

Teens, Screens and Well-Being: An Improved Approach



Amy Orben

St. Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019

To all those who supported me on this journey.

Acknowledgements

Writing an acknowledgement is a question of perspective. A question of perspective very alike the one that troubles me when conducting psychological research. Do I ‘zoom in’ and only highlight those who helped me construct the exact words that now make up my thesis? Do I look more broadly and acknowledge all those who have nurtured and challenged me over the last two decades? Do I acknowledge the circumstances and connections that interacted to help me achieve? Or do I blatantly oversimplify in order to highlight the ‘most important’ actors? Regretfully, Specification Curve Analysis cannot be applied to such complex qualitative data. You will therefore have to be content with me taking an inherently unsatisfactory middle ground.

Thank you:

To Mama and Papa, for letting me stand on my own two feet; for allowing me to argue and debate; for never questioning my passions and projects; for every mountain you forced me to climb in summer or winter; for every ounce of reassurance you gave me in difficult times; for all the love and laughter; thank you. To my Oma Marliese, for telling me not to care what other people think and for showing unending interest in my life; I miss you truly. To James, for the calm and love, for the unlimited support and the laughs we share, for being my rock; my achievements are also yours. To my family, my Oma, Opa, Tanten, Onkel, Cousinen and Cousins, for being by my side, for proving to me that blood is thicker than water and that cake is always preferred;

To my supervisors and mentors: To Robin Dunbar for his unwavering support and for giving me the freedom to explore and make my own mistakes; To Andy Przybylski for helping me when I was lost and for giving me the invaluable support to grow and learn; To Dorothy Bishop for becoming a key figure in my academic life and for showing me the value of speaking up for what I believe is right;

To the funders: To Barnardo’s for funding the last year of my DPhil; To all those who have provided me with support along the way; To the European Research Council, St. Hilda’s College, the British Psychological Society, the European Association of Social Psychology, the Jacob’s Foundation, the SRCD, PsyPAG and many more;

To the academic friends: To Sam Parsons, Sophia Crüwell and Tobias Dienlin for helping me implement crazy ideas; To Verena Heise, Matt Jacquiery, Sam Webb, Nils Reimer, Jackie Thompson, Paul Thompson, Adam Kenny, Sarah de Vries, Laura Fortunato and all the other Oxford ReproducibiliTea drinkers; To both the offline and online open science community; To the ReproducibiliTea clubs around the country and around the world; to the UK Reproducibility Network and the Cumberland Lodge community; To Daniël Lakens, Peder Isager, Anne Scheel, Leo Teokhin, Tim van der Zee and all the other Red Square Lab Members who welcomed me to Eindhoven; To Sonja Utz, Ana Levordashka and Ruoyun Lin who welcomed me to Tübingen; to Danielle Shore for being my running buddy; to Ulrik Lyngs for writing *oxforddown* that, in turn, helped me write my thesis (and Yihui Xie, Chester Ismay and John McManigle); to Pádraig Mac Carron for welcoming me to the department and to the psychology football crew he helped create;

To the institutions: To the Queen's College Community which it has been an honour to serve as College Lecturer for three great years and to the wonderful students I met along the way; To Emmanuel College for inviting me to join as a College Research Fellow post-DPhil and to all those who helped me get there including Mark Buckley, Jonathan Jones, Claire Haworth, Dianne Mitchell, Alex Walker, Rogier Kievit, Jon Simons, Kate Nation and the OxCISC research group; To my Junior Researcher Programme team and to Augustin Mutak for his love of psychometrics; to the Skoll Centre and their Leading for Impact programme; To the Science Media Centre and the other trailblazers in my research area;

To all those who put a smile on my face, who believed in me and those who believe in changing science for the better; To my amazing teachers, mentors and friends; To the choirs and organ scholars; To the doctors and massage therapists; To the housemates and gym-mates; To the college and departmental staff; To those who challenged me because they liked me – or disliked me; To the journalists; To the many inspiring conversations; And to those who I have painfully forgotten to include on this page;

Thank you to all.

Amy Orben
University of Oxford
June 2019

Abstract

Concerns about emergent technologies form a continuous cycle: they appear when a new technology surpasses a certain popularity threshold and stop once a newer technology prompts the cycle to restart. Currently, society is concerned about teenage digital technology and social media use, which is feared to negatively impact well-being. While there has been a flurry of research in the area, remarkably little consensus has been reached. I argue in this thesis that such a lack of consensus stems from a lack of methodological rigour. Using innovations put forth by psychology's recent credibility revolution, I devise an improved methodological framework for the study of emergent technologies. I then apply it to investigate digital technology and social media use's association with decreased adolescent well-being, using large-scale secondary datasets from the United Kingdom, Ireland and the United States. My first study implements a methodological approach called Specification Curve Analysis, which tackles analytical flexibility by illustrating how diverse analytical approaches can produce a wide range of numerical outcomes. Furthermore, I develop a comparison specification method to put the size of the associations found into perspective. My second study focuses on diversifying measurement and implements an explicit exploratory and confirmatory hypothesis testing framework. Lastly, my third study utilises high-quality longitudinal data to highlight the bidirectional and individual nature of the effects of interest. All studies indicate that previous systematic reviews overestimated the negative link between digital technology use and well-being. It is however still unclear whether the small, complex and inherently multivariate relationship found in this thesis should engender widespread policy change.

Contents

List of Figures	xiii
List of Tables	xxi
List of Abbreviations	xxiii
1 Introduction	1
1.1 Moral Panics	4
1.1.1 The Last Century	5
1.2 The Sisyphean Cycle of Technology Panics	7
1.2.1 I. Panic Creation	8
1.2.2 II. Political Outsourcing to Science	9
1.2.3 III. Wheel Reinvention	10
1.2.4 IV. No Progress; New Panic	12
1.3 Media Effects Research: The Common Approach	13
1.4 Teens and Screens	16
1.4.1 Screens	16
1.4.2 Teens	18
Historical Perspective	19
Sociological Perspective	19
Psychological Perspective	20
1.4.3 Well-Being	21
Definition	22
Measurement	23
Measurement Stability	24
Predictors	27
Adolescent Well-Being	28
1.4.4 Current Evidence	29
Systematic Reviews and Meta-Analyses: Digital Technologies	29

	Systematic Reviews and Meta-Analyses: Social Media	30
	Different Uses	32
	Studies: Social Media	33
	Finding Common Ground	35
	Limitations and Future Improvements	36
	Cultural Critiques	37
1.5	An Improved Approach	38
1.5.1	Enabling Research Through Openness	40
1.5.2	Improving Transparency to Decrease Biases	41
1.5.3	Splitting Exploratory and Confirmatory Research to Guide Policy	42
1.5.4	Interpreting Effect Sizes to Limit Hyperbole	43
1.6	Thesis Outline	44
2	Improving Transparency	47
2.1	Introduction	48
2.2	Methods	52
2.2.1	Datasets and Participants	52
2.2.2	Ethical Review	53
2.2.3	Measures	53
	Criterion Variables: Adolescent Well-Being	54
	Explanatory variables: Adolescent Technology Use	57
	Control variables	58
2.2.4	Analytic Approach	59
	I. Identifying Specifications	60
	II. Implementing Specifications	62
	III. Statistical Inferences	63
	IV. Comparison Specifications	64
2.2.5	Code Availability Statement	64
2.2.6	Data Availability Statement	64
2.3	Results	65
2.3.1	I. Identifying Specifications	65
2.3.2	II. Implementing Specifications	67
2.3.3	III. Statistical Inferences	73
2.3.4	IV. Comparison Specifications	75
2.4	Discussion	82

2.4.1	Limitations	84
2.5	Conclusion	85
2.6	Acknowledgements	85
3	Improving Measurement	89
3.1	Introduction	90
3.1.1	Present Research	92
3.2	Exploratory Studies	93
3.2.1	Methods	93
	Datasets and Participants	93
	Ethical Review	94
	Measures	94
	Analytic Approach	99
3.2.2	Results	101
	I. Correlations Between Retrospective Reports and Time-Use	
	Diary Estimates	101
	II. Identifying Specifications	101
	III. Implementing Specifications	102
	IV. Statistical Inferences	104
3.3	Confirmatory Study	105
3.3.1	Methods	106
	Datasets and Participants	106
	Ethical Review	106
	Measures	106
	Analytic Approach	108
3.3.2	Results	108
	I. Correlations Between Retrospective Reports and Time-Use	
	Diary Estimates	108
	II. Identifying Specifications	109
	III. Implementing Specifications	109
	IV. Statistical Inferences	109
3.4	Discussion	112
3.4.1	Limitations	115
3.5	Conclusion	115
3.6	Acknowledgements	116

4	Improving Data	117
4.1	Introduction	118
4.2	Methods	120
4.2.1	Datasets and Participants	120
4.2.2	Ethical Review	120
4.2.3	Measures	121
	Life Satisfaction	121
	Social Media Use	122
	Control Variables	122
4.2.4	Statistical Approach	123
	Random Intercept Cross-Lagged Panel Models	123
	Specification Curve Analysis	125
4.2.5	Code Availability Statement	127
4.2.6	Data Availability Statement	127
4.3	Results	128
4.4	Limitations	129
4.5	Conclusion	130
4.6	Acknowledgements	130
5	Discussion	133
5.1	Overview of Results	135
5.2	Overestimation	137
5.3	Interpretation	139
5.4	Other Limitations	141
5.5	Next Steps	143
5.5.1	The UnITED Framework of Technology Research	144
5.6	Conclusion	146
Appendices		
A	Appendix A: Improving Transparency	151
B	Appendix B: Improving Measurement	157
C	Appendix C: Improving Data	161
References		169

List of Figures

2.1	Figure showing how research papers have used different combinations of MTF measures to define depressive symptoms (blue) and self-esteem (green). This illustrates the abundance of analytical flexibility in this area. We also include Newcomb, Huba and Bentler (1986) and Rosenberg (1965) who originally devised parts of the scales.: * Note: This study split the self-esteem measures into two different scales of self-liking and self-competence. ** Note: This study measured Emotional Health and also included one item about doing things other people think is strange.	62
2.2	Results of Specification Curve Analysis of the Youth Risk and Behaviour Survey: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the x-axis represents a different combination of analytical decisions, which are displayed in the ‘dashboard’ at the bottom of the graph. The resulting standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: $\beta = -.035$ (median partial $\eta^2 = .001$, median $n = 62,297$, median standard error = .004)	69

- 2.3 Results of Specification Curve Analysis of the Monitoring the Future study: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the x-axis represents a different combination of analytical decisions, which are displayed in the ‘dashboard’ at the bottom of the graph. The resulting standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: $\beta = -.005$ (partial $\eta^2 < .001$, median $n = 78,267$, median standard error = .003) 70
- 2.4 Each panel of this figure illustrates the range of possible results found using a SCA regressing a specific technology use variable onto well-being in the MTF dataset: Panel A examines TV viewing on a weekday, Panel B examines TV viewing on a weekend, Panel C examines using the internet to get news and Panel D examines social media use. The error bars represent the standard error and the dotted line shows the median standardised regression coefficient. I highlight in red the specifications chosen by Twenge and colleagues (2017) for their analysis of the MTF dataset. The researchers used a novel combination of measures from both well-being and self-esteem scales to create a well-being scale in their study, see also Figure 2.1. They correlated this measure with technology use measures (TV viewing, using the internet for news and social media use) and included either no controls or range of controls variables. We only show Twenge et al.’s results with no controls in this visualisation due to their chosen controls being difficult to reproduce computationally. We note that our results from implementing Twenge and colleague’s specification for TV viewing are not consonant with those reported in Twenge et al. (2017). For TV viewing on a weekday the specifications chosen in Twenge et al. (2017) were at the 15th percentile of effect sizes, for TV viewing on a weekend they were at the 16th percentile. For getting news over the internet, the chosen specifications were at the 55rd percentile and for social media use the chosen values were at the 12th percentile. 71

- 2.5 Results of Specification Curve Analysis of the Millennium Cohort Study: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the x-axis represents a different combination of analytical decisions, which are displayed in the ‘dashboard’ at the bottom of the graph. The resulting standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: $\beta = -.032$ (partial $\eta^2 = .004$, median $n = 7,968$, median standard error = .010) 72
- 2.6 Results of Specification Curve Analysis of the Millennium Cohort Study split by whether control variables are included in the analysis or not: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each specification number indicates a different combination of analytical decisions. The plot then shows the outcome of the corresponding analysis (standardised regression coefficient) either including control variables (teal, median standardised $\beta = -0.005$, partial $\eta^2 = .001$, median $n = 6,566$, median standard error = .011) or not including control variables (purple, median standardised $\beta = -0.068$, partial $\eta^2 = .005$, median $n = 11,018$, median standard error = .010). The bolded parts of the line indicate analyses that did not reach significance ($p > 0.05$). The median standardised regression coefficients for analyses including or not including control variables are shown using the dashed lines and the error bars visualise the standard error. 74

- 2.7 Visualisation of the comparison specifications for MCS hypothesised to have little or no influence on well-being: bicycle use, height, handedness and wearing glasses. This graph shows Specification Curve Analyses for both the variable of interest (mean technology use) and the comparison variables; It highlights the range of possible results of a simple cross-sectional regression of the variables of interest on adolescent well-being.

Wearing glasses has the most negative association with adolescent well-being (black, median $\beta = -.061$, median $n = 7,963$, partial $\eta^2 = .005$, median standard error = .010), more negative than the association of technology use with well-being (purple, median $\beta = -.042$, median $n = 7,964$, partial $\eta^2 = .002$, median standard error = .010). Handedness (red/purple, median $\beta = -.004$, median $n = 7,972$, partial $\eta^2 < 0.001$, median standard error = .010), height of the adolescent (red, median $\beta = .065$, median $n = 7,910$, partial $\eta^2 = .005$, median standard error = .010) and whether the adolescent often rides a bicycle (yellow, median $\beta = .080$, median $n = 7,974$, partial $\eta^2 = .007$, median standard error = .010) have more positive associations with adolescent well-being than technology use does.

Panel A shows how different analytical decisions (Specifications, shown on the x-axis) lead to different statistical outcomes (Standardised Regression Coefficient, shown on the y-axis). Each line represents a different variable of interest, the error bars represent the standard error. Panel B visualises the resulting Median Standardised Regression Coefficients for those Specification Curve Analyses linking the variables of interest with adolescent well-being. 76

- 2.8 YRBS Comparison Specification Curves, split into three panels to compare the effect of technology use on wellbeing to other hypothesised positive (Panel A), neutral (Panel B) and negative factors (Panel C). The figure shows the results of 15 SCAs: illustrating the range of possible regression coefficients found when examining the association between well-being and other variables. The error bars represent the corresponding Standard Error. 78

2.9	MTF Comparison Specification Curves, split into three panels to compare the effect of technology use on wellbeing to other hypothesised positive (Panel A), neutral (Panel B) and negative factors (Panel C). The figure shows the results of 15 SCAs: illustrating the range of possible regression coefficients found when examining the association between well-being and other variables. The error bars represent the corresponding Standard Error.	79
2.10	MCS Comparison Specification Curves, split into three panels to compare the effect of technology use on wellbeing to other hypothesised positive (Panel A), neutral (Panel B) and negative factors (Panel C). The figure shows the results of 15 SCAs: illustrating the range of possible regression coefficients found when examining the association between well-being and other variables. The error bars represent the corresponding Standard Error.	81
3.1	Results of the Specification Curve Analysis of Irish, American and British datasets for the correlation between well-being and digital technology use. Grey areas denote 95% confidence intervals of regression coefficients obtained using bootstrapping. Note specifications marked in red were not statistically significant ($p > 0.05$) whereas those in black were ($p < 0.05$)	103
3.2	Results of the Specification Curve Analysis for the British data. The curve shows the correlation between digital technology use and adolescent well-being, comparing specifications including controls (teal) and no controls (purple).	113
4.1	Random-Intercept Cross Lagged Panel Model	124
4.2	Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Left Panel: between-person correlations. Middle Panel: within-person effects of social media use on life satisfaction. Right Panel: within-person effects of life satisfaction on social media use. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard depicts what gender and life satisfaction domain the specific combination of analytical decisions analysed; the resulting β is shown in the plot above (red = $p > 0.05$, black = $p < 0.05$). For graphics including all analytical decisions see Figure C.1-C.3	127

4.3	Relation between life satisfaction and social media use in female (top) and male adolescents (bottom). Left Panel: between-person correlations. Middle Panel: within-person effects of social media use on life satisfaction. Right Panel: within-person effects of life satisfaction on social media use. Small dots: results of each possible combination of theoretically defensible analytical decisions. Large dots: results of the Preferred Model (white = $p > 0.05$, black = $p < 0.05$). The life satisfaction categories examined are satisfaction with school work (school work), satisfaction with school (school), mean of the six life satisfaction items (mean), satisfaction with life (life), satisfaction with friends (friends), satisfaction with family (family) and satisfaction with appearance (appearance).	129
A.1	Specification Curve Analysis of MCS examining net effects across prespecified well-being subscales. The red squares indicate specifications that were non-significant, while the black squares represent specifications that were significant. The error bars represent the Standard Error. The dotted line indicates the median regression coefficient present in the SCA and is included to ease interpretation of the analysis.	152
A.2	Number of observations (participants) for each specification analysed in the YRBS SCA. Red dots indicate when the specification was non-significant, while black dots show significant specifications. . . .	153
A.3	Number of observations (participants) for each specification analysed in the MTF SCA. Red dots indicate when the specification was non-significant, while black dots show significant specifications. . . .	153
A.4	Number of observations (participants) for each specification analysed in the MCS SCA. Red dots indicate when the specification was non-significant, while black dots show significant specifications. . . .	154
A.5	Specification Curve Analysis of MTF, also including the few specifications that included technology use items that were only asked in conjunction with a one-item happiness question. This subset of questions was excluded from the analyses in the main text to simplify the graphical representations and enhance comprehensibility. The red areas indicate specifications that were non-significant, while the black areas represent specifications that were significant. The error bars show the Standard Error. The dotted line indicates the median Regression Coefficient present in the SCA, and is included to ease interpretation of the analysis.	155

C.1	Between-Person Correlations: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard shows all the analytical decisions.	162
C.2	Within-Person Effect of Social Media Use predicting Well-Being: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard shows all the analytical decisions.	163
C.3	Within-Person Effect of Well-Being predicting Social Media Use: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard shows all the analytical decisions.	164
C.4	Number of observations (participants) for each specification analysed in the Random Intercept Cross Lagged Panel Model. This graph shows the between-person correlation. Red dots indicate when the specification was non-significant, while black dots show significant specifications.	165
C.5	Number of observations (participants) for each specification analysed in the Random Intercept Cross Lagged Panel Model. This graph shows the within-person effect of social media use predicting well-being. Red dots indicate when the specification was non-significant, while black dots show significant specifications.	166
C.6	Number of observations (participants) for each specification analysed in the Random Intercept Cross Lagged Panel Model. This graph shows the within-person effect of well-being predicting social media use. Red dots indicate when the specification was non-significant, while black dots show significant specifications.	167

List of Tables

2.1	Items of the Strengths and Difficulties Questionnaire	56
2.2	Results of Specification Curve Analysis for the Youth Risk and Behaviour Survey (YRBS, United States), Monitoring the Future Survey (MTF, United States) and Millennium Cohort Study (MCS, United Kingdom), both overall and for different technology use variables, parent/adolescent self-report or with/without control variables. . .	67
2.3	Results of the SCA bootstrapped test for Youth Risk and Behaviour Survey (YRBS), Monitoring the Future (MTF) and Millennium Cohort Study (MCS) datasets, determining whether there is a significant association between digital technology use and well-being. For each dataset we include three different possible tests of significance, where we compare the observed results with the bootstrapped samples on the basis of (i) median effect size, (ii) the share of results with the dominant sign and (iii) the share of results with the dominant sign and a p value of less than 0.05.	
	* Note: we did not use standardised variables for this analysis and therefore the observed results can diverge from those presented in Table 2.2	75
2.4	Comparison Specification results: The table shows the size of the effect of comparison variables on adolescent-wellbeing when compared to the size of the effect of technology use (measured using the mean of technology use questions) on adolescent well-being. The values indicate how many times larger the effects of the comparison variables are in comparison to technology use when examining the Youth Risk and Behaviour Survey (YRBS), Monitoring the Future (MTF) and Millennium Cohort Study (MCS) datasets.	
	* Denotes when the effect of the comparison variable on well-being is positive, and therefore in the opposite direction to the effect of technology use.	87

3.1	Items of the PSID Children's Depression Inventory. The adolescents were asked: <i>For this next set of items, choose the number next to the statement that best describes how you have felt during the last two weeks (...)</i>	96
3.2	Results of the SCA bootstrapping tests for the Irish Dataset. * Result where both measures of significance (effect size and number of significant specifications in the dominant direction) are $p < 0.05$	104
3.3	Results of the SCA bootstrapping tests for the US Dataset. * Result where both measures of significance (effect size and number of significant specifications in the dominant direction) are $p < 0.05$	105
3.4	Results of the SCA bootstrapping tests for the UK Dataset. * Result where both measures of significance (effect size and number of significant specifications in the dominant direction) are $p < 0.05$	110
3.5	Results of the SCA bootstrapping tests for confirmatory tests . . .	112
B.1	Significance Tests for Growing up in Ireland: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.	158
B.2	Significance Tests for the Panel Study of Income Dynamics: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.	159
B.3	Significance Tests for the Millennium Cohort Study: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.	160
B.4	Significance Tests for the confirmatory hypotheses: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.	160

List of Abbreviations

CLPM		Cross-Lagged Panel Model
GDP		Gross Domestic Product
GUI		Growing Up in Ireland
m		Mean
MLR		Robust Maximum Likelihood
MCS		Millennium Cohort Study
MTF		Monitoring the Future
n		Number of participants
NHST		Null Hypothesis Significance Testing
OSF		Open Science Framework
PNAS		Proceedings of the National Academy of Sciences
PSID		Panel Study of Income Dynamics
RI-CLPM	. . .	Random Intercept Cross-Lagged Panel Model
SCA		Specification Curve Analysis
s.d.		Standard deviation
SDQ		Strength and Difficulties Questionnaire
SESOI		Smallest Effect Size of Interest
UK		United Kingdom (British)
US		United States (American)
WLSMV	. . .	Weighted Least Square Mean and Variance Adjusted Estimators
YRBS		Youth Risk and Behaviour Survey

Nullius in Verba
Take nobody's word for it
— Motto of the *Royal Society*.

1

Introduction

Contents

1.1	Moral Panics	4
1.1.1	The Last Century	5
1.2	The Sisyphean Cycle of Technology Panics	7
1.2.1	I. Panic Creation	8
1.2.2	II. Political Outsourcing to Science	9
1.2.3	III. Wheel Reinvention	10
1.2.4	IV. No Progress; New Panic	12
1.3	Media Effects Research: The Common Approach . . .	13
1.4	Teens and Screens	16
1.4.1	Screens	16
1.4.2	Teens	18
1.4.3	Well-Being	21
1.4.4	Current Evidence	29
1.5	An Improved Approach	38
1.5.1	Enabling Research Through Openness	40
1.5.2	Improving Transparency to Decrease Biases	41
1.5.3	Splitting Exploratory and Confirmatory Research to Guide Policy	42
1.5.4	Interpreting Effect Sizes to Limit Hyperbole	43
1.6	Thesis Outline	44

In 1941 Mary Preston published “Children’s Reactions to Movie Horrors and Radio Crime” in *The Journal of Pediatrics* (Preston 1941). The American paediatrician had studied hundreds of six to sixteen-year-old children and concluded that over half were severely addicted to radio and movie crime dramas, having given themselves “over to a habit-forming practice very difficult to overcome, no matter how the aftereffects are dreaded”. Most strikingly, many of them consumed these dramas “much as a chronic alcoholic does drink” (Preston 1941). Preston therefore voiced severe concerns about the children’s health and future outcomes: those who consumed more radio crime or movie dramas were more nervous and fearful, suffered from worse general health and more disturbed eating and sleep. Drawing such inferences about the radio, the emergent technology of the era, followed naturally from Preston’s opinion that the consumption of “terrifying scenes can have an inhibitory effect on the functioning of every organ in the body” (Preston 1941).

Almost 80 years later, Mary Preston’s conclusions seem like a clear exaggeration of concerns. Today many parents would welcome if their children listened to dramas on the radio instead of playing around on their phones or chatting to friends on social media. Yet this research is a striking reminder that every decade new technologies enter human lives; and that in their wake there will arrive widespread concerns about their effects on the most vulnerable in society.

Considering the context of Preston’s work more broadly, radio was experiencing an extensive growth in popularity when the study was published. In 1936 about 91% of New York homes owned a household radio, which children spent between one and three hours a day listening to (Dennis 1998). This popularity sparked concern not limited to Mary Preston’s article. A New York Times piece considered whether too much listening to the radio would harm children and lead to illnesses, since the body needed “repose” and could not “be kept up at the jazz rate forever” (Ferrari 1929; as cited in Dennis 1998). The Director of the Child Study Association of America noted how radio was worse than any media that came before because “no locks will keep this intruder out, nor can parents shift their children away from it” (Gruenberg 1935). This view was mirrored in a parenting magazine published at the time: “Here is a

device, whose voice is everywhere (...) We may question the quality of its offering for our children, we may approve or deplore its entertainments and enchantments; but we are powerless to shut it out (...) it comes into our very homes and captures our children before our very eyes” (Frank 1939; as cited in Dennis 1998).

When worries about specific technologies grip a population, it is often overlooked that such concerns form part of a constant cycle. Almost identical questions are raised about any new technology which reaches the spotlight of scientific and public attention. These are then addressed by scientists, public commentators and policy makers until a newer form of technology causes the cycle to restart. Concerns about the radio, for example, only abated once the home television became popular (Dennis 1998). This cyclical nature of technology panics was already described eloquently in 1935: “Looking backward, radio appears as but the latest of cultural emergents to invade the putative privacy of the home. Each such invasion finds the parents unprepared, frightened, resentful, and helpless. Within comparatively short member, the “movie”, the automobile, the telephone, the sensational newspaper or magazine, the “funnies,” and the cheap paper-back book have had similar effects upon the apprehensions and solitudes of parents” (Gruenberg 1935).

In 2019, one could simply replace ‘radio’ with ‘social media’ in such a statement: most parents would agree that its arrival into the home has also left them unprepared, frightened, resentful and helpless. Mirroring Preston’s concerns about the radio, academic publications and other reports now routinely liken social media to drug use (Royal Society of Public Health 2017; see commentary Przybylski and Orben 2017). These fears once again raise the spectre of vast proportions of the adolescent population becoming addicted to a new technology (Murali and George 2007) and this having diverse and far-reaching negative consequences (Greenfield 2014; see commentary Bell et al. 2015). While we are amused by previous parents’ fears of radio addiction, current concerns about smartphones and social media are shaping and influencing policy around the world. In the United Kingdom there have been various committees, inquiries and white papers about this topic (Davies et al. 2019; House of Commons Science and Technology Select Committee 2019; Viner et al.

2019; Department for Digital, Culture, Media and Sport and Secretary of State for the Home Department 2019), in Asia laws have been implemented that try to curb technology use entirely (Choi et al. 2018) and in the United States there is a movement to ban smartphones before the eighth grade (Wait until 8th 2018).

Psychology and its provision of scientific evidence play a key part in the current panic about digital technologies. This means that strict measures of methodological quality control need to be in place to ensure that the research delivered will have a positive, reasoned and scientific influence on the current technology panic and those of the future. The credibility revolution in psychology has set out many processes in which the discipline can improve the quality of its scientific endeavour (Munafò, Nosek, et al. 2017; Vazire 2018). This thesis will examine how such innovations and improvements can be implemented to produce a methodological framework that ensures psychology does not become an accomplice in exacerbating technology panics, but instead uses its power to promote scientific and robust reasoning in the face of societal change.

1.1 Moral Panics

Rapid increases in concern are not restricted to fears regarding technology but occur regularly throughout public life. Cohen defined those worries that reach widespread popularity as moral panics (S. Cohen 1972). They occur when a person, group, thing, event or other entity is perceived as challenging societal values and norms. This causes introspective ‘soul searching’ in the population and moral condemnation (Garland, 2008). The central concerns engendered are most often focused on “The Other”: a group that does not constitute the main powerholders of society. The social critics, journalists and researchers are therefore “interestingly immune” when it comes to the negative effects of the panic at hand (Grimes et al. 2008, pp. 51). They instead focus on less-powerful subgroups like children or women, for which the entity of the moral panic becomes defined as a severe threat by presenting it in a stylised and stereotypical fashion (S. Cohen 1972, pp. 28). Powerful societal

actors like editors, policy-makers, religious leaders and those occupying positions of respect speak out about the problem and what they think could be a solution. The moral panic then stays in the public mind until it “disappears, submerges or deteriorates” (S. Cohen 1972, pp. 28). Cohen therefore noted that moral panics can be triggered by a wide variety of things or people and can last a short or long time.

1.1.1 The Last Century

Moral panics about the youngest generations in society – what affects them and how they are developing – have been present for centuries. Greek philosophers already voiced concerns about the damage writing would do to society and youths’ increasing lack of respect centuries ago (Wartella and Reeves 1985; Blakemore 2019). Panics come and go: consistently reincarnated when a new technology or development gains popularity across society. A complete historical summary of moral panics is therefore out of the scope of this thesis. I will focus instead on the moral panics about new technologies of the last century, which have become powerful and recurring phenomena (Wartella and Reeves 1985).

Radio and movie dramas are only one example of the wide range of technologies that have been implicated in such technology panics. In the decade that followed the moral panic about radio, concerns about comic books were on the rise. In 1954, the psychiatrist Fredric Wertham wrote a scathing and scaremongering book called *Seduction of the Innocent*, where he voiced great concern about the effects of comic books on young people: “Slowly, and at first reluctantly, I have come to the conclusion that this chronic stimulation, temptation and seduction by comic books, both their content and their alluring advertisements of knives and guns, are contributing factors to many children’s maladjustments.” (1954, pp. 10). Society and politicians were receptive to Wertham’s hypothesis of comic books being the cause of recent rises in juvenile delinquency, even though such a trend could have also been explained by societal or cultural issues. Yet such underlying causes would have been much more difficult to address. In a New York Times article, the sociologist C. Wright Mills lauded the book as “a most commendable use of

the professional mind in the service of the public” (Wright Mills 1954; as cited in Tilley 2012). Wertham’s work was instrumental to the passing of restrictive comic book legislation that has been recognised as a major contributing factor to the demise of the genre (Tilley 2012). Yet history does not look favourably on his work. It is now known that he manipulated his evidence and exaggerated his claims while refusing to share the underlying data, privileging “his interests in the cultural elements of social psychiatry and mental hygiene at the expense of systematic and verifiable science” (Tilley 2012).

Panics about comic books, radio and innovations like new music (Sternheimer 2003), were however overshadowed by the concerns raised about television and its possible promotion of violent behaviour. A report by the US Media Task Force found that television violence was “one major contributory factor which must be considered in attempts to explain the many forms of violent behaviour that mark American society today” (Lowery and DeFleur 1988, pp. 309). Yet – again – high quality evidence is lacking. A seminal before-and-after study on children in Norwich who had previously not had access to home television found that television did not increase aggression levels (Himmelweit et al. 1961). The study also highlighted that most children watched television selectively without it dominating their lives. During times of panic, however, the provision of high-quality evidence doesn’t alleviate the worries of critics and the pressure to implement policy change. An editorial in *Pediatrics*, for example, subsequently noted that professionals need to “avoid the intellectual trap of minimizing the importance of television’s effect on child and adolescent behaviour simply because the literature does not contain straightforward, statistically validated research” (Strasburger 1989). With the advent of newer technologies, however, the moral panic about television also waned. While television, comic books, radio, or movies were once the centre of attention, the cycle of media panics has moved on, relegating these technologies to the background of societal concerns.

1.2 The Sisyphean Cycle of Technology Panics

The continual rise and fall of concern about new technologies, driven by the want to comprehend and explain their influence on society, is an age-old component of societal debate. The nature of these debates has, however, changed in the last century. While some scientific commentators played a part in the technology panics about radio or comic books (Preston 1941; Wertham 1954), most of the debate was held outside of scientific arenas. Since the technological panic about television, however, science has been progressively included in the debate (Dennis 1998). This development is unsurprising as scientists are operating in an increasingly industrialised scientific space where they are expected to solve practical problems in society (Ravetz 1971). In other words, it is now an expectation that science can provide answers to those issues that are most prominent in the public or political eye. There are also fewer areas of life where previously inherited common-sense wisdom is valued more than the evidence provided by so-called scientific experts, and the assumption is growing “that every problem, personal and social as well as natural and technical, should be amenable to solution by the application of the appropriate science” (Ravetz 1971, pp.12).

This shift alters the stakeholders central to technology panics. For better or for worse, psychology – the science most closely related to child development and parenting – now plays an integral role in what I will call the *Sisyphean cycle of technology panics*, referencing the Greek myth of King Sisyphus. King Sisyphus was condemned by the Gods to roll a boulder up a steep hill in the underworld for eternity: every time he reaches the top, it rolls back down to the bottom, forcing him to walk back and start the cycle all over again. Similarly, psychological research on technology effects is in an intricate cycle of addressing societal worries about technologies. With every new technology treated as completely separate from any technology that came before (Wartella and Reeves 1985), psychological researchers routinely address the same questions; they roll their boulder up the hill, investing effort, time and money to understand their technology’s implications,

only for it to roll down again when a novel technology is introduced. Psychology is trapped in this cycle because the fabric of moral panics has become inherently interwoven with the needs of politics, society and the scientific discipline (Grimes et al. 2008). I will outline the nature of this involvement at different stages of the Sisyphean cycle of technology panics below:

1.2.1 I. Panic Creation

Technology panics are a recurring feature of the societal landscape because of how society reacts to technological developments. The predominant approach to technological reform is *technological determinism*: the idea that (1) the technologies used by a society form basic and fundamental conditions that affect all areas of existence and (2) that when such technologies are innovated, these developments are the single most important driver of changes in said society (Leonardi 2012). Technology is therefore seen as a foundation for and agent of change, while society itself is assumed to have little power to influence the technologies themselves. “Utopian and dystopian views assume that technologies [therefore] possess intrinsic powers that affect *all* people in *all* situations the same way” (Boyd 2014, pp. 15). When the internet became increasingly popular, for example, most analyses either took a utopian or dystopian point of view (Wellman 2004; Livingstone, Mascheroni, et al. 2018).

Technological developments are rapidly linked to ongoing and complicated societal changes (Grimes et al. 2008, pp. 50), an example of which is the current concern that social media is causing observed decreases in teenage mental health (Twenge 2018). What makes arguments based on technological determinism even more powerful is that they are difficult to deny. Critics are told that society will only truly understand the impact of a certain technology when a longer time-frame is available to be examined (Leonardi 2012). Technological determinism is therefore a widespread assumption which allows panics to arise quickly, by linking technological developments to current societal changes that concern the population.

1.2.2 II. Political Outsourcing to Science

Certain political benefits arise from societal panics (Garland 2008). Politicians and policy-makers routinely embrace them as an opportunity to demonstrate their willingness to stand up to emerging technology companies and their deep concern for children and other vulnerable populations (Grimes et al. 2008). In 1954, for example, Senator Estes Kefauver headed a US House subcommittee investigation into the original Superman series, making a name for himself and ultimately running for president (Grimes et al. 2008). “Then, as now, few could resist or would deny the political dynamic fuelled by the headline potential of being opposed to violence, a champion of children, and tough on a regulated industry” (Ramey 1994). Technological panics also enable politicians and society to “deflect social reform from the much more difficult issues of racial justice, economic opportunity and educational quality” (J. A. Anderson 2008). They are therefore a welcome vehicle for steering public attention away from intractable and uncomfortable issues.

While it is in the political interest to be seen to address technology panics, it has now become common to outsource the process of finding solutions to scientific research. In the last decades, politics and parenting have increasingly turned to science as a guide for addressing difficult questions. Society now treats “scientists as experts, whose opinions are regularly sought on matters of importance and for most part accepted without question.” (Okasha 2002, pp. 121). This gives science a place in society where its specialised knowledge is used to calm fears and concerns in the general population, providing “comfort and reassurance in the face of the crucial uncertainties of the world” (Okasha 2002; Ravetz 1971, pp. 386). Any new societal concern is not just a political event, but also a challenge to the relevant science (Ravetz 1971). Outsourcing the technological panic to science by funding, commissioning and referencing research, therefore allows politicians and policy-makers to calm and reassure the population, putting the onus on academics to provide a sense of security through the production of tailored research.

On the one hand, political outsourcing is a positive for psychology: researching a technology implicated in a panic promises funding, prestige and other outcomes aligned with current scientific incentives. In the industrialised age of science, where researchers work in more precarious positions and need to find research funding to support their existence, policy's decision to fund science addressing the technological panic at hand promotes work in the area (Rubenstein 1982; Ravetz 1971). Furthermore, there is an attraction to investigating something society is inherently interested in, with the hope of ultimately helping vulnerable populations.

Psychologists therefore take the position of providing the public with research into a societal concern, rooting their work in the solution of a practical issue. This is best illustrated by researchers' introductions to their work which centrally acknowledge the public concerns (Wartella and Reeves 1985). A 1937 paper considering the radio included the following introductory sentence: "The fact that parents and teachers have persistently complained about the quality of the radio entertainment provided for children makes the need for research in this field the more urgent" (DeBoer 1937) and similar introductions will be seen at the beginning of my thesis' chapters. There is therefore a scientific shift in the area towards addressing a societal problem, rather than furthering a universal scientific theory or widely accepted research thread. Some argue that this causes additional problems. They believe that for such a scientific area to self-sustain, it now needs to uphold an aura of crisis, and keep up the face of science being necessary to cure the problem at hand (Grimes et al. 2008). By providing money, attention and prestige, technology panics therefore cater to the needs of the psychological discipline, which has to sustain itself during the 21st century. This builds a network of dependence between the public, policy and academia in the face of technological concern.

1.2.3 III. Wheel Reinvention

Yet there are also clear negatives when science becomes the provider of evidence to technological concerns. Researchers have noted the distinct similarity of research conducted to address different technological panics (Wartella and Reeves 1985).

This realisation is not new; even in 1951 researchers were complaining that for each new technological concern “we seem to go through the same stages” (Seagoe 1951). Questions about addiction to relevant technologies have been raised for radio (Preston 1941), comic books (Wertham 1954), television (Lowery and DeFleur 1988), video games (Bushman and C. A. Anderson 2002) and social media (Twenge 2018). Similarly, questions about social connection, concentration and empathy are routinely found when reading psychological commentary on a range of new technologies (Greenfield 2014). It is therefore routinely overlooked that each novel technology shares more similarities than differences with its predecessors – even though it might look completely new at first glance (Seagoe 1951). Video games, for example, were first treated as a medium distinct from television but are now commonly integrated into combined reviews of so-called screen time (Dickson et al. 2018).

This reinventing of the wheel is a symptom of an area built on the existence of a practical problem, rather than the existence of a universally accepted theoretical underpinning or research thread. While there are some common ideas about technologies like the displacement hypothesis (Przybylski and Weinstein 2017; Dienlin et al. 2018), such approaches are commonly used to explain findings without having a more influential role in shaping or progressing the work. Without an underlying paradigm to guide research, each researcher is “forced to build his field anew from its foundations” (Kuhn 1962, pp. 13). The psychological discipline examining technological panics is therefore in a Kuhnian pre-paradigmatic period of science (Kuhn 1962).

With each new technological panic, a new group of researchers begins investigating the practical problem at hand without an overarching theoretical framework or an underlying basis for scientific understanding. The pre-paradigmatic nature of the field is then “marked by frequent and deep debates over legitimate methods, problems, and standards of solution, though these serve rather to define schools than to produce agreement” (Kuhn 1962, pp. 48). This conflict about methodology and scientific standards has been evident in the fraught tone of scientific debate

about violence in video games (Elson and Ferguson 2014) and the current discussion about smartphone use (Ophir et al. 2019; Twenge 2019). The lack of an underlying theory lets different camps emerge that are in scientific disagreement with each other, leaving the quality of scientific output relatively uncontrolled (Kuhn 1962).

In pre-paradigmatic fields such as technological panics research, the wheel is therefore routinely reinvented and little progress is made due to the lack of theoretical anchors for scientific investigation and quality control. The research area struggles while the field’s “leaders and practitioners are exposed to the temptations of being accepted as consultants and experts for the rapid solution of urgent practical problems” (Ravetz 1971, pp. 400). The Sisyphean cycle of technology panics therefore restricts the researchers to addressing practical problems and internal debate, rather than building a long-lasting theoretical understanding that can shape science in the long term.

1.2.4 IV. No Progress; New Panic

Psychology – the science routinely chosen to investigate the technological panic at hand – therefore has few of the tools necessary to produce evidence quickly and effectively. The funding provided by politicians and government for the scientific field helps calm the nerves of the population but will provide little evidence to inform policy interventions. When faced with the decision about how to react to the technology at hand, policy makers therefore need to choose between three options: (1) they engage in *evidence-based policy*, and as there is not enough evidence they do not implement substantial policy change, (2) they engage in *policy-based evidence*, selecting or highlighting only the evidence that adheres with the policy they aim to implement or (3) they do not consider scientific evidence at all and instead adhere to the *precautionary principle*, implementing restrictive technology policy due to the fear of potential harm.

Many of the vocal contributors to technology panics argue for the latter approach, urging policy makers to anticipate harm even when there is no evidence that it will occur. It reverses the onus of proof so that the actor needs to show that the activity

is harmless instead of vice-versa (Kriebel et al. 2001). In the area of technology panics we are reminded of an editorial by Strasburger, which highlighted that a lack of evidence should not make paediatricians shy away from examining problematic television use (1989). Furthermore, Wertham provided a gripping analogy in his book: “Gardening consists largely in protecting plants from blight and weeds, and the same is true of attending to the growth of children. If a plant fails to grow properly because attacked by a pest, only a poor gardener would look for the cause in that plant alone. The good gardener will think immediately in terms of general precaution and spray the whole field. But with children we act like the bad gardener. We often fail to carry out elementary preventative measures, and we look for the causes in the individual child” (1954, pp. 2).

While slow scientific progress makes policy interventions difficult, sooner or later the panic will subside due to the introduction of a new technology (Dennis 1998). From the historical perspective it is clear that once a new technology is introduced, interest and concern about older technologies decreases substantially (Wartella and Reeves 1985). Public and political attention turns to the new technology, restarting the Sisyphean cycle of technology panics, while leaving behind a debate that potentially contributed little to knowledge creation, and that will be largely forgotten so that the next generation of researchers can reinvent the wheel again.

1.3 Media Effects Research: The Common Approach

It is remarkable that research conducted during different technological panics proceeds in lockstep while lacking an overarching theoretical paradigm. This is explained by researchers’ universal adherence to the same basic structures of psychological thinking and argumentation. Like in social psychology, the guiding principle for most technology research is *cognitivism* (Wartella and Reeves 1985). The effects of technologies are not treated as pure stimulus-response relationships, but as influencers of abstract and measurable cognitive factors

(e.g. violent tendencies, positive affect or attitudes). To measure the effects of technologies one can therefore examine abstract before and after measures of cognitive factors, an approach popular for almost 90 years (Thurstone 1929).

Cognitivism requires researchers to assume that cognitions, and the biological structures that underly them, are either malleable throughout life or during certain developmental periods (Grimes et al. 2008). The malleability of the brain has been shown to be substantial in both early childhood (DiPietro 2000) and adolescence (Fuhrmann et al. 2015). Long-term malleability in later life has been evidenced with respect to brain injury (Fraser et al. 2002) but the extent of malleability varies with respect to age and damage (Fashad and Kolb 2010; Lenroot and Giedd 2006; V. Anderson et al. 2005). It is also not certain whether well-being or life satisfaction are malleable over time (Lucas and Donnellan 2007). While the general assumption of malleability is readily made, it is therefore not without criticism.

Research on the effect of an emergent technology normally has four elements:

- a) it focuses on a technology, content or medium,
- b) it defines a specific audience,
- c) it examines a behaviour due to the technology, content or medium, and
- d) it provides theory or methodology to generate evidence (Grimes et al. 2008).

When initially researching a new technology both a) and b) are often broadly defined as the technology as a whole (e.g. smartphones) and a broad audience (e.g. all children) respectively. The behaviour examined (c) is the outcome which is causing concern (e.g. violence or depression), while the theoretical basis (d) is cognitivism as discussed above. This generalisation directs the early literature towards a causationist standpoint: that all members of the audience considered are affected by the new technology in the same way and that this technology is sufficient to cause long-term change (Grimes et al. 2008). This approach is similar to the *hypodermic needle model* which presupposes that every medium consumed affects behaviour in a direct, identical and non-individual way (Wartella and Reeves 1985).

