

**Table S1.** Study characteristics for each of the 89 effect sizes included in the meta-analysis.

Effect Size No.	Study Name	N	Sample Description	Females (%)	Mean Age	Depression Measure	Future Thinking Measure	Aspect of Specificity Measured	Correlation and 95% CI	Study Quality Rating
1	Addis, Hach & Tippett (2016) - D1 - FT1	48	1. Control ( <i>n</i> = 24) 2. Depressed ( <i>n</i> = 24)	0.85	24.5	1. No hist of depressive symptoms 2. Past/current MDD or >14 on BDI-II	AMT-F	Proportion of specific events	-0.53 [-0.74, -0.32]	16
2	Addis, Hach & Tippett (2016) - D1 - FT2	48	1. Control ( <i>n</i> = 24) 2. Depressed ( <i>n</i> = 24)	0.85	24.5	1. No hist of depressive symptoms 2. Past/current MDD or >14 on BDI-II	AMT-F	Detail rating	-0.17 [-0.45, 0.11]	16
3	Addis, Hach & Tippett (2016) - D2 - FT1	48	Controls & depressed	0.85	24.5	BDI-II scores	AMT-F	Proportion of specific events	-0.14 [-0.42, 0.15]	16
4	Anderson, Boland & Garner (2015) - S1	61	1. Control ( <i>n</i> = 31) 2. Dysphoric ( <i>n</i> = 30)	0.85	20.6	1. <16 on CES-D 2. >15 on CES-D	SCEFT	Proportion of specific events	0.00 [-0.25, 0.25]	12
5	Anderson, Boland & Garner (2015) - S2	53	1. Control ( <i>n</i> = 26) 2. Dysphoric ( <i>n</i> = 27)	0.60	20.4	1. <16 on CES-D 2. >15 on CES-D	E-SCEFT	Proportion of specific events	-0.43 [-0.65, -0.20]	13
6	Anderson & Evans (2015) - FT1	59	1. Control ( <i>n</i> = 28) 2. Dysphoric ( <i>n</i> = 31)	0.66	22.1	1. <16 on CES-D 2. >15 on CES-D	FET	Vividness	-0.65 [-0.80, -0.50]	14
7	Anderson & Evans (2015) - FT2	59	1. Control ( <i>n</i> = 28) 2. Dysphoric ( <i>n</i> = 31)	0.66	22.1	1. <16 on CES-D 2. >15 on CES-D	FET	Coherence	-0.64 [-0.79, -0.49]	14
8	Anderson & Evans (2015) - FT3	59	1. Control ( <i>n</i> = 28) 2. Dysphoric ( <i>n</i> = 31)	0.66	22.1	1. <16 on CES-D 2. >15 on CES-D	FET	Sensory detail	-0.68 [-0.82, -0.54]	14
9	Beaty, Seli & Schacter (2018)	72	University students	0.78	20.2	DASS-21 depression subscale scores	ES	Vividness of future thoughts	0.02 [-0.21, 0.26]	13
10	Belcher & Kangas (2014) - FT1	60	1. Control ( <i>n</i> = 30) 2. Depressed ( <i>n</i> = 30)	0.77	34.4	1. No hist MDD based on SCID 2. Current MDD based on SCID	FIT	Proportion of specific events	-0.48 [-0.68, -0.29]	14
11	Belcher & Kangas (2014) - FT2	60	1. Control ( <i>n</i> = 30) 2. Depressed ( <i>n</i> = 30)	0.77	34.4	1. No hist MDD based on SCID 2. Current MDD based on SCID	FIT	Proportion of specific goals	-0.34 [-0.57, -0.11]	14
12	Belcher & Kangas (2015) - FT1	100	General community	0.68	23.1	BDI-II scores	AMT-F	Proportion of specific events	0.08 [-0.12, 0.27]	13
13	Belcher & Kangas (2015) - FT2	100	General community	0.68	23.1	BDI-II scores	AMT-F	Proportion of specific events	0.02 [-0.18, 0.22]	13
14	Blackwell et al. (2015)	150	Depressed	0.69	35.5	BDI-II scores	PIT	Vividness	0.08 [-0.08, 0.24]	18
15	Boelen, Huntjens & van den Hout (2014)	142	University students	0.91	21.5	BDI-II scores	SCEFT-2	Proportion of specific events	-0.03 [-0.19, 0.13]	13
16	Boland, Riggs & Anderson (2018) - S1&S2	73	1. Control ( <i>n</i> = 38) 2. Moderate dysphoria ( <i>n</i> = 35)	0.77	21.2	1. No sig symptoms on CESD-R 2. Subthreshold MDE on CESD-R	FEPT	Vividness	0.00 [-0.23, 0.23]	8
17	Boland, Riggs & Anderson (2018) - S2&S3	62	1. Control ( <i>n</i> = 38) 2. High dysphoria ( <i>n</i> = 24)	0.77	21.2	1. No sig symptoms on CESD-R 2. Probable MDE on CESD-R	FEPT	Vividness	0.20 [-0.05, 0.44]	8
18	Brauer (2009) - S1&S2	40	1. Control ( <i>n</i> = 20) 2. Depressed ( <i>n</i> = 20)	1.00	21.3	1. No hist MDE based on IDD & IDD-L 2. Current MDE based on IDD	RGT	Number of concrete goals	0.41 [0.15, 0.67]	14
19	Brauer (2009) - S1&S3	40	1. Control ( <i>n</i> = 20) 2. Remitted depression ( <i>n</i> = 20)	1.00	21.9	1. No hist MDE based on IDD & IDD-L 2. Hist MDE based on IDD-L	RGT	Number of concrete goals	0.51 [0.28, 0.74]	14
20	Brauer (2012) - S1&S2	58	1. Control ( <i>n</i> = 36) 2. Depressed ( <i>n</i> = 22)	1.00	20.9	1. No hist MDE based on IDD & IDD-L 2. Current MDE based on IDD	RGT	Number of concrete goals	-0.24 [-0.49, 0.00]	16

