 (0.52)	1.45 (1.46)
<i>Knitting Demonstration</i>	1.13 (0.34)	1.13 (0.72)	1.00 (0.00)	1.00 (0.00)	1.03 (0.18)	1.58 (1.41)	1.81 (1.68)
<i>Watering Tomatoes</i>	1.81 (2.02)	1.06 (0.36)	1.06 (0.36)	1.00 (0.00)	1.03 (0.18)	1.55 (1.48)	1.74 (1.44)

1.3 Pilot Study Summary

Young and older adults self-reported their emotional experience to a wide selection of dynamic emotion-eliciting stimuli. Through the development and validation process, eight amusing films were identified, with four being assigned to the high-intensity set (*I Love Lucy*, *Police Prank*, *Baby Laughing*, *Mr Bean*), and four assigned to the low-intensity set (*Caddyshack*, *Austin Powers*, *News Prank*, *Cat Cleans Kitchen*). Eight sad films were also identified, with four being assigned to the high-intensity set (*Sophie's Choice*, *The Champ*, *My Girl*, *NeverEnding Story*), and four assigned to the low-intensity set (*My Dog Skip*, *Fresh Prince of Bel Air*, *Dangerous Minds*, *Midnight Cowboy*). Selected films were found to elicit the target emotion (e.g., amusement or sadness), with minimal levels of unrelated emotions (e.g., anger, disgust, fear, surprise) or confusion. Additionally, sets of high emotionally intense amusing and sad films were rated with greater emotion and arousal than sets of low emotionally intense amusing and sad films. Within each identified set, films did not differ on target emotion or arousal ratings. Furthermore, four neutral films were selected and validated (*Mountain Goats*, *Cooking Pasta*, *Knitting Demonstration*, *Watering Tomatoes*). Neutral films did not differ on arousal ratings, and were also rated with lower emotion and arousal than each of the amusing and sad film sets.

2. Main Study Additional Results – Preliminary Check

Table S6

Means and Standard Deviations of Zygomaticus and Corrugator Facial Muscle Reactivity to Amusing and Sad films.

Instruction	Intensity	Young Adults (n=40)				Older Adults (n=40)			
		Amuse		Sad		Amuse		Sad	
		M	SD	M	SD	M	SD	M	SD
<i>Zygomaticus muscle activity</i>									
Watch	Low	388.16	546.00	-6.71	45.14	99.85	153.70	10.09	49.61
Watch	High	748.05	868.74	-13.13	40.36	300.05	319.65	8.50	46.44
Suppress	Low	70.06	443.07	-2.68	45.82	62.48	113.69	0.44	39.38
Suppress	High	289.80	437.80	-19.16	40.44	184.26	291.78	-3.17	40.30
<i>Corrugator muscle activity</i>									
Watch	Low	43.41	94.69	64.53	66.57	-0.69	33.68	26.75	48.14
Watch	High	17.23	75.32	134.69	108.69	9.88	63.32	58.76	53.76
Suppress	Low	47.04	120.37	49.40	58.92	0.42	37.92	36.75	62.03
Suppress	High	12.93	58.59	77.34	76.89	0.41	40.03	32.97	46.18

Note. Mean (M) and standard deviations (SD). The shaded sections indicate the main muscle-emotion pairs.

3. Main Study Additional Results – Difference From Baseline or Zero

Table S7a

Differences between a Baseline Score of 0 and an Expressive Suppression Mean Score for the Percentage change in Zygomaticus Activity, as a Function of Age Group and Intensity Condition

Intensity	Expressive Suppression		One Sample <i>t</i> -test ^a		(<i>df</i> = 39)
	<i>M</i>	<i>SD</i>	<i>t</i>	<i>P</i>	<i>d</i> ^b
<i>Young Adults</i>					
Amusement (low)	226.50	443.07	3.23	.002	0.51
Amusement (high)	289.80	437.80	4.19	< .001	0.66
<i>Older Adults</i>					
Amusement (low)	62.48	113.69	3.48	.001	0.55
Amusement (high)	184.26	291.78	3.99	.001	0.63

Note. Mean scores refer to percentage change in zygomaticus 'cheek' EMG activity from baseline.

^aOne-sample *t*-test applied to determine if mean scores in facial activity differed from zero. ^bCohen's *d* = effect size.

Table S7b

Differences between a Baseline Score of 0 and an Expressive Suppression Mean Score for the Percentage change in Corrugator Activity, as a Function of Age Group and Intensity Condition

Intensity	Expressive Suppression		One Sample <i>t</i> -test ^a		(<i>df</i> = 39)
	<i>M</i>	<i>SD</i>	<i>T</i>	<i>p</i>	<i>d</i> ^b
<i>Young Adults</i>					
Sadness (low)	49.40	58.92	5.30	< .001	0.84
Sadness (high)	77.34	76.89	6.36	< .001	1.01
<i>Older Adults</i>					
Sadness (low)	36.75	62.03	3.75	.001	0.59
Sadness (high)	32.97	46.18	4.52	< .001	0.71

Note. Mean scores refer to percentage change in corrugator 'brow' EMG activity from baseline.

^aOne-sample *t* test applied to determine if mean scores in facial activity differed from zero. ^bCohen's *d* = effect size.

4. Supplement References

1. Mathuranath, P.S., et al., *A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia*. *Neurology*, 2000. **55**: p. 1613–1620.
2. Reitan, R.M., *Trail Making Test: Manual for administration and scoring*. 1992, Tucson, AZ: Reitan Neuropsychology Laboratory.
3. Nelson, H.E. and J. Willison, *National Adult Reading Test (NART): Test manual*. 2nd ed. 1991, Windsor, UK: NFER-NELSON.
4. Zigmond, A.S. and R.P. Snaith, *The Hospital Anxiety and Depression Scale*. *Acta Psychiatrica Scandinavica*, 1983. **67**: p. 361-370.
5. Watson, D., L.A. Clark, and A. Tellegen, *Development and validation of brief measures of positive and negative affect: The PANAS scales*. *Journal of Personality and Social Psychology*, 1988. **54**: p. 1063-1070.
6. Kunzmann, U. and D. Gruhn, *Age differences in emotional reactivity: the sample case of sadness*. *Psychol Aging*, 2005. **20**(1): p. 47-59.
7. Henry, J.D., et al., *Emotion experience, expression, and regulation in Alzheimer's disease*. *Psychology and Aging*, 2009. **24**: p. 252-7.
8. Gross, J.J. and R.W. Levenson, *Emotion elicitation using films*. *Cognition and Emotion*, 1995. **9**: p. 87-108.
9. Lohani, M. and D.M. Isaacowitz, *Age differences in managing response to sadness elicitors using attentional deployment, positive reappraisal and suppression*. *Cognition and Emotion*, 2014. **28**: p. 687-697.