This causationist standpoint has had many incarnations in the last century: listening to the radio causes anxiety (Preston 1941), reading comic books causes childhood maladjustment (Wertham 1954), video games cause aggression (Bushman and C. A. Anderson 2002) and smartphones and other new media cause depression (Twenge, Joiner, et al. 2017; Twenge 2018). The argument is however difficult to uphold because of its assumptions (Himmelweit et al. 1961). Firstly, if we take a social developmental approach, any slight differences in upbringing and socialisation should cause differences in cognitive structures (Grimes et al. 2008), something that has been widely recognised to be true in both psychology and neuroscience (Kolb and Gibb 2011). For a causationist standpoint to work, however, cognitive structures and development need to be identical across the population studied so that they are identically affected by the technology of interest (Grimes et al. 2008). Furthermore, the causationist standpoint assumes that the technology of interest has more influence on the outcome of interest than other aspects of a person's life (Grimes et al. 2008). Both assumptions are difficult to support in the face of widespread scientific evidence that common individual differences in the environment can substantially change a person's cognitive structures (Kolb and Gibb 2011; DiPietro 2000).

A natural progression for research in this area is therefore to examine more specific audience types. A report by Herzog published in 1941, for example, considered how age determines whether children are affected by radio (Herzog 1941; as cited in Wartella and Reeves 1985). The 1933, the Payne Fund studies investigated the effects of films and highlighted that children's reactions depended on gender, age, past experience and predispositions (Wartella and Reeves 1985). In addition to individual differences, research often also begins to take into account the different contents that can be consumed using the same technology. Blumer and Hauser, for example, found that the effects of films on delinquency depended on the film shown, as well as on the observer (Blumer and Hauser 1970).

When research progresses to consider the cues and justifications for a technology's use, it adopts a multi-process approach (Huesmann et al. 1984). This is especially the

case once researchers start considering the bidirectionality of relationships between technology use and cognitive factors. Research has, for example, shown that predispositions predict whether someone chooses to watch violent television programmes or seek out violent video games, highlighting the potential for selective exposure effects (Atkin et al. 1979; Przybylski, Ryan, et al. 2009). The psychological and cognitivist approach to investigating a new technology therefore opens up a natural progression of scientific thinking. Research starts by taking a causationist standpoint but then moves on to more complex, multi-dimensional and bidirectional approaches.

1.4 Teens and Screens

1.4.1 Screens

In this thesis I will focus on the technology panic that is currently most prominent in the western world. Concerns about digital technology use and its potential negative effects on empathy, social interaction, attention, well-being and many other important outcomes have been growing in the last decade (Greenfield 2014). While the preceding moral panic about video games is still ongoing (see Elson and Ferguson 2014), increased use of diverse screens connected to the internet has encouraged a more generalised worry about their effects on vulnerable populations (Bell et al. 2015). The shift from a fixation on a particular technology to a more general focus on ‘screen time’, can be partly explained by our increasing inability to differentiate between various forms of screen activities. Digital devices such as smartphones, tablets or laptops can now support increasingly diverse activities and contents ranging from radio and television, to gaming, reading and social media. As written by Grimes and colleagues: “In this first decade of the 21st century, we are witnessing the final erasure of the *medium* (sic) as convergence makes the distinction of motion pictures, television, and other media merely an academic exercise” (2008, pp. 41). It is therefore increasingly difficult to determine the effect of any single technology. Using a multitude of devices, 99% of UK twelve to fifteen-year-olds now go online, at an average of 20.5 hours a week (Ofcom 2019). In America, 45% of teenagers report

they are online “almost constantly” through their use of many different devices (Pew Research Centre 2018). This makes ‘screen time’ a helpful umbrella term when voicing concerns about an increasingly digital childhood. It must, however, be noted that examining time spent on digital devices is an intrinsically bad measurement of their effects (Orben, Etchells, et al. 2018; Orben and Przybylski 2019a, Chapter 3).

Since the start of my doctoral training, social media use and smartphones have become an additional focal point of technology concerns. Smartphones and tablets allow mobile and continuous engagement with digital technologies and the internet. In the United States, 95% of teenagers (aged thirteen to seventeen) report having access to or owning a smartphone (Pew Research Centre 2018). In 2018, 83% of British twelve to fifteen-year-olds owned a smartphone, and 50% owned a tablet (Ofcom 2019). Social media is often hosted on either of the two devices and encompasses “sites and services that emerged during the early 2000s, including social network sites, video sharing sites, blogging and microblogging platforms, and related tools that allow participants to create and share their own content” (Boyd 2014, pp. 6). Social media is different compared to other communication methods as it is non-directed and does not need to be topically orientated. Previous internet platforms often focused on specific topics, but social media sites downplayed the necessity of having a common topic and instead made “friendship the organizing tenant of the genre” (Boyd 2014, pp. 7).

Social media completes the erasure of the medium as it is inherently diverse and ever-changing: its content is highly individualised and can differ from person-to-person on an hour-by-hour basis. This uncontrollability poses a challenge to parents and policymakers: “Both the efficacy of regulation (of content) and governmentality (of use), the traditional comfort blankets in relation to television, are challenged by the internet because children and young people are theoretically free to roam anywhere in the places and spaces of the cyber.” (S. Brown 2005, pp. 151; France 2007).

The diversity of social media, and its inherently social nature, makes it attractive to younger generations. In the UK, 69% of twelve to fifteen-year-olds have a social

media profile (Ofcom 2019). Asked about its effect on their lives, 31% of American adolescents believe it is “mostly positive”, in comparison to 24% saying it is “mostly negative” (Pew Research Centre 2018). One of the most appreciated aspects of social media use is the ability to connect and communicate with friends and family (Pew Research Centre 2018), something that has been a key part of adolescent life for decades (Blakemore and Mills 2014). Social media creates so-called “networked publics”, spaces in society where adolescents can form social relationships and spend time together (Boyd 2014). “Teens engage with networked publics for the same reason they have always relished publics; they want to be a part of the broader world by connecting with other people and having the freedom of mobility” (Boyd 2014, pp. 10). As adolescents are less able to access public places, because they are spending more time at home, the ability to interact in such networked publics is an opportunity for them to create their own social environment in the 21st century (Boyd 2014). What makes social media concerning is therefore its sheer diversity and uncontrollability. This emergent technology provides opportunities for adolescents to construct places that allow them to socialise and connect, away from the eyes and ears of their parents and even when their use of physical public spaces is restricted.

1.4.2 Teens

Most concerns about digital technologies and social media focus on adolescents, the earliest extensive adopters of these innovations. While adolescence used to be considered the age between puberty and marriage/parenthood, the endpoints are less clear cut today. The end of adolescence is now often linked to adult responsibilities (Patton et al. 2016) or a reduction in dependency on others (Griffiths 1996). There is however no universally accepted age range or traits of adolescence. The World Health Organisation defines adolescents as ten to nineteen-year-olds, with early adolescence being between ten and fourteen years of age and late adolescence being between fifteen and nineteen years of age (Patton et al. 2016). The lack of clear age limits stands in contrast to the distinct role teenagers take on in society. Adolescents are often viewed as having distinct attributes not found in children

or adults (Seagoe 1951). But what makes an adolescent an adolescent is still relatively unclear. Historical, sociological and psychological approaches provide important insights into this stage of life.

Historical Perspective

Society's stance towards adolescents has undergone considerable fluctuations over time. A study of child portraits concluded that in the 17th century adolescents were seen as small versions of adults (Ariès 1960). Researchers have however criticised this conclusion as portraiture was only available to the most privileged and the study can therefore not generalise to other subsets of the population (France 2007; Griffiths 1996). Furthermore, adolescents seem to have been perceived as a distinct 'group' for multiple centuries. Aristotle already remarked how "youth is the age when people are most devoted to their friends" and are also "lacking in sexual self-restraint, fickle in their desires, passionate and impulsive" (Blakemore 2018b).

Throughout the 18th and 19th century, state involvement and general concern about the adolescent age group increased; concerns specifically about leisure time started to appear in the 19th and 20th century (France 2007). Adolescents were then seen as an age group that needed shielding, even though the amount of safeguarding considered necessary fluctuated over time. In the 1960s and 1970s the arrival of youth countercultures made adolescents take a more prominent place in society as consumers, while there was also a push towards more youth education and culture (Livingstone 2009). The last half century then saw more products and technologies being marketed specifically at the adolescent age group – a development still ongoing today. Looking back, it becomes evident that our definition of adolescence, and adolescents' place in our society, is still in continuous flux.

Sociological Perspective

The sociological idea of increasing individualisation in society is a concept worthy of further attention (Beck and Beck-Gernsheim 2002), especially when considering current concerns about adolescent well-being trajectories. The last decades have been

marked by this societal development: people want increasing control over their own lives. This need for individualisation is also reflected in psychological approaches like *Self-Determination Theory* (Ryan and Deci 2000). With individualisation progressing in society, individual self-fulfilment is aspired to at the highest level and the previous social orders based around the traditional family, nation states, class and ethnicity, decline in importance (Beck and Beck-Gernsheim 2002, pp. 22). This is said to affect the youngest in society. They are put under pressure to aspire to a ‘biographical project’, where they plan their lives and wilfully navigate career and lifestyle choices (Beck and Beck-Gernsheim 2002, pp. 60-61). Failures to deal with the increasing pressure or expectation to build a biographical narrative are attributed to a failure of the self. Furthermore, problems in society and “social crises appear as individual and are no longer – or only very indirectly – perceived in their social dimension”, making people responsible for their own failures (Beck and Beck-Gernsheim 2002, pp. 24). While such a sociological change is difficult to study from a psychological perspective, it is important to consider when discussing current trends and determinants of adolescent well-being.

Psychological Perspective

Some have argued that considering adolescence a separate life stage is arbitrary (Seagoe 1951). There is, however, an increasing consensus in the psychological discipline that adolescence is a “distinct period of biological, psychological and social development” (Blakemore 2019). There are three pillars of evidence: 1. There exist behaviours, not dependent on culture, that set adolescents apart from other age groups, 2. Such adolescent-type behaviours are found across species, and 3. Such adolescent-type behaviours have been documented throughout history (Blakemore and Mills 2014; Blakemore 2019). The activity most commonly associated with adolescence is social risk taking (Steinberg 2004). While risk taking might seem inherently negative, it could be evolutionarily beneficial for adolescents maturing in an unstable and changing world (Blakemore 2018a). Furthermore, adolescents are very attuned to their social peer groups (Boyd 2014) and especially affected

by social rejection (Blakemore and Mills 2014; Blakemore 2019; Blakemore 2018a). This again can be seen as something beneficial for their maturation, while it does exacerbate problems and behaviours seen as unique to that age group.

In addition to psychological changes in cognition and behaviour, there are biological developments that make adolescence a unique window of cognitive maturation (Fuhrmann et al. 2015). Such periods of biological change give reason to believe that adolescence is a distinct stage of development. During adolescence, the brain is more susceptible to social signals, when compared to the brains of adults or children (Blakemore and Mills 2014). Furthermore, the brain is undergoing long-term changes in grey and white matter, axonal myelination and synaptic pruning (Blakemore 2019). Not only in the brain, but around the body, changes in stress (i.e. glucocorticoids) and sex hormones (e.g. testosterone), controlled either through gene expression, physical changes or epigenetic mechanisms, play a key role in “sculpting the adolescent brain” (Andersen and Teicher 2008). This increased malleability is important when theorising about potential technology effects (Grimes et al. 2008), and it supports the view that adolescence is a distinct developmental period worthy of psychological attention.

1.4.3 Well-Being

Humans are motivated to live a happy life, and therefore decreases in the well-being of certain populations attract societal concern. Well-being is not only integral for mental health, but also a key predictor of productivity (Bryson et al. 2015) and general health outcomes (Steptoe et al. 2015). Yet it is much more difficult to quantify than GDP or mortality rates. Adolescent well-being levels, or rather the current decrease of adolescent well-being levels, are however at the forefront of a current panic about digital technologies and social media (The Children’s Society 2018; Twenge 2018). This thesis will therefore focus on investigating adolescents’ subjective well-being, defined as an “overall evaluation of the quality of a person’s life from her or his own perspective” (Diener, Lucas, et al. 2018). In this section I will discuss the definition of subjective well-being and how it can be measured.

I will examine its stability over the life course and whether we need to consider some of its distinct qualities when studying adolescents.

Definition

While there exists an inherent difficulty in defining what exactly well-being is (Kesebir and Diener 2008), subjective well-being is most often separated into four components (Kesebir and Diener 2008; Diener 2000):

- Life satisfaction: global judgements about one's life
- Satisfaction with certain domains: e.g. satisfaction with friends, family or school (Huebner 2004)
- Positive affect: pleasant moods and experiences
- Negative affect: negative moods and experiences

The measurement's subjective nature limits the scope of the construct: it probes quality of life from an individual's perspective, instead of trying to obtain a more general overview (Huebner 2004). The measure should therefore not be assumed to quantify overall objective well-being. Yet subjective well-being might be one of the best proxies available for measuring well-being, and its subjectivity could be a strength. "Different people likely weight different objective circumstances differently depending on their goals, their values, and even their culture. Presumably, subjective evaluations of quality of life reflect these idiosyncratic reactions to objective life circumstances in ways that alternative approaches (such as the objective-list approach) cannot" (Diener, Lucas, et al. 2018).

Subjective well-being is separable from other conceptualisations of well-being prominent in the literature. Eudaimonic well-being, for example, assumes well-being is determined by having a sense of purpose and regular positive contact with others (Ryan and Deci 2001). This form of well-being might contribute to subjective well-being measurements. For example, meaningful relationships could benefit subjective well-being (Rohrer and Lucas 2018). However, eudaimonic well-being

also falls outside of the subjective well-being conglomerate: good relationships might not be necessary for every person's happiness.

Measurement

For people's subjective judgements about well-being to be valid, the following needs to be true: 1. All experiences people have should additively sum to make some real global well-being entity, 2. These feelings of global well-being should be relatively stable and 3. People need to be able to summarise and report these feelings in an accurate manner (A. Campbell 1981). Whether the third point is true is still relatively unclear. There are both top-down and bottom-up theories trying to describe how people arrive at subjective well-being judgements. Top-down theories argue that people view their life in a certain way and this judgement determines their responses to subjective well-being questionnaires. The bottom-up view argues that people aggregate their evaluations and experiences of many diverse aspects of their lives to derive their general subjective well-being judgement (Diener, Lucas, et al. 2018).

In parallel to this distinction, it is unclear whether respondents fill out questionnaires 'experientially' (i.e. considering how they feel now) or 'evaluatively' (i.e. considering how they think they have been feeling retrospectively). At times, judgements are made too rapidly to be the result of a full memory search, hinting that responses might be heavily dependent on experiential rather than evaluative thinking (Diener, Lucas, et al. 2018). Yet if people fill out questionnaires experientially, we would expect short-term changes in the environment that affect mood to also change their well-being ratings. Yet studies have shown that mood's effect on subjective well-being judgements is minor and inconsistent (Eid and Diener 2004). Furthermore, a study of one million US residents' life satisfaction found that the weather on the day of questionnaire completion did not substantially affect well-being judgements (Lucas and Lawless 2013). As mood and day-to-day circumstances fail to pose major threats to the reliability and validity of well-being and life-satisfaction

measurements (Diener, Lucas, et al. 2018; Huebner 2004), participants are probably using a mixture of experiential and evaluative techniques to answer the questions.

Whatever way participants approach their completion, there is sufficient evidence showing that subjective well-being questionnaires are valid by traditional psychometric baselines (Diener, Lucas, et al. 2018; Diener, Emmons, et al. 1985). While participants' scores can be slightly influenced by socially desirable responding and emotional states, these influences can be discounted as non-detrimental to the measurement procedure (Diener 2000). It is however still unclear how the many different measures of subjective well-being relate to each other. The measures range from quantifying life satisfaction in various domains (Knies 2017; University of Essex, Institute for Social and Economic Research 2018) to asking about mood either on the day or over a specific past timeframe (University of Michigan. Survey Research Center 2018). As there is still little consolidation of measures, researchers expert in the area recommend that “when possible, researchers should include a broad array of measures, including both judgment-focused measures like life satisfaction and more affective measures.” (Diener, Lucas, et al. 2018). I therefore include diverse and partially overlapping scales of subjective well-being in the work completed for this thesis.

Measurement Stability

A question central to this thesis is whether subjective well-being can be effectively changed or whether it is stable throughout life. Only if well-being is malleable can media significantly affect it (Grimes et al. 2008). While there has been much disagreement about the stability of well-being, the conflicting evidence hints of the existence of a small proportion of well-being that can be changed over time.

Happiness pie hypothesis. The disagreement about stability is at its clearest when comparing the positive psychology movement to the hedonic treadmill hypothesis. Positive psychology, a popular area of psychology for over two decades, stresses that small changes to how people live their lives can increase long-term

happiness (Seligman 2002). Researchers have floated ideas like the *Happiness Pie Hypothesis*: that approximately 50% of individual differences in happiness are due to genetics, 10% are dependent on life circumstances and the remaining 40% can be influenced by the activities people decide to do in their lives (Lyubomirsky et al. 2005). This idea of 10% of happiness being controlled by your circumstances and 40% by your volition became very popular, as it fit “into a pervasive contemporary ‘feel-good’ narrative in which self-improvement is claimed to be not only desirable, but also eminently achievable” (N. J. L. Brown and Rohrer 2019). But this is an exaggeration. Researchers have criticised the Happiness Pie Hypothesis for confusing within- and between-subject variance, conflating demographic factors and life circumstances, failing to include a necessary error term and not accounting for important interactions and the uncertainty around how much genes determine happiness (N. J. L. Brown and Rohrer 2019). The Happiness Pie Hypothesis therefore overestimated the percentage of well-being that is susceptible to change through lifestyle choices by making multiple errors. While the hypothesis might be an exaggeration, however, one should note that some researchers have simulated agent-based models showing that even a small percentage of happiness malleability can cause significant differences in the happiness levels of monozygotic twins (Nes and Røysamb 2017).

The area of positive psychology, and its message that any person is the maker of their own happiness, has been increasingly criticised as well. Some believe that the narrative about real-world pursuits increasing happiness are money-making ideals lacking scientific evidence, which cause great opportunity costs for individuals, science and society (N. J. L. Brown and Rohrer 2019; Moreau et al. 2019). Moreau and colleagues stress that “overemphasizing the role of environmental factors in success may lead to failure being stigmatized, despite the fact that individual differences in many real-world endeavours may in part reflect factors that are not under people’s control” (2019).

Hedonic treadmill hypothesis. In contrast to the Happiness Pie Hypothesis, other groups of researchers have long argued that happiness might not be as malleable as often proclaimed. The *Hedonic Treadmill Hypothesis* suggests that people return to their baseline level of happiness after significant life events (Diener 2000; Brickman and D. T. Campbell 1971). The original authors of this theory found that while new paraplegics and lottery winners exhibited changes in happiness, they returned back to baseline levels in the long term (Brickman and D. T. Campbell 1971). There has been much support for this theory (Lykken and Tellegen 1996), but it has been supplemented to better address more recent research findings. The *Dynamic Equilibrium Model*, for example, now includes personality as a predictor (Diener 2000). Furthermore, people are now known not to fully return back to their original baseline of subjective well-being after an extreme life event like the loss of a spouse (Diener and Oishi 2005). Yet it is still agreed that, while life events or daily activities perturb well-being, they oftentimes have little long-term impact.

This conclusion is supported by research using a variety of methods to quantify the stability of happiness levels. It has been suggested that up to 50% of variance in life satisfaction is stable (Lykken and Tellegen 1996). A more recent study, however, has adjusted this estimate using the STARTS method, finding that the stable trait component of subject well-being is about 34-38%, while the autoregressive component is about 29-34% (Lucas and Donnellan 2007). The relatively high stability of well-being and life satisfaction could be driven by genetics (Nes and Røysamb 2017). Three meta-analyses found that the heritability of overall happiness measures is about 32-41% (Nes and Røysamb 2017). Taking what we now know about well-being's stability into account, one could conclude that some proportions of happiness can be changed by life events and recent actions, while over a third is purely stable and outside technology effects' reach. This supports the theoretical validity of hypotheses about activities like technology use affecting small proportions of well-being.

Predictors

To understand how technology might be affecting well-being outcomes, it is important to understand what activities predict increases or decreases in well-being. Social variables seem to play a large part in predicting subjective-well-being, illustrating that happiness is not an individual pursuit (Diener and Oishi 2005; Bradburn 1969). This has been reflected in seminal theories of social psychology that highlight the ‘need to belong’ as fundamental to human happiness (Ryan and Deci 2000; Baumeister and Leary 1995). Recent pre-registered and large-scale research found that those who aim to socialise to increase their well-being are most successful in reaching their goal (Rohrer and Lucas 2018). Furthermore, other research has highlighted how social connection is a core aspect of human life, both in modernity and throughout our evolutionary past (Dunbar 2018). The substantial influence of social variables on subjective well-being is found across cultures: research using the World Values Survey has highlighted that social capital is a key predictor of happiness around the world (Helliwell and Putnam 2004). When evaluating this research, we however need to keep in mind that social variables are easier to change than, for example, health variables. This could partially explain their high predictive power (Rohrer and Lucas 2018).

In addition to social circumstance, a variety of other factors predict subjective well-being. High levels of income, strong relationships and religiosity are successful predictors (Diener, Lucas, et al. 2018). Surprisingly, education and gender are less predictive – at least in adult populations (Diener, Lucas, et al. 2018). The high stability of subjective well-being and its moderate heritability also show that internal factors like personality can be predictors (Diener, Lucas, et al. 2018). While this is of no importance to this thesis, which focuses only on UK, US and Irish populations, well-being judgements’ variability between cultures should also be considered in cross-cultural work (Diener, Lucas, et al. 2018).

Lastly, it is important to note that well-being should not only be considered as an outcome – “happiness may also be functional” (Oishi et al. 2007, Chapter 3). Not

being completely happy can be a motivating factor for people to engage in activities that stimulate future achievement (Oishi et al. 2007). Happiness can therefore also be a stimulus for action in our lives, rather than just being an outcome (Nes and Røysamb 2017). We will see the importance of this notion in Chapter 4 of this thesis.

Adolescent Well-Being

While adolescents are routinely perceived as being at their healthiest, a deterioration in well-being during this time of development can have long-lasting consequences on later quality of life (Patton et al. 2016; Blakemore 2019; Andersen and Teicher 2008). Previous studies have found that the stable portion of subjective well-being is much smaller in young people (Lucas and Donnellan 2007). Furthermore, while gender is not predictive of subjective well-being in adulthood, adolescent females are twice as likely to suffer from depression than adolescent males (Andersen and Teicher 2008). There is also a higher potential for predisposed adolescents to start exhibiting the symptoms of a psychiatric disorder that will extend into their adult life (Andersen and Teicher 2008). Adolescence is therefore also a distinct stage of life when considering well-being and mental health.

Social and biological factors could underly the qualitative differences in adolescent well-being. Socially, adolescents are known to be more receptive to social rejection (Blakemore and Mills 2014), especially as self-reported quality of friends is a key predictor of future adolescent mental health (van Harmelen et al. 2017). Biologically, adolescent brains are undergoing key developmental changes and show a qualitatively different response to stress (Andersen and Teicher 2008; Eiland and Romeo 2013). Stress might also have a more negative effect on well-being because the adolescent hypothalamic-pituitary-adrenal axis is still maturing, and glucocorticoid development is not yet finalised (Romeo 2013; Eiland and Romeo 2013). Adolescence is therefore a very sensitive period for stress affecting future mental health (Fuhrmann et al. 2015). These factors combine to make adolescent well-being a distinct and influential aspect of life that can determine important long-term outcomes.

1.4.4 Current Evidence

“There is, as yet, no scientific consensus on the impact of screen-based lifestyles on the mental health of young people” (Frith 2017). Yet there have been well over 80 systematic reviews and meta-analysis published that examine this link in a range of populations (Dickson et al. 2018). This number is bound to increase further, as the production of evidence in the area is still advancing at accelerating speeds. A review of these systematic reviews and meta-analyses found that they are of varying quality standards (Dickson et al. 2018). In this literature review of the link between digital technology use and well-being, I will therefore not consider those reviews ranked as having a medium or high risk of being biased (Dickson et al. 2018). If the reviews were not ranked by Dickson and colleagues, for example if they were published in 2019, I will also include them in this review. I will however exclude those reviews specifically focused on sexting, gaming, aggressive behaviour, internet addiction or those that only examined a specific sub-population (e.g. Rice et al. 2016; B.-q. Wang et al. 2017; Mitrofan et al. 2009). It is important to note that many reviews were published using the keyword ‘sedentary behaviour’ and have therefore been routinely overlooked by the psychological or communication sciences literature.

The many competing meta-analyses and reviews regarding social media and screen use rely predominantly on cross-sectional evidence, as this makes up the vast majority of evidence in the field (Dickson et al. 2018). The quality of evidence reviewed is therefore very low (Carson et al. 2016). My literature review will begin with a review of reviews about digital technology use and psychological outcomes. I will then proceed to examining reviews specifically about social media use, before summarising particularly high-quality studies in the area. I will also discuss potential improvements and current limitations of research addressing digital technologies and their link to well-being.

Systematic Reviews and Meta-Analyses: Digital Technologies

Systematic reviews in the field have routinely been confronted with a mixture of conflicting results. If averaged, these results provide evidence for a positive

association between time spent using digital technologies (i.e. screen time) and depressive symptoms (Hoare et al. 2016). Reviews of studies on very young children found low to moderate quality evidence that TV use is linked to unfavourable outcomes (Poitras et al. 2017; LeBlanc et al. 2012). Systematic reviews examining older populations highlight that 1 in 8-12 studies find a null result, while the rest find a positive association between screen time and unfavourable psychological outcomes (Tremblay et al. 2011; Dennison et al. 2016). The relation is however not exceedingly clear. Some systematic reviews noted that a link between screen time and depressive symptoms only exists in cross-sectional and not in longitudinal studies (Liu et al. 2016). In contrast, others find that it is the longitudinal studies that report a negative or null relation (Carson et al. 2016). To make sense of such conflicting reviews, the “very low” quality of research in the area must be taken into account (Carson et al. 2016; World Health Organisation 2019). The conflicting results highlight that the evidence is still too weak to promote a uniform interpretation of the effect of interest.

The evidence base for the link between screen time and self-esteem is even weaker (Hoare et al. 2016). Just like for depression, there are many mixed results and slightly more studies find negative results (Carson et al. 2016). There has however been a randomised control trial showing that limiting television use increased self-esteem, which has been used by many systematic reviews to argue for a link (Tremblay et al. 2011). But one high-quality study on a specific intervention cannot make up for the many low-quality studies in the area that find mixed evidence.

Systematic Reviews and Meta-Analyses: Social Media

While reviews about screen time are still popular, the focus of the field has changed. The first meta-analysis about internet use and well-being outcomes was done almost a decade ago (Huang 2010). It found a small negative relation of $r = -0.05$ between using the internet and well-being, but the review’s quality was graded as low due to a high risk of bias (Dickson et al. 2018). A more recent meta-analysis of social anxiety and internet use found no correlation when examining 22 studies (Prizant-Passal

et al. 2016). The focus of scientific attention, and therefore of meta-analyses and reviews, has however shifted to social media use in the last years.

A systematic review of social media use and its links to depression, anxiety and distress highlights that this research literature is also conflicting (Keles et al. 2019; Verduyn, Ybarra, et al. 2017). Furthermore, the evidence is low-quality and cross-sectional in nature (McCrae et al. 2017; Frost and Rickwood 2017). Reviews have found small correlations between social media use and depressive symptoms (Frost and Rickwood 2017; Verduyn, Ybarra, et al. 2017) that (if numerically provided) range from $r = 0.11$ (Yoon et al. 2019) and $r = 0.13$ (McCrae et al. 2017) to $r = 0.17$ (Vahedi and Zannella 2019). Another meta-analysis found no significant relationship between social media use and well-being ($r < -0.01$, Hancock et al. 2019). Yet when this meta-analysis only examined studies of adolescents, this correlation did rise to levels similar to those found in other meta-analyses ($r = -0.07$). This was also the case when the meta-analysis solely examined studies that related anxiety or depression to social media use ($r = 0.11$, Hancock et al. 2019). The associations between social media use and well-being therefore range from about $r = -0.15$ to $r = -0.10$. It is however still unclear what such a small effect tells us about well-being outcomes and whether policy and parenting should be adapted in the light of such evidence. This is especially the case as social media use is inherently linked in complex ways with other aspects of life and it therefore should not be surprising to find only a small correlation present. This question will be discussed further in the introduction and other chapters of this thesis.

It is also important to note that other reviews have highlighted positive effects of social media. Some find that social media increases well-being, social communication, social support, social capital, authentic self-presentation and social connectedness while decreasing loneliness – even though these reviews routinely note that other studies have found exactly the opposite (Erfani and Abedin 2018). One review concluded that those users who go to Facebook to promote social support and connection show lower levels of depressive symptoms (Frost and Rickwood 2017). Other meta-analyses have also found that social media use increases social support

(Liu et al. 2016) and that online media use increases perceived social resources ($r = 0.12$, Domahidi 2018). One way to explain such a conflict is that different outcomes were examined. To arrive at an overarching conclusion, it might be necessary to differentiate the emotional and social outcomes of social media use (Bayer et al. 2018). Social media might have a negative effect on emotional outcomes (e.g. mood or depression), but a positive effect on social outcomes (e.g. social connectedness). Yet even when examining the same outcome, positive and negative results can coexist because effects of social media can vary across users and time frames: it is therefore likely “that some users experience positive outcomes while others (and possibly the same users at different points in time) experience deleterious outcomes” (Frost and Rickwood 2017).

Different Uses

Different uses and utilisations of social media might therefore be important to consider in order to obtain a better understanding of social media effects (Burke, Marlow, et al. 2010). One major distinction is that between active and passive use, with active use representing activities like chatting, messaging and liking while passive use including activities like browsing newsfeeds, profiles or scrolling through photos and news items (Ellison et al. 2007). Researchers have hypothesised that active use increases social capital and connectedness, therefore positively affecting well-being, while passive use increases upward social comparisons and envy, in turn decreasing well-being (Verduyn, Ybarra, et al. 2017). Studies have found that active use increases bonding social capital and decrease loneliness, while passive use doesn’t have such positive outcomes (Burke, Marlow, et al. 2010). Experimental and experience sampling studies support this idea by highlighting that passive use decreases well-being, potentially by increasing envy (Verduyn, Lee, et al. 2015). It is therefore important to differentiate between active and passive uses of social media. Yet it is important to note that the results are still not clear cut. A study of 10,557 Facebook users whose Facebook data were examined for three months prior to them filling out a questionnaire, found that active Facebook use did not

influence well-being. Only direct communication with close friends and family was linked to positive results (Burke and Kraut 2016).

When considering different uses of social media, one also needs to examine the style of a user's online self-presentation. A qualitative synthesis of 21 observational studies examining Facebook self-presentation and mental health outcomes found that inauthentic self-presentation was related to low self-esteem and high social anxiety. More authentic or positive self-presentation was associated with increased levels of self-esteem and social support (Twomey and O'Reilly 2017). A two-wave longitudinal study found that people who were more authentic on their profile reported higher positive affect and life satisfaction, and lower negative affect six months later (Reinecke and Trepte 2014). In addition to active and passive use, a person's self-presentation might therefore be an important factor to consider in order to understand the link between social media use and well-being.

Studies: Social Media

There have been a variety of experimental and longitudinal studies that are worth mentioning when reviewing the evidence around social media use and psychological effects. Many experimental studies have asked participants to refrain from using social media. They often find inconclusive effects, that however suggest a positive effect between limiting social media use and well-being. A study showed that those participants told to refrain from using Facebook for five days exhibit lower cortisol levels but also decreased life satisfaction (Vanman et al. 2018). In another study, those participants asked not to go on Facebook for a week showed increased life satisfaction, especially if they were heavy users (Tromholt 2016). In contrast, a study asked undergraduates to limit their social media use to 10 minutes per day or continue as normal: both the experimental and the control group showed decreases in anxiety and fear of missing out, but only the experimental group showed additional decreases in loneliness and depression (Hunt et al. 2018). A more extensive study of 2,897 participants where one group was told to deactivate Facebook for four weeks, found that the experimental group showed small increases in well-being measured

retrospectively. There were however no changes in the well-being measures collected by experience sampling or loneliness reports (Allcott et al. 2019).

‘Facebook detox’ studies therefore find inherently conflicting results. Such conflicts could be the result of the studies’ low quality. Many experimental designs did not limit all social media use and most studies found it difficult to obtain good levels of participant compliance (Vanman et al. 2018; Allcott et al. 2019; Tromholt 2016). Furthermore, there is a potential for bias in participant selection: those potential participants who are not as reliant on social media to obtain positive outcomes might be more likely to take part in studies asking for them to give up social media.

There are also many longitudinal and experience sampling studies examining social media use and well-being. Some have found negative results on outcomes like life satisfaction (Kross et al. 2013). Others have found that those who communicate more frequently on social media are more satisfied with life (Dienlin et al. 2017) or have more positive emotions (Wenninger et al. 2019). In contrast, other studies found no association between social media use and life satisfaction (Utz and Breuer 2016) or depression (Jelenchick et al. 2013). Interestingly, effects might be dependent on the longitudinal time frame considered in the study: it was found that posting a status update increased positive affect after 10 minutes but not after 30 minutes or two weeks (Bayer et al. 2018).

It is important to also note a recent trend in the study of social media. There has been increased interest in and publication of cross-sectional results based of large-scale epidemiological datasets (Twenge, Joiner, et al. 2017; Twenge, Martin, et al. 2018; Twenge, Spitzberg, et al. 2019; Booker et al. 2018; Kelly et al. 2019; Khouja et al. 2019). While such studies help diversify the participant pool and give a larger sample size, they come with their own problems and some have been criticised for overemphasising negligible correlations (Ophir et al. 2019). In the mixed landscape of screen time research, they do not provide evidence with much potential to move the field forward.

Finding Common Ground

While the research area is filled with conflicting findings based on cross-sectional evidence, there is however some common ground. Many studies and meta-analyses find a small negative association between social media use and well-being of about $r = -0.15$ to $r = -0.10$. Correlations and observed effects in this ballpark have been shown in meta-analytic studies considering anxiety and depressive outcomes (e.g. McCrae et al. 2017; Yoon et al. 2019; Vahedi and Zannella 2019; Hancock et al. 2019), but have also been found in longitudinal research (Kross et al. 2013; Frison and Eggermont 2017; Reinecke, Meier, et al. 2018; Bayer et al. 2018) and experimental work (Allcott et al. 2019). As mentioned above, it is still unclear what such a range of effects can tell us about well-being and how it is affected by social media use. This is because there are a range of third factors that can influence both variables, and there have been sources of bias not addressed properly in a literature that is largely cross-sectional and exploratory.

The same kind of effect size has, however, also been found bidirectionally: for social media use decreasing well-being and well-being decreasing social media use (J.-L. Wang et al. 2018). The importance of bidirectional effects is clearly evident, but the results remain unclear. An early group of experimental and correlational studies found that while disconnection drives the use of Facebook, connection results from Facebook use (Sheldon et al. 2011). This does not fall in line with those studies finding negative relations in both directions (J.-L. Wang et al. 2018; Frison and Eggermont 2017; Aalbers et al. 2018), only in the direction of social media use decreasing well-being (Kross et al. 2013) or only in the direction of loneliness leading to Facebook use (Song et al. 2014). It is therefore clear that more work considering bidirectional effects needs to be completed before true effects become evident. To start finding common ground, research therefore needs to increase transparency, while doing more to interpret the size and importance of effects and highlight their bidirectionality.

Limitations and Future Improvements

The low quality and conflicting state of the literature highlights many areas of the research field that could be improved further. On the one hand there needs to be an increased focus on individual differences. This would be helped by the study of more diverse and rigorously recruited samples (Erfani and Abedin 2018). More studies should also account for factors like gender or age. While age is not a routine focus of studies (Hancock et al. 2019), gender has been shown to be a predictive factor in recent work (Frison and Eggermont 2016; Twenge, Joiner, et al. 2017). While the predominant line of reasoning is that girls are more negatively affected by social media than boys (Twenge, Martin, et al. 2018), one meta-analysis found that screen time is linked to higher depressive symptoms and lower self-esteem in boys, but not in girls (Tremblay et al. 2011), and other studies found screen time effects to be independent of gender (Hoare et al. 2016; Frison and Eggermont 2017). It is however important to note that social media and general screen time are two different activities and can therefore have different gender effects.

“Ultimately, our findings demonstrate the lack of a uniform overall ‘Facebook effect’ on individuals, and illustrate the need to build temporal and spatial components into future research on Facebook and the wider social media ecosystem.” (Bayer et al. 2018). It is therefore important to conduct more longitudinal work (Dickson et al. 2018; Carson et al. 2016; Carson et al. 2016; Frost and Rickwood 2017) with more diverse time frames (Bayer et al. 2018) ranging from short-term experience sampling (Aalbers et al. 2018) to long-term annual studies (J.-L. Wang et al. 2018).

It also needs to be noted that there has been increasing discontent about the measurement practices used in the area. Researchers argue that there are now the psychometric tools available to move away from measuring self-reported screen time (Andrews et al. 2015; Wilcockson et al. 2018; Ellis 2019; Ellis et al. 2019), which is known to be a flawed measure of media effects (Scharkow 2016). Better measurement could lead to more exact and consistent results in the literature. More exact tracking would also allow different types of social media and technology use

to be examined in more nuanced and diverse ways, distinguishing different activities and timings of use. This would support important processes of triangulation (Munafò and Davey Smith 2018). Furthermore, it would enable researchers to home in on possible non-linear dose-response relationships between technology use and psychological outcomes, which have been shown in previous work (Hoare et al. 2016; Przybylski and Weinstein 2017; Hoare et al. 2016). This review therefore suggests multiple areas of potential improvements ranging from better measurement to more longitudinal work and a keener eye for individual differences.

Cultural Critiques

A variety of other criticisms have been voiced about research into digital technologies and their effects. Commentators have noted that blaming social media and digital technologies for societal concerns is easy to do, but can lead public conversation astray: “All too often, it is easier to focus on the technology than on the broader systemic issues that are at play because technical changes are easier to see” (Boyd 2014, pp. 16). The broad popularity of smartphones makes them a clearly visible scapegoat, while increasing inequality, austerity and deficiencies in children mental health services are less easy to spot. Current fears about digital technologies can therefore be “dangerously misguided when we consider all the real, though not as sexy, issues that get pushed out of the headlines in favour of media fears. Poverty, family violence, child abuse and neglect, and the lack of quality education and health care are problems that merit public attention way before media culture” (Sternheimer 2003, pp. 3). Yet it should be noted that while there having been rapid changes in societal life (e.g. austerity measures in the UK), such developments are much more long-term and gradual than technological developments. Furthermore, important aspects of a child’s life, like their risk of maltreatment, have been improving over the last decades (Degli Esposti et al. 2018). While it is easier to pinpoint the blame for negative developments on technologies, however, technology panics are used as a political tool to 1. take attention away from other societal issues and 2. to allow political figures to profile themselves. The media also profits

from them: news consumers feel more obliged to read or watch scaremongering coverage, because not doing so might be detrimental to their, or their children's, safety (Altheide 2002). The resulting widespread media coverage then makes the concerns spread more widely throughout the public conversation.

Furthermore, technology panics might be overemphasising the effect of technologies, in the face of other life factors: "It often seems as though media researchers presume that people have none but a life in the media. People hang out, hook up, and live in face-to-face relationships every day" (Grimes et al. 2008, pp. 69). Grimes and colleagues add that the resulting "findings are banal (is that so . . . people can learn and be influenced by the media?), a scientizing of the obvious. Its primary effect has been to advance disciplinary interests and individual careers. Its unintended consequence has been to deflect attention from the hard issues of social reform by promising a quick regulatory fix" (2008, pp. 91)

1.5 An Improved Approach

Is the psychological study of digital technologies and social media predestined to be subsumed into the ever-continuing Sisyphean cycle of technology panics? Will the research outputs be forgotten and ignored once a new technology gains widespread societal traction? Parts of this introduction paint a bleak picture of the research area and its interplay with policy and the public. Some of the following chapters will also uncover further quality control issues stifling progress in the field. Yet there are multiple factors which make the study of digital technologies and social media an important and influential research area.

Firstly, we are only on the cusp of a digital technology revolution that will change society for decades to come. The potential for digital technology and social media to alter the most basic features of society easily surpasses that of previous innovations like comic books, radio and television. It is still unclear how such changes will affect us and future generations. This makes it especially important for scientists to determine how such technologies are already affecting society today:

not just to increase our knowledge of what is to come, but also to help shape future policy approaches and technology design. Psychology therefore has the potential to play a crucial and important role in ensuring that the digital revolution will have a net positive effect on society.

Secondly, the last decade has seen psychology in a time of crisis and reform (Vazire 2018; Munafò, Nosek, et al. 2017). Discontent about past methods and scientific approaches (Simmons et al. 2011; Open Science Collaboration 2015) has led to the development of many ideas aimed at improving the discipline’s scientific processes. We are therefore at the beginning of a so-called *credibility revolution* (Vazire 2018). As the field of technology panics research is affected by quality control issues and lacks a methodological framework, current initiatives to improve the quality and efficiency of psychological research could provide an impetus for improvement. Better quality of evidence in the field would allow science to effectively address the technology questions of the future, while it would also limit the opportunity costs and financial burdens created by current technology panics.

Thirdly, science is now facing an ever-accelerating technological revolution that is making research on novel technologies a race against time (Valkenburg and Piotrowski 2017). Innovations to improve research efficiency could therefore provide an important boost to psychology’s ability to address those technological innovations introduced to society at accelerating speeds. As Lewis Carroll once wrote: “My dear, here we must run as fast as we can, just to stay in place. And if you want to get somewhere else you must run at least twice as fast as that” (Carroll 1871; as cited in Valkenburg and Piotrowski 2017).

Taking these three factors to heart, my thesis will develop and implement a new methodological framework to improve, accelerate and innovate the provision of scientific evidence during technology panics. Such work has the potential to better policy approaches and inform changes that will ultimately benefit society. While the methodological framework I introduce and develop during this thesis cannot halt the Sisyphean cycle of technology panics, it can improve the research output considerably. This would ensure that the public and policy are better informed

about the effects technologies are having on the population. In total, my thesis will consider five potential innovations: speeding up research through openness, improving transparency to decrease biases, splitting exploratory and confirmatory research to guide policy and interpreting effect sizes to limit hyperbole.

1.5.1 Enabling Research Through Openness

Making data and code openly accessible has the potential to accelerate scientific progress, while providing an additional impetus to improve research quality. Accelerating scientific progress without compromising on quality is especially important in times of increasing technological innovation rates (Valkenburg and Piotrowski 2017). Since the publication of the 6th edition of the APA manual in 2009, researchers are required to share their data on request for up to five years after publication (American Psychological Association 2010). This is however not common practice in the field (Houtkoop et al. 2018; Ceci 1988). A reluctance to share research data has been linked to weaker evidence and more errors in statistical reporting (Wicherts et al. 2011). When asked about the barriers to sharing data, 77% of researchers highlighted the fear of alternative analyses exposing invalid conclusions. Furthermore, 73% feared discoveries of errors in data (Houtkoop et al. 2018). While posing barriers to widespread adoption, these fears highlight how openness would benefit the scientific community by facilitating better quality control (Stark 2018). Naturally not all data can, or should, be shared: a discussion of ethical and privacy issues is however out of the scope of this thesis (see Gilmore et al. 2018). But even when data need to be kept private, the provision of analysis code can still provide many benefits – exposing errors and allowing researchers to reuse work to speed up progress in the field (as noted in Orben and Przybylski 2019b). Openness would therefore allow research to progress at greater speeds, providing evidence faster and adapting quicker to changes in technologies.

1.5.2 Improving Transparency to Decrease Biases

Flexibility in how researchers analyse and report their data is an ingrained and substantive problem throughout science. Simmons and colleagues showed that “undisclosed flexibility in data collection and analysis allows presenting anything as significant”, even the finding that undergraduates get physically younger when listening to specific music (2011). The extent of analytical flexibility is best explained by the *garden of forking paths* analogy (Gelman and Loken 2014). Any researcher needs to make multiple decisions when analysing their data (e.g. what outliers to exclude, what control variables to add, etc.). When making these decisions while analysing their data, they can unconsciously or consciously choose those data analysis methods that lead them towards the result that they were expecting or hoping to find. Researchers wander through this garden of forking paths of analytical choices, but only report the one analytical pathway they took to obtain their final result. Looking back, they often believe this analytical path is the only one they would have taken all along. But the garden of forking paths shows that there are oftentimes thousands of equally respectable ways of analysing the data, which were not reported and might have provided very different results. The ability to choose the preferred analytical path, so-called *researcher degrees of freedom*, enables studies to find physically impossible results (Simmons et al. 2011). They also allow researchers to report (false-) positive results in their work (Bishop 2019). The power of researcher degrees of freedom has been compellingly demonstrated in recent crowdsourced analyses where multiple statistical teams analysed the same dataset to answer the same research question, producing a wide range of possible end-results (Silberzahn et al. 2018; see also MAPS Project 2019).