Effect Size No.	Study Name	N	Sample Description	Females (%)	Mean Age	Depression Measure	Future Thinking Measure	Aspect of Specificity Measured	Correlation and 95% CI	Study Quality Rating
21	Brauer (2012) - S1&S3	63	1. Control ( <i>n</i> = 36) 2. Remitted depression ( <i>n</i> = 27)	1.00	20.8	1. No hist MDE based on IDD & IDD-L 2. Hist MDE based on IDD-L	RGT	Number of concrete goals	0.17 [-0.08, 0.41]	16
22	Di Simplicio et al. (2016)	51	1. Control ( <i>n</i> = 26) 2. Depressed ( <i>n</i> = 25)	0.65	42.9	1. No hist DSM-IV disorder 2. MDE based on DSM-IV & >8 on HAM-D	PIT	Vividness	0.22 [-0.05, 0.48]	15
23	Dickson & Bates (2006)	34	1. Control ( <i>n</i> = 17) 2. Dysphoric ( <i>n</i> = 17)	0.76	22.0	1. <10 on BDI 2. 16-29 on BDI	FET	Specificity of events	-0.80 [-0.92, -0.68]	11
24	Dickson & MacLeod (2004) - S1 - FT1	55	1. Control ( <i>n</i> = 30) 2. Depressed ( <i>n</i> = 25)	0.33	16.4	1. 1-7 on BDI & BAI 2. 14-36 on BDI & 2-9 on BAI	Goals Task	Specificity of goals	-0.86 [-0.93, -0.79]	14
25	Dickson & MacLeod (2004) - S1 - FT2	55	1. Control ( <i>n</i> = 30) 2. Depressed ( <i>n</i> = 25)	0.33	16.4	1. 1-7 on BDI & BAI 2. 14-36 on BDI & 2-9 on BAI	Plans Task	Specificity of plans	-0.85 [-0.93, -0.78]	14
26	Dickson & MacLeod (2004) - S2 - FT1	60	1. Control ( <i>n</i> = 30) 2. Depressed & anxious ( <i>n</i> = 30)	0.52	16.4	1. 1-7 on the BDI & BAI 2. 14-36 on the BDI & 14-40 on BAI	Goals Task	Specificity of goals	-0.73 [-0.85, -0.61]	14
27	Dickson & MacLeod (2004) - S2 - FT2	60	1. Control ( <i>n</i> = 30) 2. Depressed & anxious ( <i>n</i> = 30)	0.52	16.4	1. 1-7 on the BDI & BAI 2. 14-36 on the BDI & 14-40 on BAI	Plans Task	Specificity of plans	-0.80 [-0.89, -0.71]	14
28	Dickson & Moberly (2013) - FT1	45	1. Control ( <i>n</i> = 24) 2. Depressed ( <i>n</i> = 21)	0.69	34.3	1. No DSM-IV diagnosis & <14 on BDI-II 2. MDE based on DSM-IV & >13 on BDI-II	Goals Task	Specificity of goals	-0.51 [-0.73, -0.29]	13
29	Dickson & Moberly (2013) - FT2	45	1. Control ( <i>n</i> = 24) 2. Depressed ( <i>n</i> = 21)	0.69	34.3	1. No DSM-IV diagnosis & <14 on BDI-II 2. MDE based on DSM-IV & >13 on BDI-II	GET	Specificity of explanations for goals	-0.43 [-0.67, -0.19]	13
30	Driessen (2017)	32	University students	0.81	21.7	BDI-II scores	AI	Number of episodic details	0.06 [-0.29, 0.41]	11
31	Emmons (1992) - S1	40	University students	0.70	20.0	BDI-II scores	PSL	Specificity of goals	-0.19 [-0.49, 0.11]	10
32	Emmons (1992) - S2	48	University students	0.73	20.0	BDI-II scores	PSL	Specificity of goals	-0.31 [-0.57, -0.05]	10
33	Feldman & Hayes (2005) - S1 - D1	318	University students	0.65	18.8	MASQ (depression subscale) scores	MMAP	Plan rehearsal	-0.06 [-0.17, 0.05]	12
34	Feldman & Hayes (2005) - S1 - D2	324	University students	0.65	18.8	MASQ (depression subscale) scores	MMAP	Plan rehearsal	-0.17 [-0.28, -0.06]	12
35	Feldman & Hayes (2005) - S2 - D1	99	University students	0.65	21.0	IDD scores	MMAP	Plan rehearsal	-0.20 [-0.39, -0.01]	14
36	Gamble (2015) - FT1	34	General community	0.76	24.7	BDI-II scores	SCT	Number of episodic details	-0.10 [-0.43, 0.24]	12
37	Gamble (2015) - FT2	34	General community	0.76	24.7	BDI-II scores	SCT	Vividness	-0.19 [-0.52, 0.14]	12
38	Gamble (2015) - FT3	34	General community	0.76	24.7	BDI-II scores	SCT	Spatial coherence	-0.06 [-0.40, 0.28]	12
39	Gamble (2015) - FT4	34	General community	0.76	24.7	BDI-II scores	SCT	Quality judgment	-0.16 [-0.49, 0.18]	12
40	Hach, Tippett & Addis (2016) - D1	48	1. Control ( <i>n</i> = 24) 2. Depressed ( <i>n</i> = 24)	0.83	25.2	1. No hist of depressive symptoms 2. Past/current MDD or >14 on BDI-II	AI	Number of episodic details	-0.27 [-0.53, 0.00]	13
41	Hach, Tippett & Addis (2016) - D2	48	Controls & depressed	0.83	25.2	BDI-II scores	AI	Number of episodic details	0.08 [-0.21, 0.36]	13