Researcher degrees of freedom are hugely influential and can increase the false positive rate in a discipline, especially when there are cognitive biases and widespread pressures to publish positive results (Wagenmakers, Wetzels, et al. 2012; Bishop 2019). This has negative consequences for science as “research findings that do not replicate are worse than fairy tales; with fairy tales the reader is at least aware that the work is fictional” (Wagenmakers, Wetzels, et al. 2012). In the

last years, the problem of analytical flexibility has been increasingly recognised as substantial: previously “everyone knew it was wrong [to do flexible data analysis], but they thought it was wrong the way it is wrong to jaywalk [... but now we recognise] it was wrong the way it is wrong to rob a bank.” (Simmons et al. 2018). Researchers have therefore been advocating for more transparent disclosures of analytical pathways (Simmons et al. 2018; Stark 2018; Bishop 2019).

Preregistration and Registered Reports have been suggested as two possible ways to combat analytical flexibility (Wagenmakers, Wetzels, et al. 2012; Chambers 2013; Chambers 2014; van’t Veer and Giner-Sorolla 2016). Preregistration entails registering the process of data analysis before accessing the data, meaning that the researcher decides on their path through the garden of forking paths before the data can bias their choice. Registered Reports further aim to remove publication bias, by providing peer review before data collection and giving a paper ‘in principle acceptance’ at a journal before the results are known (Chambers 2014). They have already been successfully applied to the research considering new media and technologies (Elson and Przybylski 2017). Such initiatives have shown the potential for transparent research to better inform policy, the public and academia. Transparency therefore has the potential to hugely benefit the provision of evidence about new technologies.

1.5.3 Splitting Exploratory and Confirmatory Research to Guide Policy

Preregistration and Registered Reports allow researchers to clearly distinguish between exploratory and confirmatory work (Wagenmakers, Wetzels, et al. 2012; Elson and Przybylski 2017). The difference between exploratory and confirmatory evidence is important to highlight, especially when the evidence routinely informs policy decisions (House of Commons Science and Technology Select Committee 2019). When implementing conventional statistics, a hypothesis test should only be used once on a dataset; it is however extremely common that multiple analyses are run on data until a favourable result is found (Wagenmakers, Wetzels, et al. 2012;

Simmons et al. 2011). Pre-registration – specifying the analysis plan before the data are available to the researcher – avoids the Texas sharp shooter fallacy where the shooter fires his bullets before painting on the targets afterwards, therefore achieving so-called perfect performance even though this was clearly not the case. While some have criticised strictly confirmative approaches as limiting creativity (S. Scott 2013), others disagree and believe that differentiating exploratory and confirmatory work channels scientific creativity into more fruitful pursuits (Wagenmakers, Dutilh, et al. 2018). Some go as far as arguing that policy should only be based on confirmatory research, while exploratory research should be solely used to inform future studies and theorising. As policy change can have a substantial impact on society, it is a compelling argument that only gold-standard confirmatory evidence should be used as its foundation.

1.5.4 Interpreting Effect Sizes to Limit Hyperbole

With many psychological fields becoming increasingly big-data driven (Lang 2013), new approaches for communicating effect sizes will also become important. Using large-scale datasets is often wrongly equated to doing better quality research (for discussion see Orben and Przybylski 2019b). The problems associated with large sample sizes are best explained by Trout’s 1994 metaphor: a study with too low power (a too small sample size) is like trying to catch fish with a too large meshed net, you will not catch many fish and if you do, the fish will be uncharacteristically large. If you increase your power and sample size, your smaller meshed net will catch more of the fish you want (Trout 1994). Taking this analogy one step further, however, it is also true that if your sample size becomes even larger, the higher power will provide you with such a tightly meshed net that you will also catch tiny correlational dirt and not just the fish you were looking for. Because larger datasets can ‘catch’ increasingly smaller effects (Orben and Lakens 2019; Meehl 1967), statistical significance ceases to be a good indicator of practical significance.

This problem has been evident in recent research where statistically significant results were reported, but they were so small they should have little to no effect

on our actions in the real world (Kramer et al. 2014). To ensure that minute, but statistically significant, effects are not over-reported, researchers have suggested defining a Smallest Effect Size of Interest: the smallest possible effect that will be reported as practically ‘significant’ in a study (Lakens, Scheel, et al. 2018). Defining such a value is however very difficult (Anvari and Lakens 2019) and depends on the perspective that one takes about what populations will be affected (Rose 2008). It is however increasingly clear that effective communication of effect sizes will become crucial for both academia and policy in times of increasingly large-scale data research.

1.6 Thesis Outline

This thesis will devise a new methodological framework for the study of teenagers and their use of digital technologies. To do so it will apply and develop multiple methodological and statistical innovations that have the potential to increase the quality, efficiency and transparency of research in the field. Taking into account the many areas of reform highlighted by psychology’s credibility revolution (Vazire 2018; Crüwell et al. 2018), it will outline an improved approach for current and future work on emergent technologies.

In Chapter 2, I focus on how bias in data analysis, *p*-hacking and exaggeration of effect sizes can be understood, estimated and their impact minimised in the research area. The chapter introduces Specification Curve Analysis as a way to add transparency and illustrate how the selection of different analytical choices can lead to the production of skewed results. Furthermore, I detail a method I devised called comparison specifications to put the effect sizes of important relations into a novel perspective. Together, the approaches make the inherent uncertainty surrounding research results more transparent, allowing academia, policy and the public to better judge results’ relative importance.

In Chapter 3, I focus on the measurement issues afflicting the quantification of digital technology effects by focusing on diversifying measurement. I analyse three datasets that include time-use diaries and retrospective self-report measures of

digital technology use. I also use an explicit exploratory and confirmatory hypothesis testing framework to introduce a more confirmatory approach to the research area. This chapter therefore focuses on additional methodological innovations, while further developing the statistical approaches pioneered in Chapter 2.

Chapter 4 concludes the thesis by detailing one of the first high-quality and transparent longitudinal analyses considering the role of social media in the lives of adolescents. The research examines social media use and its long-term effect on life satisfaction using Random Intercepts Cross-Lagged Panel Modelling executed in a Specification Curve Analysis framework, allowing me to disentangle within- and between-person effects. In this study, I also consider gender differences for the first time. Just like the previous two chapters, the work is fully accessible online with open code for all the analyses reported. This should help accelerate error detection and future scientific progress.

The thesis therefore provides an overview of an improved methodological framework that can address some of the major problems holding back research in the area of digital technology use and its effect on adolescent well-being. It takes initial steps towards ultimately addressing parts of the Sisyphean cycle of technology panics, while future-proofing a research field which is already struggling to keep up with an increasingly accelerating and influential digital technology revolution.

This chapter is based on the published work:

Orben, A., & Przybylski, A. K. (2019). *The association between adolescent well-being and digital technology use*. *Nature Human Behaviour*, 3(2), 173.

2

Improving Transparency

Contents

2.1	Introduction	48
2.2	Methods	52
2.2.1	Datasets and Participants	52
2.2.2	Ethical Review	53
2.2.3	Measures	53
2.2.4	Analytic Approach	59
2.2.5	Code Availability Statement	64
2.2.6	Data Availability Statement	64
2.3	Results	65
2.3.1	I. Identifying Specifications	65
2.3.2	II. Implementing Specifications	67
2.3.3	III. Statistical Inferences	73
2.3.4	IV. Comparison Specifications	75
2.4	Discussion	82
2.4.1	Limitations	84
2.5	Conclusion	85
2.6	Acknowledgements	85

The most recent empirical evidence supporting the idea that regular use of digital technologies by teenagers negatively impacts their psychological well-being is largely based on secondary analyses of large-scale social datasets. Though these datasets provide a valuable resource for highly powered investigations, their many variables and observations are often explored with an analytic flexibility that marks small effects as statistically significant, thereby leading to potential false positives and conflicting results. Here I address these methodological challenges by applying Specification Curve Analysis across three large-scale social datasets ($n_{\text{tot}} = 355,358$) to rigorously examine correlational evidence for digital technology affecting adolescents. The association I find between digital technology use and adolescent well-being is negative but small, explaining at most 0.4% of the variation in well-being. Taking the broader context of the data into account suggests that these effects are probably too small to warrant policy change.

2.1 Introduction

The idea that digital devices and the internet have an enduring influence on how humans develop, socialize, and thrive is a compelling one (Bell et al. 2015). As the time young people spend online has doubled in the past decade (Ofcom 2019), the debate about whether this shift negatively impacts children and adolescents is becoming increasingly heated (Steers 2016). A number of professional and governmental organizations have therefore called for more research into digital screen time (House of Commons Science and Technology Select Committee 2019; British Youth Council 2017), which has led to household panel surveys (Johnston et al. 2016; Kann et al. 2016) and large-scale social datasets adding measures of digital technology use to those already assessing psychological well-being (University of London 2017). Unfortunately, findings derived from the cross-sectional analysis of these datasets are conflicting; in some cases negative associations between digital technology use and well-being are found (Etchells et al. 2016; Twenge, Joiner, et al. 2017), often receiving much attention even when correlations are small. Yet other results are mixed (Parkes et al. 2013) or contest previously found negative effects

when re-analysing identical data (Ferguson 2018). A high-quality pre-registered analysis of UK adolescents found that moderate digital engagement does not correlate with well-being, but very high levels of usage possibly have small negative associations (Ferguson 2017; Przybylski and Weinstein 2017).

There are at least three reasons why the inferences behavioural scientists draw from large-scale datasets might produce divergent findings. First, these datasets are mostly collected in collaboration with multidisciplinary research councils and are characterized by a battery of items meant to be completed by postal survey, face-to-face or telephone interview (Johnston et al. 2016; Kann et al. 2016; University of London 2017). Though research councils engage in public consultations (National Health Service 2016), the pre-tested or validated scales common in clinical, social or personality psychology are often abbreviated or altered to reduce participant burden (U.S. Department of Health and Human Services, Health Resources and Services Administration and Bureau 2014; Livingstone, Haddon, et al. 2011). Scientists wishing to make inferences about digital technology's effects using these data need to take numerous decisions about how to analyse, combine and interpret the measures. Taking advantage of these valuable datasets is therefore fraught with many subjective analytical decisions, which can lead to high numbers of researcher degrees of freedom (Silberzahn et al. 2018). With nearly all decisions taken after the data are known, they are not apparent to those reading the published paper highlighting only the final analytical pathway (Gelman and Loken 2014; Simmons et al. 2011).

The second possible explanation for conflicting patterns of effects found in large-scale datasets is rooted in the scale of the data analysed. Compared to the laboratory- and community-based samples typical of behavioural research (mostly $< 1,000$, Marszalek et al. 2011), large-scale social datasets feature high numbers of participant observations (ranging from $n = 5,000$ to $n = 5,000,000$, Johnston et al. 2016; Kann et al. 2016; University of London 2017). This means very small covariations (e.g. r 's $< .01$) between self-report items will result in compelling evidence for rejecting the null hypothesis at α levels typically interpreted as statistically significant by behavioural scientists (i.e. p 's $< .05$). As minute correlations are arguably

omnipresent in self-report data (Meehl 1990a; Meehl 1990b; Orben and Lakens 2019), such correlations are therefore more often labelled statistically significant.

Thirdly, it is important to note that most datasets are cross-sectional and therefore only provide correlational evidence, making it difficult to pinpoint causes and effects (see Chapter 4 for more detail). Thus, large-scale datasets are simultaneously attractive and problematic for researchers, peer reviewers and the public. They are a resource for testing behavioural theories at scale but are, at the same time, inherently susceptible to false positives and significant-but-minute effects using the α levels traditionally employed in psychology.

Given that digital technology’s impact on child well-being is a topic of widespread scientific debate among those studying human behaviour (Bell et al. 2015) and has real-world implications (House of Commons Science and Technology Select Committee 2019), it is important for researchers to make the most of existing large-scale dataset investments. This makes it necessary to employ transparent and robust analytic practices, which recognize that the measures of digital technology use and well-being in large-scale datasets may not be well-matched to specific research questions. Further, behavioural scientists must be transparent about how the hundreds of variables and many thousands of observations can quickly branch out into gardens of forking paths with millions, and in some cases billions of analysis options (Gelman and Loken 2014). This risk is compounded by a reliance on statistical significance, i.e. using $p < .05$, to demarcate ‘true’ effects. Unfortunately, the large number of participants in these designs means small effects are easily publishable and, if positive, garner outsized press and policy attention (Ferguson 2018).

As large-scale secondary datasets are increasingly available freely online, it is not possible to convincingly document a scientist’s ignorance of the data before analysis (Chambers 2013; Munafò, Nosek, et al. 2017; van’t Veer and Giner-Sorolla 2016), making hypothesis preregistration untenable as a general solution to the problem of subjective analytical decisions. Specification Curve Analysis (SCA, Simonsohn et al. 2015) provides a promising alternative (see 2.2.4 for further explanation). SCA is a

tool for mapping the sum of theory-driven analytic decisions that could have been justifiably taken when analysing quantitative data. Researchers demarcate every possible analytical pathway and then calculate the results of each one. Instead of reporting a handful of analyses in their paper, they report all results of all theoretically defensible analyses. This approach has been published in the previous literature (Simonsohn et al. 2015; Rohrer, Egloff, et al. 2017).

Because of the substantial disagreements within the literature, the extent to which adolescents' screen-time may actually be impacting their psychological well-being remains unclear. The present chapter addresses this gap in our understanding by relying on large-scale data paired with a conservative analytic approach to provide a more transparent and clearly contextualized test of the association between screen use and well-being.

To this end, three large-scale exemplar datasets (Monitoring the Future, the Youth Risk and Behaviour Survey and the Millennium Cohort Study) from the US and the UK were selected to highlight the particular strengths and weaknesses of drawing general inferences from large-scale social data and how they can be reconceptualised by SCA (Johnston et al. 2016; Kann et al. 2016; University of London 2017). Further, I tackle the problem of significant-but-minimal effects in large-scale social data by using the abundance of questions in each dataset to compute comparison specifications; I directly compare the effects of digital technology to the effects of other activities on psychological well-being (e.g. sleep, eating breakfast, illicit drug use), using extant literatures and psychological theory as a guide. This allows me to simultaneously examine the impact of adolescent technology use against real-world benchmarks while modelling and accounting for analytic flexibility.

2.2 Methods

2.2.1 Datasets and Participants

This chapter’s analysis pipeline spans three nationally-representative datasets from the US and the UK (Johnston et al. 2016; Kann et al. 2016; University of London 2017), encompassing a total of 355,358, predominately twelve- to eighteen-year-old, adolescents surveyed between the years of 2007 and 2016. These datasets were selected because they feature measures of adolescents’ psychological well-being, digital technology use, and have been the focus of secondary data analysis to study digital technology effects [Parkes et al. (2013); Twenge, Joiner, et al. (2017); Twenge, Martin, et al. (2018); Kelly2019]. Furthermore, the American datasets were analysed in a paper that I evaluated in a blogpost at publication, and which initially motivated me to do this research (Twenge, Joiner, et al. 2017; Orben 2017).

Two datasets are based on samples collected in the United States. The first, the Youth Risk and Behaviour Survey (YRBS, Kann et al. 2016) launched in 1990, is a biennial survey of adolescents that reflects a nationally-representative sample of students attending secondary schools in the US (years 9-12). The resulting sample from the YRBS was collected from 2007 to 2015 and included 37,402 girls and 37,412 boys, ranging in age from “12 years or younger” to “18 years or older” (*median* = 16, *s.d.* = 1.24). The second US dataset, Monitoring the Future (MTF, Johnston et al. 2016), launched in 1975 and is an annual nationally-representative survey of approximately 50,000 American adolescents in Grades 8, 10 and 12. While the survey includes adolescents in Grade 12, many of the key items of interest cannot be correlated in their survey, and therefore their data were not included in my analysis. The resulting sample from the MTF was collected from 2008 to 2016, and included 136,190 girls and 132,482 boys, though the exact age of individual respondents was removed from the dataset by study coordinators during anonymization.

The UK dataset under analysis was the Millennium Cohort Study (MCS, University of London 2017), a prospective study collected in the UK; it follows a specific cohort of children born between September 2000 and January 2001.

These data were not previously analysed by the studies that motivated this work, but were added because I see these data as particularly high in quality. This is mostly due to its inclusion of pre-tested measures and extensive documentation, highlighting good data collection and project management practices. The dataset has an over-representation of minority groups and disadvantaged areas due to clustered stratified sampling. Data in this sample is provided by caregivers as well as adolescent participants. In my analysis, I only include data from the primary caregivers and adolescent respondents. The sample under analysis from the MCS was comprised of 5,926 girls and 5,946 boys who ranged in age from 13 to 15 ($m = 13.77$, $s.d. = .45$) and 10,605 primary caregivers.

While the omnibus sample of adolescents is 355,358 teenagers in total, it is important to note that the sample sizes of the analyses are often smaller, in some cases by an order of magnitude or more. This is due to missing values, but also because in questionnaires like MTF teenagers only answered a subset of questions. More information about what questions were asked together in MTF can be found on the Open Science Framework (OSF; Table 1, <http://dx.doi.org/10.17605/OSF.IO/87QGM>)

2.2.2 Ethical Review

Ethical review and approval for data collection for YRBS was conducted and granted by the CDC Institutional Review Board. The University of Michigan Institutional Review Board oversees MTF. Ethical review and approval for the MCS is monitored by the UK National Health Service, London, Northern, Yorkshire and South-West Research Ethics Committees.

2.2.3 Measures

This study focuses on measures of both digital technology use and psychological well-being. Prior to performing the analysis, all three datasets were reviewed, noting the variables of theoretical interest in each with respect to human behaviour and effects of technology engagement. Some questions have been modified with successive waves

of data collection. In all cases these changes are relatively minor and are noted in the OSF repository (Table 2, <http://dx.doi.org/10.17605/OSF.IO/87QGM>). In my ongoing analyses I use the questionnaires in many different constellations and therefore refrain from including reliability measurements.

Criterion Variables: Adolescent Well-Being

All datasets contained a wide range of different questions that concern the adolescents' psychological well-being and functioning. I reversed select measures so that they are all in the same direction, with higher scores indicating higher well-being.

Adolescents were asked five items related to mental health and suicidal ideation in the YRBS. Three of the items were on a yes-no scale: *"During the past 12 months did you ever:"*, *"feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities"*, *"seriously consider attempting suicide"*, *"make a plan about how you would attempt suicide"*. There was also one measure asking *"During the 12 months, how many times did you actually attempt suicide?"* (1 = *"0 times"* to 5 = *"6 or more times"*). I recoded this measure to 1 = 0 times and 0 = 1 or more times. Furthermore, there was one question about *"if you attempted suicide during the past 12 months, did any attempts result in an injury, poisoning, or overdose that had to be treated by a doctor or nurse?"*, (1 = *"I did not attempt suicide during the past 12 months"*, 2 = *"Yes"*, 3 = *"No"*). I recoded this measure so 1 = did not attempt suicide or did not need to be treated by doctor or nurse, 0 = needed to be treated by doctor or nurse.

In MTF, participants were asked one of two subsets of self-report questions. The first tranche of participants was asked thirteen questions about their mental health: twelve measures uniquely asked to this subset, and one measure completed by all participants in the survey. The twelve items asked only to this subset included a four-item depressive symptoms scale which studies state to be "similar to those on the Center for Epidemiologic Studies Depression Scale (Maslowsky et al. 2014): five-point scale (1 = *"disagree"* to 5 = *"agree"*), *"Life often seems meaningless"* (reverse coded), *"I enjoy life as much as anyone"*, *"The future often seems hopeless"*

(reverse coded), *“It feels good to be alive”*. The study also included the Rosenberg Self-Esteem scale which was presented on a similar five-point scale (Maslowsky et al. 2014): *“I take a positive attitude towards myself”*, *“Sometimes I think that I am no good at all”* (reverse coded), *“I feel I am a person of worth, on an equal plane with others”*, *“I am able to do things as well as most other people”* and *“I feel I do not have much to be proud of”* (reverse coded). There were, furthermore, two additional negative self-esteem items added by the MTF survey administrators to ensure there were an equal amount of positively and negatively coded self-esteem items: *“I feel that I can’t do anything right”* (reverse coded), *“I feel that my life is not very useful”* (reverse coded). I do not include *“Sometimes I am no good at all”* as a measure in my SCA, as it could not be clearly attributed to a questionnaire. Participants of all subsets were additionally asked a single life satisfaction measure: *“Taken all things together, how would you say things are these days – would you say you’re very happy (3, recoded to 2), pretty happy (2, recoded to 1) or not to happy these days (1, recoded to 0)?”*

There are two kinds of psychological well-being indicators included in the MCS: (1) those filled out by the cohort members, and (2) those completed by their primary caregivers. To measure wellbeing, MCS participants were asked how they felt about *“your school work”*, *“the way you look”*, *“your family”*, *“your friends”*, *“the school you go to”*, *“your life as a whole”* (1 = completely happy, 7 = not at all happy, scale subsequently reversed). They were also asked five items from an abbreviated self-esteem measure (Robins et al. 2001): *“On the whole, I am satisfied with myself”*, *“I feel I have a number of good qualities”*, *“I am able to do things as well as most other people”*, *“I am a person of value”*, *“I feel good about myself”* (1 = strongly agree to 4 = strongly disagree, scale subsequently reversed). Finally, they were asked twelve questions tapping subjective affective states and general mood (Angold et al. 1995): *“For each question please select the answer which reflects how you have been feeling or acting in the past two weeks:”*, *“I felt miserable or unhappy”*, *“I didn’t enjoy anything at all”*, *“I felt so tired I just sat around and did nothing”*, *“I was very restless”*, *“I felt I was no good any more”*, *“I cried a lot”*, *“I found*

Table 2.1: Items of the Strengths and Difficulties Questionnaire

Question	Not True	Somewhat True	Certainly True
Considerate of other people's feelings			
Restless, overactive, cannot stay still for long			
Often complains of headaches, stomach-aches or sickness			
Shares readily with other children (treats, toys, pencils etc.)			
Often has temper tantrums or hot tempers			
Rather solitary, tends to play alone			
Generally obedient, usually does what adults request			
Many worries, often seems worried			
Helpful if someone is hurt, upset or feeling ill			
Constantly fidgeting or squirming			
Has at least one good friend			
Often fights with other children or bullies them			
Often unhappy, down-hearted or tearful			
Generally like by other children			
Easily distracted, concentration wanders			
Nervous or clingy in new situations, easily loses confidence			
Kind to younger children			
Often lies or cheats			
Picked on or bullied by other children			
Often volunteers to help others (parents, teachers, other children)			
Thinks things out before acting			
Steals from home, school or elsewhere			
Gets on better with adults than with other children			
Many fears, easily scared			
Sees tasks through to the end, good attention span			

it hard to think properly or concentrate”, “I hated myself”, “I was a bad person”, “I felt lonely”, “I thought nobody really loved me”, “I thought I could never be as good as other kids” and “I did everything wrong” (1 = not true, 2 = sometimes, 3 = true, scale subsequently reversed).

Primary caregivers completed the Strengths and Difficulties Questionnaire (SDQ) (Goodman et al. 2000), a well-validated measure of psychosocial functioning, for each adolescent cohort member they took care of (Table 2.1). The SDQ has been used extensively in school, home, and clinical settings with adolescents from a wide range of social, ethnic, and national backgrounds (Desai et al. 1989). It includes 25 questions, five each about prosocial behaviour, hyperactivity or inattention, emotional symptoms, conduct problems and peer relationship problems.

Explanatory variables: Adolescent Technology Use

The YRBS dataset included two seven-point technology use questions. One YRBS technology use question was about the use of electronic devices, *“On an average school day, how many hours do you play video or computer games or use a computer for something that is not school work? (Count time spent on things such as Xbox, PlayStation, an iPad or other tablet, a smartphone, texting, YouTube, Instagram, Facebook or other social media)”*, (1 = *“I do not play video or computer games or use a computer for something that is not school work”* to 7 = *“5 or more hours per day”*). The other question asked *“on an average school day, how many hours do you watch TV?”*, (1 = *“I do not watch TV on an average school day”* to 7 = *“5 or more hours per day”*).

The MTF asked a variety of technology use questions. As the questionnaire was split into six parts (with each participant only filling in one part), some questions were filled out by one subset of adolescents, while other questions were filled out by another. One subset answered questions about frequency of social media use and getting information about news from the internet (five-point scale): *“How often do you do each of the following? Visit social networking websites (like Facebook)”* and *“How often do you use each of the following to get information about news and current events? The internet.”*. Both questions were coded on a five-point scale (1 = never to 5 = almost every day). Furthermore, these participants were asked two seven-point questions about frequency of watching TV on the weekend and weekday: *“How much TV do you estimate you watch on an average weekend/weekday?”*, seven-point scale (1 = none to 7 = nine or more hours). Another group of MTF participants were asked seven hourly measures of technology use on a nine-point scale (1 = none to 9 = 40 hours or more). The questions asked about using the internet, playing electronic games, texting on a cell phone, calling on a cell phone, using social media, video chatting and using computers for school. There are, therefore, a total of eleven technology use measures that can be used when analysing the MTF dataset.

In the MCS, the participants were asked five questions concerning technology use. The MCS included four eight-point questions about hours per weekday spent “*watching television programmes or films*”, spent “*playing electronic games on a computer or games systems*”, “*spent using the internet*” at home and spent “*on social networking or messaging sites or Apps on the internet*” (1 = none to 8 = 7 hours or more). There was also one yes-no measure about whether participants own a computer: “*do you have a computer . . . of your own?*”.

Control variables

Mirroring previous studies analysing data from the MCS (University of London 2017), I included sociodemographic factors and maternal characteristics as control variables (i.e. covariates) in my analyses. These include mother’s ethnicity, education, employment and psychological distress (using the K6 Kessler Scale) which have previously been found to influence child well-being in studies analysing large-scale data (Desai et al. 1989; Kiernan and Mensah 2009), including analyses of the MCS (Mensah and Kiernan 2010). I also included equivalised household income, whether the biological father is present and number of adolescent’s siblings in the household, as these household factors have also been found to affect adolescent well-being (Kiernan and Mensah 2009). Furthermore, I included parental behavioural factors such as closeness to parents and the amount of time the primary caretaker spends with the adolescent (Thomson et al. 1994; The Children’s Society and Barnardo’s 2018). Addressing previous reports of their influence on child well-being, I additionally used parent reports of any adolescent’s long-term illness, and the adolescent’s own negative attitudes towards school as control variables (Cadman et al. 1987; The Children’s Society and Barnardo’s 2018). Finally, I included the primary caretaker’s word activity score as a measure of current cognitive ability, to control for other environmental factors that could influence child well-being (Parkes et al. 2013).

For YRBS and MTF I included all variables part of the respective questionnaires that conceptually mirrored those control variables utilized in the MCS. For YRBS I included the adolescent’s ethnicity. For MTF I included ethnicity, number of

siblings, mother’s education level, whether the mother has a job, the adolescent’s enjoyment of school, predicted school grade and whether they feel like they can talk with their parents about problems.

2.2.4 Analytic Approach

SCA was developed by Simonsohn and colleagues to find a (partial) solution for the problem that *“Empirical results often hinge on data analytic decisions that are simultaneously defensible, arbitrary, and motivated”* (2015). The method is closely related to the concept of Multiverse Analysis which was introduced at a similar time by Sara Steegen, Andrew Gelman and colleagues (Steegen et al. 2016). In their initial paper, Simonsohn and colleagues focused on a study published in Proceedings of the National Academy of Sciences (PNAS) that put forth the argument that hurricanes with feminine names cause more deaths, mainly due to the fact that they are perceived as less dangerous (Jung et al. 2014). While this paper reported one analytical method, there were many ways in which the study’s data could have been analysed. This was down to analytical decisions (i.e. researcher degrees of freedom) including what storms to include, how to operationalise whether hurricane names are feminine or not, what control variables to use, what type of regression to use and what effects to examine. Instead of analysing only one model, which incorporates only one choice for each of these decisions, a SCA analyses all theoretically defensible ‘specifications’: with a specification being the unique combination of theoretically defensible analytical choices that could be used to answer the research question of interest.

Naturally the inclusion of specifications in a SCA analysis – while as transparent as possible – is still influenced by biases, prior experiences and computational limits. SCA is therefore no panacea for the problem of researcher degrees of freedom, but it represents an improvement on the often intransparent reporting of how specific – conscious or unconscious – analytical decisions taken during data analysis can help researchers obtain results that agree with their hypotheses (Gelman and Loken 2014). *“In sum, specification-curve is an imperfect solution*

to the problem of selective reporting, but it is less imperfect than the alternatives we are aware of.” (Simonsohn et al. 2015)

In their paper, Simonsohn and colleagues report 1,728 reasonable specifications for the PNAS paper and analyse all of them. They find that most of the specifications have results of the sign that the original authors found (female names cause more deaths). They, however, also find that some effects were in the opposite direction and that only few were statistically significant overall. After plotting the results it was then possible to test whether, when taking into account all the possible specifications, the results found are inconsistent with results expected when the null hypothesis is true. To do so, the authors used a permutation technique to simulate datasets where the null hypothesis is known to be true. I note that this permutation technique can be used when analysing experimental data, but not when testing correlational data, like the data presented in this thesis. Using this method, Simonsohn and colleagues find that the original data were not significantly different from data simulated where there is no connection between hurricane femininity and deadliness. Therefore, their analysis paints a very different picture than the original study.

This chapter implements a SCA examining the correlation between my explanatory (digital technology engagement) and criterion variables (psychological well-being) using the three-step SCA approach outlined by Simonsohn and colleagues (2015) and applied in a recent paper by Rohrer and colleagues (2017). I also add a fourth step in order to aid interpretability of my results in the context of large-scale data.

I. Identifying Specifications

The first step taken was to identify all the possible analysis pathways that could be used to relate technology use and adolescent well-being. Due to the complexity of the original data, I decided to use simple linear regression modelling to draw inferences about technology associations. This left three key analytical decisions: (1) How to measure well-being, (2) How to measure technology use, and (3) How to include control variables.

There are a wide variety of questions and questionnaires relating to well-being in each dataset. Many of these items, even if partitioned questionnaires reflecting a specific construct, have been selectively reported over the years. It is noteworthy that researchers have not been consistent and have instead engaged in picking and choosing within and between questionnaires (see Figure 2.1). These analytic decisions have produced many different possibilities for combining and analysing such measures, making the pre-specified constructs more of an accessory for publication than a guide for analysis. Any combination of the mental health indicators is therefore included in this chapter's SCA: The measures by themselves, the mean of the measures in pairs of two, the mean of the measures in threes etc. up to the mean of all measures. The Appendix additionally presents an SCA which includes only pre-specified well-being questionnaires for MCS (Figure A.1) and I take this more conservative approach in my other chapters. However, I decided against using only pre-specified well-being questionnaires in this chapter as it would not allow for comparisons of my SCAs to results of previous work that has selectively combined questions from various questionnaires (Twenge, Joiner, et al. 2017).

For MCS, I also included a decision of whether to use well-being questions answered by cohort members or those answered by their caregivers. I do not combine the two. For YRBS, I included an additional analytical decision of whether to take the mean of the five dichotomous well-being measures, or whether to code each participant as "1" who answered yes to one or more of the questions, as this has been done in previous analyses of the data (Twenge, Joiner, et al. 2017).

The next analytical decision is what technology use variables to include, where I include all questions concerning technology use in the questionnaires, and their mean, as done by previous studies (Twenge, Joiner, et al. 2017). The last analytical decision taken is whether or not to include control variables in the models. Because of the sheer size of these datasets, there is a combinatorial explosion of different control variable combinations that could be used in each regression. I therefore analysed regressions either without control variables or with a pre-specified set

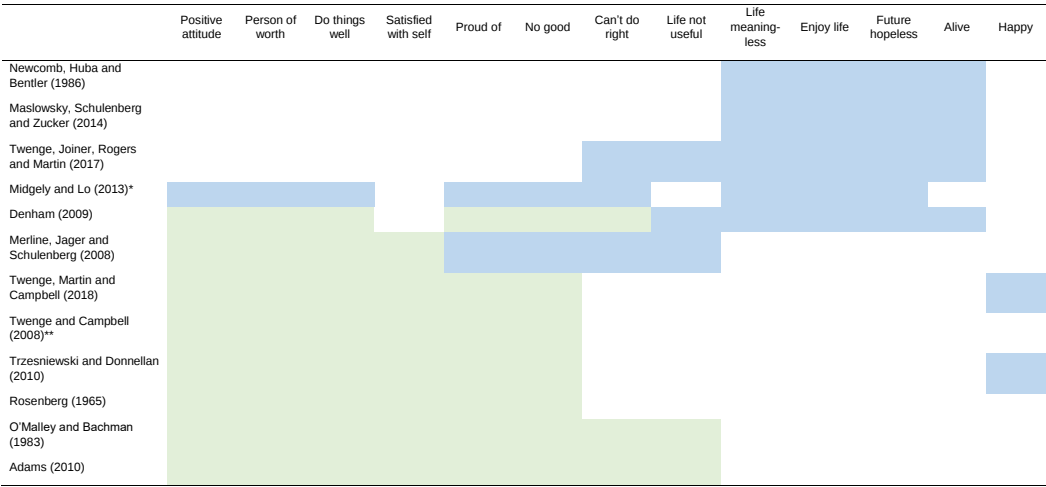


Figure 2.1: Figure showing how research papers have used different combinations of MTF measures to define depressive symptoms (blue) and self-esteem (green). This illustrates the abundance of analytical flexibility in this area. We also include Newcomb, Huba and Bentler (1986) and Rosenberg (1965) who originally devised parts of the scales.: * Note: This study split the self-esteem measures into two different scales of self-liking and self-competence. ** Note: This study measured Emotional Health and also included one item about doing things other people think is strange.

of control variables based on a literature review concerning child well-being and digital technology use (Parkes et al. 2013).

When examining the distributions of the data, many of the variables are highly skewed (e.g. the 5-item technology use measures in MTF) or questionably linear (e.g. 3-item happiness measure in MTF). I opted to treat these variables as continuous so that my analyses and results would be directly comparable with those of previous studies (Twenge, Joiner, et al. 2017; Twenge, Martin, et al. 2018). Data distribution was assumed to be normal throughout the analysis but is not formally tested for each specification.

II. Implementing Specifications

Next, for each specification defined I ran the appropriate regression, and noted the standardised β of technology uses' association with psychological well-being, the corresponding two-sided p value and the partial η^2 calculated using the R *heplots*

package. Listwise deletion for missing data was used as this was more efficient in terms of computational time. This assumes that data are missing completely at random, which could easily not be the case. For example, a child's health, academic performance or socioeconomic background could change its probability of completing the questionnaire fully and is likely to bias estimates. It is therefore important to note that this is a potential source of bias, possibly changing the nature or strength of associations found.

To make the results easily interpretable, the specifications were ranked and plotted in terms of ascending standardised β (e.g. in Figure 2.2). The median standardised β of all the possible specifications provides a general overview of the effect size, and is indicated by a dashed horizontal line in the figures. Each dot on the top part of the graph represents a result (β) of a different specification shown on the x-axis. Tracing down vertically from each dot you can read the bottom 'dashboard' plot. Each row represents a different analytical decision and if a dot is present in that row the specification which produced the result above incorporated that analytical decision. The dashboard therefore allows me to visualise what analytical decisions influence the results of the SCA.

III. Statistical Inferences

It is then possible to test whether, when considering all the possible specifications, the results found are inconsistent with results when the null hypothesis is true (i.e. that technology use and adolescent well-being are unrelated). To do so, a bootstrapping technique put forth by Simonsohn and colleagues (2015) was implemented, creating data where the null hypothesis is true by forcing the null on the data. This technique is more complex than the permutation technique Simonsohn and colleagues used to analyse their experimental data (2015). To create my null dataset, the β -coefficient of the variable of interest from the full regression model multiplied by the x-variable (technology use) was subtracted from the y-variable (well-being). This created a new set of data points that were then used as the new y-variable, creating datasets where the null hypothesis was known to be true. Participants were then drawn at random

– with replacement – from this null dataset, creating bootstrapped null samples on which a new SCA model is run. This was done 500 times. Once I obtained 500 bootstrapped SCAs, where I knew the null hypothesis was true, I examined whether the median effect size in the original SCA was significantly different to the median effect size in the bootstrapped SCAs. To do so, I divided the number of bootstrapped datasets that have larger median effect sizes than the original SCA by the total number of bootstraps to find the p value of this test. I repeat this test focusing also on the share of results with the dominant sign, and also the share of statistically significant results with the dominant sign (Simonsohn et al. 2015).

IV. Comparison Specifications

Lastly, these analyses were supplemented with a comparison specifications section, putting into context the effects found in the SCA. To do so, I performed a literature review to choose four variables in each dataset that should be positively correlated with psychological well-being, four variables that should be negatively correlated with psychological well-being and four that should have no or little association with psychological well-being. A SCA was run relating each of the variables chosen, and the mean of the technology use variables present in the dataset, with adolescent well-being. These methods provide a way for researchers to transparently, openly and robustly analyse large-scale datasets to produce research that accurately depicts associations found in the data for both the academy and the public.

2.2.5 Code Availability Statement

Intermediate analysis files and a live version of the analysis code can be found on the OSF (<http://dx.doi.org/10.17605/OSF.IO/PHF8V>).

2.2.6 Data Availability Statement

The data that support the findings of this study are available from the Centre for Disease Control and Prevention (YRBS), Monitoring the Future (MTF) and the UK Data Service (MCS) but restrictions apply to the availability of these data, which

were used under license for the current study, and so are not publicly available. Data are however available from the relevant third-party repository after agreement to their terms of usage. Information about data collection and questionnaires can be found on the OSF (<http://dx.doi.org/10.17605/OSF.IO/PHF8V>).

2.3 Results

2.3.1 I. Identifying Specifications

I first identified the main analytical decisions that needed to be taken when regressing digital technology use on adolescents' psychological well-being in each dataset. In particular I included the following analytical decisions in the analysis:

- YRBS
 - *Operationalizing adolescent well-being*: Mean of any possible combination of five items concerning mental health and suicidal ideation
 - *Operationalizing technology use*: Two questions concerning electronic device use and TV use, or the mean of these questions
 - *Which control variables to include*: Either include control variables or not
 - *Other specifications*: Either take the mean of the dichotomous well-being measures, or code all cohort members who answered 'yes' to one or more as '1' and all others as '0'
- MTF
 - *Operationalizing adolescent well-being*: Mean of any possible combination of 13 items concerning depression, happiness and self-esteem
 - *Operationalizing technology use*: Eleven technology use measures concerning the internet, electronic games, mobile phone use, social media use and computer use, or the mean of all these questions
 - *Which control variables to include*: Either include control variables or not

- MCS
 - *Operationalizing adolescent well-being*: Mean of any possible combination of 24 questions concerning well-being, self-esteem and feelings (completed by cohort members), or mean of any possible combination of 25 questions from the Strengths and Difficulties Questionnaire (completed by caregivers)
 - *Operationalizing technology use*: Five questions concerning TV use, electronic games, social media use, owning a computer and using the internet at home, or the mean of all these questions
 - *Which control variables to include*: Either include control variables or not
 - *Other specifications*: Use well-being measures completed by cohort members or those completed by their caregivers

372 justifiable specifications for YRBS, 40,966 plausible specifications for MTF, and a total of 603,979,752 defensible specifications for MCS were identified. Although more than 600 million specifications might seem high, this number is best understood in relation to the total possible iterations of dependent (6 analysis options) and independent variables ($22^4 + 22^5 - 2$ analysis options) and whether control variables are included or not (2 analysis options). The number rises even higher to 2.5 trillion specifications for MCS if any combination of control variables (2^{12} analysis options) is included. Given this, and to reduce computational time, I selected 20,004 specifications for MCS. To do so, I included specifications of all well-being items by themselves (24 items for cohort members and 25 for caregivers), the mean of previously used questionnaires for the cohort members (3 questionnaires: self-esteem, mood and depression), and any combinations of measures found in the previous literature for the SDQ filled out by caregivers (8 variants: conduct problems, hyperactivity, peer relationship problems, prosociality, emotional problems, emotional problems and peer relationship problems, conduct problems and hyperactivity, total). I then supplemented these specifications with other randomly selected combinations of the well-being measures (806 for cohort members and 801 for caregivers).

Table 2.2: Results of Specification Curve Analysis for the Youth Risk and Behaviour Survey (YRBS, United States), Monitoring the Future Survey (MTF, United States) and Millennium Cohort Study (MCS, United Kingdom), both overall and for different technology use variables, parent/adolescent self-report or with/without control variables.

Dataset	Median beta of Specification Curve Analysis	Median partial eta-squared of Specification Curve Analysis	Median n	Median Standard Error
YRBS				
Complete Specification Curve Analysis	-0.035	0.001	62297	0.004
Electronic Device Use Only	-0.071	0.005	62368	0.004
TV Use Only	-0.012	<.001	62352	0.004
With Control Variables Only	-0.034	0.001	61525	0.004
Without Control Variables Only	-0.035	0.001	62638	0.004
MTF				
Complete Specification Curve Analysis	-0.005	<.001	78267	0.003
Social Media Use Only	-0.031	0.001	102963	0.003
TV Viewing On Weekend Only	0.008	0.001	115738	0.003
Using Internet for News Only	-0.002	<.001	115580	0.003
TV Viewing on Weekday Only	0.002	<.001	115783	0.003
With Control Variables Only	0.001	<.001	72525	0.003
Without Control Variables Only	-0.013	<.001	117560	0.003
MCS				
Complete Specification Curve Analysis	-0.032	0.004	7968	0.010
Own a Computer Only	-0.003	0.011	7973	0.010
Weekday Electronic Games Only	0.013	<.001	7977	0.010
Hours of Social Media Use Only	-0.056	0.009	7972	0.010
TV Viewing on Weekday Only	-0.043	0.003	7971	0.010
Use of Internet of Home Only	-0.07	0.006	7975	0.010
Caregiver-Report Well-Being Only	<.001	0.003	7893	0.010
Adolescent-Report Well-Being Only	-0.046	0.008	8857	0.010
With Control Variables Only	-0.005	0.001	6566	0.011
Without Control Variables Only	-0.068	0.005	11018	0.010

2.3.2 II. Implementing Specifications

After noting down all specifications, the result of every possible combination of these specifications was computed for each dataset. The standardised β coefficient for technology uses' association with well-being was then plotted for each specification. The number of participants analysed for each specification can be found in Figure A.2-A.4, the median standardised β , n , partial η^2 and standard error can be found in Table 2.2. For YRBS, the median association of technology use with adolescent well-being was $\beta = -.035$ (median partial $\eta^2 = .001$, median $n = 62,297$, median standard error = .004, see Figure 2.2). From the figure one can discern the analytical choices that influence the size of this effect. When using electronic device use as the independent variable in the model, the effects were more negative (median β

= -.071, median partial $\eta^2 = .005$, median $n = 62,368$, median standard error = .004), while when including TV use in the model the effects were less negative and sometimes become non-significant (median $\beta = -.012$, median partial $\eta^2 < .001$, median $n = 62,352$, median standard error = .004). Even though YRBS does not have high quality control variables, including them yielded a smaller effect size for the relations of interest (*controls*: median $\beta = -.034$, median partial $\eta^2 = .001$, median $n = 61,525$, median standard error = .004; *no controls*: median $\beta = -.035$, median partial $\eta^2 = .001$, median $n = 62,638$, median standard error = .004).

For the MTF data, a median standardised β of -.005 was observed (median partial $\eta^2 < .001$, median $n = 78,267$, median standard error = .003), a value which fell into the non-significant range of the justifiable specifications (see Figure 2.3). This result was surprising, as MTF had the highest number of observations, making it difficult for even small associations to be flagged as non-significant using traditional α thresholds (i.e., $p < .05$). In Figure 2.3, and my bootstrapping test, I do not include the few specifications of the participants that only filled in one well-being measure (to see the SCA of all participants, see Figure A.5). From Figure 2.3 it is again possible to discern that even controls of lower standard made the association either less negative or even positive (no controls: median $\beta = -.013$, median partial $\eta^2 < .001$, median $n = 117,560$, median standard error = .003; controls: median $\beta = .001$, median partial $\eta^2 < .001$, median $n = 72,525$, median standard error = .003). TV viewing on the weekend had a median positive association with well-being of $\beta = .008$ (median partial $\eta^2 = .001$, median $n = 115,738$, median standard error = .003), while social media use had a median negative association with well-being of $\beta = -.031$ (median partial $\eta^2 = .001$, median $n = 102,963$, median standard error = .003), though the effect was small suggesting that technology use operationalised in these terms accounts for less than 0.1% of the observed variability in well-being. Using the internet for news and TV viewing on a weekday showed mainly very small median associations, $\beta = -.002$ (median partial $\eta^2 < .001$, median $n = 115,580$, median standard error = .003) and $\beta = .002$ (median partial $\eta^2 < .001$, median $n = 115,783$, median standard error = .003) respectively.

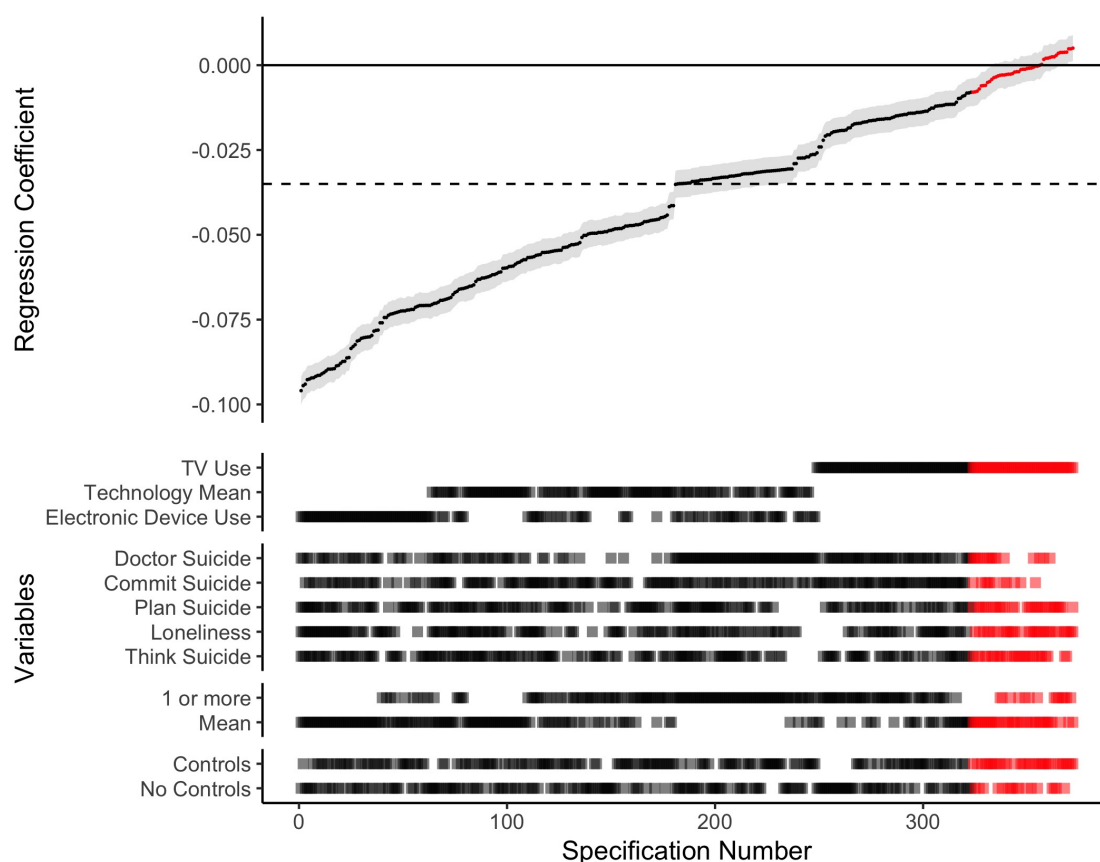


Figure 2.2: Results of Specification Curve Analysis of the Youth Risk and Behaviour Survey: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the x-axis represents a different combination of analytical decisions, which are displayed in the ‘dashboard’ at the bottom of the graph. The resulting standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: $\beta = -.035$ (median partial $\eta^2 = .001$, median $n = 62,297$, median standard error = .004)

As previous studies have addressed the association between technology use and well-being using the same dataset (Twenge, Joiner, et al. 2017), I include Figure 2.4 which shows how these study’s specifications influence their reported results.

Lastly, results from MCS, the highest quality dataset I examined, were interesting because the literature provided us with control variables based on extant theory (Parkes et al. 2013) and convergent data from adolescent and caregiver reports. In these data I found a median β of technology use’s association with wellbeing of $\beta = -.032$ (median partial $\eta^2 = .004$, median $n = 7,968$, median standard error =

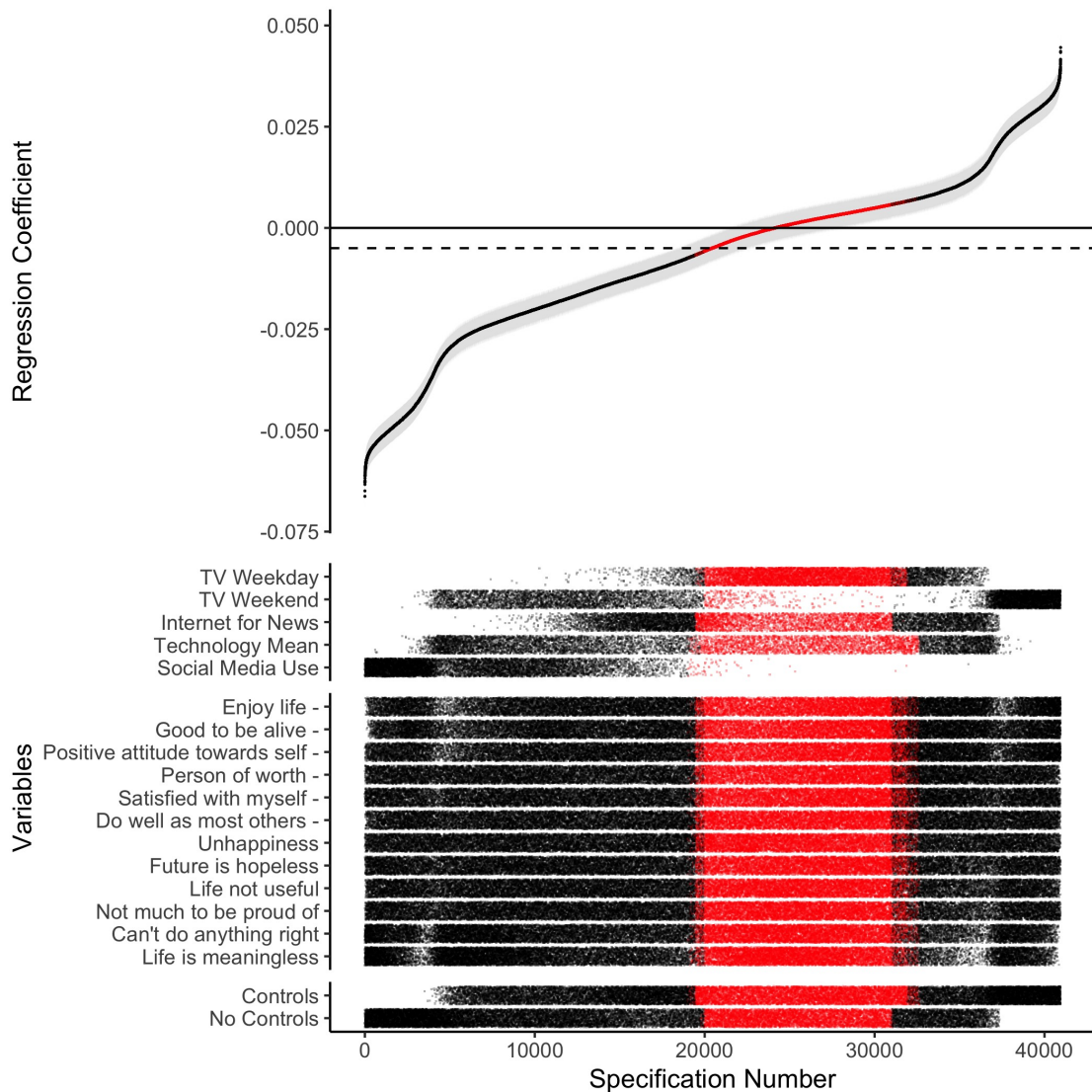


Figure 2.3: Results of Specification Curve Analysis of the Monitoring the Future study: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the x-axis represents a different combination of analytical decisions, which are displayed in the ‘dashboard’ at the bottom of the graph. The resulting standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: $\beta = -.005$ (partial $\eta^2 < .001$, median $n = 78,267$, median standard error = .003)

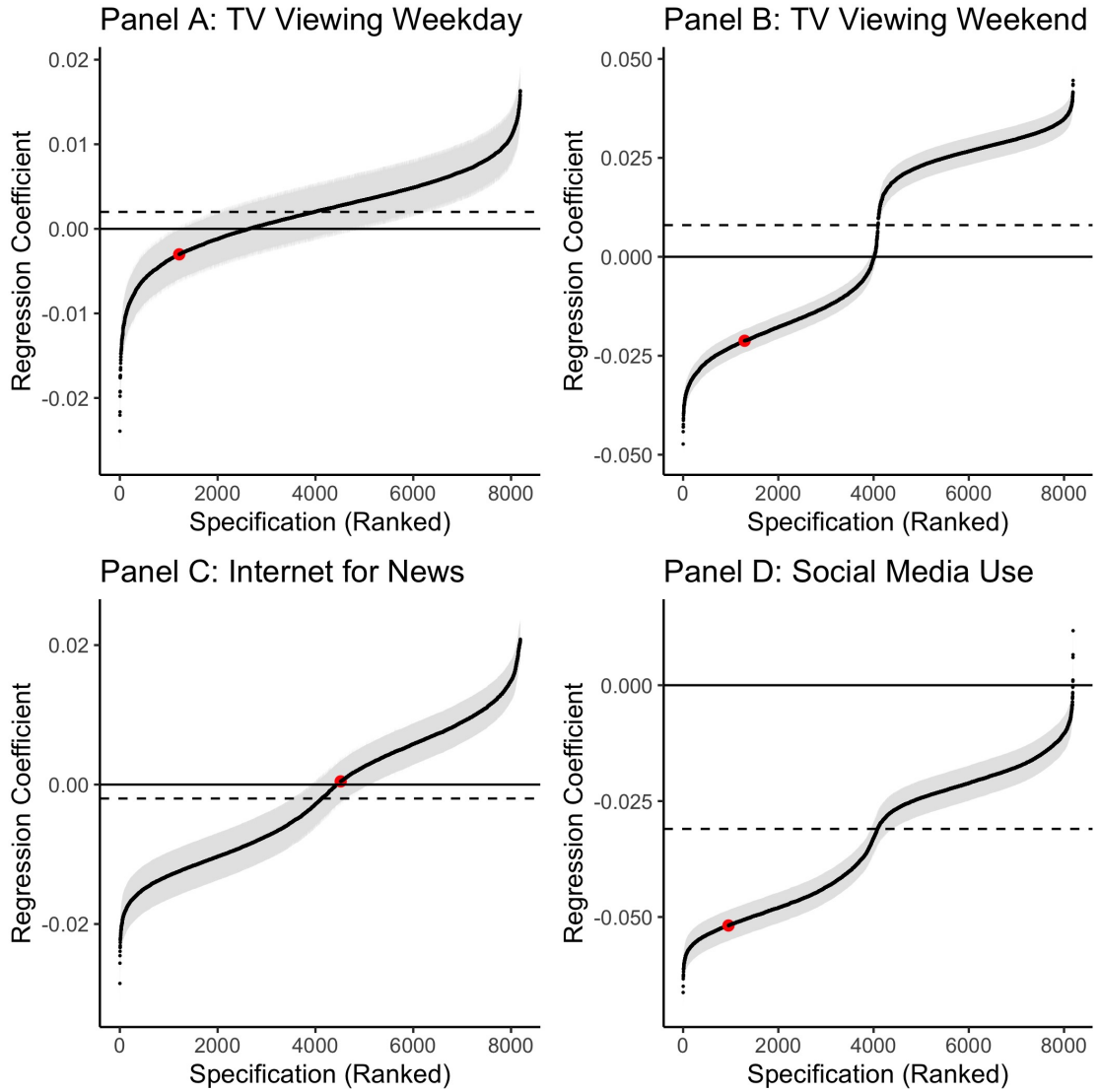


Figure 2.4: Each panel of this figure illustrates the range of possible results found using a SCA regressing a specific technology use variable onto well-being in the MTF dataset: Panel A examines TV viewing on a weekday, Panel B examines TV viewing on a weekend, Panel C examines using the internet to get news and Panel D examines social media use. The error bars represent the standard error and the dotted line shows the median standardised regression coefficient. I highlight in red the specifications chosen by Twenge and colleagues (2017) for their analysis of the MTF dataset. The researchers used a novel combination of measures from both well-being and self-esteem scales to create a well-being scale in their study, see also Figure 2.1. They correlated this measure with technology use measures (TV viewing, using the internet for news and social media use) and included either no controls or range of controls variables. We only show Twenge et al.’s results with no controls in this visualisation due to their chosen controls being difficult to reproduce computationally. We note that our results from implementing Twenge and colleague’s specification for TV viewing are not consonant with those reported in Twenge et al. (2017). For TV viewing on a weekday the specifications chosen in Twenge et al. (2017) were at the 15th percentile of effect sizes, for TV viewing on a weekend they were at the 16th percentile. For getting news over the internet, the chosen specifications were at the 55rd percentile and for social media use the chosen values were at the 12th percentile.

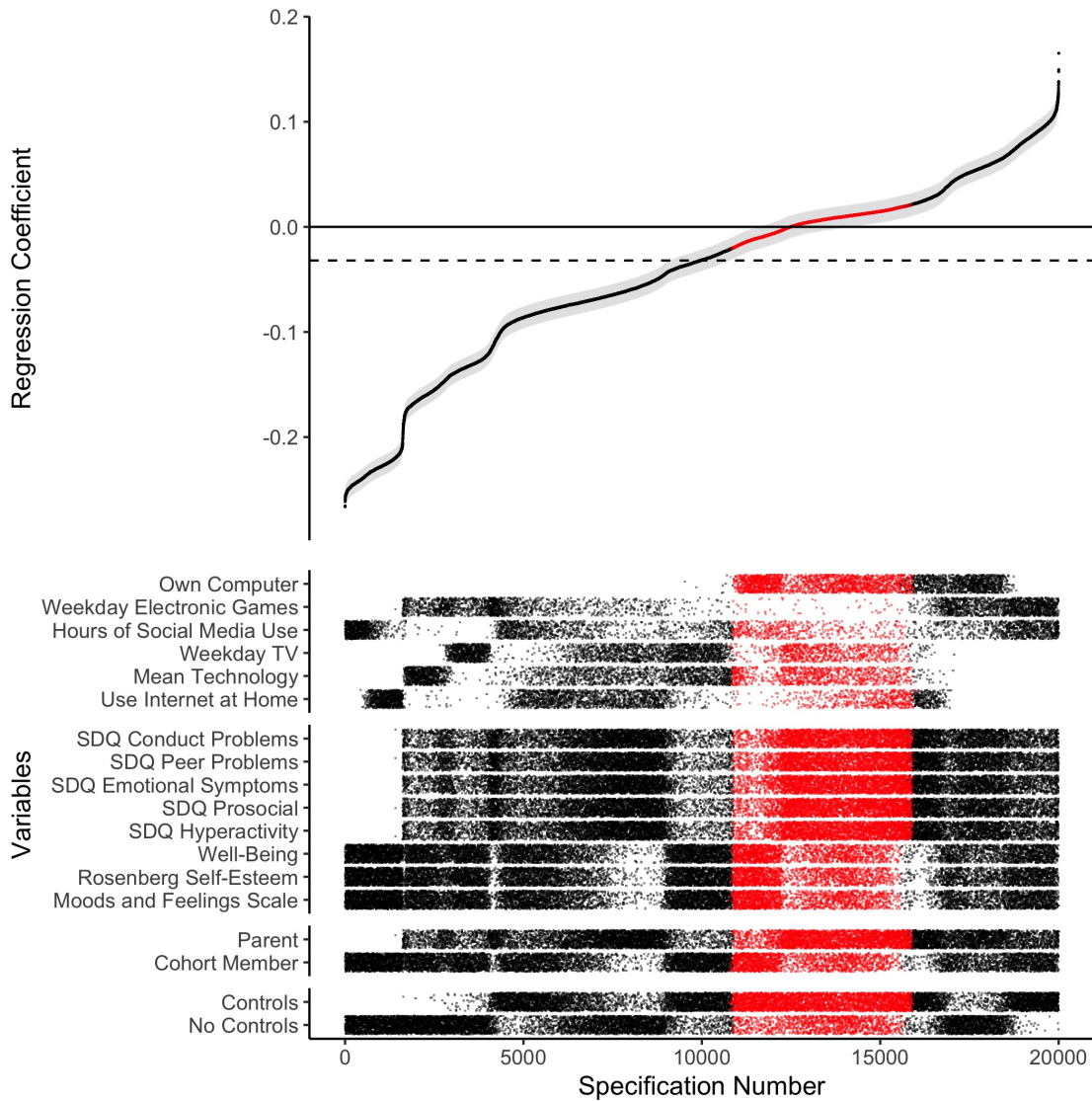


Figure 2.5: Results of Specification Curve Analysis of the Millennium Cohort Study: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the x-axis represents a different combination of analytical decisions, which are displayed in the ‘dashboard’ at the bottom of the graph. The resulting standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: $\beta = -.032$ (partial $\eta^2 = .004$, median $n = 7,968$, median standard error = .010)

.010, see Figure 2.5). Across the board, if using well-being measures completed by the caregivers, the median association was less negative or more positive (median $\beta < .001$, median partial $\eta^2 = .003$, median $n = 7,893$, median standard error = .010), while the opposite was in evidence when considering well-being measures completed by the adolescent (median $\beta = -.046$, median partial $\eta^2 = .008$, median $n = 8,857$, median standard error = .010). This pattern of shared covariation speaks to the idea that correlations between technology use and well-being might be rooted in common method variance, as one single informant fills out well-being and technology measures and the association might be driven by other common factors.

To further address the importance of control variables, I plot separate specification curves for MCS analyses with and without control variables (see Figure 2.6). The association for the uncorrected models had a median β of -.068 (median partial $\eta^2 = .005$, median $n = 11,018$, median standard error = .010). In contrast, the corrected models only found a median β of technology use regressed on wellbeing of -.005 (median partial $\eta^2 = .001$, median $n = 6,566$, median standard error = .011). Additional SCAs using only pre-specified questionnaires are presented in Figure A.1.

2.3.3 III. Statistical Inferences

The SCAs showed that there is a small negative association between technology use and well-being, but it is not possible to make many statistical inferences because the specifications are not part of the same model and are therefore not independent. A bootstrapping technique was therefore used to run 500 SCA tests on resampled data, where it is known that the null hypothesis is true. Results presented in Table 2.3 indicate that the effects found were highly significant for all three datasets, and all three measures of significance included in my bootstrapped tests. For the three datasets, there was no SCA analysing bootstrapped samples which resulted in a larger median effect size than the median effect size of the original SCA ($p = 0.00$, original effect sizes: YRBS median $\beta = -.040$, MTF median $\beta = -.007$, MCS median $\beta = -.045$). Furthermore, there was no bootstrapped SCA with more total or statistically significant specifications of the dominant sign than the original SCA

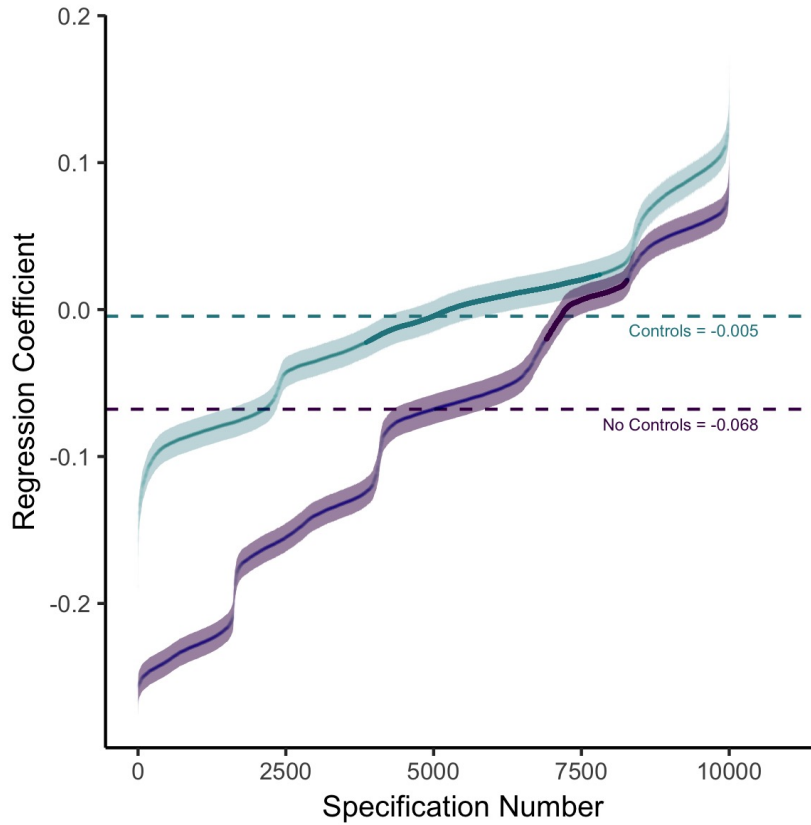


Figure 2.6: Results of Specification Curve Analysis of the Millennium Cohort Study split by whether control variables are included in the analysis or not: Specification Curve Analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each specification number indicates a different combination of analytical decisions. The plot then shows the outcome of the corresponding analysis (standardised regression coefficient) either including control variables (teal, median standardised $\beta = -0.005$, partial $\eta^2 = .001$, median $n = 6,566$, median standard error = .011) or not including control variables (purple, median standardised $\beta = -0.068$, partial $\eta^2 = .005$, median $n = 11,018$, median standard error = .010). The bolded parts of the line indicate analyses that did not reach significance ($p > 0.05$). The median standardised regression coefficients for analyses including or not including control variables are shown using the dashed lines and the error bars visualise the standard error.

Table 2.3: Results of the SCA bootstrapped test for Youth Risk and Behaviour Survey (YRBS), Monitoring the Future (MTF) and Millennium Cohort Study (MCS) datasets, determining whether there is a significant association between digital technology use and well-being. For each dataset we include three different possible tests of significance, where we compare the observed results with the bootstrapped samples on the basis of (i) median effect size, (ii) the share of results with the dominant sign and (iii) the share of results with the dominant sign and a p value of less than 0.05.

* Note: we did not use standardised variables for this analysis and therefore the observed results can diverge from those presented in Table 2.2

Dataset	Observed Result	p value
YRBS		
Median effect size	-0.040	0.00
Share of results with dominant sign	356	0.00
Share of results with dominant sign & $p < .05$	323	0.00
MTF		
Median effect size	-0.007	0.00
Share of results with dominant sign	24149	0.00
Share of results with dominant sign & $p < .05$	19636	0.00
MCS		
Median effect size	-0.045	0.00
Share of results with dominant sign	12481	0.00
Share of results with dominant sign & $p < .05$	10857	0.00

(share of specifications with dominant sign $p = 0.00$; original number: YRBS = 356, MTF = 24,164, MCS = 12,481; share of statistically significant specifications with dominant sign $p = 0.00$; original number: YRBS = 323, MTF = 19,649, MCS = 10,857). This result provides evidence that digital technology use and adolescent well-being could be negatively related at above chance levels in my data.

2.3.4 IV. Comparison Specifications

To put the results of the SCAs into perspective with respect to the broader context of human behaviour as measured in these datasets, I compare specification curves

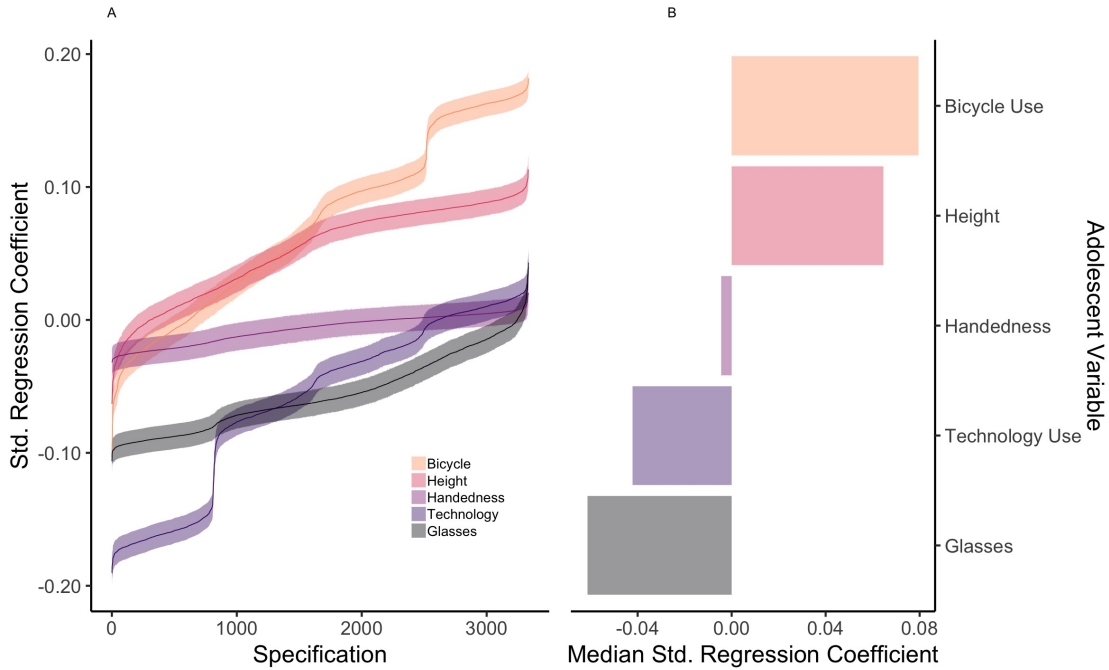


Figure 2.7: Visualisation of the comparison specifications for MCS hypothesised to have little or no influence on well-being: bicycle use, height, handedness and wearing glasses. This graph shows Specification Curve Analyses for both the variable of interest (mean technology use) and the comparison variables; It highlights the range of possible results of a simple cross-sectional regression of the variables of interest on adolescent well-being.

Wearing glasses has the most negative association with adolescent well-being (black, median $\beta = -.061$, median $n = 7,963$, partial $\eta^2 = .005$, median standard error = .010), more negative than the association of technology use with well-being (purple, median $\beta = -.042$, median $n = 7,964$, partial $\eta^2 = .002$, median standard error = .010). Handedness (red/purple, median $\beta = -.004$, median $n = 7,972$, partial $\eta^2 < 0.001$, median standard error = .010), height of the adolescent (red, median $\beta = .065$, median $n = 7,910$, partial $\eta^2 = .005$, median standard error = .010) and whether the adolescent often rides a bicycle (yellow, median $\beta = .080$, median $n = 7,974$, partial $\eta^2 = .007$, median standard error = .010) have more positive associations with adolescent well-being than technology use does.

Panel A shows how different analytical decisions (Specifications, shown on the x-axis) lead to different statistical outcomes (Standardised Regression Coefficient, shown on the y-axis). Each line represents a different variable of interest, the error bars represent the standard error. Panel B visualises the resulting Median Standardised Regression Coefficients for those Specification Curve Analyses linking the variables of interest with adolescent well-being.

for the mean of the technology use variables in each dataset to other associations that have been shown to relate, or are hypothesised not to relate, to adolescent mental health: binge drinking, smoking marijuana, being bullied, getting into fights, smoking cigarettes, being arrested, perceived weight, eating potatoes, having asthma, drinking milk, going to the movies, religion, listening to music, doing homework, cycling, height, wearing glasses, handedness, eating fruit, eating vegetables, getting enough sleep and eating breakfast. For results see Table 2.4 at the end of the chapter and Figures 2.7-2.10).

For YRBS (Figure 2.8), the association of mean technology use with well-being (median $\beta = -.049$, median $n = 62,166$, partial $\eta^2 = .002$, median standard error = .004) was exceeded by well-being's association with being bullied (median $\beta = -.212$, median $n = 50,066$, partial $\eta^2 = .044$, median standard error = .004), getting into fights (median $\beta = -.179$, median $n = 62,106$, partial $\eta^2 = .031$, median standard error = .004), binge drinking (median $\beta = -.144$, median $n = 62,010$, partial $\eta^2 = .021$, median standard error = .004), smoking marijuana (median $\beta = -.132$, median $n = 62,361$, partial $\eta^2 = .018$, median standard error = .004), having asthma (median $\beta = -.066$, median $n = 60,863$, partial $\eta^2 = .004$, median standard error = .004) and perceived weight (median $\beta = -.050$, median $n = 62,752$, partial $\eta^2 = .002$, median standard error = .004). There is a smaller negative association for eating potatoes (median $\beta = -.042$, median $n = 61,912$, partial $\eta^2 = .002$, median standard error = .004), eating vegetables (median $\beta = -.013$, median $n = 62,034$, partial $\eta^2 < .001$, median standard error = .004) and eating fruit (median $\beta = -.005$, median $n = 62,436$, partial $\eta^2 < .001$, median standard error = .004). There is a smaller positive association for drinking milk (median $\beta = .014$, median $n = 60,021$, partial $\eta^2 < .001$, median standard error = .004). Lastly, there is a larger positive association for eating breakfast (median $\beta = .116$, median $n = 34,010$, partial $\eta^2 = .013$, median standard error = .006) and getting enough sleep (median $\beta = .150$, median $n = 56,552$, partial $\eta^2 = .022$, median standard error = .004).

For the MTF (Figure 2.9), I compare the association of mean technology use with well-being (median $\beta = -.006$, median $n = 102,186$, partial $\eta^2 < .001$, median

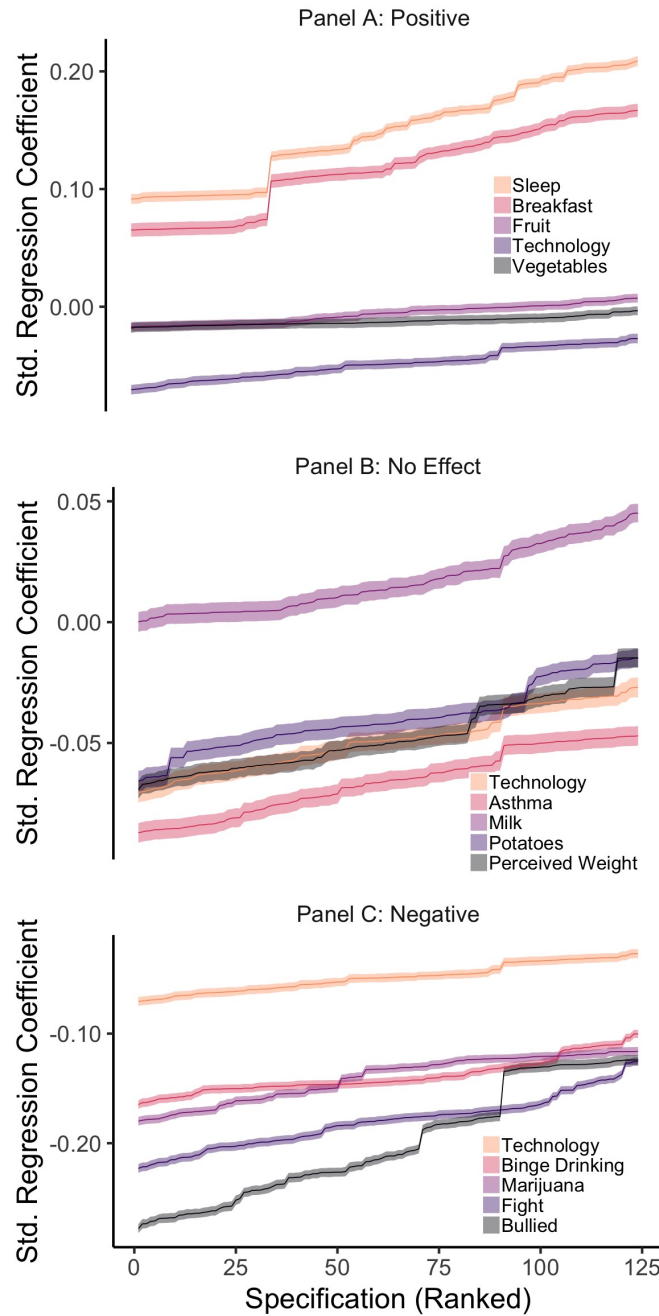


Figure 2.8: YRBS Comparison Specification Curves, split into three panels to compare the effect of technology use on wellbeing to other hypothesised positive (Panel A), neutral (Panel B) and negative factors (Panel C). The figure shows the results of 15 SCAs: illustrating the range of possible regression coefficients found when examining the association between well-being and other variables. The error bars represent the corresponding Standard Error.

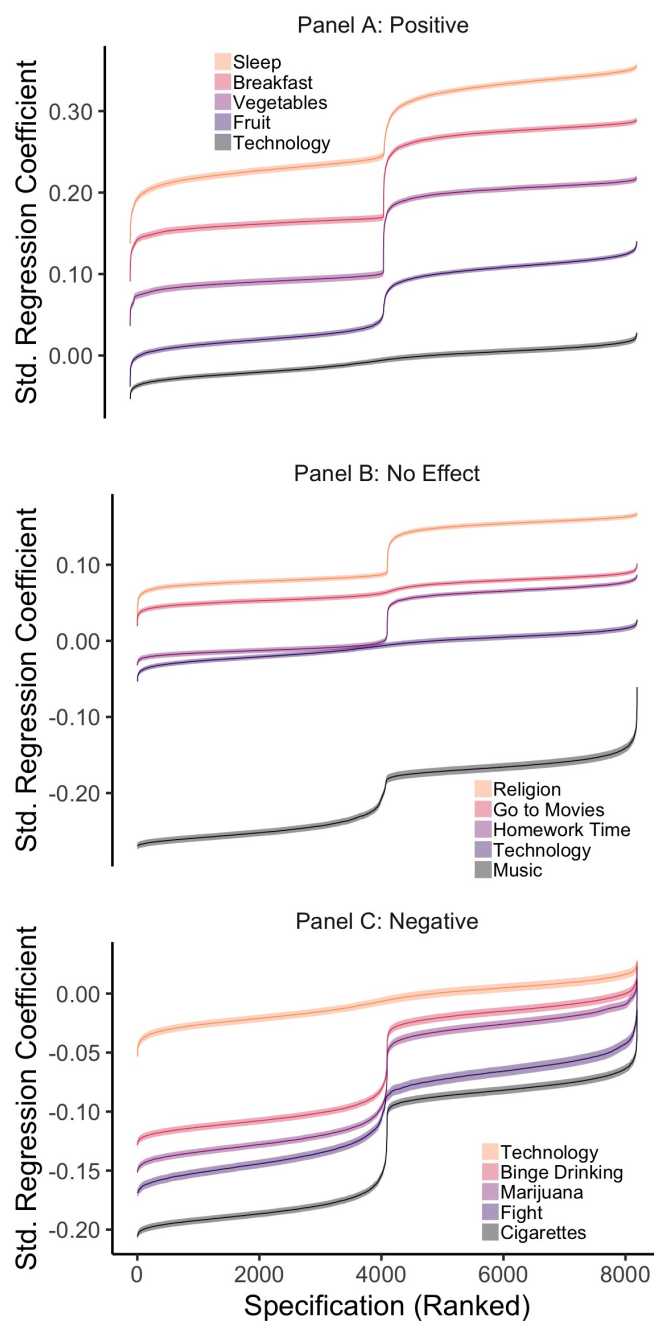


Figure 2.9: MTF Comparison Specification Curves, split into three panels to compare the effect of technology use on wellbeing to other hypothesised positive (Panel A), neutral (Panel B) and negative factors (Panel C). The figure shows the results of 15 SCAs: illustrating the range of possible regression coefficients found when examining the association between well-being and other variables. The error bars represent the corresponding Standard Error.

standard error = .003) to the variables I hypothesised *a priori* to have no association: going to the movies (median β = .064, median n = 115,943, partial η^2 = .005, median standard error = .003), time spent on homework (median β = .020, median n = 115,225, partial η^2 = .001, median standard error = .003), attending religious services (median β = .091, median n = 89,453, partial η^2 = .010, median standard error = .003) and listening to music (median β = -.182, median n = 49,514, partial η^2 = .035, median standard error = .005) all had larger effects. I also examined those comparison variables I hypothesised to have a more positive association: eating breakfast (median β = .170, median n = 62,330, partial η^2 = .034, median standard error = .004), eating fruit (median β = .053, median n = 115,334, partial η^2 = .003, median standard error = .003), sleep (median β = .246, median n = 61,903, partial η^2 = .070, median standard error = .004), and eating vegetables (median β = .115, median n = 62,072, partial η^2 = .014, median standard error = .004). Lastly, I looked at those variables that I hypothesised to have a more negative association: binge drinking (median β = -.045, median n = 107,994, partial η^2 = .002, median standard error = .003), fighting (median β = -.087, median n = 62,683, partial η^2 = .008, median standard error = .004), smoking marijuana (median β = -.056, median n = 113,611, partial η^2 = .003, median standard error = .003) and smoking cigarettes (median β = -.103, median n = 113,424, partial η^2 = .012, median standard error = .003).

For MCS (Figure 2.10), mean technology use (median β = -.042, median n = 7,964, partial η^2 = .002, median standard error = .010) was compared to amount of sleep (median β = .070, median n = 7,954, partial η^2 = .005, median standard error = .010), eating fruit (median β = .056, median n = 7,960, partial η^2 = .004, median standard error = .010), eating breakfast (median β = .140, median n = 7,964, partial η^2 = .025, median standard error = .010) and eating vegetables (median β = .064, median n = 7,949, partial η^2 = .005, median standard error = .010) that have *a priori* hypothesised positive associations; being arrested (median β = -.041, median n = 7,908, partial η^2 = .002, median standard error = .011), being bullied (median β = -.208, median n = 7,898, partial η^2 = .048, median standard error = .010), binge

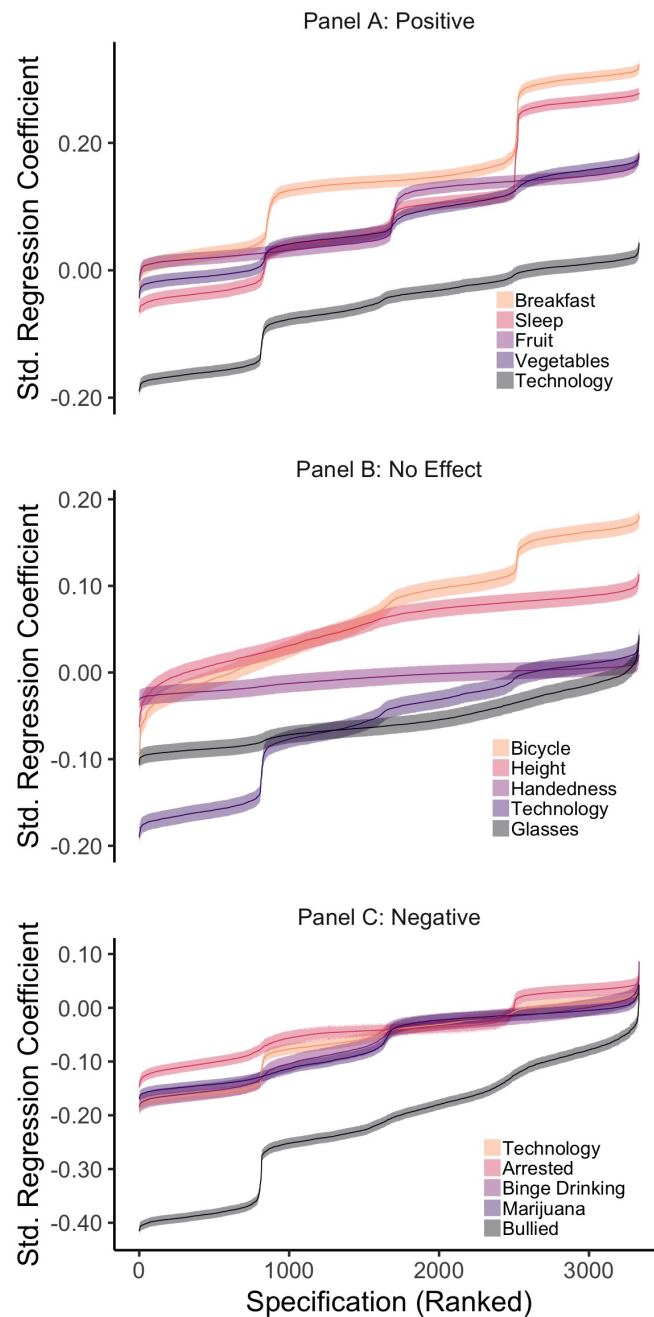


Figure 2.10: MCS Comparison Specification Curves, split into three panels to compare the effect of technology use on wellbeing to other hypothesised positive (Panel A), neutral (Panel B) and negative factors (Panel C). The figure shows the results of 15 SCAs: illustrating the range of possible regression coefficients found when examining the association between well-being and other variables. The error bars represent the corresponding Standard Error.

drinking (median $\beta = -.043$, median $n = 3,656$, partial $\eta^2 = .002$, median standard error = .015) and smoking marijuana (median $\beta = -.048$, median $n = 7,903$, partial $\eta^2 = .003$, median standard error = .010) that have *a priori* hypothesised negative associations; wearing glasses (median $\beta = -.061$, median $n = 7,963$, partial $\eta^2 = .005$, median standard error = .010), being left-handed (median $\beta = -.004$, median $n = 7,972$, partial $\eta^2 < 0.001$, median standard error = .010), bicycle use (median $\beta = .080$, median $n = 7,974$, partial $\eta^2 = .007$, median standard error = .010) and height (median $\beta = .065$, median $n = 7,910$, partial $\eta^2 = .005$, median standard error = .010) that have no *a priori* hypothesised associations (also see Figure 2.7).

2.4 Discussion

The possibility that adolescents' digital technology use has a negative impact on psychological well-being is an important question worthy of rigorous empirical testing. While previous research in this area has equated findings derived from large-scale social data with empirical robustness, the present research highlights deep-seated problems associated with drawing strong inferences from such analyses. To provide a robust and transparent investigation of the association of digital technology use with adolescent well-being, I implemented Specification Curve Analysis (SCA) with comparison specifications using three large-scale datasets from the US and UK.

While I find that digital technology use has a small negative association with adolescent well-being, this finding is best understood in terms of other human behaviours captured in these large-scale social datasets. When viewed in the broader context of the data, it becomes clear that the outsized weight given to digital screen time in scientific and public discourse might not be merited on the basis of the available evidence. For example, in all three datasets the associations found for both smoking marijuana and bullying have larger negative associations with adolescent well-being (2.7x and 4.3x respectively for YRBS) than technology use does. Positive antecedents of well-being are equally illustrative; simple actions like getting enough sleep and regularly eating breakfast have much more positive

associations with well-being than the average impact of technology use (ranging from 1.7x to 44.2x more positive in all datasets). Neutral factors provide perhaps the most useful context to judge technology engagement effects: the association of well-being with regularly eating potatoes was nearly as negative as the association with technology use (0.9x, YRBS) and wearing glasses was more negatively associated with well-being (1.5x, MCS).

With this in mind, the evidence simultaneously suggests technology effects might be statistically significant but so minimal that they hold little practical value. The nuanced picture these results provide are in line with previous psychological and epidemiological research suggesting the associations between digital screen time and adolescent outcomes are not as simple as many might think (Parkes et al. 2013; Przybylski and Weinstein 2017). This work therefore puts previous work that used the YRBS and MTF to highlight technology use as a potential culprit for decreasing adolescent well-being (Twenge, Joiner, et al. 2017) into perspective, showing the range of possible analytical results and comparison specifications. The finding that the association between digital technology use and well-being is much smaller than previously put forth has extensive implications for stakeholders and policy-makers considering monetary investments into decreasing technology use in order to increase adolescent well-being (Department of Health and Social Care 2018).

Importantly, the small negative associations diminish even further when proper and pre-specified control variables, or caretaker responses about adolescent well-being, are included in the analyses. This finding underlines the importance of considering high-quality control variables, *a priori* specification of effect sizes of interest, and a critical evaluation of the role that common method variance may play when mapping the effect of digital technology use on adolescent well-being (Ferguson 2009). It is not enough to rely on statistical power to improve scientific endeavour, large-scale social data analysis harbours its own challenges for statistical inference and scientific progress.

This investigation therefore highlights two intrinsic problems confronting behavioural scientists using large-scale social data. First, large numbers of ill-defined

variables necessitate researcher flexibility, potentially exacerbating the garden of forking paths problem: for some datasets analysed there were more than a trillion different ways to operationalize a simple regression (Gelman and Loken 2014). Second, high numbers of observations render minutely small associations significant through the default NHST lens (Lakens and Evers 2014). With these challenges in mind, my approach, grounded in SCA and including comparison specifications presents a promising solution, so that behavioural scientists can build accurate and practically actionable representations of effects found in large-scale datasets. Overall, the findings place popular worries about the putative links between technology use and mental health indicators into context. They underscore the need for open and impartial reporting of small correlations derived from large-scale social data.

2.4.1 Limitations

My analyses, however, do not provide a definite answer to whether digital technology impacts adolescent well-being. Firstly, it is important to note that using most large-scale datasets one can only examine cross-sectional correlations links, and it is therefore unclear what is driving effects where present. I know very little about whether more technology use might cause lower well-being, whether lower well-being might cause more technology use or whether a third confounding factor underlies both (see Chapter 4). As I am examining something inherently complex, the likelihood of unaccounted factors affecting both technology use and wellbeing is high. It is therefore possible that the associations I document, and those that previous authors have documented, are spurious.

For the sake of simplicity and comparison, simple linear regressions were used in this study, overlooking the fact that the relationship of interest is probably more complex, non-linear, or hierarchical (Przybylski and Weinstein 2017). Many measures used were also of low quality, non-normal, heterogenous, or outdated, limiting the generalisability of the study's inferences. As self-report digital technology measures are known to be noisy (Scharkow 2016), this could have also led to the association of technology use with well-being being diminished due

to low-quality measurement (addressed in Chapter 3). Lastly, I used NHST to interpret significance, which is problematic when using such extensive data. To improve, partnerships between research councils and behavioural scientists to better measurement, and pre-registering of analyses plans, will be crucial.