Effect Size No.	Study Name	N	Sample Description	Females (%)	Mean Age	Depression Measure	Future Thinking Measure	Aspect of Specificity Measured	Correlation and 95% CI	Study Quality Rating
42	Hadley & MacLeod (2010)	82	Members of charity group	0.51	51.0	HADS depression subscale scores	MEPGAP	Proportion of specific goals	0.06 [-0.15, 0.28]	14
43	Hallford (2018)	158	1. Control ( <i>n</i> = 79) 2. Dysphoric ( <i>n</i> = 79)	0.48	30.8	1. <5 on PHQ-9 2. >10 on PHQ-9	PCAFT	Vividness/detail	0.35 [ 0.21, 0.49]	12
44	Holmes et al. (2008) - S1	78	1. Control ( <i>n</i> = 45) 2. Dysphoric ( <i>n</i> = 33)	0.61	22.6	1. <7 on BDI 2. >13 on the BDI	PIT	Vividness	0.00 [-0.22, 0.23]	14
45	Holmes et al. (2008) - S2	126	General community	0.61	22.6	BDI-II scores	PIT	Vividness	-0.05 [-0.23, 0.12]	14
46	Ji, Holmes, MacLeod, Murphy (2018)	42	University students	0.60	21.6	BDI-II scores	TUTT	Vividness	0.00 [-0.31, 0.31]	16
47	Jumentier et al. (2018) - S1 - FT1	49	General community; older	0.53	69.4	CES-D scores	AMT-F	Perceptual index	-0.10 [-0.38, 0.18]	15
48	Jumentier et al. (2018) - S2 - FT1	51	General community; middle-aged	0.51	39.5	CES-D scores	AMT-F	Perceptual index	0.03 [-0.25, 0.31]	15
49	Jumentier et al. (2018) - S1 - FT2	49	General community; older	0.53	69.4	CES-D scores	AMT-F	Proportion of specific events	-0.12 [-0.40, 0.16]	15
50	Jumentier et al. (2018) - S2 - FT2	51	General community; middle-aged	0.51	39.5	CES-D scores	AMT-F	Proportion of specific events	-0.10 [-0.37, 0.17]	13
51	King et al. (2011)	44	1. Control ( <i>n</i> = 22) 2. Depressed ( <i>n</i> = 22)	0.66	46.6	1. No sub-threshold psychiatric illness 2. MDD based on DSM-IV	AI	Number of episodic details	-0.58 [-0.78, -0.38]	14
52	Lang et al. (2012) - D1	26	Depressed general community	0.78	28.5	BDI-II scores	PIT	Vividness	0.06 [-0.33, 0.45]	14
53	Lang et al. (2012) - D2	26	Depressed general community	0.78	28.5	HAM-D scores	PIT	Vividness	0.05 [-0.34, 0.44]	13
54	Lapp & Spaniol (2017) - S1 - FT1	33	University students	0.85	20.9	DASS-21 depression subscale scores	AI	Vividness	0.21 [-0.12, 0.54]	13
55	Lapp & Spaniol (2017) - S1 - FT2	33	University students	0.85	20.9	DASS-21 depression subscale scores	AI	Number of episodic details	0.07 [-0.28, 0.41]	13
56	Lapp & Spaniol (2017) - S2 - FT1	33	General community	0.88	72.1	DASS-21 depression subscale scores	AI	Vividness	-0.43 [-0.71, -0.15]	13
57	Lapp & Spaniol (2017) - S2 - FT2	33	General community	0.88	72.1	DASS-21 depression subscale scores	AI	Number of episodic details	-0.12 [-0.46, 0.23]	14
58	Lopez-Perez, Deepröse & Hanoch (2018)	7	Prisoners in med. security prison	0.00	50.1	CES-D scores	PIT	Vividness	-0.09 [-0.88, 0.70]	10
59	MacLeod & Croyley (1995)	54	University students	0.63	19.4	BDI-II scores	SGT	Proportion of general events	-0.38 [-0.61, -0.15]	14
60	Macrynika et al. (2017)	261	University students	0.82	20.3	BDI-II scores	FOF	Engagement/vividness	0.02 [-0.10, 0.14]	14
61	Marsh et al. (2018)	107	General community	0.81	32.0	HADS depression subscale scores	AMT-F	Number of specific future episodes	-0.22 [-0.40, -0.04]	14
62	McGowan et al. (2017) - FT1	258	University students	0.62	19.4	BDI-II scores	TS	Concreteness during worry	-0.12 [-0.24, 0.00]	14

Effect Size No.	Study Name	N	Sample Description	Females (%)	Mean Age	Depression Measure	Future Thinking Measure	Aspect of Specificity Measured	Correlation and 95% CI	Study Quality Rating
63	McGowan et al. (2017) - FT2	258	University students	0.62	19.4	BDI-II scores	TS	Concreteness during anticipatory processing	-0.16 [-0.28, -0.04]	14
64	Morina et al. (2011)	56	1. Control ( <i>n</i> = 32) 2. Depressed ( <i>n</i> = 24)	0.61	39.9	1. No MDD/anx based on DSM-IV/ HADS 2. MDD & no anxiety based on DSM-IV	PIT	Vividness	-0.33 [-0.56, -0.09]	12
65	Murphy et al. (2015)	75	General community	0.57	67.1	BDI-II scores	PIT	Vividness	0.00 [-0.22, 0.23]	13
66	Oettingen, Meyer & Portnow (2016)	148	University students	0.64	20.0	BDI-II scores	FF	Vividness	0.13 [-0.03, 0.29]	15
67	Parlar et al. (2016)	41	1. Control ( <i>n</i> = 20) 2. Depressed & trauma expo. ( <i>n</i> = 21)	0.51	39.0	1. No hist of psychiatric illness or hist trauma 2. MDD based on DSM-IV & hist trauma	AI	Number of episodic details	-0.29 [-0.58, -0.01]	13
68	Pile & Lau (2018)	367	Secondary school students	0.53	13.7	CDI scores	PIT	Vividness	0.00 [-0.11, 0.10]	14
69	Plimpton, Patel & Kvavilashvili (2015) - FT1	27	1. Control ( <i>n</i> = 13) 2. Dysphoric ( <i>n</i> = 14)	0.68	33.2	1. <10 on BDI 2. >15 on the BDI	TQ	Vividness	0.09 [-0.29, 0.47]	16
70	Plimpton, Patel & Kvavilashvili (2015) - FT2	27	1. Control ( <i>n</i> = 13) 2. Dysphoric ( <i>n</i> = 14)	0.68	33.2	1. <10 on BDI 2. >15 on the BDI	TQ	Proportion of specific future thoughts	-0.32 [-0.67, 0.02]	16
71	Ranger (2018) - FT1	54	1. Control ( <i>n</i> = 29) 2. Dysphoric ( <i>n</i> = 25)	0.93	21.0	1. <10 on PHQ-9 2. >9 on PHQ-9	PCQ	Vividness	0.00 [-0.27, 0.27]	14
72	Ranger (2018) - FT2	54	1. Control ( <i>n</i> = 29) 2. Dysphoric ( <i>n</i> = 25)	0.93	21.0	1. <10 on PHQ-9 2. >9 on PHQ-9	AI	Number of episodic details	-0.12 [-0.38, 0.15]	14
73	Robinaugh, Lubin, Babic & McNally (2013)	169	General community	0.78	21.0	CES-D scores	SCEFT	Proportion of specific events	0.18 [0.03, 0.33]	9
74	Stevens et al. (2017) - S1&S2	78	1. Control ( <i>n</i> = 56) 2. Dysphoric ( <i>n</i> = 22)	0.72	19.2	1. <14 on BDI 2. >19 on BDI	FOT	Concreteness of feared outcome	-0.19 [-0.40, 0.03]	11
75	Stevens et al. (2017) - S3&S4	110	1. Anxious ( <i>n</i> = 55) 2. Dysphoric & anxious ( <i>n</i> = 55)	0.72	19.2	1. <14 on BDI & >5.7 on GAD-Q-IV 2. >19 on BDI & >5.7 on GAD-Q-IV	FOT	Concreteness of feared outcome	0.17 [-0.01, 0.36]	11
76	Stevens et al. (2017) - S1	56	University students; controls	0.72	19.2	BDI-II scores	FOT	Concreteness of feared outcome	-0.14 [-0.40, 0.12]	11
77	Stevens et al. (2017) - S2	22	University students; dysphoric	0.72	19.2	BDI-II scores	FOT	Concreteness of feared outcome	0.02 [-0.41, 0.45]	11
78	Stevens et al. (2017) - S3	55	University students; anxious	0.72	19.2	BDI-II scores	FOT	Concreteness of feared outcome	-0.02 [-0.29, 0.25]	11
79	Stevens et al. (2017) - S4	55	University students; dysphoric & anxious	0.72	19.2	BDI-II scores	FOT	Concreteness of feared outcome	-0.08 [-0.34, 0.19]	11
80	Stöber (2000)	70	University students	0.66	18.6	BDI-13 scores	PIT	Vividness, detailed, speed (combined)	-0.12 [-0.35, 0.11]	12
81	Szöllösi et al. (2015) - FT1	183	General community	0.74	25.0	BDI-13 scores	MCQ	Vividness	-0.04 [-0.18, 0.11]	10
82	Szöllösi et al. (2015) - FT2	183	General community	0.74	25.0	BDI-13 scores	MCQ	Clear/distinct time	0.01 [-0.14, 0.16]	10
83	Szöllösi et al. (2015) - FT3	183	General community	0.74	25.0	BDI-13 scores	MCQ	Clear/distinct location	-0.06 [-0.21, 0.08]	10

Effect Size No.	Study Name	N	Sample Description	Females (%)	Mean Age	Depression Measure	Future Thinking Measure	Aspect of Specificity Measured	Correlation and 95% CI	Study Quality Rating
84	Szöllösi et al. (2015) - FT4	183	General community	0.74	25.0	BDI-13 scores	MCQ	Visual details	-0.01 [-0.16, 0.14]	10
85	Szöllösi et al. (2015) - FT5	183	General community	0.74	25.0	BDI-13 scores	MCQ	Sound details	0.01 [-0.14, 0.15]	10
86	Szöllösi et al. (2015) - FT6	183	General community	0.74	25.0	BDI-13 scores	MCQ	Accuracy of imagining	0.04 [-0.11, 0.18]	10
87	Tyser, Scott, Readdy & McCrea (2014)	145	Adolescents	0.53	14.0	CDI scores	PSL	Specificity of goals	-0.04 [-0.20, 0.12]	14
88	Williams et al. (1996) - S1	48	General community & hospital patients	0.67	34.0	BDI-II scores	AMT-F	Proportion of specific events	-0.30 [-0.56, -0.04]	14
89	Williams et al. (1996) - S2	24	Overdose patients	0.67	34.0	BDI-II scores	AMT-F	Proportion of specific events	-0.22 [-0.61, 0.17]	14

<sup>1</sup> More information about each study, including coding of all moderator variables, sources of data used to calculate effect sizes, and notes on any difficult decisions during data extraction and coding, can be downloaded from [osf.io/a6q5y](https://osf.io/a6q5y).

<sup>2</sup> Samples are described separately (i.e., groups 1 and 2) for studies that featured a categorical group design.