2.5 Conclusion

Whether they are collected as part of multi-lab projects or research council funded cohort studies, large-scale social datasets are an increasingly important part of the research infrastructure available for psychologists wanting to study emergent technologies. On balance, I am optimistic these investments provide an invaluable tool for studying technology effects in adolescents. To realise this promise, I firmly believe researchers must ground their work and debate in open and robust practices. In the quest for high power, I caution scientists studying emergent technology effects to understand the intrinsic limitations of large-scale data and to implement approaches that guard against researcher degrees of freedom. While preregistration might be implausible for analyses of open large-scale social data, methodologies like Specification Curve Analyses provide solutions that don't only support robust statistical inferences, but also provide a comprehensive way to report the effects found for academia, policy and the public.

2.6 Acknowledgements

This chapter is based on the published work *Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3(2), 173.*

The National Institute on Drug Abuse provided funding for MTF conducted at the Survey Research Centre in the Institute for Social Research, University of Michigan; YRBS was collected by the Centres for Disease Control and Prevention; The Centre for Longitudinal Studies, UCL Institute of Education collected MCS

and the UK Data Archive/UK Data Service provided the data; They bear no responsibility for its aggregation, analysis, or interpretation.

Thank you to U. Simonsohn, N. K. Reimer and N. Weinstein for their valuable input and J. M. Rohrer, U. Simonsohn, J. P. Simmons and L. D. Nelson for code provision. I also acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this research: <http://dx.doi.org/10.5281/zenodo.22558>.

Table 2.4: Comparison Specification results: The table shows the size of the effect of comparison variables on adolescent-wellbeing when compared to the size of the effect of technology use (measured using the mean of technology use questions) on adolescent well-being. The values indicate how many times larger the effects of the comparison variables are in comparison to technology use when examining the Youth Risk and Behaviour Survey (YRBS), Monitoring the Future (MTF) and Millennium Cohort Study (MCS) datasets.

* Denotes when the effect of the comparison variable on well-being is positive, and therefore in the opposite direction to the effect of technology use.

Factor	Comparison Specifications	YRBS	MTF	MCS
Negative Factors	Binge drinking	2.95x	8.10x	1.02x
	Marijuana	2.70x	10.09x	1.14x
	Bullying	4.33x	—	4.92x
	Getting into fights	3.65x	15.58x	—
	Cigarettes	—	18.47x	—
	Being arrested	—	—	0.96x
Neutral Factors	Perceived weight	1.02x	—	—
	Potatoes	0.86x	—	—
	Asthma	1.34x	—	—
	Milk	0.28x*	—	—
	Going to Movies	—	11.51x*	—
	Religion	—	16.29x*	—
	Music	—	32.68x	—
	Homework	—	3.57x*	—
	Cycling	—	—	1.88x*
	Height	—	—	1.53x*
	Glasses	—	—	1.45x
	Handedness	—	—	0.10x
Positive Factors	Fruit	0.11x	9.49x*	1.32x*
	Vegetables	0.27x	20.63x*	1.52x*
	Sleep	3.06x*	44.23x*	1.65x*
	Breakfast	2.37x*	30.55x*	3.32x*

This chapter is based on the published work:

Orben, A., & Przybylski, A. K. (2019). *Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies*. *Psychological Science*.

3

Improving Measurement

Contents

3.1	Introduction	90
3.1.1	Present Research	92
3.2	Exploratory Studies	93
3.2.1	Methods	93
3.2.2	Results	101
3.3	Confirmatory Study	105
3.3.1	Methods	106
3.3.2	Results	108
3.4	Discussion	112
3.4.1	Limitations	115
3.5	Conclusion	115
3.6	Acknowledgements	116

The current level of psychological evidence for digital technology use decreasing adolescent well-being is far removed from the certainty voiced by many commentators. There is little clear-cut evidence that ‘screen time’ decreases adolescent well-being, with most psychological results based on single-country, exploratory studies that rely on inaccurate but popular self-report measures of digital technology use. This chapter, encompassing three nationally representative large-scale datasets from Ireland, the United States and the United Kingdom ($n_{tot} = 17,247$ after data exclusions) and including improved time-use diary measures of digital technology use, uses both exploratory and confirmatory study designs to introduce methodological and analytical improvements to a growing psychological research area. I find little evidence for substantial negative associations between digital technology use – measured either throughout the day or particularly before bedtime – and adolescent well-being.

3.1 Introduction

As digital technologies are an increasingly integral part of daily life for many, concerns about their use have become common (Bell et al. 2015). Yet there is still little consensus as to whether and, if so, how digital technology use affects psychological well-being; results of studies have been mixed and inconclusive, with associations – when found – often being small (Etchells et al. 2016; Orben and Przybylski 2019b; Parkes et al. 2013; Przybylski and Weinstein 2017; Smith et al. 2018).

Most previous work, including the thesis’ other two data chapters, considers the amount of time spent using digital devices or certain technological platforms, as the primary determinant of positive or negative technology effects (Neuman 1988; Przybylski and Weinstein 2017). It is therefore imperative that this work incorporates high quality assessments of such ‘screen time’. Yet, with the vast majority of studies relying on retrospective self-report scales, research indicates that there is good reason to believe this is not the case (Robinson 1985; Scharrow 2016). On the one hand, people are not skilled at perceiving the time they spend engaging in specific activities (Grondin 2010). On the other hand, there are also a

myriad of additional reasons why people fail to give accurate retrospective self-report judgements (Boase and Ling 2013; Schwarz and Oyserman 2001).

Recent work has demonstrated that only one third of participants provide accurate judgements when asked about their weekly internet use, while 42% overestimate and 26% underestimate their usage time (Scharkow 2016). Inaccuracies vary systematically as a function of actual digital engagement (Vanden Abeele et al. 2013; Wonneberger and Irazoqui 2017): Heavy internet users tend to underestimate the amount of time they spend online, while infrequent users over-report this behaviour (Scharkow 2016). Both these trends have been replicated in subsequent studies (Araujo et al. 2017). There are therefore substantial and endemic issues regarding the majority of current research investigating digital technology use and its effects.

Direct tracking of screen time and digital activities on the device-level is a promising approach for addressing this measurement problem (Andrews et al. 2015; David et al. 2018), yet the method comes with technical issues (G. Miller 2012) and is still limited to small samples (Junco 2013). Given the importance of rapidly gauging the impact of digital technology use on well-being, other approaches for measuring the phenomena that can be implemented more widely are needed for psychological science to progress.

To this end, a handful of recent studies have applied experience sampling methodology, asking participants specific technology-related questions throughout the day (Verduyn, Lee, et al. 2015), or after specific bouts of digital technology use (Masur 2019). The method is complemented by studies using time-use diaries, which ask participants to recall what activities they were engaged in during pre-specified days, building a detailed picture of the participants' daily life (Hanson et al. 2010). As most time-use diaries ask participants to recount small time windows (e.g. every 10 minutes), they facilitate the summation of total time spent engaging with digital technologies and allow for investigation into the time of day that these activities occur. Time-use diaries could therefore extend and complement more commonly used self-report measurement methodology. Yet work

using these promising time-use diary measures has focused mainly on single smaller datasets, was not pre-registered, and has not examined psychological well-being as a consequence of digital technology use.

Specifically, the use of time-use diaries allows us to examine how digital technology use before bedtime affects both sleep quality and duration. Researchers have postulated that, by promoting continued availability and fears of missing out, social media platforms can decrease the amount of time adolescents sleep (H. Scott et al. 2018). Previous research found negative effects when adolescents engage with digital technologies 30 minutes (Levenson et al. 2017), one hour (Harbard et al. 2016) and two hours (Orzech et al. 2016) before bedtime. This could be due to delayed bedtimes (Cain and Gradisar 2010; Orzech et al. 2016) or difficulties in relaxing after engaging in stimulating technology use (Harbard et al. 2016).

3.1.1 Present Research

In this chapter, I focus on the relations between digital technology use and psychological well-being using both time-use diaries and retrospective self-report data obtained from adolescents of three different countries: Ireland, the United States and the United Kingdom. Across all datasets my aim is to determine the direction, magnitude, and statistical significance of relations between adolescent digital technology use and psychological well-being, with a particular focus on the effects of digital technology use before bedtime. In order to clarify the mixed literature and provide high generalizability and transparency, the research uses the first two studies to extend a general research question concerning the link between digital technology use and well-being into specific hypotheses, that are then tested in a third confirmatory study. More specifically, I use Specification Curve Analysis to identify promising links in my two exploratory studies, generating informed data- and theory-based hypotheses. The robustness of these hypotheses is then evaluated in a third study using a preregistered confirmatory design. Subjecting the results from the first two studies to the highest methodological standards of testing, this research

aims to shed further light on whether digital technology use has reliable, measurable and substantial associations with the psychological well-being of adolescents.

3.2 Exploratory Studies

3.2.1 Methods

Datasets and Participants

Data from two nationally representative datasets collected in Ireland and the United States were used to explore the plausible links between psychological well-being and digital technology use, generating hypotheses for subsequent testing. I selected both datasets because they were large in comparison to normal social psychological research datasets (total $n = 5,363$; Ireland: $n = 4,573$, United States: $n = 790$ after data exclusions), nationally representative, and have harmonized well-being and time-use diary measurements. Because technology use changes so rapidly, only the most recent wave of time-use diaries was analysed so that the data reflects the current state of digital technology use.

The first dataset under analysis was Growing up in Ireland (GUI, Williams et al. 2009). In my study, I focused on the GUI child cohort that tracked 5,023 nine-year-olds, recruited via random sampling of primary schools. The wave of interest took place between August 2011 and March 2012 and includes 2,514 boys and 2,509 girls, mostly aged thirteen (4,943 thirteen-year-olds, 24 twelve-year-olds and 56 fourteen-year-olds). The time-use diaries were completed on a day individually allocated by the head office (either weekend or weekday) after the primary interview of both children and their caretakers. After data exclusions, 4,573 adolescents were included in the study.

Collected between 2014 and 2015, the second dataset of interest was the United States Panel Study of Income Dynamics (PSID, University of Michigan. Survey Research Center 2018), including 741 girls and 767 boys. It encompassed participants of a variety of different age groups: 108 eight-year-olds, 100 nine-year-olds, 110 ten-year-olds, 89 eleven-year-olds, 201 twelve-year-olds, 213 thirteen-year-olds, 190

fourteen-year-olds, 186 fifteen-year-olds, 165 sixteen-year-olds, 127 seventeen-year-olds and 19 did not provide an age. I only selected those 790 participants between the ages of twelve and fifteen, to match the age ranges in the other datasets used. The sample was collected by involving all children in households already interviewed by the PSID who descended from either the original families recruited in 1968 or the 1997 new immigrant family sample. Those participants in the child supplement that were selected to receive an in-home visit, were asked to complete two time-use diaries on randomly assigned days (one on a weekday and one on a weekend day).

Ethical Review

The Research Ethics Committee of the Health Research Board in Ireland gave ethical approval to the GUI study. The University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board reviews the PSID annually to ensure its compliance to ethical standards.

Measures

This paper examines a variety of well-being and digital technology use measures. While each dataset included a range of well-being questionnaires, I only considered those measures present in at least one of the exploratory datasets (i.e. the GUI and PSID) and in the dataset used for my confirmatory study. This included the popular Strengths and Difficulties Questionnaire (SDQ) completed by caretakers (present in GUI and the confirmatory study), and two questionnaires filled out by adolescents: (1) A well-being questionnaire for each dataset, including the Short Mood and Feelings Questionnaire for the GUI and the confirmatory study and the Children's Depression Inventory present in the PSID and (2) The Rosenberg Self-Esteem scale (present in the PSID and the confirmatory study).

Adolescent well-being. The first measure of adolescent well-being considered was the SDQ completed by the Irish participants' primary caretakers (Goodman et al. 2000). This measure of psychosocial functioning has been widely used and validated

in school, home and clinical contexts (Goodman 1997). It includes 25 questions, five each about prosocial behaviour, hyperactivity or inattention, emotional symptoms, conduct problems and peer relationship problems (0 = not true, 1 = somewhat true, 2 = certainly true, scale subsequently reversed, see Table 2.1).

The second measure of adolescent well-being was an abbreviated version of the Rosenberg Self-Esteem Scale completed by US participants (Robins et al. 2001). This was a five item measure that asked: *“How much do you agree or disagree with the following statement?:”* *“On the whole, I am satisfied with myself”*, *“I feel like I have a number of good qualities”*, *“I am able to do things as well as most other people”*, *“I am a person of value”* and *“I feel good about myself”*. The participants answered these questions on a four item Likert scale from strongly agree (4) to strongly disagree (1).

Third, for the Irish dataset I included the Short Mood and Feelings Questionnaire as a negative indicator of well-being. The adolescent participants answered questions about how they felt or acted in the past two weeks using a three-level Likert Scale from true (1) to not true (3). Items included *“I felt miserable or unhappy”*, *“I didn’t enjoy anything at all”*, *“I felt so tired I just sat around and did nothing”*, *“I was very restless”*, *“I felt I was no good any more”*, *“I cried a lot”*, *“I found it hard to think properly or concentrate”*, *“I hated myself”*, *“I was a bad person”*, *“I felt lonely”*, *“I thought nobody really loved me”*, *“I thought I could never be as good as other kids”* and *“I did everything wrong”*. I subsequently reversed item scores so that they instead measured adolescent well-being.

Finally, for the US sample, I included the Children’s Depression Inventory as a measure of adolescent well-being. The participants were asked to think about the last two weeks and select a sentence that best described their feelings (see Table 3.1). The sentences are very similar to the twelve questions about subjective affective states and general mood asked in the confirmatory dataset detailed later in the paper.

Table 3.1: Items of the PSID Children’s Depression Inventory. The adolescents were asked: *For this next set of items, choose the number next to the statement that best describes how you have felt during the last two weeks (...)*

Option 1	Option 2	Option 3
I am sad once in a while	I am sad many times	I am sad all the time
Nothing will ever work out for me	I am not sure if things will work out for me	Things will work out for me
I do most things O.K	I do many things O.K	I do everything wrong
I hate myself	I do not like myself	I like myself
I feel like crying everyday	I feel like crying many days	I feel like crying once in a while
Things bother me all the time	Things bother me many times	Things bother me once in a while
I look O.K	There are some bad things about my looks	I look ugly
I do not feel alone	I feel alone many times	I feel alone all the time
I have plenty of friends	I have some friends, but I wish I had more	I do not have any friends
Nobody really loves me	I am not sure if anybody loves me	I am sure that somebody loves me

Adolescent digital technology use. The study included two varieties of digital technology use measures: Retrospective self-report measures of digital technology use, and estimates derived from time-use diaries. Details regarding these measures varied for each dataset due to differences in the questionnaires and time diaries used. For all datasets, I removed those participants who filled out a time-use diary during a weekday that was not term or school time, furthermore if the participants went to bed after midnight (after the time-use diary is concluded) I coded them as going to bed at midnight.

Retrospective reports. The Irish dataset included three questions asking participants to think of a normal weekday during term-time and estimate: *“How many hours do you spend watching television, videos or DVDs?”*, *“How much time do you spend using the computer (do not include time spent using computers*

in school)?” and “How much time do you spend playing video games such as PlayStation, X-box, Nintendo, etc.?” Participants could answer in hours and minutes, yet this was recoded by the survey administrator into a 13-level scale. I took the mean of these measures to obtain a general digital technology use measure. In the US dataset adolescents were asked: *“In the past 30 days, how often did you use a computer or other electronic device (such as a tablet or smartphone):”* *“For school work done at school or at home”, “For these types of online activities (visiting a newspaper or news-related website; watch or listen to music, videos, TV shows or movies; follow topics or people that interest you on websites, blogs, or social media sites (like Facebook, Instagram or Twitter), not including following or interacting with friends or family online)”*, *“To play games”* and *“For interacting with others”*. Participants answered using a five-point Likert scale ranging from every day (5) to never (1). For the US data I took the mean of these four items to obtain a general digital technology use measure.

Time-use diaries. The study focused on five discrete measures from the participants’ self-completed time-use diaries: (1) Whether the participants reported engaging with any digital technologies, (2) How much time they spent doing so, and (3) Whether they did so two hours, (4) One hour and (5) 30 minutes before going to bed. I separated these numerical measures for weekend and weekday, resulting in a total of ten different variables.

Each time-use diary, although harmonized by study administrators, was administered and coded slightly differently. The Irish dataset contained 21 pre-coded activities that participants could select for each 15-minute period. They included the four categories I then aggregated into my digital technology use measure: *“Using the internet/emailing (including social networking, browsing etc.)”*, *“Playing computer games (e.g. PlayStation, PSP, X-Box or Wii)”*, *“Talking on the phone or texting”* or *“watching TV, films, videos or DVDs”*. In the US dataset, participants (or their caretakers) could report their activities freely, including primary and secondary activities, duration and where the activity occurred. Research

assistants coded them afterwards. There were 13 codes I aggregated in my digital technology use measure including lessons in using a computer or other electronic device, playing electronic games, other technology based recreational activities, communication using technology social media, texting, uploading or creating internet content, non-specific work with technology like installing software or hardware, photographic processing and other activities involving a computer or electronic device. I aggregated these measures and do not include them in my analyses separately because there were too few people who scored on any one coded variable.

Time-use diary activity measures commonly have high positive skew: many participants do not note down the activity at all, while only a few report spending much time on the activity. It is common practice to address this by splitting the time-use variable into two measures: The first reflecting *participation* and the second reflecting amount of participation (i.e. the *time spent* doing this activity, Hammer 2012; Rohrer and Lucas 2018). Participation is a dichotomous variable representing whether a participant reported engaging in the activity on a given day, whereas time spent is a continuous variable considering the amount of technology use for those participants who reported doing the activity.

In addition to including these two different measures – for weekends and weekdays separately – I also created six measures of technology use before bedtime. These measures were dichotomous, indicating whether the participant had used technology in the specified time interval. These time intervals were 30 minutes, 1 hour and 2 hours before bed on a weekend day and weekday separately.

Control variables. Minimal control variables were incorporated in these exploratory analyses to prevent spurious correlations or conditional associations complicating my hypothesis generating process: gender and age for both Irish and US datasets.

Analytic Approach

To examine the correlation between digital technology use and well-being, I use the SCA approach proposed by Simonsohn and colleagues (Simonsohn et al. 2015) and detailed in Chapter 2. SCA enables researchers to implement many possible analytical pathways and interpret them as one entity, respecting that the garden of forking paths allows for many different data analysis options which should be taken into account in scientific reporting (Gelman and Loken 2014). Because the aim for these analyses was to generate informed data- and theory-driven hypotheses, to then test in a later confirmatory study, the analyses consisted of four steps.

I. Correlations between retrospective reports and time-use diary estimates. The first analytical step was to examine the correlations between retrospective self-report and time-use diary measures of digital technology use, to gauge whether they are measuring similar or removed concepts. This was done to inform later interpretations of the SCA and to give valuable insights to researchers about such widely used measures.

II. Identifying specifications. I then decided which theoretically defensible specifications to include in the SCA. While this was done *a priori* for all studies, it was only pre-registered specifically for the confirmatory study. The three main analytical choices addressed in the SCA were: (1) How to measure well-being, (2) How to measure digital technology use and (3) Whether to include statistical controls or not (see Section 3.2.2). There were three different possible measures of well-being included in the exploratory datasets: (1) The Strengths and Difficulties Questionnaire, (2) The reversed Children’s Depression Inventory or Short Mood and Feelings Scale, and (3) The Rosenberg Self-Esteem scale. There were 11 possible measures of digital technology use, including the retrospective self-report measure and the time-use diary measures separated for weekend day or weekday (participation, time spent, technology use <2 hours, <1 hour and <30 minutes

before bedtime). Lastly, there was a choice of whether to include controls in the subsequent analyses or not.

III. Implementing Specifications. Taking each specification in turn, I ran a linear regression to obtain the standardized regression coefficient (β) linking digital technology use measurements to well-being outcomes. To do so, I first used the various digital technology use measures to predict the specific well-being questionnaires identified in the study. The regression either included control variables or not, depending on the specification. I noted the resulting standardized regression coefficient, the corresponding p value and the partial r^2 . I also ran 500 bootstrapped models of each SCA to obtain the 95% confidence intervals around the standardized regression coefficient and the effect size measure.

The specifications were then ranked by their regression coefficient and plotted in a specification curve, where the spread of the associations is most clearly visualized. The bottom panel of the specification curve plot illustrates what analytical decisions lead to what results, creating a tool for mapping out the, too often invisible, garden of forking paths (for example, see Figure 3.1).

IV. Statistical Inferences. Bootstrapped models were implemented to examine whether the associations evident in the calculated specifications were significant (Orben and Przybylski 2019b; Simonsohn et al. 2015). I was particularly interested in the different measures and time of digital technology use, so ran a separate significance test for each technology use measure. My bootstrapped approach was necessary, because the specifications do not meet the independency assumption of conventional statistical testing. Like in Chapter 2, I created datasets where I knew the null hypothesis was true and examined the median β and number of significant specifications in the dominant direction (the sign of the majority of the specifications) they produced. I used these two significance measures as proposed by Simonsohn and colleagues, but do not report the number of specifications in the dominant direction – a significance measure also proposed by the authors and used

in Chapter 2 – as the nature of the data meant these tests did not give an accurate overview of the data (they can be found in the Appendix, Table B.1-B.4). It was possible to calculate whether the amount of significant specifications or median β s found in the original dataset were ‘surprising’, that is whether less than 5% of the null-hypothesis datasets had more significant specifications in the dominant direction or more extreme median β s than the original dataset.

To create the datasets where the null-hypothesis was true, I extracted the standardised regression coefficient of interest (β), multiplied it by the technology use measure and subtracted it from the well-being measure. I then used these values as my dependent well-being variable, in a dataset where I now know the effect not to be present. I then ran 500 bootstrapped SCAs using this data. As the bootstrapping operation was repeated 500 times, it was possible to examine whether each bootstrapped dataset (where the null hypothesis was known to be true) had more significant specifications or more extreme median β s than the original dataset. To obtain the p value of the bootstrapping test, I divided the number of bootstraps with more significant specifications in the dominant direction or more extreme median β s than the original dataset by the overall number of bootstraps.

3.2.2 Results

I. Correlations Between Retrospective Reports and Time-Use Diary Estimates

For Irish adolescents, the correlation of measures relating to digital technology use, operationalized using the time-use diary estimate (prior to dichotomization into participation and time spent) and retrospective self-report measurement, was small ($r = 0.18$). For American adolescents, the correlations relating time-use diary measures on a weekday ($r = 0.08$) and weekend day ($r = 0.05$) to self-report digital technology use measurements were small as well.

II. Identifying Specifications

I identified 44 specifications each for the Irish and US datasets.

- Irish dataset
 - *Operationalizing adolescent well-being*: Strength and Difficulties Questionnaire; Short Mood and Feelings Questionnaire (Well-being)
 - *Operationalizing digital technology use*: Retrospective self-report measure; Time-use diary measures (weekday and weekend separately: participation, time spent, < 2 hours before bedtime, < 1 hour before bedtime, < 30 minutes before bedtime)
 - *Which control variables to include*: Either include control variables or not
- American dataset
 - *Operationalizing adolescent well-being*: Children’s Depression Inventory (Well-being); Rosenberg Self-Esteem Scale
 - *Operationalizing digital technology use*: Retrospective self-report measure; Time-use diary measures (weekday and weekend separately: participation, time spent, < 2 hours before bedtime, < 1 hour before bedtime, < 30 minutes before bedtime)
 - *Which control variables to include*: Either include control variables or not

III. Implementing Specifications

After all analytical pathways specified in the previous step were implemented, it was evident that there were significant specifications present in both datasets (Figure 3.1, left and middle panel). Some specifications showed significant negative associations ($k = 16$), though there was a larger proportion of non-significant specifications present ($k = 72$). No statistically significant specifications were positive. Specifications using retrospective self-report digital technology use measures resulted in the largest negative associations in the Irish data. I did not find this trend in the US data, possibly due to the restricted range of response anchors connected to their self-report digital technology use measures.

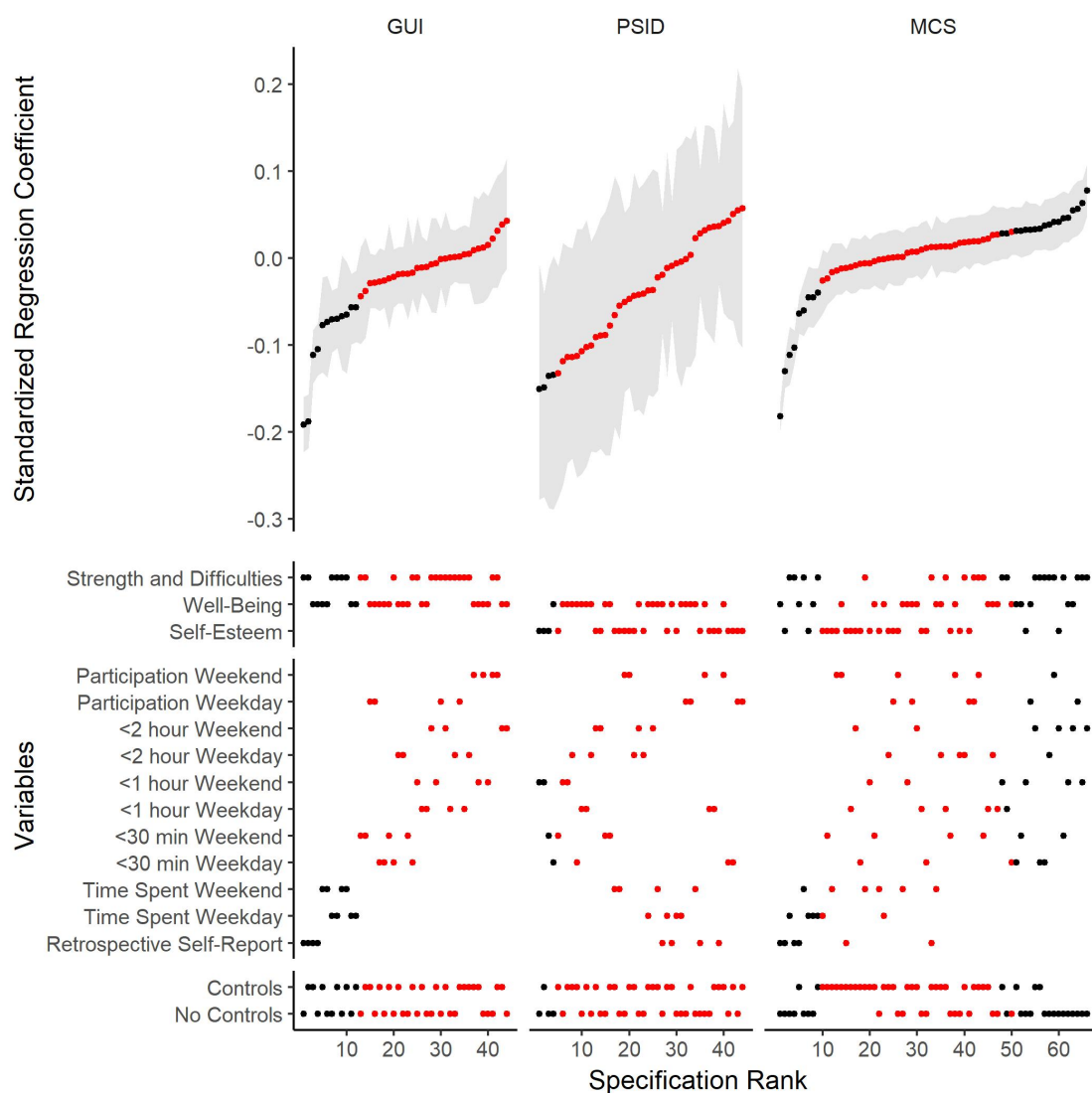


Figure 3.1: Results of the Specification Curve Analysis of Irish, American and British datasets for the correlation between well-being and digital technology use. Grey areas denote 95% confidence intervals of regression coefficients obtained using bootstrapping. Note specifications marked in red were not statistically significant ($p > 0.05$) whereas those in black were ($p < 0.05$)

Table 3.2: Results of the SCA bootstrapping tests for the Irish Dataset. * Result where both measures of significance (effect size and number of significant specifications in the dominant direction) are $p < 0.05$

Technology Measure	Median Point Estimate		Sig. Results in Predominant Direction	
	beta	p value	number	p value
Participation: Weekend	0.02 (-0.02, 0.05)	0.31	0	1.00
Participation: Weekday	-0.01 (-0.03, 0.01)	0.16	0	1.00
<2 hours: Weekend	0.02 (-0.02, 0.02)	0.27	0	1.00
<2 hours: Weekday	-0.01 (-0.02, 0.01)	0.39	0	1.00
<1 hour: Weekend	0.00 (-0.01, 0.05)	0.91	0	1.00
<1 hour: Weekday	0.00 (-0.04, 0.00)	0.65	0	1.00
<30 minutes: Weekend	-0.03 (-0.06, 0.00)	0.05	0	1.00
<30 minutes: Weekday	-0.02 (-0.03, 0.04)	0.01	0	1.00
Time spent: Weekend	-0.07* (-0.10, -0.04)	0.00*	4*	0.00*
Time spent: Weekday	-0.06* (-0.08, -0.04)	0.00*	4*	0.00*
Self-report	-0.15* (-0.17, -0.13)	0.00*	4*	0.00*

IV. Statistical Inferences

Using bootstrapped null models, I found significant correlations between digital technology use and psychological well-being in both the Irish and American datasets (Table 3.2 and 3.3). I count those correlations as significant that find significant effects both for the median β and number of significant tests in the dominant direction.

There was a significant correlation between retrospective self-report digital technology use (median $\beta = -.15$, $p < .001$; number of sig. results in dominant direction (# sig results) = 4/4, $p < .001$) and adolescent well-being in the Irish dataset. There were also negative associations for some of the time-use diary measures: Notably time spent using digital technologies on a weekday (median $\beta = -.07$, $p < .001$; # sig result = 4/4, $p < .001$) and on a weekend (median $\beta = -.06$, $p < .001$; # sig result = 4/4, $p < .001$). In the American dataset I found significant associations only for digital technology use 1 hour before bedtime on a weekend day (median $\beta = -.13$, $p < .001$; # sig result = 2/4, $p = .001$). There were no significant associations of retrospective self-reported digital technology use. Taking this pattern of results as a whole, I derived a series of promising data- and theory-driven hypotheses to test in a confirmatory study.

Table 3.3: Results of the SCA bootstrapping tests for the US Dataset. * Result where both measures of significance (effect size and number of significant specifications in the dominant direction) are $p < 0.05$

Technology Measure	Median Point Estimate		Sig. Results in Predominant Direction	
	beta	p value	number	p value
Participation: Weekend	-0.01 (-0.07, 0.05)	0.47	0	1.00
Participation: Weekday	0.03 (-0.05, 0.10)	0.32	0	1.00
<2 hours: Weekend	-0.07 (-0.11, 0.05)	0.04	0	1.00
<2 hours: Weekday	-0.07 (-0.14, 0.00)	0.08	0	1.00
<1 hour: Weekend	-0.13* (-0.15, 0.00)	0.00*	2*	0.01*
<1 hour: Weekday	-0.03 (-0.12, 0.04)	0.77	0	1.00
<30 minutes: Weekend	-0.11 (-0.18, -0.03)	0.00	1	0.30
<30 minutes: Weekday	-0.03 (-0.20, -0.06)	0.81	1	0.20
Time spent: Weekend	-0.04 (-0.11, 0.05)	0.24	0	1.00
Time spent: Weekday	-0.01 (-0.09, 0.05)	0.74	0	1.00
Self-report	0.01 (-0.03, 0.05)	0.40	0	1.00

3.3 Confirmatory Study

From the two exploratory studies detailed above, and from previous literature about the negative effects of technology use during the week on well-being (Harbard et al. 2016; Levenson et al. 2017; Owens 2014), I derived five specific hypotheses concerning digital technology use and psychological well-being. My aim was to evaluate the robustness of these hypotheses in a third representative adolescent cohort. To this end, I pre-registered my data analysis plan on the Open Science Framework (<http://dx.doi.org/10.17605/OSF.IO/WRH4X>), focusing on the data collected as part of the Millennium Cohort Study (University of London 2017), prior to the date that the data were made available to researchers. My hypotheses were fivefold:

H1: Higher retrospective reports of digital technology use would correlate with lower observed levels of adolescent well-being

H2: Total time spent engaging with digital technologies, derived from time-use diary measures, would correlate with lower observed levels of adolescent well-being

H3: Digital technology use 30 minutes before bedtime on weekdays, derived from time-use diary measures, would correlate with lower observed levels of adolescent well-being

H4: Digital technology use 1 hour before bedtime on weekdays, derived from time-use diary measures, would correlate with lower observed levels of adolescent well-being

H5: In models without controls (detailed below) the negative association will be more pronounced (i.e. will have a larger absolute value) than in models with controls

3.3.1 Methods

Datasets and Participants

The focus of the confirmatory analyses was the longitudinal Millennium Cohort Study (MCS), which followed a UK cohort of young people born between September 2000 and January 2001 (University of London 2017). The survey of interest was administered in 2015 and included responses by 11,884 adolescents and their caregivers. It encompassed 5,931 girls and 5,953 boys: 2,864 thirteen-year-olds, 8,860 fourteen-year-olds and 160 fifteen-year-olds. Using clustered stratified sampling, it oversampled minorities and participants living in disadvantaged areas. Each adolescent completed two (one weekend day and one weekday), paper, web-based or app-based time-use diaries, within ten days of the main interviewer visit.

Ethical Review

The UK National Health Service (NHS) London, Northern, Yorkshire and South-West Research Ethics Committees gave ethical approval for data collection.

Measures

Adolescent well-being. In addition to the Strengths and Difficulties questionnaire completed by the caretaker, the Rosenberg Self-Esteem scale was used, as was an abbreviated version of the short form Mood and Feelings Questionnaire (Angold et al. 1995). This measure instructed participants: *“For each question please select the answer which reflects how you have been feeling or acting in the past two weeks:”*, *“I felt miserable or unhappy”*, *“I didn’t enjoy anything at all”*, *“I felt so tired I just sat around and did nothing_”*, *“I was very restless”*, *“I felt I was no good any more”*, *“I cried a lot”*, *“I found it hard to think properly or concentrate”*, *“I hated myself”*, *“I was a bad person”*, *“I felt lonely”*, *“I thought nobody really loved me”*, *“I thought*

I could never be as good as other kids” and *“I did everything wrong”* (1 = “not true,” 2 = “sometimes,” 3 = “true,” scale subsequently reversed).

Adolescent technology use. Like the US and Irish datasets, the UK dataset included a retrospective self-report digital technology use items. The mean of the four questions concerned with hours per weekday the adolescent spent *“watching television programmes or films”*, *“playing electronic games on a computer or games systems”*, *“using the internet”* at home and *“on social networking or messaging sites or Apps on the internet”* (1 = “none” to 8 = “7 hours or more”) was taken.

Using the time-use diary data, digital technology use measures were derived in line with the approach used for the exploratory datasets. Participants could select certain activity codes for each 10-minute time slot (except in the app-based time diary where there were 1-minute slots): Five of these activity codes were used in the aggregate measure of digital technology use: *“answering emails, instant messaging, texting”*, *“browsing and updating social networking sites”*, *“general internet browsing, programming”*, *“playing electronic games and Apps”* and *“watching TV, DVDs, downloaded videos”*.

Control variables. The control variables detailed in the analysis plan were chosen using previous studies as a template (Orben and Przybylski 2019b; Parkes et al. 2013). In both these studies and in the present analysis, a range of sociodemographic factors and maternal characteristics, including the child’s sex and age, mother’s education, ethnicity, psychological distress (K6 Kessler Scale) and employment were included as control variables. These factors also included household-level variables like household income, number of siblings present, whether the father was present, closeness to parents, the time the primary caretaker could spend with the children. Finally, there were adolescent-level variables that included reports of long-term illness and negative attitudes towards school. To control for the caretaker’s current cognitive ability, I also included the primary caretaker’s score

on a word activity task where they were presented with a list of target words and needed to choose synonyms from a corresponding list.

Analytic Approach

In broad strokes, the confirmatory analytical pathway followed the approach used to examine the exploratory datasets. I included bootstrapped models of all variables, like in my exploratory analyses, but also extended these to examine the specific pre-registered hypotheses. I adapted my pre-registered analysis plan to run simple regressions instead of structural equation models to allow me to implement my significance testing analyses. Furthermore, after submitting my pre-registration I decided to analyse bootstrapped SCAs to obtain confidence intervals and to run two-sided hypothesis tests, rather than one-sided tests, as they are more informative for the reader.

In the pre-registration I also specified a Smallest Effect Size of Interest (SESOI, Lakens, Scheel, et al. 2018), a concept proposed to avoid the problematic over-interpretation of significant but minimal associations, which are becoming increasingly common in large-scale studies of technology use outcomes (Ferguson 2009; Orben and Przybylski 2019b). In line with Ferguson, a correlation coefficient SESOI of $r = .10$ (95% CI .099 to .101) was pre-registered. In other words, digital technology use associations that explained less than 1% (i.e. $r^2 < 0.01$) of well-being outcomes were judged, *a priori*, as being too modest in practical terms to be worthy of extended scientific discussion.

3.3.2 Results

I. Correlations Between Retrospective Reports and Time-Use Diary Estimates

The correlation between self-report digital technology use and time-use diary measures of digital technology use was in line with the Irish data ($r = .18$ for both weekdays and weekend days). This was higher and more consistent than what

was observed in the US data, as the retrospective self-report response options in the British and Irish data were of better quality.

II. Identifying Specifications

I identified 66 specifications for the UK dataset (22 more than for Irish or US data) because there were three different measures of psychological well-being, 11 digital technology use measures and the decision whether to include control variables or not.

- *Operationalizing adolescent well-being:* Strength and Difficulties Questionnaire; Short Mood and Feelings Questionnaire (Well-being), Rosenberg Self-Esteem Scale
- *Operationalizing digital technology use:* Retrospective self-report measure; Time-use diary measures (weekday and weekend separately: participation, time spent, < 2 hours before bedtime, < 1 hour before bedtime, < 30 minutes before bedtime)
- *Which control variables to include:* Either include control variables or not

III. Implementing Specifications

The specification results are plotted in the right panel of Figure 3.1. In contrast to the exploratory analyses, both significant positive and negative associations between digital technology use and well-being were in evidence. As was the case for the exploratory data, retrospective self-report measures consistently showed the most negative correlations. Digital technology use before bedtime measures showed significant positive associations ($k = 18$) while not showing any negative associations.

IV. Statistical Inferences

Mirroring the approach used to analyse the exploratory datasets (see Table 3.4), I found that there was a significant negative correlation between retrospective self-report digital technology use and adolescent well-being (median $\beta = -.08$, $p < .001$; # sig result = 4/6, $p < .001$). There was also a negative association of time spent engaging with digital technologies on a weekday (median $\beta = -.04$, p

Table 3.4: Results of the SCA bootstrapping tests for the UK Dataset. * Result where both measures of significance (effect size and number of significant specifications in the dominant direction) are $p < 0.05$

Technology Measure	Median Point Estimate		Sig. Results in Predominant Direction	
	beta	p value	number	p value
Participation: Weekend	0.01 (-0.01, 0.02)	0.28	1	0.32
Participation: Weekday	0.02* (0.00, 0.04)	0.01*	2*	0.02*
<2 hours: Weekend	0.04* (0.00, 0.03)	0.00*	4*	0.00*
<2 hours: Weekday	0.02 (0.00, 0.03)	0.01	1	0.27
<1 hour: Weekend	0.03* (0.02, 0.05)	0.00*	4*	0.00*
<1 hour: Weekday	0.02 (0.01, 0.04)	0.02	1	0.27
<30 minutes: Weekend	0.02* (0.00, 0.03)	0.01*	2*	0.01*
<30 minutes: Weekday	0.03* (0.02, 0.50)	0.00*	3*	0.00*
Time spent: Weekend	0.00 (-0.02, 0.01)	0.66	1	0.25
Time spent: Weekday	-0.04* (-0.06, -0.02)	0.00*	4*	0.00*
Self-report	-0.08* (-0.10, -0.07)	0.00*	4*	0.00*

$< .001$; # sig result = 4/6, $p < .001$). There were, however, significant positive associations for other time-use diary measures of digital technology use, including participation with digital technologies on a weekday (median $\beta = .02$, $p = .010$; # sig result = 2/6, $p = .020$), digital technology use 30 minutes before bedtime on a weekday (median $\beta = .03$, $p < .001$; # sig result = 3/6, $p < .001$) and weekend day (median $\beta = .02$, $p = .010$; # sig result = 2/6, $p = .010$), digital technology use 1 hour before bedtime on a weekend (median $\beta = .03$, $p < .001$; # sig result = 4/6, $p < .001$) and digital technology use 2 hours before bedtime on a weekend (median $\beta = .04$, $p < .001$; # sig result = 4/6, $p < .001$).

Hypothesis 1: Retrospective self-reported digital technology use and psychological well-being. Because I find that retrospective self-report digital technology use significantly decreases adolescent well-being (median $\beta = -.08$ [-.10, -.07], $p < .001$; # sig result = 4/6, $p < .001$), using a two-sided bootstrapped test, my first hypothesis is supported (see Table 3.5). The median partial r^2 value (partial $r^2 = .008$ [.006, .011]) is below my Smallest Effect Size of Interest (SESOI; i.e., $r = -.10$) detailed in the preregistered analysis plan. It must, however, be noted that the 95% confidence interval extends above the SESOI.

Hypothesis 2: Time spent engaging with digital technologies and psychological well-being. I examined general time spent engaging with digital technologies, both on a weekend and weekday, using time-use diary measures and one-sided bootstrapped tests. A significant negative association (median $\beta = -.02$ $[-.04, -.01]$, $p < .001$; # sig result = 5/12, $p < .001$) was in evidence. The direction and significance of this correlation was in line with the registered hypothesis, yet this association was also smaller than the prespecified SESOI (partial $r^2 = .001$ $[-.000, .002]$). Again, the 95% confidence interval of the effect size falls above the SESOI.

Hypothesis 3: Technology use 30 min before bedtime on a weekday and psychological well-being. Results focusing on digital technology use 30 minutes before bed, using a two-sided bootstrapped test, indicated that this hypothesis was not confirmed (median $\beta = .03$ $[-.01, .05]$, $p < .001$; # sig result = 3/6, $p < .001$), as the effect was in the opposite direction.

Hypothesis 4: Technology use 1 hour before bedtime on a weekday and psychological well-being. Models examining the association of digital technology use 1 hour before bedtime on a weekday found no effect in the hypothesized negative direction (median $\beta = .02$ $[-.01, .04]$, $p = .02$; # sig result = 1/6, $p = .27$), therefore not supporting the fourth hypothesis.

Hypothesis 5: Comparing models that do and do not account for confounding variables when testing the relation between digital technology use and psychological well-being. Lastly, the effect of including controls in my confirmatory models was evaluated. Figure 3.2 presents two different specification curves, one including and one excluding controls. Visual inspection of the models shows that those with controls exhibit less extreme negative associations. This is supported by the difference in the median associations (controls $r = .026$, no controls $r = .001$). A one-sided paired t-test comparing the correlation coefficients found using specifications with controls to those found when not including controls indicates a non-significant association ($t_{32} = 0.26$, $p = .79$). This result does not

Table 3.5: Results of the SCA bootstrapping tests for confirmatory tests

Technology Measure	Significance Measure	Observed [95% CI]	Partial r squared values [95% CI]	p value
Self-Report	Effect Size	-0.08 [-0.10, -0.07]	0.008 [0.006, 0.011]	0.00
Time Spent	Sig Results	4		0.00
	Effect Size	-0.02 [-0.04, -0.01]	0.001 [0.000, 0.002]	0.00
< 30 Minutes Weekday	Sig Results	5		0.00
	Effect Size	0.03 [0.01, 0.05]	0.001 [0.000, 0.003]	0.00
< 1 Hour Weekday	Sig Results	3		0.00
	Effect Size	0.02 [0.01, 0.04]	0.001 [0.000, 0.003]	0.02
	Sig Results	1		0.27

support the hypothesis that correlations present when not including controls in the model are more negative than when controls are included.