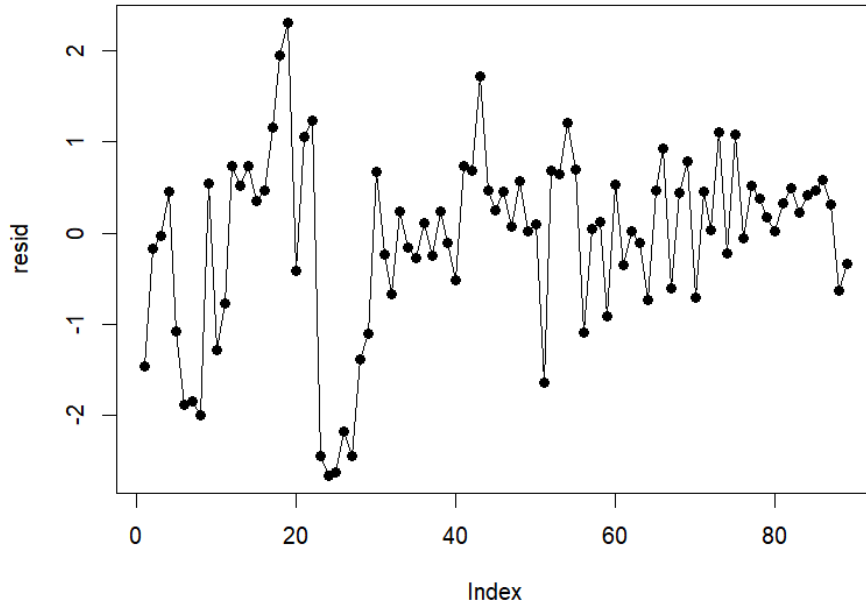
<sup>3</sup> Depression measures: BDI: Beck Depression Inventory; BDI-13: Beck Depression Inventory-13 Item Version; BDI-II: Beck Depression Inventory-Second Edition; CDI: Childhood Depression Inventory; CES-D: Centre for Epidemiological Studies Depression Scale; CESD-R: Centre for Epidemiological Studies Depression Scale-Revised; DASS-21: Depression Anxiety and Stress Scale-21 Item Version; DSM: Diagnostic and Statistical Manual of Mental Disorders; HADS: Hospital Anxiety and Depression Scale; HAM-D: Hamilton Depression Scale; IDD: Inventory to Diagnose Depression; IDD-L: Inventory to Diagnose Depression-Lifetime Version; MASQ: Mood and Anxiety Symptoms Questionnaire; PHQ-9: Patient Health Questionnaire-9 Item Version.

<sup>4</sup> Future thinking measures: AI: Autobiographical Interview (Future Version); AMT-F: Autobiographical Memory Test (Future Version); ES: Experience Sampling (unnamed task); E-SCEFT: Sentence Completion for Events in the Future Test (with emotional valence); FEPT: Future Events Prediction Task; FET: Future Events Task; FF: Future Fantasies; FIT: Future Imagining Test; FOF: Future-Oriented Fantasies; FOT: Feared Outcomes Task; GET: Goal Explanation Task; MCQ -F: Memory Characteristics Questionnaire (Future version); MEPGAP: Measure for Eliciting Positive Future Goals and Plans; MMAP: Measure of Mental Anticipatory Processes; PCAFT: Phenomenological Characteristics of Autobiographical Future Thinking; PCQ: Phenomenological Characteristics of Future Thoughts; PIT: Prospective Imagery Task; PPAI: Personal Project Analysis Inventory; PSL: Personal Strivings Listing; RGT: Revised Goals Task; SCEFT: Sentence Completion for Events in the Future Test; SCT: Scene Construction Task; SGT: Specific Generation Task; TS: Thought Samples; TQ: Thoughts Questionnaire; TUTT: Task Unrelated Thinking Task.

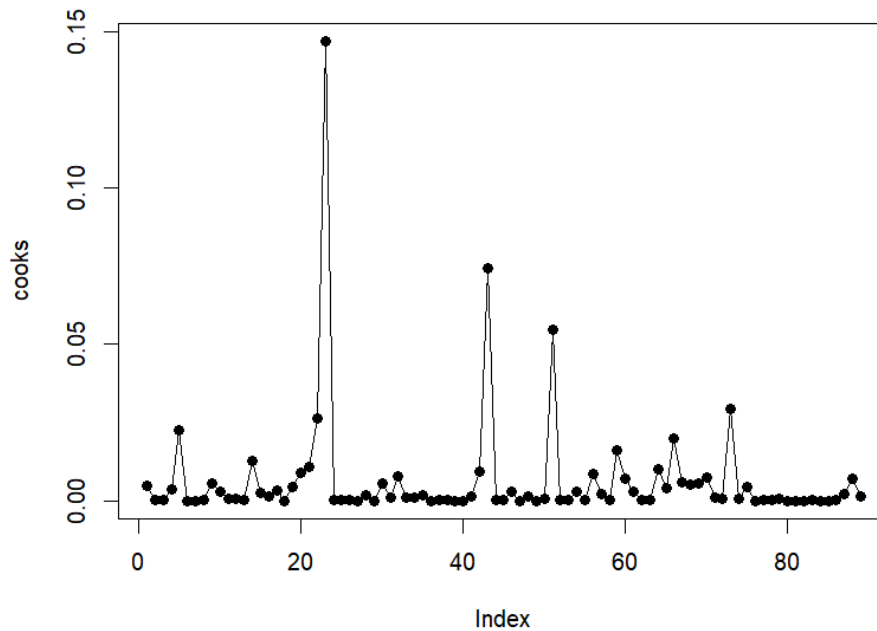
**Table S2.** Results for study quality ratings across the 46 included studies, made using the Checklist for Measuring Quality (Everaert et al., 2017) adapted from Downs and Black (1998).

<b>Item</b>	<b>Question</b>	<b>Mean</b>	<b>SD</b>
1	Hypotheses, aims, objectives clearly identified	0.98	0.14
2	Primary outcomes clearly described in intro/methods	1.00	0.00
3	Participant characteristics clearly described	0.88	0.33
4	Subjects asked is representative	0.98	0.14
5	Subjects participating is representative	0.16	0.37
6	Participants recruited from the same population	0.71	0.46
7	Participants recruited within the same time window	0.18	0.39
8	Tasks and measures clearly described	1.00	0.00
9	Main outcome measures used are valid and reliable	0.94	0.24
10	Participant engagement with the experimental task assessed	0.49	0.51
11	Consideration of principal confounders	1.29	0.82
12	Appropriate use of statistical tests to assess main outcomes	0.98	0.14
13	Main findings of the study clearly described	1.00	0.00
14	Estimates reported for random variability of main outcomes	0.94	0.24
15	Actual probability values reported	0.49	0.51
16	Withdrawals and drop-outs reported (i.e., numbers and/or reasons)	0.37	0.53
17	If any results based on data-dredging, this is made clear	0.61	0.49
18	Power analysis reported	0.12	0.33

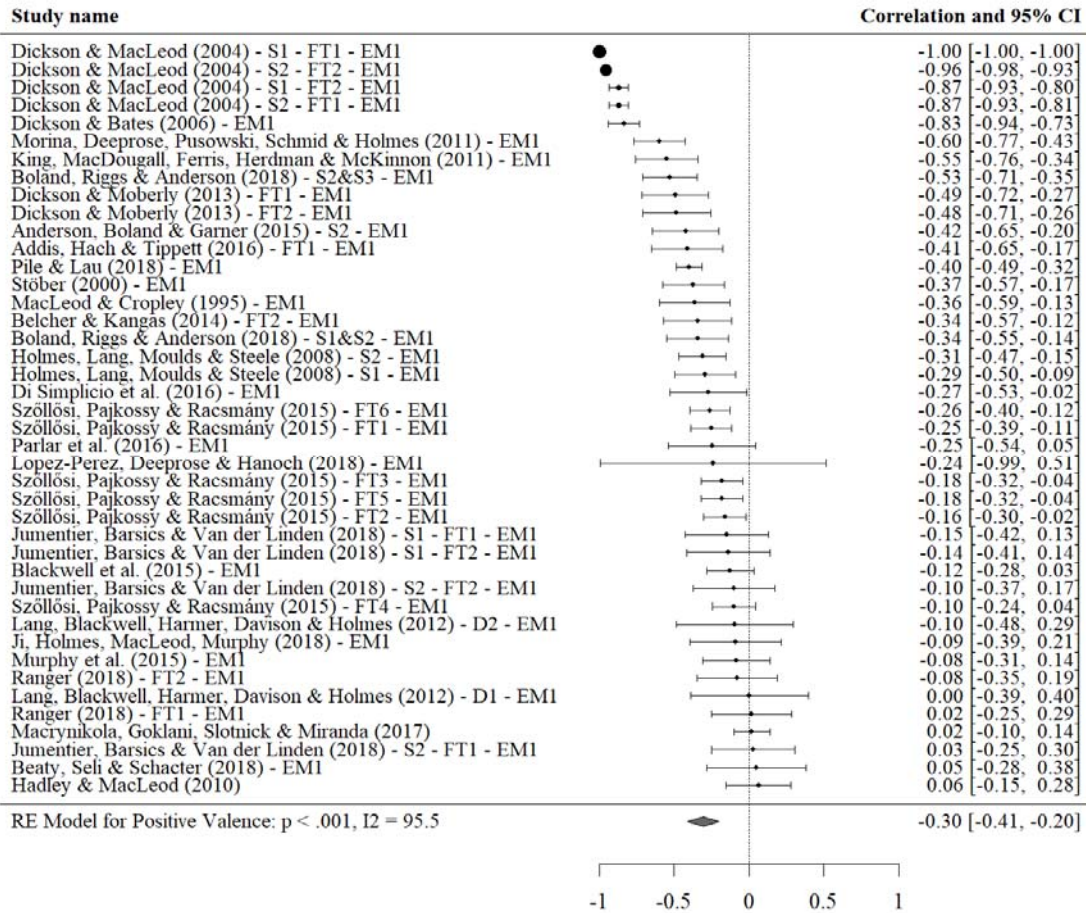
<sup>1</sup> The possible range for item scores is 0-1, except for item 11 (possible range: 0-2).



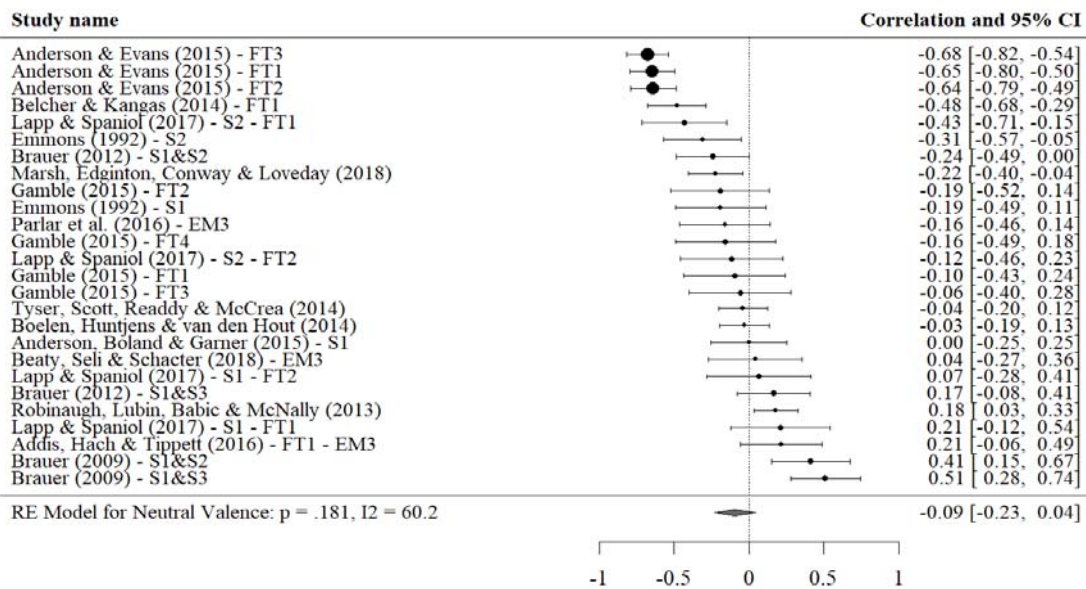
**Figure S1.** Standardised residuals of each effect size ( $k = 89$ ) included in the main random effects model. Outliers were pre-defined as effect sizes with standardised residuals greater than 3; none met this threshold and so none were excluded from the primary analyses.



**Figure S2.** Cook's distance ( $D_i$ ) scores for each effect size included in the main random effects model ( $k = 89$ ). This was an exploratory (unplanned) assessment of outliers. Based on the typical rule of thumb that effect size with  $D_i$  scores greater than 3 times the mean  $D_i$  may have an excessive influence on the model, seven effect sizes would have been classified as outliers: Anderson, Boland & Garner (2015) – S2; Di Simplicio et al. (2016); Dickson & Bates (2006); Hallford (2018), King, MacDougall, Ferris, Herdman & McKinnon (2011); Oettingen, Meyer & Portnow (2016); and Robinaugh, Lubin, Babic & McNally (2013).