3.4 Discussion

Because technologies are inherently embedded in our social and professional lives, research concerning digital technology use and its effects on adolescent well-being is under increasingly intense scientific, public and policy scrutiny. It is therefore essential that the psychological evidence contributing to the available literature be of the highest possible standard. There are however considerable problems, ranging from measurement issues, to lack of transparency, little confirmatory work and over-interpretation of miniscule effect sizes (Orben and Przybylski 2019b). Only a few studies regarding technology effects have used a preregistered confirmatory framework (Elson and Przybylski 2017; McCarthy et al. 2016; Przybylski and Weinstein 2017); No large-scale cross-national work has tried to move away from retrospective self-report measures to gauge time spent engaged with digital technologies, yet it has been evident for years that such self-report measures are inherently problematic (Scharkow 2016; Schwarz and Oyserman 2001). Until these three facts are reconciled

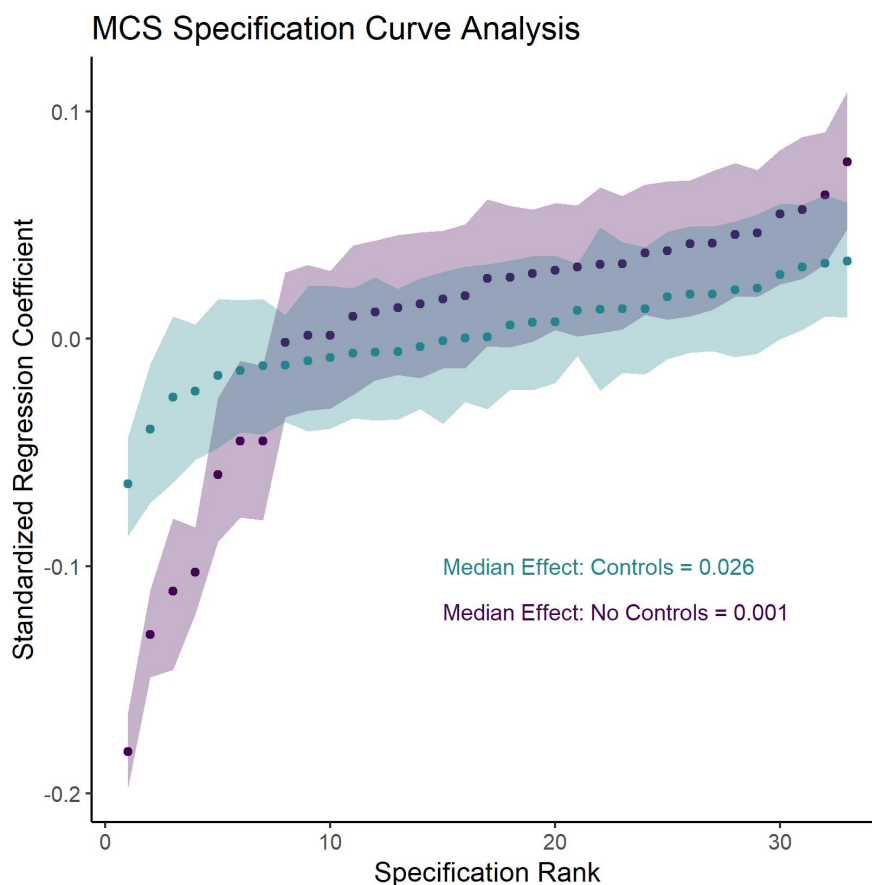


Figure 3.2: Results of the Specification Curve Analysis for the British data. The curve shows the correlation between digital technology use and adolescent well-being, comparing specifications including controls (teal) and no controls (purple).

in the literature, exploratory studies wholly dependent on retrospective accounts will command an outsized share of public attention (Orben 2017).

This chapter marks a novel contribution to the psychological study of technology in a variety of ways. First, it introduces a new measurement of screen time, implements rigorous and transparent approaches to statistical testing and explicitly separates hypothesis generation from hypothesis testing. Given the practical and reputational stakes for psychological science, I firmly argue that this methodological framework should be the new baseline for those wanting to make scientific claims about the effects of digital technology use on human behaviour, development, and well-being.

Second, just like the second chapter, this chapter finds little substantive sta-

tistically significant and negative associations between digital technology use and well-being in adolescents. The most negative associations were found when both self-reported technology use and well-being measures were used, which could be a result of common method variance or noise found in such large-scale questionnaire data. Where statistically significant, associations were smaller than my pre-registered cut off for a practically significant effect, even though some 95% CI crossed this pre-registered threshold. In other words, though there were some small statistically significant negative associations, they fell below the prespecified SESOI of a correlation coefficient of $r = .10$: Digital technology use explained less than 1% of variance in adolescent well-being. This is in line with results from previous research showing that the association between digital technology use and wellbeing often falls below or near this threshold (Ferguson 2009; Orben and Przybylski 2019b; Twenge, Joiner, et al. 2017; Twenge, Martin, et al. 2018). I argue that these effects are therefore potentially too small to merit substantial scientific discussion (Lakens, Scheel, et al. 2018).

This supports previous research and the results of Chapter 2 which show that there is a small significant negative association between technology use and well-being, which – when compared to other activities in an adolescent’s life – is miniscule (Orben and Przybylski 2019b). Whether smaller effects, even when extremely small, are important, is up for debate, as technology use affects the large majority of the population (Rose 2008). Yet this evidence can be interpreted against calls for extreme approaches to decreasing adolescent technology use.

Third, this chapter was also one of the first pieces of research to examine whether digital technology use before bedtime is especially detrimental to adolescent psychological well-being. Public opinion seems to suggest that using digital technologies immediately before bed should be more harmful for teens than screen time spread throughout the day. My exploratory and confirmatory analyses provided very mixed effects: Some were negative while others were positive or inconclusive. This chapter therefore highlights that technology use before bedtime might not

be inherently harmful to psychological well-being, even though this is a well-worn idea in both media and public debates.

3.4.1 Limitations

While I aim to implement the best possible analyses of the research questions posed in this paper, there are issues intrinsic to the data that need to be noted. First, time-use diaries as a method for measuring technology use are not inherently problem-free. It could be that the reflexive or brief uses of technology concurrent with other activities are not properly recorded by this method. Likewise, I cannot ensure that all days that were under analysis were representative. To address both issues, one would need to holistically track technology use across multiple devices over multiple days, though doing this with a population representative cohort would be extremely resource intensive (Wilcockson et al. 2018). Second, it is important to note that the time-use diary and well-being measures were not collected on the same occasion. Because the well-being measures inquire about feelings in general, instead of just about the feelings on the specific day of questioning, the study assumed that the correlation between both measures still holds as it reflects links between exemplar days and general experiences. Finally, it bears mentioning that the study is correlational, and the directionality of effects cannot, and should not, be inferred from the data.

3.5 Conclusion

Until they are displaced by a new technological innovation, digital technologies will remain a fixture of human experience. Psychological science can provide a powerful tool for quantifying the association between digital technology use and adolescent well-being, yet it routinely fails to supply the robust, objective and replicable evidence necessary to support its findings. As the influence of psychological science on policy and public opinion increases, so must we raise our standards of evidence. This chapter proposes and applies multiple methodological

and analytical innovations to set a new standard for quality of psychological research on digital contexts. Granular technology engagement metrics, large-scale data, using specification curve analysis to generate hypotheses and preregistration for hypothesis testing should form the basis of a future methodological framework (Wagenmakers, Wetzels, et al. 2012). To retain the influence and trust we often take for granted as a psychological research community, robust and transparent research practices will need to become the norm – not the exception.

3.6 Acknowledgements

This chapter is based on the published work *Orben, A., & Przybylski, A. K. (2019). Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. Psychological Science.*

The National Institutes of Health (R01-HD069609/R01-AG040213), and National Science Foundation (SES-1157698/1623684) supported PSID. The Department of Children and Youth Affairs funded Growing Up in Ireland, carried out by the Economic and Social Research Institute and Trinity College Dublin. The Centre for Longitudinal Studies, UCL Institute of Education collected MCS and the UK Data Archive/UK Data Service provided the data; They bear no responsibility for its analysis or interpretation. I also thank J. M. Rohrer for providing open access code, on which parts of my analyses are based.

This chapter is based on the published work:

Orben, A., Dienlin, T. & Przybylski, A. K. (2019). Social Medias Enduring Effect on Adolescent Life Satisfaction. Proceedings of the National Academy of Sciences of the United States of America.

4

Improving Data

Contents

4.1	Introduction	118
4.2	Methods	120
4.2.1	Datasets and Participants	120
4.2.2	Ethical Review	120
4.2.3	Measures	121
4.2.4	Statistical Approach	123
4.2.5	Code Availability Statement	127
4.2.6	Data Availability Statement	127
4.3	Results	128
4.4	Limitations	129
4.5	Conclusion	130
4.6	Acknowledgements	130

Although concerns about adolescent social media use and its influence on life satisfaction are mounting, our empirical understanding of the phenomenon is still predominantly informed by cross-sectional research. Reliable inferences about longer-term effects therefore remain elusive. This is the case, in part, because estimating effects observed within the same person over time is inherently different than probing differences in effects between people at one timepoint. Much of the current literature erroneously infers the former using the latter. In this chapter I use large-scale representative panel data to disentangle such between-person and within-person effects. To this end I find negative cross-sectional associations between social media use and life satisfaction and less consistent smaller reciprocal longitudinal relations. Gender plays a decisive role in this dynamic, as there are only few significant cross-sectional or longitudinal relations in adolescent boys. In contrast, in adolescent girls there are small but statistically significant reciprocal within-person effects 1) Increases in life satisfaction predict lower social media use one year later and 2) Increases in social media use predict lower life satisfaction one year later. This suggests that social media is not, in of itself, a robust predictor of life satisfaction across the adolescent population. Instead, its effects are nuanced, reciprocal, and contingent on analytic judgements made by researchers. Research therefore would benefit from an increased focus on the motives and modes adolescent girls adopt when using social media to fundamentally advance the understanding of social media's wider influence on the adolescent population.

4.1 Introduction

Does the increasing amount of time adolescents devote to social media negatively affect their satisfaction with life? Set against the rapid pace of technological innovation, this simple question has grown into a pressing concern for scientists, caregivers, and policymakers. Research, however, has not kept pace (Bell et al. 2015; Valkenburg and Piotrowski 2017). Focussed on cross-sectional relations, scientists have few means of parsing longitudinal effects from artefacts introduced by common statistical modelling methodologies (Hamaker et al. 2015). Furthermore,

as seen in Chapter 2, the volume of data under analysis, paired with unchecked analytical flexibility, enables selective research reporting, biasing the literature towards finding statistically significant effects (Gelman and Loken 2014). It is important to note that, when compared to other factors affecting youth well-being, the purported links between digital technology use and well-being are still extremely small (Orben and Przybylski 2019a; Orben and Przybylski 2019b). Nevertheless, these modest trends are routinely overinterpreted by those under increasing pressure to rapidly craft evidence-based policies. This is especially the case for social media effects, as social media use is increasingly being highlighted as the kind of digital technology use which causes most concern in the broader population (House of Commons Science and Technology Select Committee 2019). This chapter will therefore focus exclusively on social media use.

Throughout this thesis, and in the scientific literature more generally, our understanding of digital technology effects has been predominately shaped by analyses of cross-sectional associations between technology or social media use measures and self-reported adolescent outcomes. Highly-popular studies highlight modest negative correlations (Twenge, Joiner, et al. 2017), but many of their conclusions are problematic (Ophir et al. 2019). It is not possible to assume that any reported *between-person associations* – comparing different people at the same timepoint – translate into *within-person effects* – tracking an individual, and what affects them, over time (Hamaker et al. 2015). Drawing this flawed inference risks misinforming the public or changing policy on the basis of unsuitable evidence. This chapter therefore uses improved longitudinal data and modelling techniques to disentangle within- and between-person relations linking adolescent social media use and life satisfaction.

4.2 Methods

4.2.1 Datasets and Participants

To examine the diverse links between adolescent social media use and life satisfaction I analysed an eight wave, large-scale, and nationally representative panel dataset (UK Understanding Society, the UK Household Panel Survey, University of Essex, Institute for Social and Economic Research 2018). It entails 12,672 UK adolescents between ten and fifteen years of age and was collected as part of an annual longitudinal panel study sampling 40,000 households in the United Kingdom (England, Scotland, Wales and Northern Ireland, University of Essex, Institute for Social and Economic Research 2018). It is important to note that the number of participants for any of the analyses run varied by age and whether full or imputed data were used (range, $n = 539$ to 5,492; median, $n = 1,699$). The sample of households consists of a General Population Sample, an Ethnic Minority Boost Sample, a General Population Comparison sample, and samples previously included in the predecessor of Understanding Society, the British Household Panel Survey. Understanding Society data collection is typically done via face-to-face interview, with trained interviewers visiting households personally.

The youth questionnaire is administered on paper and each adolescent member of a sampled household is interviewed and re-interviewed every year until they graduate into the adult survey at age 16. This study uses seven waves of collected data from 2009 until 2017. The data collection takes place over a 24-month period for each wave, so the annual waves overlap and I determine each wave from when it was first collected. In total, the dataset includes 29,155 individual responses, with each year having between 3,466 and 5,020 responses. Participants had a mean age of 12.54 (s.d. = 1.69), and included 5,881 girls and 5,983 boys.

4.2.2 Ethical Review

The unique and extensive data collection is funded by the UK Economic and Social Research Council and other governmental departments with ethical approval

from the University of Essex Ethics Committee (provided in 2007, 2010, 2013, 2014, 2015 and 2016, Knies 2017). In particular the University of Essex Ethics Committee approved all data collection on the Understanding Society main study and innovation panel waves, including asking consent for all data linkages except to health records. Requesting consent for health record linkage was approved at Wave 1 by the National Research Ethics Service (NRES) Oxfordshire REC A (08/H0604/124), at BHPS Wave 18 by the NRES Royal Free Hospital & Medical School (08/H0720/60) and at Wave 4 by NRES Southampton REC A (11/SC/0274). Approval for the collection of biosocial data by trained nurses in Waves 2 and 3 of the main survey was obtained from the National Research Ethics Service (Understanding Society – UK Household Longitudinal Study: A Biosocial Component, Oxfordshire A REC, Reference: 10/H0604/2).

4.2.3 Measures

Adolescents and their caretakers filled out a variety of questionnaires in each wave, some questionnaires were only completed on every second or third wave. For the current chapter, I focus on those questionnaires completed by participants in each wave, which included measures of life satisfaction and social media use. The code for data recoding can be found in the detailed online documentation (<http://dx.doi.org/10.17605/OSF.IO/4XP3V>).

Life Satisfaction

Life satisfaction was measured using six items. The response format was graphical: Individual options were presented as either frowning or smiling faces; the lower the number, the happier the face. The exact questions were as follows: *“The next few questions are about how you feel about different aspects of your life. The faces express various types of feelings. Below each face is a number where 1 is completely happy and 7 is not at all happy. Please tick the box which comes closest to expressing how you feel about each of the following things: (a) Your school work?; (b) Your appearance?; (c) Your family?; (d) Your friends?; (e) The school you go to?; (f)*

Which best describes how you feel about your life as a whole?”. I take the mean of the six items to create another *mean satisfaction* measure and reverse the scale so that a higher score denoted higher life satisfaction.

Social Media Use

The social media use variable was more complicated to code as there was a mistake on the side of Understanding Society, who I found were inconsistent in how they directed children through the study in different years. In total there were two questions about social media use. One was dichotomous: *“Do you have a social media profile or account on any sites or apps?”*, 1 = *Yes*, 2 = *No*. The other was a 5-point scale: *“How many hours do you spend chatting or interacting with friends through a social web-site like [bebo, facebook, myspace] on a normal school day?”*, 1 = *none*, 2 = *less than an hour*, 3 = *1-3 hours*, 4 = *4-6 hours*, and 5 = *7 or more hours*.

In some waves of the study, children who recorded that they do not use social media on the dichotomous measure were automatically coded as *none* on the 5-item scale and sometimes this was not the case. As there was no perfect solution for remedying this error on side of the Understanding Society team, I implemented a recoding approach. If an adolescent said they do not own a social media account on the dichotomous measure or they do not use social media to interact with friends on the 5-item measure I coded them as scoring the lowest score of ‘1’, for the rest of the participants I took their 5-item measure score that quantifies how much time they spend interacting socially online.

Control Variables

As control variables, I included age, maternal employment (0 = *unemployed*, 1 = *employed*), number of children of mother, maternal ethnic background (0 = *white*, 1 = *non-white*), socializing of mother with adolescent (*“How often do you and your [child/children] spend time together on leisure activities or outings outside the home such as going to the park or zoo, going to the movies, sports or to have a picnic?”*; 1 = *never or rarely* to 6 *almost every day*), support given by the family (*“Do you*

feel supported by your family, that is the people who live with you?"; 1 = *I feel supported by my family in most or all of the things I do*, 2 = *I feel supported by my family in some of the things I do*, 3 = *I do not feel supported by my family in the things I do*), and maternal depression ("How often have you felt downhearted and depressed?"; 1 = *none of the time* to 5 *all of the time*).

4.2.4 Statistical Approach

This chapter combines Random Intercept Cross-Lagged Panel Models (Hamaker et al. 2015) with the Specification Curve Analysis framework I adopted in the previous two chapters (Simonsohn et al. 2015).

Random Intercept Cross-Lagged Panel Models

Using Random Intercept Cross-Lagged Panel Models (RI-CLPM) on longitudinal data I can distinguish between-person from within-person variance when quantifying the relations between social media use and life satisfaction (Hamaker et al. 2015). Differentiating these two sources of variance is important when wanting to investigate true directional effects; it isn't uncommon to find that between-person and within-person relations are in opposite directions (Dietvorst et al. 2018). To illustrate this, one can imagine investigating people's typing speed and how this relates to the number of typing errors they make. When comparing participants with one another (i.e. at the between-person level), the relationship between typing speed and typing errors is *negative*: Those participants who are more skilled and can type faster also commit fewer mistakes. When examining each person individually (i.e. at the within-person level), however, this relationship is *positive*: Typing faster leads to more typing mistakes. This is also known as Simpson's paradox (Kievit et al. 2013; Simpson 1951). Random intercept models provide a valuable statistical tool to differentiate between such within- and between-person effects (Dienlin et al. 2018). Their increasing use in the literature has highlighted how traditional cross-lagged panel approaches have often delivered false results (Dietvorst et al. 2018).

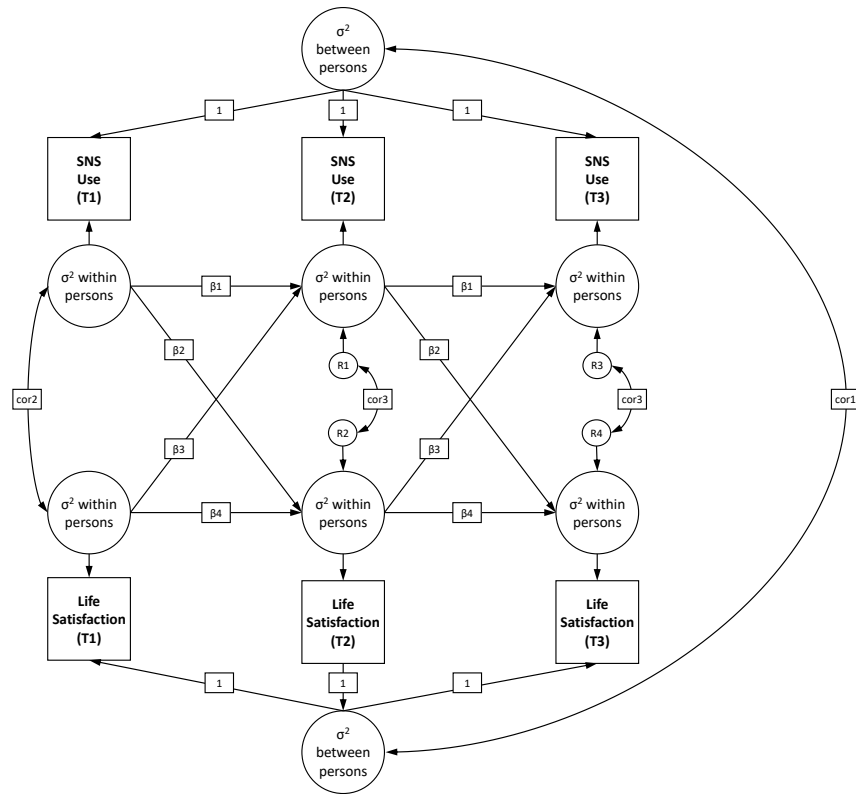


Figure 4.1: Random-Intercept Cross Lagged Panel Model

In this chapter, I specified a RI-CLPM as proposed by Hamaker and colleagues (see Figure 4.1, 2015). I constrained longitudinal effects to be equal across all waves. Doing so provides a measure of the effect of social media use on life satisfaction (and vice versa) that is: (a) More robust (the information of several years are aggregated) and, (b) More comprehensive (one single measure is produced). This procedure builds on the assumption that the effect of social media use on life satisfaction is time invariant, that it does not change from say year 2010 to 2012.

Furthermore, I examined how much variance in the outcome is due to between-person differences versus within-person changes, as with too little variance a RI-CLPM would be unsuitable to the data. Precisely, if there is no within-person variance, there are no causal effects; likewise, if there is only within-person variance or no between-person variance, one could simply run a regular Cross Lagged Panel Model (CLPM). I therefore explicitly calculated the Intra Class Correlation coefficients, which estimate how much variance is due to aggregation on the second

level (i.e. in this case on the person level). I found that 30% of the variance in social media use is due to between-person differences. As a result, it is important to run CLPMs with random intercepts, to be able to partial out that variance. At the same time, 70% of the variance in social media use results from within-person changes, which leaves ample room for within-person effects. Similarly, depending on the exact life satisfaction domain analysed, I find that 27-43% of variance is due to between-person differences.

Specification Curve Analysis

The SCA approach in this chapter is composed of three analytical steps that provide an analytic and conceptual frame for me to implement all theoretically defensible analysis options, with the goal of exhaustively examining the research question of interest. It is important to note that in contrast to the previous two chapters, the RI-CLPMs chosen in this chapter do not allow for statistical testing of SCA results (e.g. via permutation testing or bootstrapping), I therefore refrain from including such statistical conjectures.

When determining what specifications to include in my SCA, I needed to consider two major dimensions. The first dimension refers to *content*: For example, one can control for different sociodemographic variables in statistical analyses. The second dimension refers to *methodology*: For example, one can analyse the same research question using either three waves of data, which would maximise the number of participants, or five waves of data, which would increase the robustness of the estimate. As one can deliberately confound SCAs by including defensible but still unlikely or sub-par analyses, thereby diluting the effect, I have specifically aimed to design equally defensible specifications, which all reach the gold-standard of data analysis aspired to.

I implemented the following specifications in my SCA:

- *Life satisfaction measure*: I analysed either the mean of all life satisfaction subdimensions (“mean satisfaction”) or the subdimensions individually (7 specifications)

- *Number of waves*: I built my model either on data of all participants who had successfully completed three, four or five subsequent waves (3 specifications)
- *Control variables*: I either included no control variables, each of the aforementioned control variables individually, or all control variables (9 specifications)
- *Modelling*: I either included MLR modelling treating the variables as continuous or WLSMV modelling that does not assume normal distribution and is therefore suited for modelling ordinal data (2 specifications)
- *Missing data*: I either used original data or imputed missing data using predictive mean matching (2 specifications)
- *Gender*: I either examined males and females separately or all the data together (3 specifications)

Overall, I ran 2,268 different specifications. Nonetheless, later in the chapter I highlight a specific specification in detail (the *Preferred Model*) to interpret results in more detail. The Preferred Model is the combination of analytical decisions that I believe is the best approach to model the data because of the data's characteristics, and the methodological and theoretical literature. This allows me to compare effects (e.g. examine gender differences) and highlight practical and statistical significance in a simpler framework than when taking a whole SCA approach. While I still present the range of results found in the complete SCA to provide information about the effects' variability in the light of analytical changes, I focus on a single result in complicated parts of the paper. I do not use a Preferred Model in Chapter 2 and 3 as the analysis and interpretation of these pieces of research was simpler and I did not compare effects across conditions like gender.

My Preferred Model uses the following analytical decisions I believe represent the most robust ways of modelling the data:

1. I use data from participants who completed four waves: Examining only three waves would be the absolute minimum necessary for our models (Hamaker et al. 2015) and examining more would decrease participant numbers, thereby reducing power

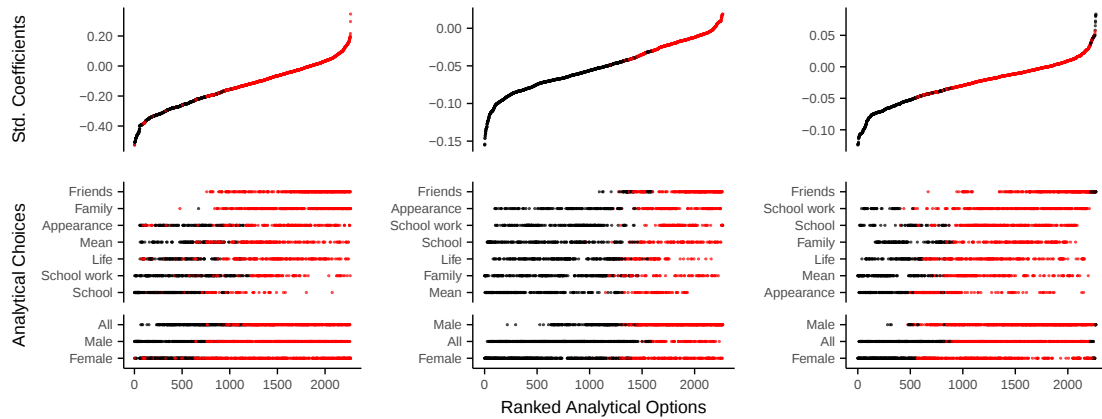


Figure 4.2: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Left Panel: between-person correlations. Middle Panel: within-person effects of social media use on life satisfaction. Right Panel: within-person effects of life satisfaction on social media use. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard depicts what gender and life satisfaction domain the specific combination of analytical decisions analysed; the resulting β is shown in the plot above (red = $p > 0.05$, black = $p < 0.05$). For graphics including all analytical decisions see Figure C.1-C.3

2. I include all relevant control variables
3. I use WLSMV to account for ordinal scales
4. I implement predictive mean matching to impute missing data (Little 1988)
5. I consider female and male participants separately

4.2.5 Code Availability Statement

Intermediate analysis files and a live version of the analysis code can be found on the Open Science Framework (<http://dx.doi.org/10.17605/OSF.IO/4XP3V>).

4.2.6 Data Availability Statement

The data that support the findings of this study are available from the UK Data Service (Knies 2017).

4.3 Results

I first examined between-person associations (Figure 4.2, Panel 1), addressing the question: *Do adolescents using more social media show different levels of life satisfaction compared to adolescents using less?* Across all operationalisations, the median cross-sectional correlation was negative ($\psi = -0.13$), an effect judged as small by behavioural scientists (J. Cohen 1992). Next, I examined the within-person effects of social media use on life satisfaction (Figure 4.2, Panel 2) and of life satisfaction on social media use (Figure 4.2, Panel 3), asking the questions: *Does an adolescent using social media more than they do on average drive subsequent changes in life satisfaction?* and *To what extent is the relation reciprocal?* Both median longitudinal effects were trivial in size (social media predicting life satisfaction: $\beta = -0.05$, life satisfaction predicting social media use: $\beta = -0.02$). The number of participants in each analysis can be found in Figures C.4-C.6.

When examining the range of possible specifications, the importance of gender became apparent: Only 16% of significant models in all three panels arose from male data. Across most models (Figure 4.2) the median between-person relation and within-person effects appeared more negative for females, hinting that gender plays an under-explored role in the influence of social media. I therefore conducted a focused analysis of my Preferred Model informed by the extant social media effects literature, statistical best practices, and the data characteristics.

Although some significant between-person correlations for both genders were in evidence (Figure 4.3, Panel 1), males showed only two significant longitudinal within-person effects: Social media predicted tenuous decreases in satisfaction with life and mean satisfaction (unstandardised bs = -0.08 to -0.04; standardised β s = -0.07 to -0.05; Figure 4.3, Panel 2). For females, however, social media was a predictor of slightly decreased life satisfaction across all domains excepting satisfaction with appearance (unstandardised bs = -0.13 to -0.05; standardised β s = -0.09 to -0.04; Figure 4.3, Panel 2). Furthermore, all domains of life satisfaction, excepting

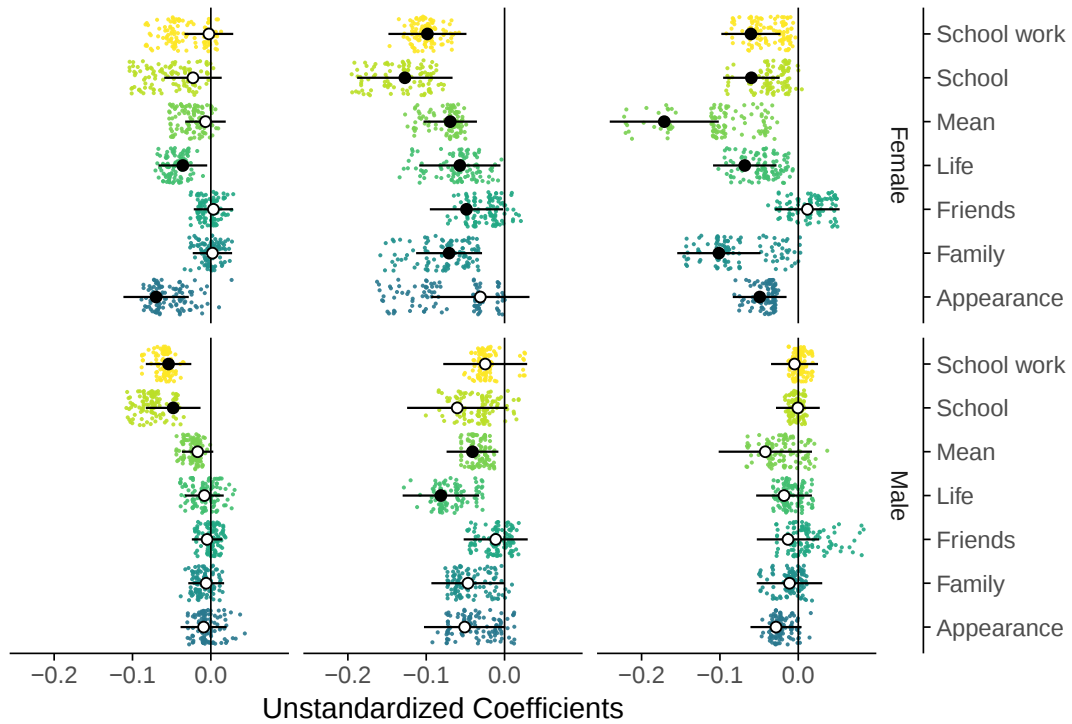


Figure 4.3: Relation between life satisfaction and social media use in female (top) and male adolescents (bottom). Left Panel: between-person correlations. Middle Panel: within-person effects of social media use on life satisfaction. Right Panel: within-person effects of life satisfaction on social media use. Small dots: results of each possible combination of theoretically defensible analytical decisions. Large dots: results of the Preferred Model (white = $p > 0.05$, black = $p < 0.05$). The life satisfaction categories examined are satisfaction with school work (school work), satisfaction with school (school), mean of the six life satisfaction items (mean), satisfaction with life (life), satisfaction with friends (friends), satisfaction with family (family) and satisfaction with appearance (appearance).

satisfaction with friends, predicted slightly reduced social media use (unstandardised $bs = -0.17$ to -0.05 ; standardised $\beta s = -0.11$ to -0.07 ; Figure 4.3, Panel 3).

4.4 Limitations

However, some caution is warranted: When comparing both genders the effects' confidence intervals overlap, and the reduction in significant effects in males alone is not evidence that the effects are substantial for females (Gelman and Stern 2006) – especially as they are very small in size. Importantly, the yearly interval between measurements in my dataset might not be optimal for testing reciprocal social

media effects, underlining how no single study can capture the full causal picture. I also highlight that, as detailed in Chapter 3, self-report measures only partially reflect the objective time adolescents spend engaging with social media (Scharkow 2016), yet they form the foundation of technological assessments included in the best quality datasets informing vital research in this area today.

4.5 Conclusion

The relations linking social media use and life satisfaction are therefore more nuanced than previously assumed: They are inconsistent, possibly contingent on gender, and vary substantively depending on how the data are analysed. Most effects are tiny – arguably trivial; where best statistical practices are followed, they are not significant in more than half of models. Yet some effects are worthy of further exploration and replication: There might be small reciprocal within-person effects in females, with increases in life satisfaction predicting slightly lower social media use, and increases in social media use predicting tenuous decreases in life satisfaction.

With the unknowns of social media effects still substantially outnumbering the knowns, it is critical that independent scientists, policymakers, and industry researchers cooperate more closely. Scientists must embrace circumspection, transparency and robust ways of working that safeguard against bias and analytical flexibility. Doing so will afford parents and policymakers with the reliable insights they need on a topic most often characterized by unfounded media hype. Finally, and most importantly, social media companies must support independent research by sharing granular user engagement data and participating in large-scale team-based open science. Only then will we truly unravel the complex constellations of effects shaping young people in the digital age.

4.6 Acknowledgements

This chapter is based on the published work *Orben, A., Dienlin, T. & Przybylski, A. K. (2019). Social Medias Enduring Effect on Adolescent Life Satisfaction.*

Proceedings of the National Academy of Sciences of the United States of America.

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service.

5

Discussion

Digital technologies are progressively shaping the lives that we live, and their power to do so will only increase in the foreseeable future. High quality scientific evidence is necessary to understand how these changes will affect us and our society. It would also ensure that critical stakeholders such as governments, regulators, designers, programmers, parents and digital technology users are provided with the information necessary to make informed decisions. Yet psychological research into how these new technologies influence us has a uniquely short-term and problem-based focus. It therefore lacks a methodological framework and proper quality control mechanisms. It is apparent from studying past technology panics that research in the area routinely fails to efficiently deliver answers to important and divisive research questions (Grimes et al. 2008).

These issues impact the current panic about digital technologies and adolescent well-being (Pew Research Centre 2018). Research in the area is clearly stagnant: something illustrated by the over 80 competing meta-analyses and systematic reviews about digital technology use's effect on psychological outcomes (Dickson et al. 2018). Yet the field remains popular. Large amounts of money and time have been invested to support technology research (House of Commons Science and Technology Select Committee 2019; Davies et al. 2019), and there is promise of more

funding to come. To date these investments have not engendered much research progress. After a decade of investigation, academics are still debating the answers to acutely simple and similar research questions, providing space for non-scientific commentators to dominate a public conversation that is searching for answers (Bell et al. 2015). Throughout this thesis I have been uncovering why this is the case, and what can be done to improve our approach to emergent technology research.

In this thesis' introduction I addressed the political, societal and academic factors that allow technology panics to repeatedly reincarnate and examined the reasons behind the lacklustre research progress in the field. In a 'Sisyphean cycle of technology panics' the lack of clear methodological and theoretical frameworks forces new groups of scientists to build their research area from scratch for each novel technology that gains popularity and needs to be considered. With the prospect of technological innovation accelerating in future (Valkenburg and Piotrowski 2017), psychology must provide better research at faster rates in order to play a constructive role in addressing, questioning and probing our inherent concerns about new technologies. If it fails to innovate, psychology might be left at the side-lines, becoming a mere accomplice to current and future technology panics.

This thesis therefore set out to provide a gold-standard methodological framework to guide research efforts in the area. It introduced novel approaches of interest to both the specific research field (Scharkow 2019) and academia more broadly (Nature 2019; Rai 2019). In writing this thesis, I aimed to make research on the psychological effects of digital technologies fit for purpose. The resulting work has already contributed substantially to the current understanding of digital technology and social media effects.

In this discussion, I will summarise these methodological improvements and highlight what my research allows us to conclude about digital technology and social media use's influence on adolescent well-being. I will then move on to consider the lack of a theoretical framework in the study of technology panics, and how this could be addressed to truly break the Sisyphean cycle of technology panics. Proposing improved approaches, this thesis therefore rewrites the broken rule-book currently

guiding research addressing emergent technologies. This kind of synthesis, which brings together improved methods and societally meaningful research questions, presents a way forward for psychological study in the area. This innovation will not only be of importance to academics, but also to funders, government officials and media organisations: both now and in the future.

5.1 Overview of Results

This thesis introduced more transparent analytical methods, diverse measurement and better data to the important and conflicting debate about adolescents and their relationship with digital technology and social media. It therefore set a new gold-standard for scientific evidence in the field and contributed to our current understanding of technology effects. In all, I analysed six high-quality datasets from the United Kingdom, Ireland and the United States to determine the extent to which the previous reviews' approximations of digital technology and social media use's associations with well-being of around $r = -0.15$ to $r = -0.10$ are accurate.

Chapter 2 introduced Specification Curve Analysis to improve the transparency of data analysis in the area. Instead of reporting one analysis pathway, it instead maps the results of all theoretically defensible analysis pathways available due to the garden of forking paths (Gelman and Loken 2014). While this presents no panacea for researcher degrees of freedom, it is a substantial improvement. Applied across three datasets, I found a wide range of possible associations between digital technology use and well-being. Researchers could have written thousands of different scientific articles about digital technology use and well-being using the same dataset, showing anything from overwhelmingly negative to substantially positive relations. In the three different datasets examined, the median correlations found linking well-being and digital technology use ranged from $r = -0.04$ to $r = -0.01$. The relations became less negative if only specifications with control variables were included. To put the effect sizes into perspective, I created and applied a new method called comparison specifications which contrasts the association between digital technology

use and well-being to the associations of other activities with well-being. I found that wearing glasses was 1.45 times more negatively associated with adolescent well-being than digital technology use, demonstrating that the associations reported were exceedingly small. This conclusion was supported by other comparisons with activities such as eating potatoes or attributes such as height.

To provide a more transparent overview of effects I also focused on measurement. Chapter 3 diversified the measurements of digital technology use currently found in large-scale scientific work by including both time-use diary and retrospective measures of the activity. These measures did not correlate to a great degree, giving further weight to the concerns that the quality of measurement in the area is low (Scharkow 2016). To examine the relation between digital technology use and well-being, Chapter 3 applied the same Specification Curve Analysis framework as Chapter 2, but also explicitly differentiated between exploratory and confirmatory tests. The study was one of the first to examine digital technology use before bed-time and its association with well-being. The inconclusive evidence found for such a link, highlighted how removed some of the public debate is from the current state of scientific evidence.

Adolescent well-being was most negatively correlated with the retrospective digital technology use measures routinely used throughout the research area ($r = -0.15$ to $r = 0.01$). Like in Chapter 2, the range of correlations found neared zero (and even found positive relations in some of the datasets examined), and the most negative median correlation was close to the correlations reported in previous systematic reviews. Yet examining other measures of digital technology use provided different results. Well-being was less correlated with total time spent engaging with digital technologies, a measure calculated by summing the time on digital technologies reported in the time-use diaries, both on weekdays ($r = -0.06$ to $r = -0.01$) and weekends ($r = 0.00$ to $r = 0.06$).

Chapter 4 conceptually replicated and extended the inquiry pursued in previous chapters by differentiating within- and between-person relations through the analysis of a multi-wave longitudinal dataset. This allowed me to focus on the

potential bidirectionality of social media effects. The cross-sectional between-person association between social media use and life satisfaction was $r = -0.13$. This effect size was in line with results of previous reviews, which predominantly synthesised cross-sectional evidence. Such evidence however cannot provide an indication about whether a person increasing or decreasing their social media use will experience changes in well-being in the long-term. Chapter 4 provided evidence that such within-person longitudinal effects are much smaller than their between-person cross-sectional counterparts. An increase in a participant's social media use in one year, compared to their own average, predicted a small decrease in their life satisfaction one year later ($\beta = -0.05$). A decrease in a person's life satisfaction in one year, compared to their own average, predicted an increase in social media use one year later ($\beta = -0.02$). Yet the effects are small, and the inclusion of control variables decreased the effect sizes found. The study also highlighted a possible gender difference, with girls showing more significant specifications than boys.

5.2 Overestimation

This thesis' research therefore provided evidence that previous studies possibly overestimated the association between digital technology use and well-being. Firstly, the cross-sectional relations found when averaging the results of Specification Curve Analyses were smaller than those summarised by previous review papers. While review paper estimates of the association were routinely in the range of $r = -0.15$ to $r = -0.10$, my studies found effect sizes around about $r = -0.10$ to $r = -0.05$. In other words, the strength of the association I found linking technology use to well-being outcomes using rigorous methodological approaches was roughly one third smaller than the size suggested by the literature. These differences could be due to the more transparent analytical practices used; especially as previous papers have been shown to implement certain analytical decisions that result in a larger than average effect of interest (Twenge, Joiner, et al. 2017). Such exhibits of biased practices also discredit the argument that the effect sizes found using Specification Curve Analysis were only less negative because 'worse' analytical

practices were used. The results found in my thesis therefore question whether systematic reviews and meta-analyses really provide the best quality evidence in the area: they are only as good as the sum of their parts.

Secondly, Chapter 3 showed that retrospective measures of digital technology use are more negatively correlated with well-being than other measures devised from time-use diaries. The vast majority of the current research relies solely on such retrospective self-report measures, and if they do overestimate associations, this would create a systematic bias in the literature. It is necessary to note here, however, that time-use diaries have not undergone validation as a measure of digital technology use. This should however not quiet concerns about the field's current over-reliance on a single type of digital technology measure that has routinely been shown to be biased (Wilcockson et al. 2018). As these measures might also be overestimating the negative associations between digital technology use and well-being, my thesis highlighted the immediate need for the research field to improve its measurement approaches.

Finally, in Chapter 4, I found that the between-person cross-sectional relations between social media use and life satisfaction were more negative than the within-person longitudinal effects of social media use on life satisfaction and vice versa. Again, as most evidence in the area is based on cross-sectional data, this indicates that past work could have overestimated the link between social media use and well-being outcomes. Furthermore, taking into account the bidirectionality of effects in psychological research is crucial (Kievit et al. 2013), especially as the use of observational data does not allow for a clear differentiation of cause and effect (Rosenbaum 2017). Current cross-sectional results, which can only tell us about between-person relations, are often incorrectly interpreted in a longitudinal framework, misleading large proportions of the literature and the public. With psychology increasingly moving towards the use of random-intercept models that can differentiate within- and between-person effects, research on technologies should adopt such methodological innovations rapidly to improve its inferences.

5.3 Interpretation

Taken as a whole, the results derived from the investigations detailed in this thesis suggest that previous systematic reviews overestimated the negative association between digital technology use and adolescent well-being. The associations found in this thesis hint at a smaller negative correlation, which nears zero insofar as high-quality control variables are included or when within-person effects are examined. This raises two questions: 1) is this small association obtaining an outsized share of public attention? and 2) is the size of the association proof that we do not have to worry about new technologies like social media affecting teenagers?

The first question is more difficult to answer. While I used various methods to examine whether a statistically significant effect was practically significant, judging what effect sizes are important is still fraught with difficulty. While comparison specifications (Chapter 2), Smallest Effect Sizes of Interest (Lakens, Scheel, et al. 2018, Chapter 3) and traditionally used cut-offs (J. Cohen 1992, Chapter 4) supplied three windows into interpreting the importance of various effects, they could not provide the whole picture. Furthermore, the picture was not overarchingly clear (Funder and Ozer 2019). While I found that wearing glasses was more negatively associated with well-being than digital technology use in Chapter 2, the same data found that activities like smoking marijuana were only 1.14 times more negatively associated with well-being than digital technology use as well. It is inherently difficult to specify what effect sizes are important, and there is still no universally agreed method to approach this issue. If a whole population is affected, should we care about a tiny effect (Rose 2008)? Policy and behavioural change are costly and time consuming, so should we not focus on those effects that are most substantial (Ferguson 2009)?

As little work has been done in the area, my approaches are open to criticism and conflicting interpretations. Further research could help provide the necessary guidance to judge the effect sizes found. For example, one method to determine a smallest effect size of interest might be to examine the smallest possible difference in

the outcome that participants would notice (Bauer-Staeb et al. 2019). For example, I could implement such a method for the median effects found analysing the MCS dataset in Chapter 3, assuming that well-being needs to decrease by 0.50 standard deviations for participants to take note. Adolescents who reported using technology would need to report 63 hours and 31 minutes more technology use a day in their time-use diaries to perceive such a decrease. The above calculation is based on the median of calculated effect sizes, however, if we only consider the specification with the maximum effect size, the time an adolescent needs to spend on technology to experience the relevant decline in well-being decreases to 11 hours and 14 minutes. The 0.50 standard deviation in change is often referenced as a cutoff for those effects that participants would become subjectively aware of (G. A. Miller 1956; Norman et al. 2003). Yet it has not been directly tested on the questionnaires and populations used in this study. Furthermore, it is still up for debate whether smaller effects, even when not noticeable, are still important because digital technology is used by a large majority of the population (Rose 2008). Such problems are evident throughout psychology, as multiple efforts are currently being made to help researchers when choosing smallest effect sizes of interest (Anvari and Lakens 2019), interpreting effects sizes (Funder and Ozer 2019) and visualizing their effects (Ho et al. 2019).

Whether the effect sizes found are practically significant or not, this thesis does not provide evidence that adolescents are not affected by digital technology or social media use. Instead, it highlights that when all uses of digital technology and all adolescent users are averaged across a large population there is little to no relation between digital technology use, social media use and long-term well-being outcomes. Yet, as seen in Chapter 4, that does not preclude that there isn't a certain subset of adolescents that are affected – either positively or negatively – by digital technology and social media use. The research area therefore needs to invest more time into researching individual differences, highlighting those that are most affected by digital technologies, instead of trying to average across a very diverse pool of adolescents. Preliminary research has highlighted, for example, that the negative effects of digital technology use on psychological well-being might be

greatest for those adolescents least privileged in society (Odgers 2018) or adolescent girls (Twenge, Joiner, et al. 2017). Furthermore, there might be certain uses of technology that are more harmful than others (Burke, Marlow, et al. 2010), so averaging across all types of digital technology or social media use might hide such effects. This can only be addressed once detailed trace data stored by technology companies become accessible to academics, something discussed further in the next section. This thesis therefore provides the necessary transparent and robust evidence to show that the relation between time spent using digital technologies and adolescent well-being is smaller than previously expected, and that the research field needs to redirect its research questions to find more coherent results. In particular, research should start taking into account the complex interplay of risk factors, technology uses and individual differences that makes each adolescent unique.

5.4 Other Limitations

It is important to acknowledge some further limitations of the work presented in this thesis before examining potential next steps. While most substantial limitations were noted throughout the last chapters, I would like to refocus attention on measurement.

How to quantify concepts represents an enduring problem to the psychological discipline. This thesis utilized high-quality secondary datasets that enabled me to make more robust inferences without having to find the millions of pounds required to collect the necessary data (Understanding Society 2018). Yet the nature of secondary data meant that some measures were not harmonized across datasets, even when impressive amounts of harmonization were achieved in cases like the time use diaries analysed in Chapter 3. The digital technology use measures were the most diverse. They queried different technology uses, over different time frames, and in different years, while technology could have developed or changed in the meantime. Previous reviews have however used similar data to make overarching observations about emergent technologies (Dickson et al. 2018), and the aim of this thesis was to do the same. I therefore decided to strike a balance by using the

best quality secondary data available but acknowledging the need to exert care in generalizing between the variety of technology measures present.

The nature of measurements available was also a limiting factor in my research. While Chapter 3 introduced more diverse procedures to address the low-quality of technology use measures available in the area, this thesis still predominantly relied on self-report questionnaires to measure digital technology use. Such measures are known to have inherent limitations and biases (Andrews et al. 2015; Wilcockson et al. 2018). Furthermore, they constrain the research questions that I was able to ask; limiting my investigation to examining time spent using digital technologies. This puts questions about specific uses of digital technologies out of the scope of this thesis, even though they represent promising avenues for research. My colleagues and I have therefore called for more collaboration between researchers, policy makers and technology companies throughout my doctorate. Such collaborations would need to be preceded by an in-depth discussion with diverse stakeholders about the ethics and transparency of digital technology trace data sharing. Once agreed and accessed, however, such resources would allow academics to answer crucial research questions still untouched by the scientific field.

Lastly, I note throughout this thesis that high quality control variables are important and lessen the negative associations found between digital technology use and well-being. The inclusion of such control variables allowed me to account for an adolescent's environment, to ultimately determine a more direct relationship between digital technology use and well-being. This is because environmental factors can affect both variables of interest. Research has shown that disadvantaged teenagers use more digital technology (Pew Research Centre 2018) and also that they score worse on well-being questionnaires (The Children's Society 2018). We would therefore expect a negative correlation between digital technology use and well-being even without any underlying causal structure linking the two (Odgers 2018). Moving forward, improved experimental and causal modelling approaches will be necessary to further understand this relationship (Pearl and Mackenzie 2018). While backed up by evidence, some of the control variables I used can be

disagreed upon, for example adolescents' closeness to the primary caretaker or the time the primary caretaker spends with the adolescent. With little theory to inform control variable inclusion, such variables can either be seen as important third-factors that are necessary to control for or outcomes that if controlled for will confound results. Currently my solution is transparency: I report both the results with and without controls and provide the code and materials necessary for researchers to repeat and alter my analyses. My methods and approaches therefore try to address the limitations I have detailed throughout this thesis, but many of those limitations represent problems in the psychological discipline that still need to be addressed more broadly.

5.5 Next Steps

While this thesis does much to increase the efficiency and transparency of research into emergent technologies, it does not provide a solution, in and of itself, to fully disrupt the continuous Sisyphean cycle of technology panics. The cycle is fuelled by more than researcher degrees of freedom and publication biases. For this reason, the solution requires more than improved methodological frameworks. Academics have little power to change the governmental or scientific system, so ideas around reforming political and academic incentives are out of the scope of this thesis. The development of better theories and theoretical approaches could also lessen the impact of the Sisyphean cycle of technology panics. A strong theoretical base would make it possible to integrate research on older technologies into more current research considering recent technological developments. New technologies would challenge previous theories, requiring them to be revised but not replaced with a whole new theoretical framework. It would be a momentous task, however, to reverse-engineer a complete theoretical framework for an established research area. Furthermore, as research into technological innovations is mainly funded to solve a concrete practical problem, such a long-term theoretical task will be difficult to find the necessary political and academic backing for (Grimes et al. 2008).

Another more viable approach is to streamline and focus current psychological research from the outset, so that the work done to investigate emergent technologies has enough nuance to provide cohesive and replicable outcomes. Such a framework for technology research would ensure that work is produced at the highest possible standard and that the research area learns from the issues that have plagued it in the past. It would allow research to provide the robust raw empirical insights needed to develop meaningful theory and build a more cohesive field, focusing on the depth and not the breadth of claims (Kaelin Jr 2017). With a more detailed focus, it would also open up opportunities to develop new methodological approaches to promote triangulation, a process that could substantially improve inferences in the area (Munafò and Davey Smith 2018). The creation of such a framework would be an initiative worthy of large-scale collaboration, and could in time attract the attention of funders, media organisations and governments.

I have spent the last year of my DPhil reflecting on what form such a framework could take. I will detail a preliminary version of such a framework below. My *UnITED framework for technology research* outlines aspects that should be considered in all future research and funding applications investigating emergent technologies. It therefore also provides an outline for what research should be conducted next to address the concerns about digital technologies and social media.

5.5.1 The UnITED Framework of Technology Research

Research on emergent technologies should consider the following factors:

1. *Unique use*: Research should not investigate the effects of a technology as a whole (e.g. digital technology use or social media use) but should focus instead on a unique use that sets the specific technology of interest apart. It should also consider the mechanisms necessary for this unique use to affect the outcome of interest. For example, if interested in social media use's effect on well-being, research should focus on an aspect or feature of social media use that makes it unique (for example, the non-direct nature of

communication, Altman and Taylor 1973). Furthermore, if using a theory like the displacement hypothesis to explain a possible mechanism linking social media use and decreased well-being, research needs to consider what activity social media use can be displacing.

2. *Individual*: Research should specify a specific population when making claims about emergent technology effects. Furthermore, the work should clearly demarcate whether it is examining within-person or between-person effects. In the process, the researcher should highlight how they attempted to control for other individual factors that could affect both the technology use measure and the outcome of interest.
3. *Time frame*: In a longitudinal study, the time frame of the effects measured should be specified and the theoretical impact should be discussed. A study should clearly communicate whether, for example, it examined the changes predicted by social media one minute after use, rather than one year after use, and why such a time scale is theoretically plausible.
4. *Effect size*: The study should report the size of the effect or association that it investigated and provide an interpretation of what the size means for stakeholders. The key question is: statistical significance aside, why should I believe this effect is important? While this thesis used frameworks like comparison specifications or smallest effect sizes of interest, there are multiple other approaches to providing such information.
5. *Direction*: Cross-sectional studies on emergent technologies should acknowledge that media effects are inherently bidirectional, and that correlational work is unable to decipher which direction the effects of interest are in. Longitudinal work should interpret and highlight the bidirectional paths of the effects of interest.

If implemented, this framework could improve the quality of research and substantiate the depth of understanding about social media and digital technology

effects. The UNITED framework sets out areas that need to be addressed to ultimately promote the production of robust evidence capable of influencing and informing academia, the public and policy. The framework is not the only component that needs to change to stop the Sisyphean cycle of technology panics, but its implementation would add another dimension to research done in the area. Paired with a widespread adoption of the methodological framework presented in this thesis, it has the potential to support a more critical and scientific probing of technology effects in the psychological discipline. This would move psychology away from being an accomplice of technology panics and allow it to become an integral scientific actor that provides debates about emerging technologies with robust and improved scientific evidence.

5.6 Conclusion

Science's aim is ultimately to find truth. Researchers are expected to do their best to understand complexity and they are expected to sort the meaningful signals of nature from the background of statistical noise. Ever too often, the psychological study of emergent technologies loses sight of this ultimate goal. There are biases and inclinations that all humans share, which make us cautious about new technologies and their potential to change society. There are opportunities to sell books, give talks, be seen and heard, not just by academia but by society as a whole. There is a need for researchers to sustain themselves in a time of scientific industrialisation and increasing precariousness. The lack of a methodological framework in this research area – created on the basis of a societal problem, not a scientific theory – means there is little support and structure in place. Scientific progress is slow, and the research output produced is routinely conflicting. This makes providing clear advice to policy and the public difficult. There is however little impetus for the field to reflect about its own methodology and the apparent problems with it. Yet “the first principle [of science] is that you must not fool yourself – and you are the easiest person to fool” (Feynman 1974).

To not fool ourselves, and to ensure that psychology does not become an accomplice to a never-ending Sisyphean cycle of technology panics, the research area has to recognise the need for change. Psychology's credibility and reproducibility revolution has established multiple methodological improvements that are applied and developed in this thesis to make research more efficient and transparent. This improved methodological framework can advance the way we investigate emergent technologies. Through their transparency and openness, the studies presented in this thesis have progressed our understanding about adolescents and digital technologies, having already informed and shaped policy proposals (House of Commons Science and Technology Select Committee 2019; Davies et al. 2019). Yet their focus on presenting a new gold-standard approach to the research area could prove even more impactful. If adapted further, a revised methodological framework for the discipline could empower psychology to steer predictable public concerns about emergent technologies into a more productive and evidence-based future.

Appendices



Appendix A: Improving Transparency

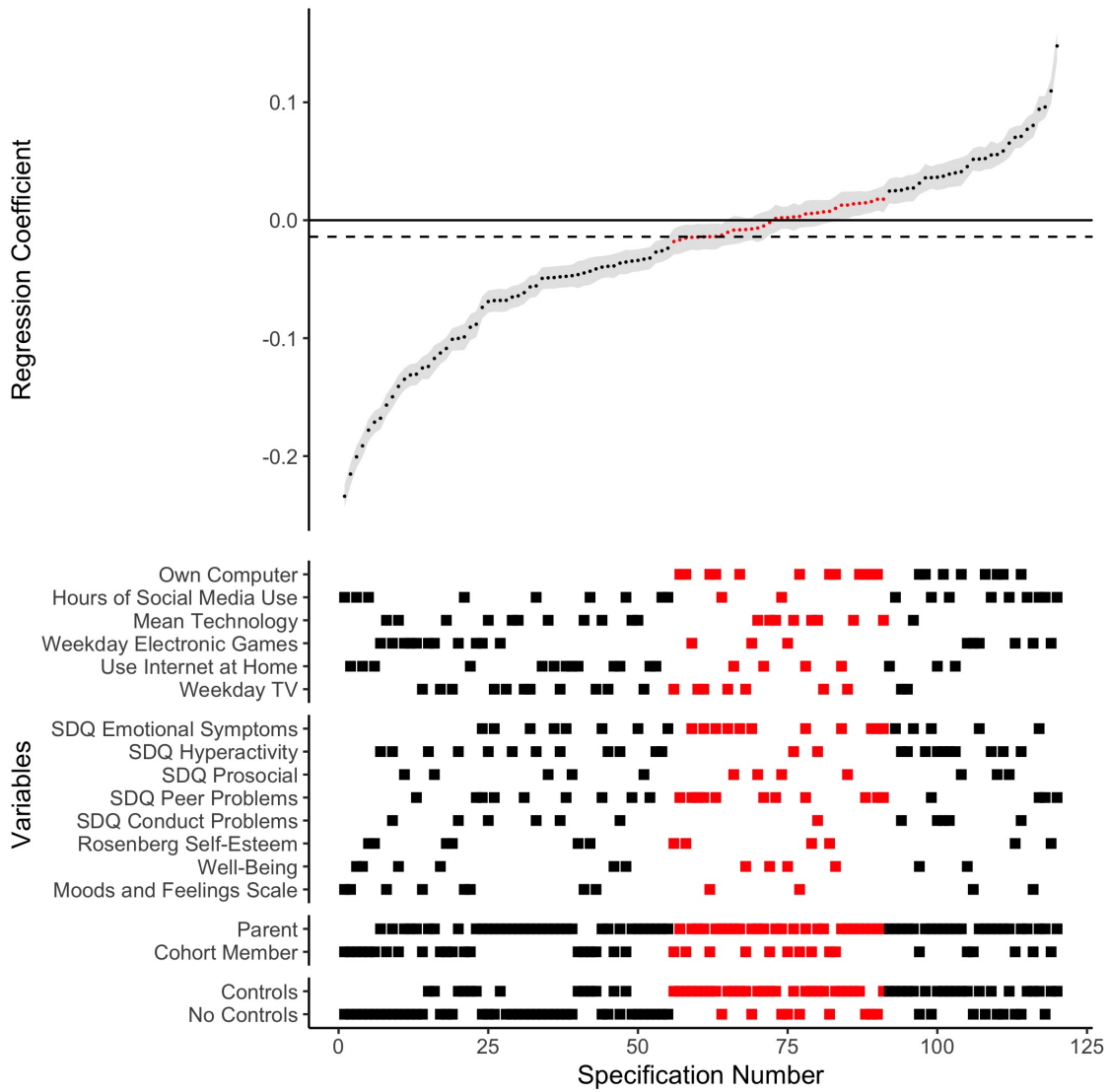


Figure A.1: Specification Curve Analysis of MCS examining net effects across prespecified well-being subscales. The red squares indicate specifications that were non-significant, while the black squares represent specifications that were significant. The error bars represent the Standard Error. The dotted line indicates the median regression coefficient present in the SCA and is included to ease interpretation of the analysis.

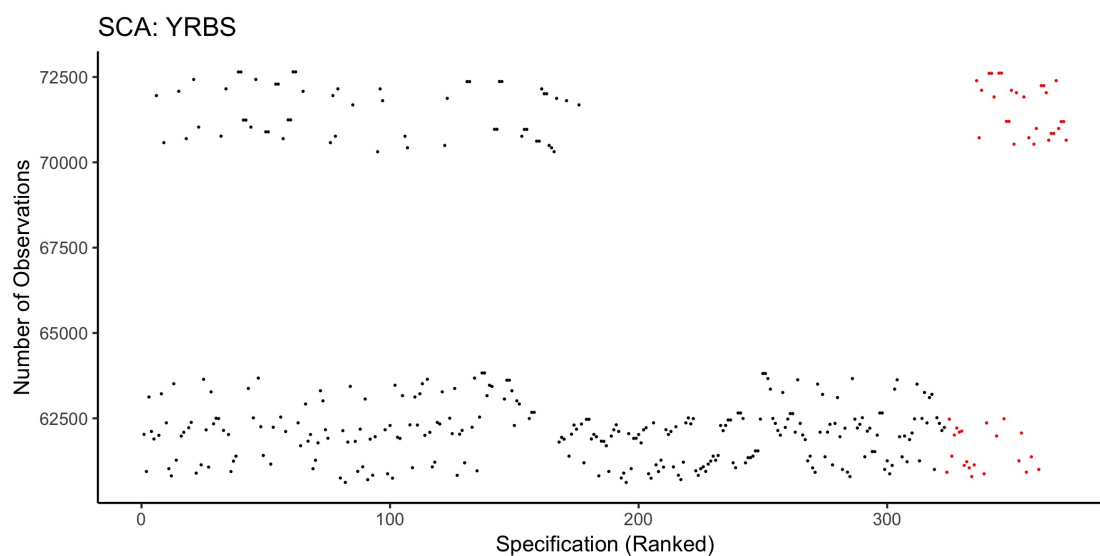


Figure A.2: Number of observations (participants) for each specification analysed in the YRBS SCA. Red dots indicate when the specification was non-significant, while black dots show significant specifications.

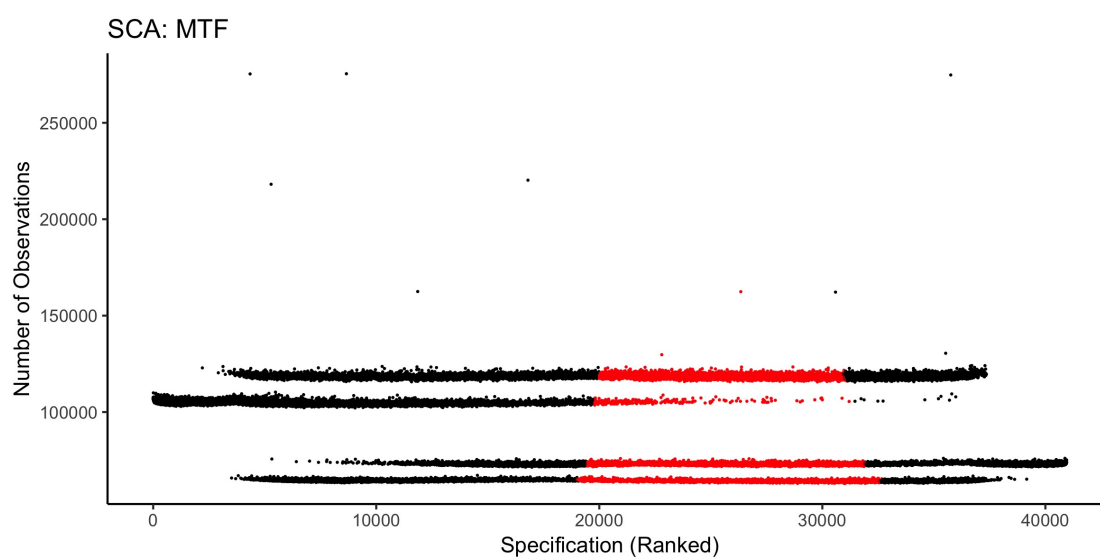


Figure A.3: Number of observations (participants) for each specification analysed in the MTF SCA. Red dots indicate when the specification was non-significant, while black dots show significant specifications.

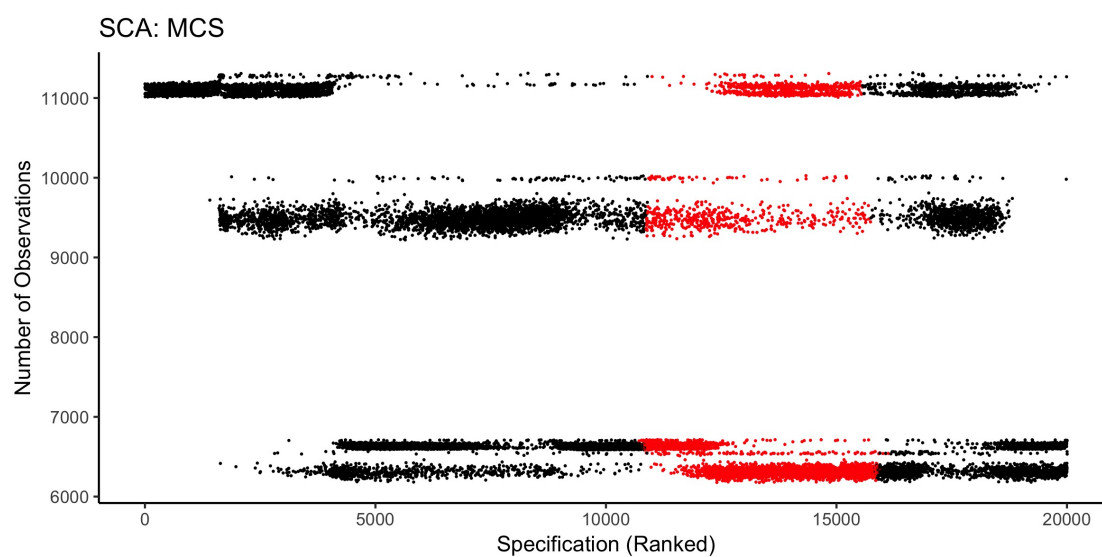


Figure A.4: Number of observations (participants) for each specification analysed in the MCS SCA. Red dots indicate when the specification was non-significant, while black dots show significant specifications.

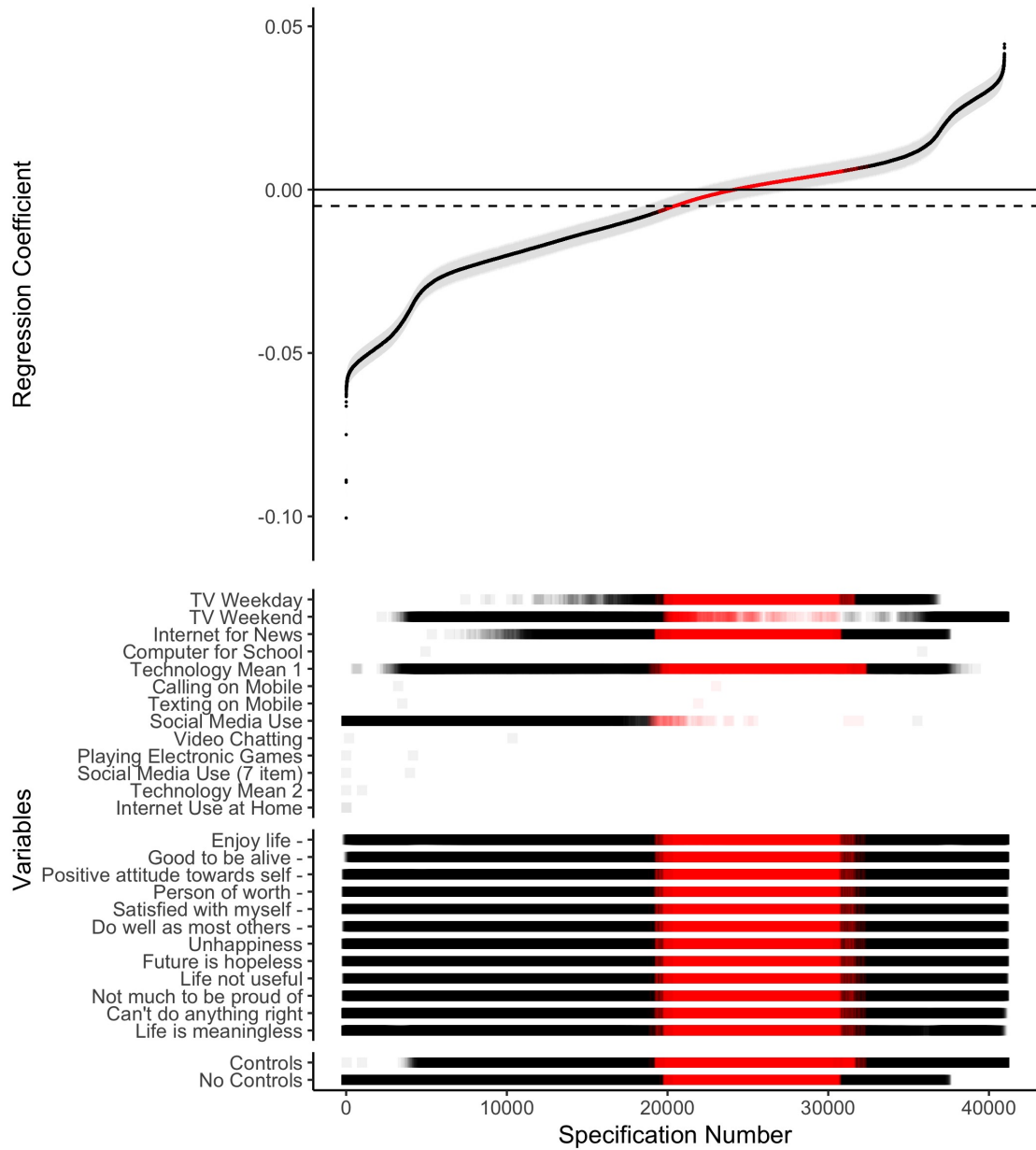


Figure A.5: Specification Curve Analysis of MTF, also including the few specifications that included technology use items that were only asked in conjunction with a one-item happiness question. This subset of questions was excluded from the analyses in the main text to simplify the graphical representations and enhance comprehensibility. The red areas indicate specifications that were non-significant, while the black areas represent specifications that were significant. The error bars show the Standard Error. The dotted line indicates the median Regression Coefficient present in the SCA, and is included to ease interpretation of the analysis.

B

Appendix B: Improving Measurement

Table B.1: Significance Tests for Growing up in Ireland: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.

Technology Measure	# results in dominant direction	p value
Self-Report	4	0.13
Participation (Weekday)	3	0.68
Participation (Weekend)	4	0.11
Time Spent (Weekday)	4	0.13
Time Spent (Weekend)	4	0.13
< 30 min (Weekday)	4	0.13
< 30 min (Weekend)	4	0.12
< 1 hour (Weekday)	2	1.00
< 1 hour (Weekend)	2	1.00
< 2 hours (Weekday)	2	1.00
< 2 hours (Weekend)	2	1.00

Table B.2: Significance Tests for the Panel Study of Income Dynamics: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.

Technology Measure	# results in dominant direction	p value
Self-Report	2	1.00
Participation (Weekday)	3	0.63
Participation (Weekend)	2	1.00
Time Spent (Weekday)	4	0.14
Time Spent (Weekend)	3	0.63
< 30 min (Weekday)	2	1.00
< 30 min (Weekend)	4	0.12
< 1 hour (Weekday)	2	1.00
< 1 hour (Weekend)	4	0.11
< 2 hours (Weekday)	4	0.13
< 2 hours (Weekend)	4	0.12

Table B.3: Significance Tests for the Millennium Cohort Study: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.

Technology Measure	# results in dominant direction	p value
Self-Report	5	0.24
Participation (Weekday)	6	0.03
Participation (Weekend)	4	0.73
Time Spent (Weekday)	6	0.04
Time Spent (Weekend)	4	0.67
< 30 min (Weekday)	5	0.23
< 30 min (Weekend)	4	0.73
< 1 hour (Weekday)	5	0.21
< 1 hour (Weekend)	5	0.21
< 2 hours (Weekday)	6	0.03
< 2 hours (Weekend)	5	0.23

Table B.4: Significance Tests for the confirmatory hypotheses: the number of results in the dominant direction for the original SCAs and the corresponding p value of the significant test.

Measure	# results in dominant direction	p value
Self-Report	5/6	0.24
Time Spent	5/6	0.04
< 30 Minutes on Weekday	5/6	0.23
< 1 Hour on Weekday	5/6	0.21

C

Appendix C: Improving Data

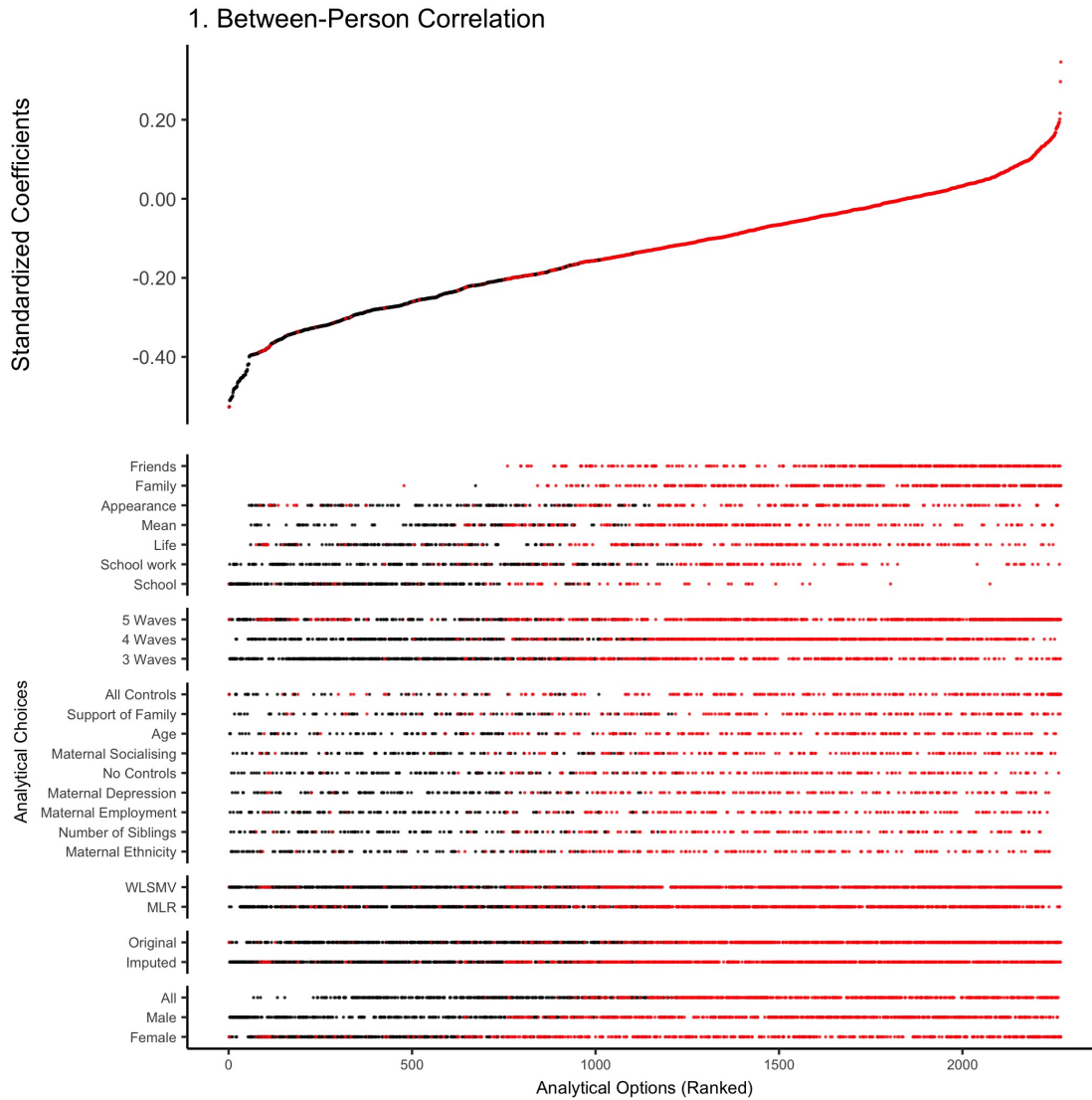


Figure C.1: Between-Person Correlations: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard shows all the analytical decisions.

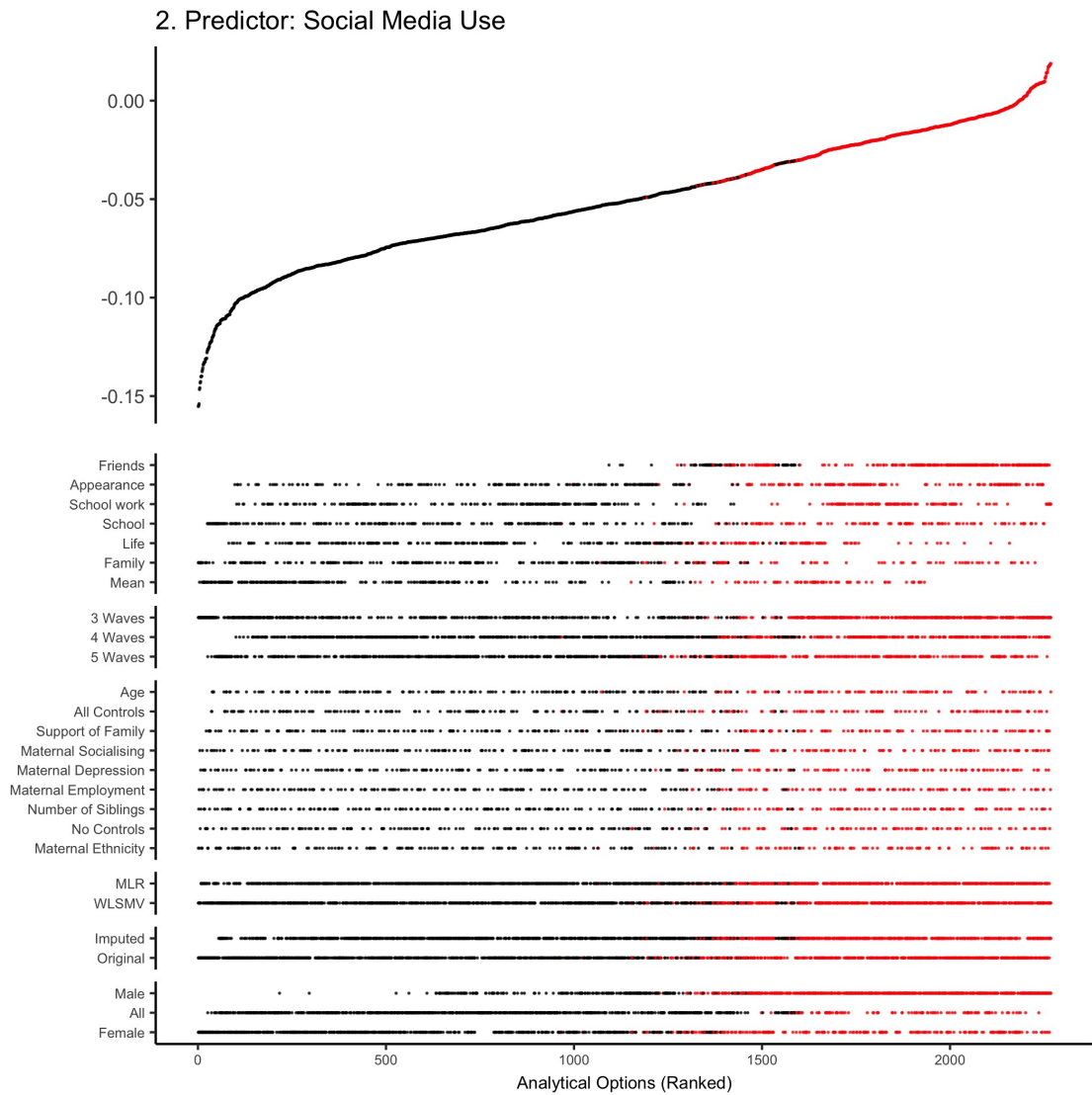


Figure C.2: Within-Person Effect of Social Media Use predicting Well-Being: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard shows all the analytical decisions.

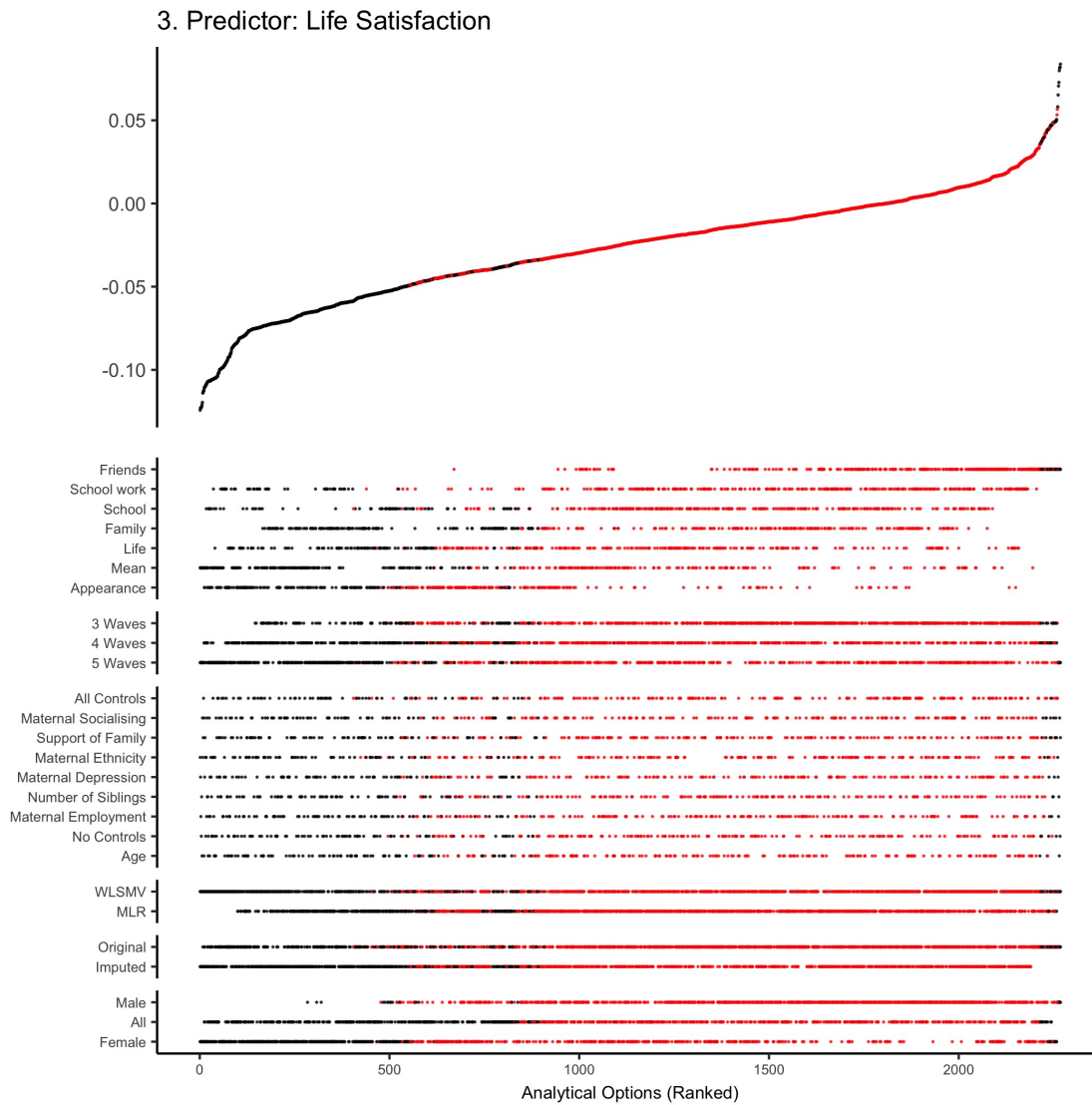


Figure C.3: Within-Person Effect of Well-Being predicting Social Media Use: Results of a Random-Intercept Cross Lagged Panel Model Specification Curve Analysis relating social media use and life satisfaction. Each point on the x-axis represents a different combination of analytical decisions (i.e. life satisfaction domain, gender, number of waves, estimator, data imputation and control variables). The dashboard shows all the analytical decisions.

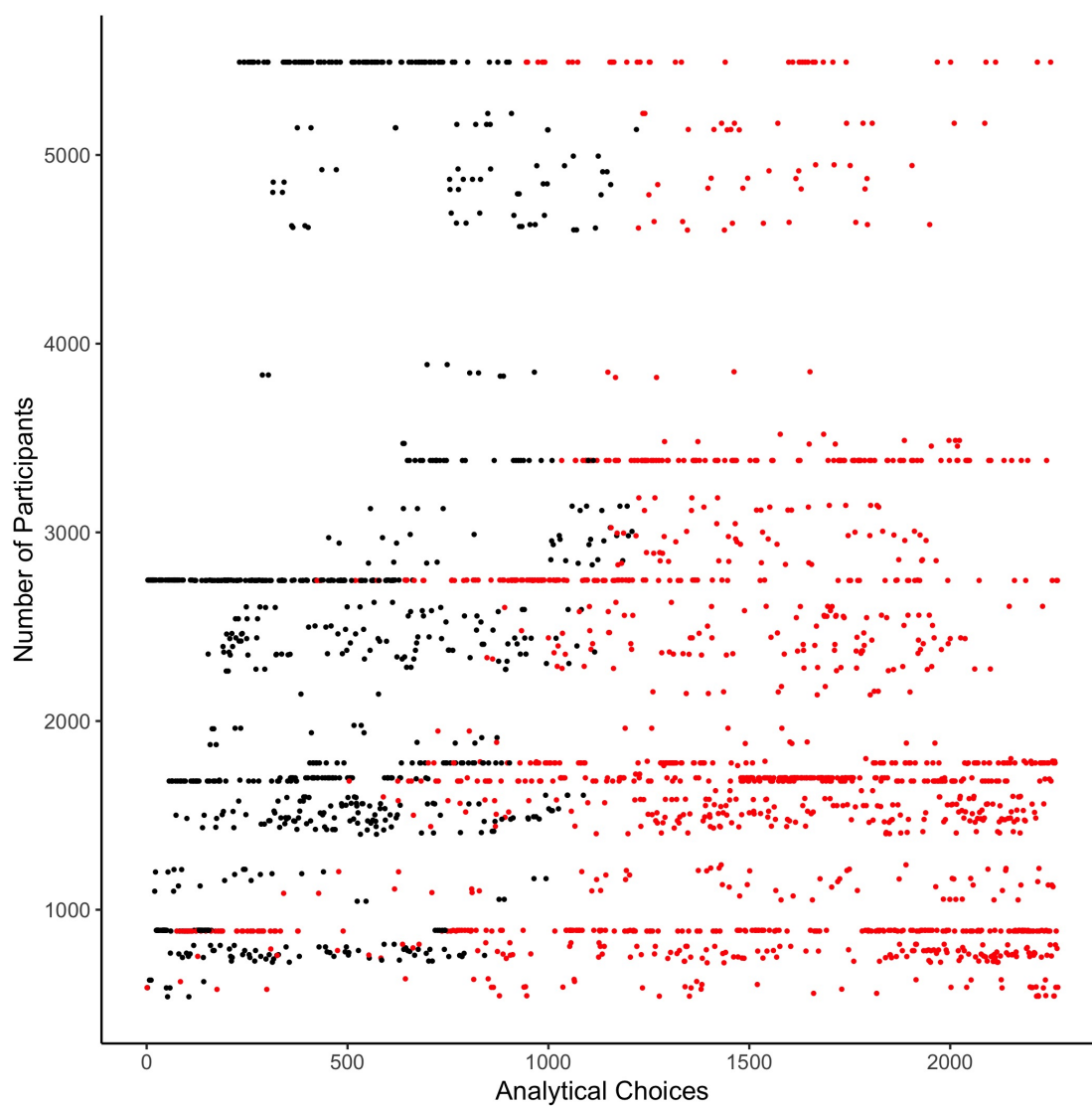


Figure C.4: Number of observations (participants) for each specification analysed in the Random Intercept Cross Lagged Panel Model. This graph shows the between-person correlation. Red dots indicate when the specification was non-significant, while black dots show significant specifications.

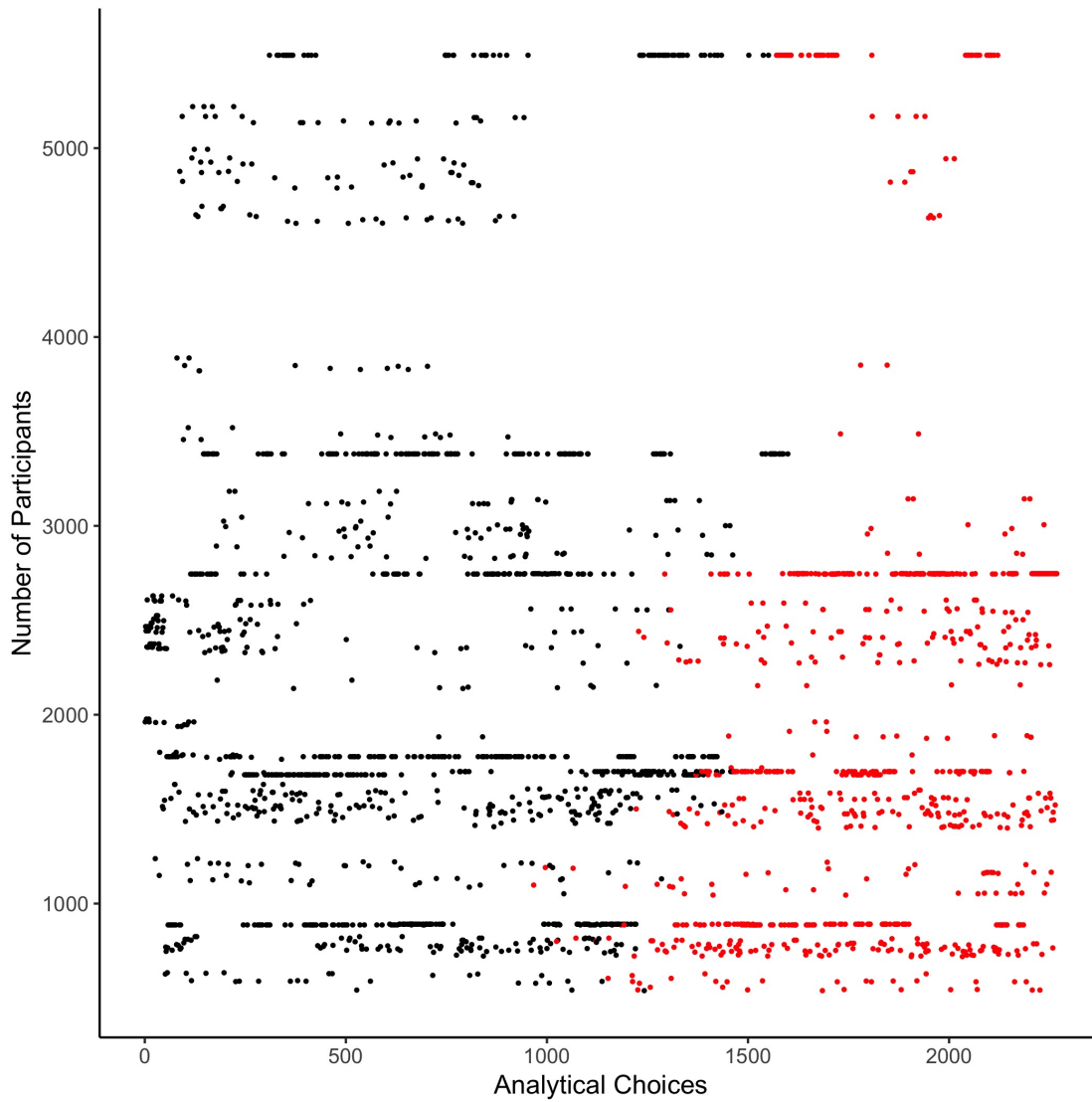


Figure C.5: Number of observations (participants) for each specification analysed in the Random Intercept Cross Lagged Panel Model. This graph shows the within-person effect of social media use predicting well-being. Red dots indicate when the specification was non-significant, while black dots show significant specifications.