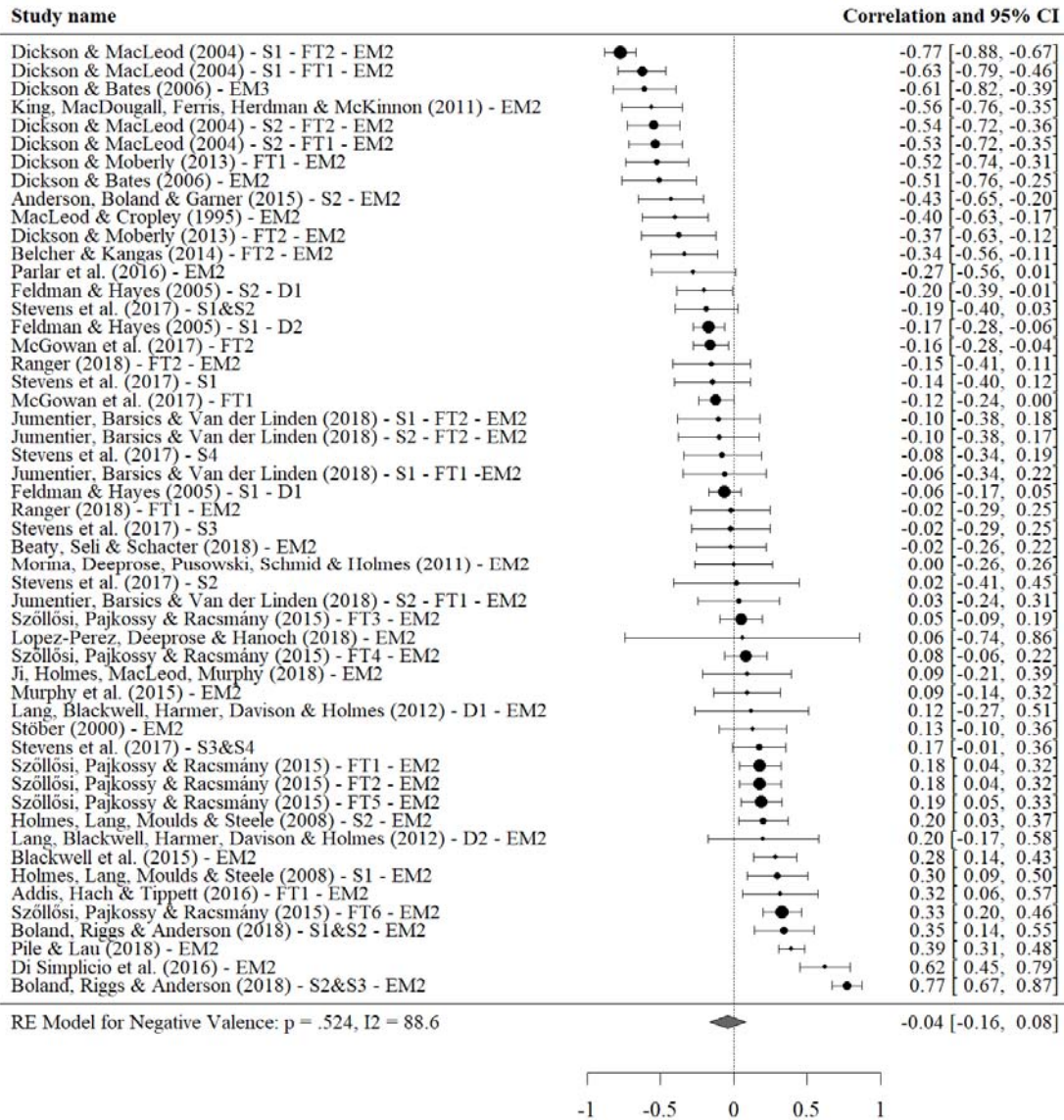


**Figure S3.** Forest plot for the effect sizes ( $k = 42$ ) arising from prompts tailored to elicit positive future thinking or future thinking retrospectively rated as positive.

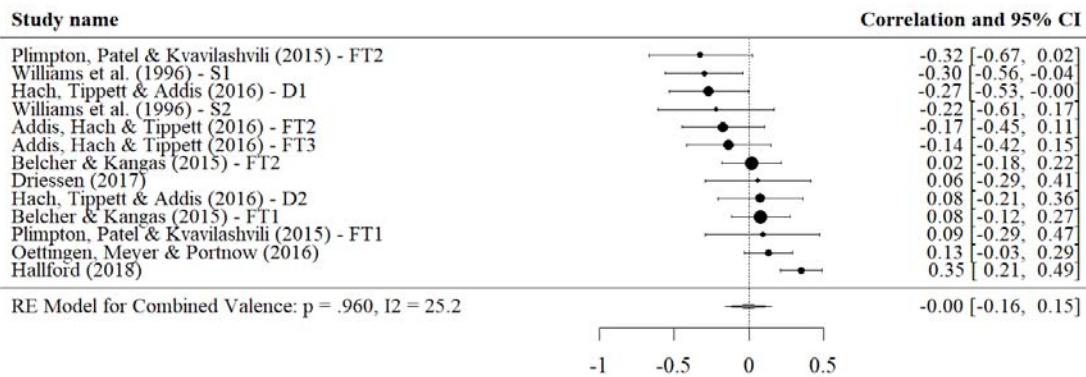


**Figure S4.** Forest plot for the effect sizes ( $k = 26$ ) arising from prompts tailored to elicit neutral future thinking or future thinking of no particular emotional valence.

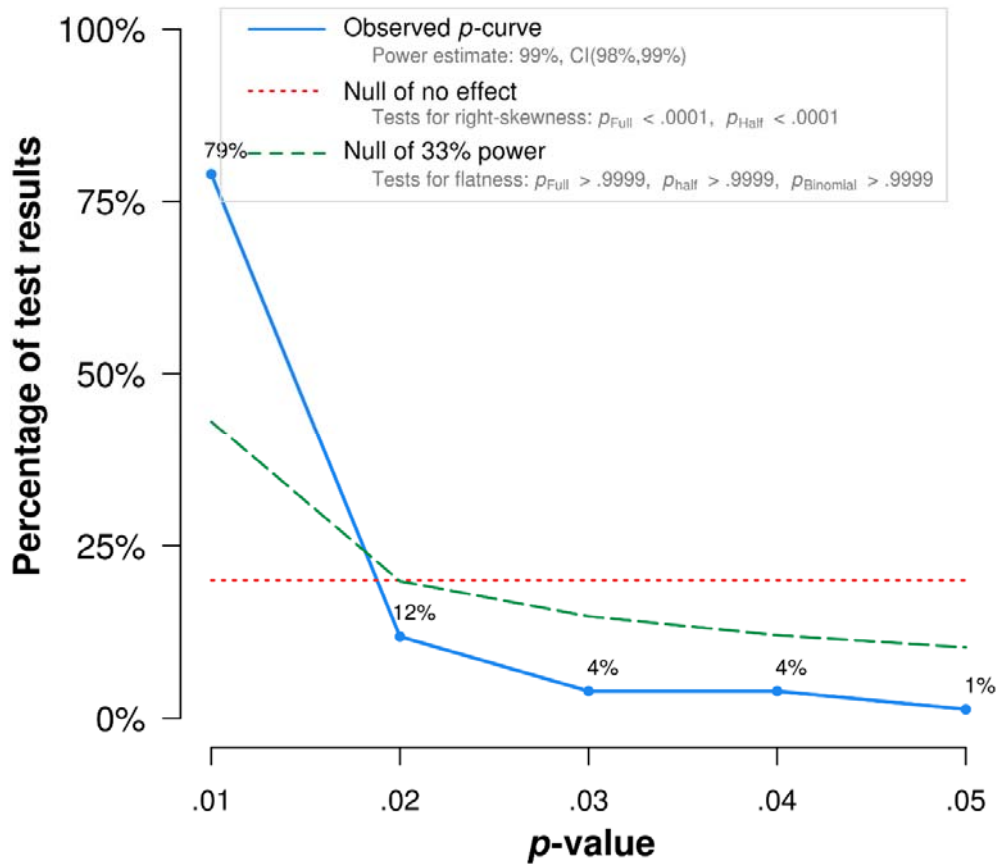




**Figure S5.** Forest plot for the effect sizes ( $k = 52$ ) arising from prompts tailored to elicit negative future thinking or future thinking retrospectively rated as negative.

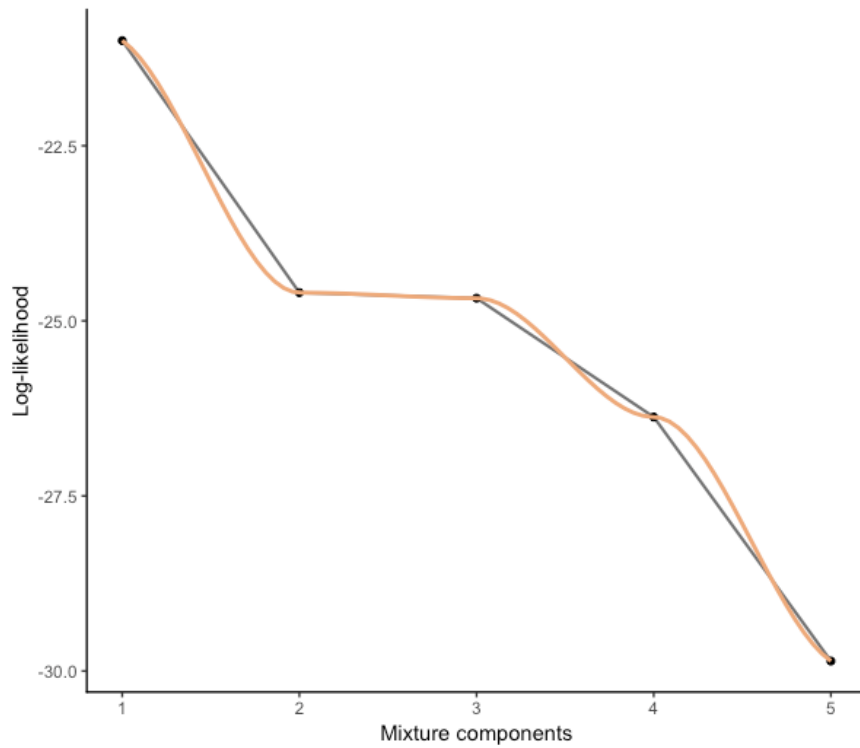


**Figure S6.** Forest plot for the effect sizes ( $k = 13$ ) reported when positively and negatively valenced future thinking were not reported individually; i.e., only combined (summary) effect sizes were reported.



Note: The observed  $p$ -curve includes 76 statistically significant ( $p < .05$ ) results, of which 71 are  $p < .025$ . There were 97 additional results entered but excluded from  $p$ -curve because they were  $p > .05$ .

**Figure S7.**  $P$ -curve of included effect sizes that were statistically significant (generated from [p-curve.com](http://p-curve.com)). The right-skew of the observed  $p$ -curve (blue) suggests there is no evidence the included effect sizes were subject to publication and/or reporting bias. All effect sizes, including those specific to the within-study moderator variable *emotional valence*, were entered, hence total  $k$  was 173 rather than 89 as reported in primary analyses.



**Figure S8.** Results of mixture model analysis (Moreau & Corballis, 2019) to explore whether the effect sizes likely belong to one or more underlying distribution(s). Out of five possible scenarios ranging from one to five distribution modes, the expectation–maximization algorithm showed the highest log-likelihood for the single-component solution, indicating that only one component should be retained (black line). The orange curve shows locally weighted smoothing. These results speak in favour of being able to meaningfully combine the effect sizes for meta-analysis, despite the high level of heterogeneity.