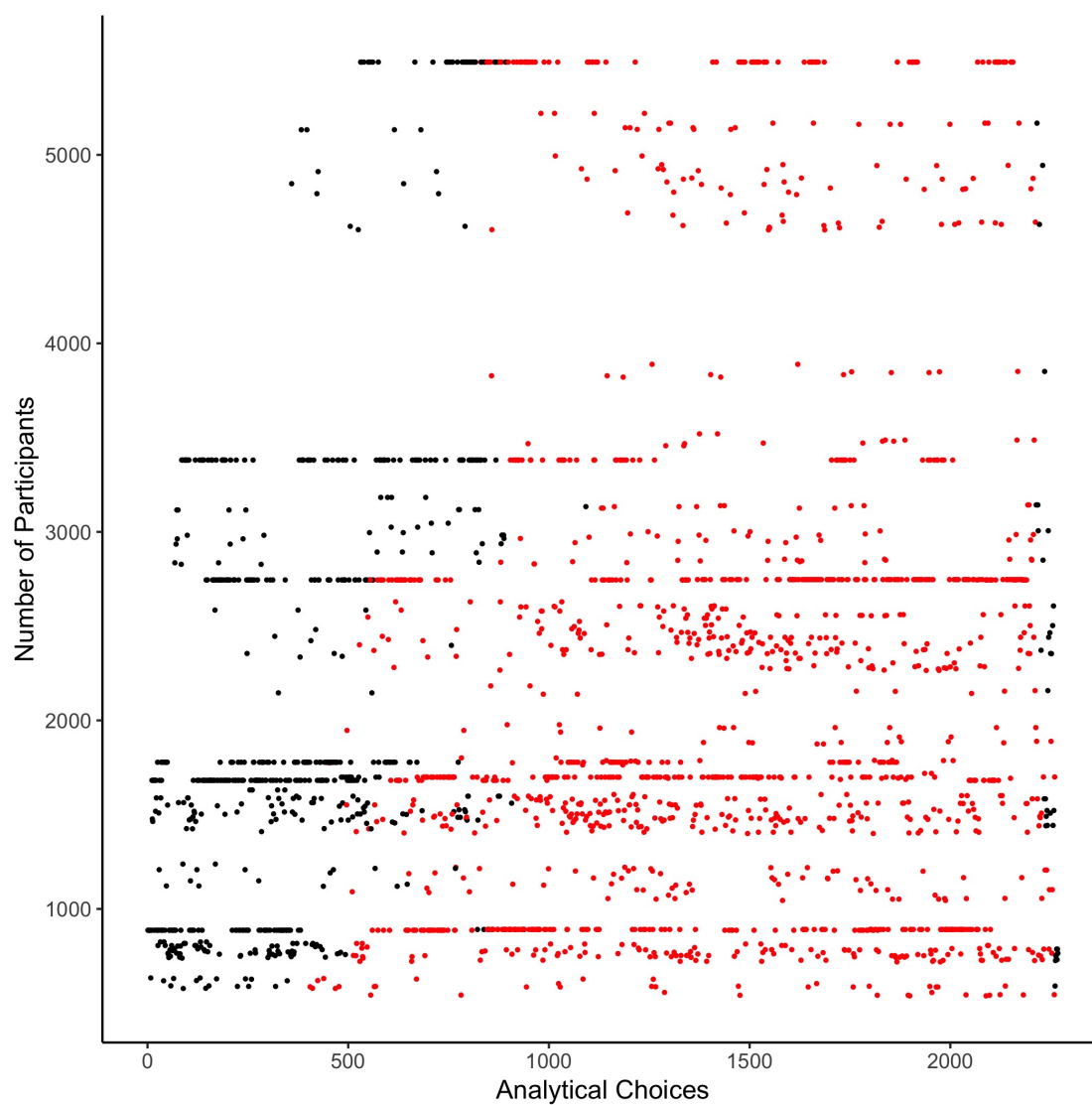


Figure C.6: Number of observations (participants) for each specification analysed in the Random Intercept Cross Lagged Panel Model. This graph shows the within-person effect of well-being predicting social media use. Red dots indicate when the specification was non-significant, while black dots show significant specifications.

References

- Aalbers, George et al. (2018). "Social media and depression symptoms: A network perspective." In: *Journal of Experimental Psychology: General*.
- Allcott, Hunt et al. (2019). "The Welfare Effects of Social Media". URL: <https://web.stanford.edu/~gentzkow/research/facebook.pdf>.
- Altheide, David L. (2002). *Creating Fear: News and the Construction of Crisis*. Transaction Publishers, p. 223.
- Altman, Irwin and Dalmás A. Taylor (1973). *Social Penetration: The Development of Interpersonal Relationships*. New York, NY: Holt, Rinehart and Winston.
- American Psychological Association (2010). *Publication manual of the American Psychological Association*. 6th Edition. American Psychological Association, p. 272.
- Andersen, Susan L. and Martin H. Teicher (2008). "Stress, sensitive periods and maturational events in adolescent depression". In: *Trends in Neurosciences* 31.4, pp. 183–191.
- Anderson, James A. (2008). "The Production of Media Violence and Aggression Research: A Cultural Analysis". In: *American Behavioral Scientist* 51.8, pp. 1260–1279.
- Anderson, Vicki et al. (2005). "Functional plasticity or vulnerability after early brain injury?" In: *Pediatrics* 116.6, pp. 1374–82.
- Andrews, Sally et al. (2015). "Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use". In: *PloS ONE* 10.10, e0139004.
- Angold, Adrian et al. (1995). "Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents." In: *International Journal of Methods in Psychiatric Research* 5.4, pp. 237–249.
- Anvari, Farid and Daniël Lakens (2019). "Using Anchor-Based Methods to Determine the Smallest Effect Size of Interest". URL: <https://psyarxiv.com/syp5a/>.
- Araujo, Theo et al. (2017). "How Much Time Do You Spend Online? Understanding and Improving the Accuracy of Self-Reported Measures of Internet Use". In: *Communication Methods and Measures* 11.3, pp. 173–190.
- Ariès, Philippe (1960). *Centuries of Childhood*. Harmondsworth: Penguin Books Ltd.
- Atkin, Charles et al. (1979). "Selective exposure to televised violence". In: *Journal of Broadcasting* 23.1, pp. 5–13.
- Bauer-Staeb, Clarissa et al. (2019). *The Minimal Clinically Important Difference on the Generalised Anxiety Disorder Scale (GAD- 7)*. URL: <https://researchportal.bath.ac.uk/en/publications/the-minimal-clinically-important-difference-on-the-generalised-an>.
- Baumeister, Roy F. and Mark R. Leary (1995). "The need to belong: Desire for interpersonal attachments as a fundamental human motivation." In: *Psychological Bulletin* 117.3, pp. 497–529.
- Bayer, Joseph et al. (2018). "Facebook in context(s): Measuring emotional responses across time and space". In: *New Media & Society* 20.3, pp. 1047–1067.

- Beck, Ulrich and Elisabeth Beck-Gernsheim (2002). *Individualization: Institutionalized individualism and its social and political consequences*. SAGE, p. 221.
- Bell, Vaughan, Dorothy V. M. Bishop, and Andrew K. Przybylski (2015). "The debate over digital technology and young people". In: *BMJ* 351, h3064.
- Bishop, Dorothy V. M. (2019). "Rein in the four horsemen of irreproducibility". In: *Nature* 568, pp. 435–435.
- Blakemore, Sarah-Jayne (2018a). "Avoiding Social Risk in Adolescence". In: *Current Directions in Psychological Science* 27.2, pp. 116–122.
- (2018b). *Inventing ourselves: the secret life of the teenage brain*. Random House, p. 240.
- (2019). "Adolescence and mental health." In: *Lancet* 393.10185, pp. 2030–2031.
- Blakemore, Sarah-Jayne and Kathryn L. Mills (2014). "Is Adolescence a Sensitive Period for Sociocultural Processing?" In: *Annual Review of Psychology* 65.1, pp. 187–207.
- Blumer, Herbert and Philip M. Hauser (1970). *Movies, delinquency, and crime*. Arno Press, p. 233.
- Boase, Jeffrey and Rich Ling (2013). "Measuring Mobile Phone Use: Self-Report Versus Log Data". In: *Journal of Computer-Mediated Communication* 18, pp. 508–519.
- Booker, Cara L., Yvonne J. Kelly, and Amanda Sacker (2018). "Gender differences in the associations between age trends of social media interaction and well-being among 10-15 year olds in the UK". In: *BMC Public Health* 18.1, p. 321.
- Boyd, Danah (2014). *It's complicated: the social lives of networked teens*. Yale University Press, p. 281.
- Bradburn, Norman M. (1969). *The structure of psychological well-being*. Oxford: Aldine.
- Brickman, P. and D. T. Campbell (1971). "Hedonic Relativism and Planning the Good Science". In: *Adaptation-Level Theory: A Symposium*. Ed. by Mortimer Herbert Appley. Academic Press, pp. 287–302.
- British Youth Council (2017). *A Body Confident Future*. Tech. rep. British Youth Council: Youth Select Committee 2017. URL: <http://www.byc.org.uk/wp-content/uploads/2017/11/Youth-Select-Committee-A-Body-Confident-Future.pdf>.
- Brown, Nicholas J. L. and Julia M. Rohrer (2019). "Easy as (Happiness) Pie? A Critical Evaluation of a Popular Model of the Determinants of Well-Being". In: *Journal of Happiness Studies*, pp. 1–17.
- Brown, Sheila (2005). *Understanding Youth and Crime*. Maidenhead: Open University Press.
- Bryson, Alex, John Forth, and Lucy Stokes (2015). "Does Worker Wellbeing Affect Workplace Performance?" URL: https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Ddid=2655044.
- Burke, Moira and Robert E. Kraut (2016). "The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength". In: *Journal of Computer-Mediated Communication* 21.4, pp. 265–281.
- Burke, Moira, Cameron Marlow, and Thomas Lento (2010). "Social network activity and social well-being". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1909–1912.
- Bushman, Brad J. and Craig A. Anderson (2002). "Violent video games and hostile expectations: A test of the general aggression model". In: *Personality and Social Psychology Bulletin* 28.12, pp. 1679–1686.

- Cadman, David et al. (1987). "Chronic Illness, Disability, and Mental and Social Well-Being: Findings of the Ontario Child Health Study". In: *Pediatrics* 79.5, pp. 805–813.
- Cain, Neralie and Michael Gradisar (2010). "Electronic media use and sleep in school-aged children and adolescents: A review". In: *Sleep Medicine* 11, pp. 735–742.
- Campbell, Angus (1981). *The sense of well-being in America: Recent patterns and trends*. New York: McGraw-Hill.
- Carroll, Lewis (1871). *Through the Looking-Glass*. Macmillan Publishers Limited.
- Carson, Valerie et al. (2016). "Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update". In: *Applied Physiology, Nutrition, and Metabolism* 41.6 (Suppl. 3), S240–S265.
- Ceci, Stephen J. (1988). "Scientists' Attitudes toward Data Sharing". In: *Science, Technology, & Human Values* 13.1-2, pp. 45–52.
- Chambers, Christopher D. (2013). "Registered Reports: A new publishing initiative at Cortex". In: *Cortex* 49.3, pp. 609–610.
- (2014). *The changing face of psychology*. URL: <https://www.theguardian.com/science/head-quarters/2014/jan/24/the-changing-face-of-psychology>.
- Choi, Jiyun et al. (2018). "Effect of the Online Game Shutdown Policy on Internet Use, Internet Addiction, and Sleeping Hours in Korean Adolescents." In: *The Journal of Adolescent Health* 62.5, pp. 548–555.
- Cohen, Jacob (1992). "A power primer." In: *Psychological Bulletin* 112.1, pp. 155–159.
- Cohen, Stanley (1972). *Folk Devils and Moral Panics: The Creation of the Mods and Rockers*. MacGibbon and Kee Ltd, p. 201.
- Crüwell, Sophia et al. (2018). "8 Easy Steps to Open Science: An Annotated Reading List". URL: <https://psyarxiv.com/cfzyx>.
- David, Meredith E., James A. Roberts, and Brett Christenson (2018). "Too Much of a Good Thing: Investigating the Association between Actual Smartphone Use and Individual Well-Being". In: *International Journal of Human-Computer Interaction* 34.3, pp. 265–275.
- Davies, Sally C. et al. (2019). *United Kingdom Chief Medical Officers' commentary on 'Screen-based activities and children and young people's mental health and psychosocial wellbeing: a systematic map of reviews'*. Tech. rep. Department of Health and Social Care. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment%7B%5C_%7Ddata/file/777026/UK%7B%5C_%7DCMO%7B%5C_%7Dcommentary%7B%5C_%7Don%7B%5C_%7Dscreentime%7B%5C_%7Dand%7B%5C_%7Dsocial%7B%5C_%7Dmedia%7B%5C_%7Dmap%7B%5C_%7Dof%7B%5C_%7Dreviews.pdf.
- DeBoer, John J. (1937). "The determination of children's interests in radio drama". In: *Journal of Applied Psychology* 21.4, pp. 456–463.
- Degli Esposti, Michelle, David K. Humphreys, and Lucy Bowes (2018). "Are long-term trends in child maltreatment decreasing in England and Wales? A time-series analysis of official record data, 1858–2016". In: *The Lancet* 392, S28.
- Dennis, Paul M. (1998). "Chills and Thrills: Does Radio Harm Our Children? The Controversy over Program Violence During the Age of Radio". In: *Journal of the History of the Behavioural Sciences* 34.1, pp. 33–50.

- Dennison, Michelle, Susan B. Sisson, and Amanda Morris (2016). "Obesogenic behaviours and depressive symptoms in children: a narrative literature review". In: *Obesity Reviews* 17.8, pp. 735–757.
- Department for Digital, Culture, Media and Sport and Secretary of State for the Home Department (2019). *Online Harms White Paper*. Tech. rep. London: HM Government.
- Department of Health and Social Care (2018). *Matt Hancock warns of dangers of social media on children's mental health*. URL: <https://www.gov.uk/government/news/matt-hancock-warns-of-dangers-of-social-media-on-childrens-mental-health> (visited on 11/08/2018).
- Desai, Sonalde, P. Lindsay Chase-Lansdale, and Robert T. Michael (1989). "Mother or Market? Effects of Maternal Employment on the Intellectual Ability of 4-Year-Old Children". In: *Demography* 26.4, pp. 545–561.
- Dickson, Kelly et al. (2018). *Screen-based activities and children and young people's mental health and psychosocial wellbeing: a systematic map of reviews*. Tech. rep. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. URL: <http://eppi.ioe.ac.uk/>.
- Diener, Ed (2000). "Subjective Well-Being: The Science of Happiness and a Proposal for a National Index". In: *American Psychologist* 55.1, pp. 34–43.
- Diener, Ed, Robert A. Emmons, et al. (1985). "The Satisfaction With Life Scale". In: *Journal of Personality Assessment* 49.1, pp. 71–75.
- Diener, Ed, Richard E. Lucas, and Shigehiro Oishi (2018). "Advances and Open Questions in the Science of Subjective Well-Being". In: *Collabra: Psychology* 4.1, p. 15.
- Diener, Ed and Shigehiro Oishi (2005). "The Nonobvious Social Psychology of Happiness". In: *Psychological Inquiry* 16.4, pp. 162–167.
- Dienlin, Tobias, Philipp K. Masur, and Sabine Trepte (2017). "Reinforcement or Displacement? The Reciprocity of FtF, IM, and SNS Communication and Their Effects on Loneliness and Life Satisfaction". In: *Journal of Computer-Mediated Communication* 22.2, pp. 71–87.
- (2018). *A longitudinal analysis of the privacy paradox using RI-CLPM*. Paper presented at the ICA's 68th Annual Conference, Prague, Czech Republic.
- Dietvorst, Evelien et al. (2018). "Adolescent Perceptions of Parental Privacy Invasion and Adolescent Secrecy: An Illustration of Simpson's Paradox". In: *Child Development* 89.6, pp. 2081–2090.
- DiPietro, Janet A. (2000). "Baby and The Brain: Advances in Child Development". In: *Annual Review of Public Health* 21.1, pp. 455–471.
- Domahidi, Emese (2018). "The Associations Between Online Media Use and Users' Perceived Social Resources: A Meta-Analysis". In: *Journal of Computer-Mediated Communication* 23.4, pp. 181–200.
- Dunbar, Robin I. M. (2018). "The Anatomy of Friendship". In: *Trends in Cognitive Sciences* 22.1, pp. 32–51.
- Eid, Michael and Ed Diener (2004). "Global Judgments of Subjective Well-Being: Situational Variability and Long-Term Stability". In: *Social Indicators Research* 65.3, pp. 245–277.
- Eiland, Lisa and Russell D. Romeo (2013). "Stress and the developing adolescent brain." In: *Neuroscience* 249, pp. 162–71.

- Ellis, David A. (2019). “Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors”. In: *Computers in Human Behavior* 97, pp. 60–66.
- Ellis, David A. et al. (2019). “Do smartphone usage scales predict behavior?” In: *International Journal of Human-Computer Studies* 130, pp. 86–92.
- Ellison, Nicole, Charles Steinfield, and Cliff Lampe (2007). “The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites”. In: *Journal of Computer-Mediated Communication* 12.4, pp. 1143–1168.
- Elson, Malte and Christopher J. Ferguson (2014). “Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives, and a debate gone astray”. In: *European Psychologist* 19.1, pp. 33–46.
- Elson, Malte and Andrew K. Przybylski (2017). “The Science of Technology and Human Behavior: Standards, Old and New”. In: *Journal of Media Psychology* 29, pp. 1–7.
- Erfani, Seyedezahra Shadi and Babak Abedin (2018). “Impacts of the use of social network sites on users’ psychological well-being: A systematic review”. In: *Journal of the Association for Information Science and Technology* 69.7, pp. 900–912.
- Etchells, Peter J. et al. (2016). “Prospective Investigation of Video Game Use in Children and Subsequent Conduct Disorder and Depression Using Data from the Avon Longitudinal Study of Parents and Children”. In: *PLOS ONE* 11.1, e0147732.
- Fashad, Nemati and Bryan Kolb (2010). “Motor cortex injury has different behavioral and anatomical effects in early and late adolescence”. In: *Behavioural Neuroscience* 124.5, pp. 612–622.
- Ferguson, Christopher J. (2009). “An effect size primer: A guide for clinicians and researchers.” In: *Professional Psychology: Research and Practice* 40.5, pp. 532–538.
- (2017). “Everything in Moderation: Moderate Use of Screens Unassociated with Child Behavior Problems”. In: *Psychiatric Quarterly* 88.4, pp. 797–805.
- (2018). “The problem of false positives and false negatives in violent video game experiments”. In: *International Journal of Law and Psychiatry* 56, pp. 35–43.
- Ferrari, Robert (1929). *Do Radio Noises Cause Illness?*
- Feynman, Richard P. (1974). *Cargo Cult Science*. California. URL: <http://calteches.library.caltech.edu/51/2/CargoCult.htm>.
- France, Alan (2007). *Understanding Youth in Late Modernity*. Maidenhead: Open University Press.
- Frank, Josette (1939). *These Children’s Programs!*
- Fraser, Chris et al. (2002). “Driving Plasticity in Human Adult Motor Cortex Is Associated with Improved Motor Function after Brain Injury”. In: *Neuron* 34.5, pp. 831–840.
- Frison, Eline and Steven Eggermont (2016). “Exploring the Relationships Between Different Types of Facebook Use, Perceived Online Social Support, and Adolescents’ Depressed Mood”. In: *Social Science Computer Review* 34.2, pp. 153–171.
- (2017). “Browsing, Posting, and Liking on Instagram: The Reciprocal Relationships Between Different Types of Instagram Use and Adolescents’ Depressed Mood”. In: *Cyberpsychology, Behavior, and Social Networking* 20.10, pp. 603–609.
- Frith, Emily (2017). *Social media and children’s mental health: a review of the evidence*. Tech. rep. Education Policy Institute. URL: https://epi.org.uk/wp-content/uploads/2018/01/Social-Media%7B%5C_%7DMental-Health%7B%5C_%7DEPI-Report.pdf.

- Frost, Rachel L. and Debra J. Rickwood (2017). "A systematic review of the mental health outcomes associated with Facebook use". In: *Computers in Human Behavior* 76, pp. 576–600.
- Fuhrmann, Delia, Lisa J. Knoll, and Sarah-Jayne Blakemore (2015). "Adolescence as a Sensitive Period of Brain Development". In: *Trends in Cognitive Sciences* 19.10, pp. 558–566.
- Funder, David C. and Daniel J. Ozer (2019). "Evaluating Effect Size in Psychological Research: Sense and Nonsense". In: *Advances in Methods and Practices in Psychological Science* 2.2, pp. 156–168.
- Garland, David (2008). "On the concept of moral panic". In: *Crime, Media, Culture: An International Journal* 4.1, pp. 9–30.
- Gelman, Andrew and Eric Loken (2014). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Tech. rep. New York, NY. URL: http://www.stat.columbia.edu/%7B~%7Dgelman/research/unpublished/p%7B%5C_%7Dhacking.pdf%7B%5C%7D5Cnhttp://doi.apa.org/getdoi.cfm?doi=10.1037/a0037714.
- Gelman, Andrew and Hal Stern (2006). "The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant". In: *The American Statistician* 60.4, pp. 328–331.
- Gilmore, Rick O., Joy Lorenzo Kennedy, and Karen E. Adolph (2018). "Practical Solutions for Sharing Data and Materials From Psychological Research". In: *Advances in Methods and Practices in Psychological Science* 1.1, pp. 121–130.
- Goodman, Robert (1997). "The Strengths and Difficulties Questionnaire: A Research Note". In: *Journal of Child Psychology and Psychiatry* 38, pp. 581–586.
- Goodman, Robert et al. (2000). "Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample." In: *The British Journal of Psychiatry* 177, pp. 534–539.
- Greenfield, Susan (2014). *Mind Change: How digital technologies are leaving their mark on our brains*. Ebury Digital.
- Griffiths, Paul (1996). *Youth and authority: formative experiences in England, 1560-1640*. Clarendon Press, p. 459.
- Grimes, Tom, James A. Anderson, and Lori Bergen (2008). *Media violence and aggression: Science and ideology*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Grondin, Simon (2010). "Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions". In: *Attention, Perception, & Psychophysics* 72.3, pp. 561–582.
- Gruenberg, Sidonie Matsner (1935). "Radio and the Child". In: *The Annals of the American Academy of Political and Social Science* 177.1, pp. 123–128.
- Hamaker, Ellen L., Rebecca M. Kuiper, and Raoul P. P. P. Grasman (2015). "A critique of the cross-lagged panel model". In: *Psychological Methods* 20.1, pp. 102–116.
- Hammer, Bernhard (2012). "Statistical Models for Time Use Data: An Application to Housework and Childcare Activities Using the Austrian Time Use Surveys from 2008 and 1992". PhD thesis. University of Vienna.
- Hancock, Jeffrey T. et al. (2019). "Social Media Use and Psychological Well-Being: A Meta-Analysis". In: *International Communications Association*.
- Hanson, Trudy L. et al. (2010). "Cell Phones, Text Messaging, and Facebook: Competing Time Demands of Today's College Students". In: *College Teaching* 59.1, pp. 23–30.

- Harbard, Emily et al. (2016). "What's Keeping Teenagers Up? Prebedtime Behaviors and Actigraphy-Assessed Sleep over School and Vacation". In: *Journal of Adolescent Health* 58, pp. 426–432.
- Helliwell, John F. and Robert D. Putnam (2004). "The social context of well-being." In: *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 359.1449, pp. 1435–46.
- Herzog, Herta (1941). *Children and Their Leisure Time Listening to the Radio*. Tech. rep. New York: Radio Council on Children's Radio Programs.
- Himmelweit, Hilde T., Abraham Naftali Oppenheim, and Pamela Vince (1961). *Television and the child: an empirical study of the effect of television on the young*. Oxford: Oxford University Press, p. 522.
- Ho, Joses et al. (2019). "Moving beyond P values: data analysis with estimation graphics". In: *Nature Methods*, p. 1.
- Hoare, Erin et al. (2016). "The associations between sedentary behaviour and mental health among adolescents: a systematic review". In: *International Journal of Behavioral Nutrition and Physical Activity* 13.1, p. 108.
- House of Commons Science and Technology Select Committee (2019). *Impact of social media and screen-use on young people's health*. Tech. rep. London: House of Commons, pp. 1–92. URL: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf>.
- Houtkoop, Bobby Lee et al. (2018). "Data Sharing in Psychology: A Survey on Barriers and Preconditions". In: *Advances in Methods and Practices in Psychological Science* 1.1, pp. 70–85.
- Huang, Chiungjung (2010). "Internet Use and Psychological Well-being: A Meta-Analysis". In: *Cyberpsychology, Behavior, and Social Networking* 13.3, pp. 241–249.
- Huebner, E. Scott (2004). "Research on Assessment of Life Satisfaction of Children and Adolescents". In: *Quality-of-Life Research on Children and Adolescents. Social Indicators Research Series, Vol. 23*. Ed. by A. Dannerbeck et al. Springer, Dordrecht, pp. 3–33.
- Huesmann, L. Rowell, Kirsti Lagerspetz, and Leonard D. Eron (1984). "Intervening variables in the TV violence-aggression relation: Evidence from two countries". In: *Developmental Psychology* 20.5, pp. 746–775.
- Hunt, Melissa G. et al. (2018). "No More FOMO: Limiting Social Media Decreases Loneliness and Depression". In: *Journal of Social and Clinical Psychology* 37.10, pp. 751–768.
- Jelenchick, Lauren A., Jens C. Eickhoff, and Megan A. Moreno (2013). "'Facebook Depression?' Social Networking Site Use and Depression in Older Adolescents". In: *Journal of Adolescent Health* 52.1, pp. 128–130.
- Johnston, Lloyd D. et al. (2016). *Monitoring the Future: A Continuing Study of American Youth (8th- and 10th-Grade Surveys)*. URL: <http://doi.org/10.3886/ICPSR36799.v1>.
- Junco, Reynol (2013). "Comparing actual and self-reported measures of Facebook use". In: *Computers in Human Behavior* 29, pp. 626–631.
- Jung, Kiju et al. (2014). "Female hurricanes are not deadlier than male hurricanes". In: *Proceedings of the National Academy of Sciences* 111.34, E3496–E3496.
- Kaelin Jr, William G. (2017). "Publish houses of brick, not mansions of straw". In: *Nature* 545.7655, pp. 387–387.

- Kann, Laura et al. (2016). *Youth Risk Behavior Surveillance - United States, 2015*. Tech. rep. 6, pp. 1–174. URL: <http://www.cdc.gov/mmwr/volumes/65/ss/ss6506a1.htm>.
- Keles, Betul, Niall McCrae, and Annmarie Grealish (2019). “A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents”. In: *International Journal of Adolescence and Youth*.
- Kelly, Yvonne et al. (2019). “Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study”. English. In: *EClinicalMedicine* 6, pp. 59–68.
- Kesebir, Pelin and Ed Diener (2008). “In Pursuit of Happiness: Empirical Answers to Philosophical Questions”. In: *Perspectives on Psychological Science* 3.2, pp. 117–125.
- Khouja, Jasmine N et al. (2019). “Is screen time associated with anxiety or depression in young people? Results from a UK birth cohort”. In: *BMC Public Health* 19.1, p. 82.
- Kiernan, Kathleen E. and Fiona K. Mensah (2009). “Poverty, Maternal Depression, Family Status and Children’s Cognitive and Behavioural Development in Early Childhood: A Longitudinal Study”. In: *Journal of Social Policy* 38.4, pp. 569–588.
- Kievit, Rogier A. et al. (2013). “Simpson’s paradox in psychological science: a practical guide”. In: *Frontiers in Psychology* 4, p. 513.
- Knies, Gundi (2017). *Understanding Society: Waves 1-7, 2009-2016 and harmonised British Household Panel Survey: Waves 1-18, 1991-2009, User Guide*. Colchester, UK: University of Essex.
- Kolb, Bryan and Robbin Gibb (2011). “Brain plasticity and behaviour in the developing brain.” In: *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 20.4, pp. 265–76.
- Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock (2014). “Experimental evidence of massive-scale emotional contagion through social networks.” In: *Proceedings of the National Academy of Sciences of the United States of America* 111.24, pp. 8788–90.
- Kriebel, David et al. (2001). “The precautionary principle in environmental science.” In: *Environmental Health Perspectives* 109.9, pp. 871–876.
- Kross, Ethan et al. (2013). “Facebook Use Predicts Declines in Subjective Well-Being in Young Adults”. In: *PLoS ONE* 8.8. Ed. by Cédric Sueur, e69841.
- Kuhn, Thomas S. (1962). *The structure of scientific revolutions*. London: The University of Chicago Press, p. 217.
- Lakens, Daniël and Ellen R. K. Evers (2014). “Sailing From the Seas of Chaos Into the Corridor of Stability: Practical Recommendations to Increase the Informational Value of Studies”. In: *Perspectives on Psychological Science* 9.3, pp. 278–292.
- Lakens, Daniël, Anne M. Scheel, and Peder M. Isager (2018). “Equivalence Testing for Psychological Research: A Tutorial”. In: *Advances in Methods and Practices in Psychological Science* 1.2, pp. 259–269.
- Lang, Annie (2013). “Discipline in Crisis? The Shifting Paradigm of Mass Communication Research”. In: *Communication Theory* 23, pp. 10–24.
- LeBlanc, Allana G. et al. (2012). “Systematic review of sedentary behaviour and health indicators in the early years (aged 0–4 years)”. In: *Applied Physiology, Nutrition, and Metabolism* 37.4, pp. 753–772.
- Lenroot, Rhoshel K. and Jay N. Giedd (2006). “Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging”. In: *Neuroscience & Biobehavioral Reviews* 30.6, pp. 718–729.

- Leonardi, Paul M. (2012). *Car crashes without cars: lessons about simulation technology and organizational change from automotive design*. MIT Press, p. 334.
- Levenson, Jessica C. et al. (2017). "Social media use before bed and sleep disturbance among young adults in the United States: A nationally representative study". In: *Sleep* 40.9, pp. 1–7.
- Little, Roderick J. A. (1988). "Missing-Data Adjustments in Large Surveys". In: *Journal of Business & Economic Statistics* 6.3, pp. 287–296.
- Liu, Mingli, Lang Wu, and Shuqiao Yao (2016). "Dose-response association of screen time-based sedentary behaviour in children and adolescents and depression: a meta-analysis of observational studies." In: *British Journal of Sports Medicine* 50.20, pp. 1252–1258.
- Livingstone, Sonia M. (2009). *Children and the Internet: great expectations, challenging realities*. Polity, p. 301.
- Livingstone, Sonia M., Leslie Haddon, et al. (2011). *Technical Report and User Guide: The 2010 EU Kids Online Survey*. Tech. rep. LSE, London: EU Kids Online, pp. 1–101.
- Livingstone, Sonia M., Giovanna Mascheroni, and Elisabeth Staksrud (2018). "European research on children's internet use: Assessing the past and anticipating the future". In: *New Media & Society* 20.3, pp. 1103–1122.
- Lowery, Shearon. and Melvin L. DeFleur (1988). "Violence and Media: The Turbulent Sixties". In: *Milestones in Mass Communication Research*. New York: Longman Publishers USA, p. 415.
- Lucas, Richard E. and M. Brent Donnellan (2007). "How stable is happiness? Using the STARTS model to estimate the stability of life satisfaction". In: *Journal of Research in Personality* 41.5, pp. 1091–1098.
- Lucas, Richard E. and Nicole M. Lawless (2013). "Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments." In: *Journal of Personality and Social Psychology* 104.5, pp. 872–84.
- Lykken, David and Auke Tellegen (1996). "Happiness Is a Stochastic Phenomenon". In: *Psychological Science* 7.3, pp. 186–189.
- Lyubomirsky, Sonja, Laura King, and Ed Diener (2005). "The Benefits of Frequent Positive Affect: Does Happiness Lead to Success?" In: *Psychological Bulletin* 131.6, pp. 803–855.
- MAPS Project (2019). *MAPS: mapping the analytical paths of a crowdsourced data analysis*.
- Marszalek, Jacob M. et al. (2011). "Sample Size in Psychological Research over the Past 30 Years". In: *Perceptual and Motor Skills* 112.2, pp. 331–348.
- Maslowsky, Julie, John Schulenberg, and Robert Zucker (2014). "Influence of Conduct Problems and Depressive Symptomatology on Adolescent Substance Use: Developmentally Proximal Versus Distal Effects". In: *Developmental Psychology* 50.4, pp. 1179–1189.
- Masur, Philipp K. (2019). *Situational Privacy and Self-Disclosure*. 1st ed. Springer International Publishing.
- McCarthy, Randy J. et al. (2016). "Does playing video games with violent content temporarily increase aggressive inclinations? A pre-registered experimental study". In: *Journal of Experimental Social Psychology* 67, pp. 13–19.

- McCrae, Niall, Sheryl Gettings, and Edward Purcell (2017). "Social Media and Depressive Symptoms in Childhood and Adolescence: A Systematic Review". In: *Adolescent Research Review* 2.4, pp. 315–330.
- Meehl, Paul E. (1967). "Theory-Testing in Psychology and Physics: A Methodological Paradox". In: *Philosophical Transactions of the Royal Society. Series B, Biological sciences* 34.2, pp. 103–115.
- (1990a). "Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It". In: *Psychological Inquiry* 1.2, pp. 108–141.
- (1990b). "Why Summaries of Research on Psychological Theories are Often Uninterpretable". In: *Psychological Reports* 66, pp. 195–244.
- Mensah, Fiona K. and Kathleen E. Kiernan (2010). "Maternal general health and children's cognitive development and behaviour in the early years: findings from the Millennium Cohort Study". In: *Child: Care, Health and Development* 37.1, pp. 44–54.
- Miller, Geoffrey (2012). "The Smartphone Psychology Manifesto". In: *Perspectives on Psychological Science* 7.3, pp. 221–237.
- Miller, George A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information". In: *Psychological Review* 101.2, pp. 343–352.
- Mitrofan, Oana, Moli Paul, and Nicholas Spencer (2009). "Is aggression in children with behavioural and emotional difficulties associated with television viewing and video game playing? A systematic review". In: *Child: Care, Health and Development* 35.1, pp. 5–15.
- Moreau, David, Brooke N. Macnamara, and David Z. Hambrick (2019). "Overstating the Role of Environmental Factors in Success: A Cautionary Note". In: *Current Directions in Psychological Science* 28.1, pp. 28–33.
- Munafò, Marcus R. and George Davey Smith (2018). "Robust research needs many lines of evidence". In: *Nature* 553.7689, pp. 399–401.
- Munafò, Marcus R., Brian A. Nosek, et al. (2017). "A manifesto for reproducible science". In: *Nature Human Behaviour* 1.1, pp. 1–21.
- Murali, Vijaya and Sanju George (2007). "Lost online: an overview of internet addiction". In: *Advances in Psychiatric Treatment* 13.1, pp. 24–30.
- National Health Service (2016). *Health and Wellbeing of 15-year-olds in England – What About YOUTH? (WAY): Survey Consultation Outcomes Paper*. Tech. rep. Health & Social Care Information Centre, pp. 1–40. URL: <https://digital.nhs.uk/data-and-information/areas-of-interest/public-health/what-about-youth-study>.
- Nature (2019). "Screen time: how much is too much?" In: *Nature* 565.7739, pp. 265–266.
- Nes, Ragnhild Bang and Espen Røysamb (2017). "Happiness in Behaviour Genetics: An Update on Heritability and Changeability". In: *Journal of Happiness Studies* 18.5, pp. 1533–1552.
- Neuman, Susan B. (1988). "The Displacement Effect: Assessing the Relation between Television Viewing and Reading Performance". In: *Reading Research Quarterly* 23.4, pp. 414–440.
- Norman, Geoffrey R., Jeff A. Sloan, and Kathleen W. Wyrwich (2003). "Interpretation of Changes in Health-related Quality of Life". In: *Medical Care* 41.5, pp. 582–592.
- Odgers, Candice (2018). "Smartphones are bad for some teens, not all". In: *Nature* 554.7693, pp. 432–434.

- Ofcom (2019). *Children and parents: media use and attitudes report 2018*. Tech. rep. Ofcom, pp. 1–17. URL: <https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2018>.
- Oishi, Shigehiro, Ed Diener, and Richard E. Lucas (2007). “The Optimum Level of Well-Being: Can People Be Too Happy?” In: *Perspectives on Psychological Science* 2.4, pp. 346–360.
- Okasha, Samir (2002). *Philosophy of science: a very short introduction*. Oxford: Oxford University Press, p. 144.
- Open Science Collaboration (2015). “Estimating the reproducibility of psychological science”. In: *Science* 349.6251, aac4716.
- Ophir, Yaakov, Yuliya Lipshits-Braziler, and Hananel Rosenberg (2019). “New-Media Screen Time is Not (Necessarily) Linked to Depression: Comments on Twenge, Joiner, Rogers, and Martin (2018)”. In: *Clinical Psychological Science*.
- Orben, Amy (2017). “Social Media and Suicide: A Critical Appraisal”. In: *Medium*.
- Orben, Amy, Peter J. Etchells, and Andrew K. Przybylski (2018). *Three problems with the debate around screen time*.
- Orben, Amy and Daniël Lakens (2019). “Crud (Re)defined”. In: URL: <https://psyarxiv.com/96dpy>.
- Orben, Amy and Andrew K. Przybylski (2019a). “Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies”. In: *Psychological Science* 30.5, pp. 682–696.
- (2019b). “The association between adolescent well-being and digital technology use”. In: *Nature Human Behaviour* 3.2, pp. 173–182.
- Orzech, Kathryn M. et al. (2016). “Digital media use in the 2h before bedtime is associated with sleep variables in university students”. In: *Computers in Human Behavior* 55, pp. 43–50.
- Owens, Judith (2014). “Insufficient Sleep in Adolescents and Young Adults: An Update on Causes and Consequences”. In: *Pediatrics* 134.3, e921–e932.
- Parkes, Alison et al. (2013). “Do television and electronic games predict children’s psychosocial adjustment? Longitudinal research using the UK Millennium Cohort Study”. In: *Archives of Disease in Childhood* 98.5, pp. 341–348.
- Patton, George C. et al. (2016). “Our future: a Lancet commission on adolescent health and wellbeing.” In: *The Lancet* 387.10036, pp. 2423–78.
- Pearl, Judea and Dana Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Penguin, p. 418.
- Pew Research Centre (2018). *Teens, Social Media & Technology 2018*. Tech. rep.
- Poitras, Veronica J. et al. (2017). “Systematic review of the relationships between sedentary behaviour and health indicators in the early years (0–4 years)”. In: *BMC Public Health* 17.S5, p. 868.
- Preston, Mary I. (1941). “Children’s reactions to movie horrors and radiocrime”. In: *The Journal of Pediatrics* 19.2, pp. 147–168.
- Prizant-Passal, Shiri, Tomer Shechner, and Idan M. Aderka (2016). “Social anxiety and internet use – A meta-analysis: What do we know? What are we missing?” In: *Computers in Human Behavior* 62, pp. 221–229.
- Przybylski, Andrew K. and Amy Orben (2017). *Social media is nothing like drugs, despite all the horror stories*.

- Przybylski, Andrew K., Richard M. Ryan, and C. Scott Rigby (2009). "The Motivating Role of Violence in Video Games". In: *Personality and Social Psychology Bulletin* 35.2, pp. 243–259.
- Przybylski, Andrew K. and Netta Weinstein (2017). "A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations Between Digital-Screen Use and the Mental Well-Being of Adolescents". In: *Psychological Science* 28.2, pp. 204–215.
- Rai, Tage S. (2019). "Social media use and mental health". In: *Science* 364.6446. Ed. by Caroline Ash and Jesse Smith, pp. 1147.3–1148.
- Ramey, Carl R. (1994). "In the Battle Over TV Violence, The Communications Act Should Be Cheered, Not Changed!" In: *Federal Communications Law Journal* 42.2, pp. 349–355.
- Ravetz, Jerome R. (1971). *Scientific Knowledge and its Social Problems*. London: Oxford University Press, p. 449.
- Reinecke, Leonard, Adrian Meier, et al. (2018). "The Relationship Between Trait Procrastination, Internet Use, and Psychological Functioning: Results From a Community Sample of German Adolescents." In: *Frontiers in psychology* 9, p. 913.
- Reinecke, Leonard and Sabine Trepte (2014). "Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication". In: *Computers in Human Behavior* 30, pp. 95–102.
- Rice, Emma S. et al. (2016). "Social media and digital technology use among Indigenous young people in Australia: a literature review". In: *International Journal for Equity in Health* 15.
- Robins, Richard W., Holly M. Hendin, and Kali H. Trzesniewski (2001). "Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale." In: *Personality and Social Psychology Bulletin* 27.2, pp. 151–161.
- Robinson, John P. (1985). "The Validity and Reliability of Diaries versus Alternative Time Use Measures". In: *Time, Goods and Well-Being*. Ed. by Francis Thomas Juster and Frank P. Sanford. Ann Arbor, MI: Institute for Social Research, University of Michigan, pp. 33–62.
- Rohrer, Julia M., Boris Egloff, and Stefan C. Schmukle (2017). "Probing Birth-Order Effects on Narrow Traits Using Specification-Curve Analysis". In: *Psychological Science* 28.12, pp. 1821–1832.
- Rohrer, Julia M. and Richard E. Lucas (2018). "Only so Many Hours: Correlations between Personality and Daily Time Use in a Representative German Panel". In: *Collabra: Psychology* 4.1, pp. 1–10.
- Romeo, Russell D. (2013). "The Teenage Brain". In: *Current Directions in Psychological Science* 22.2, pp. 140–145.
- Rose, Geoffrey (2008). *Rose's Strategy of Preventive Medicine*. Revised Ed. Oxford, UK: Oxford University Press.
- Rosenbaum, Paul R. (2017). *Observation and Experiment: An Introduction To Causal Inference*. Harvard University Press, p. 374.
- Royal Society of Public Health (2017). *#StatusOfMind*. Tech. rep. URL: <https://www.rsph.org.uk/our-work/campaigns/status-of-mind.html>.
- Rubenstein, Eli A. (1982). "Introductory Comments". In: *Television and behavior: Ten years of scientific progress and implications for the eighties*. Ed. by David Pearl, Lorraine Bouthilet, and Joyce Lazar. Rockville, MD: Rockville, Md., pp. 104–107.

- Ryan, Richard M. and Edward L. Deci (2000). "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being". In: *American Psychologist* 55.1, pp. 68–78.
- (2001). "On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being". In: *Annual Review of Psychology* 52.1, pp. 141–166.
- Scharkow, Michael (2016). "The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data". In: *Communication Methods and Measures* 10.1, pp. 13–27.
- (2019). *Getting more information out of the specification curve*. URL: <https://underused.org/2019-01-spec-curve>.
- Schwarz, Norbert and Daphna Oyserman (2001). "Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction". In: *American Journal of Evaluation* 22.2, pp. 127–160.
- Scott, Holly, Stephany M. Biello, and Heather Cleland (2018). "Identifying drivers for bedtime social media use despite sleep costs: the adolescent perspective".
- Scott, Sophie (2013). *Pre-registration would put science in chains*.
- Seagoe, May V. (1951). "Children's Television Habits and Preferences". In: *The Quarterly of Film Radio and Television* 6.2, pp. 143–153.
- Seligman, Martin E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. New York, NY: Free Press.
- Sheldon, Kennon M., Neetu Abad, and Christian Hinsch (2011). "A two-process view of Facebook use and relatedness need-satisfaction: Disconnection drives use, and connection rewards it." In: *Journal of Personality and Social Psychology* 100.4, pp. 766–775.
- Silberzahn, Raphael et al. (2018). "Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results". In: *Advances in Methods and Practices in Psychological Science* 1.3, pp. 337–356.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant". In: *Psychological Science* 22.11, pp. 1359–1366.
- (2018). "False-Positive Citations". In: *Perspectives on Psychological Science* 13.2, pp. 255–259.
- Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson (2015). "Specification Curve: Descriptive and inferential statistics on all reasonable specifications".
- Simpson, Edward H. (1951). "The interpretation of interaction in contingency tables". In: *Journal of the Royal Statistical Society. Series B, Methodological* 13.2, pp. 238–241.
- Smith, Sven, Christopher J. Ferguson, and Kevin Beaver (2018). "A longitudinal analysis of shooter games and their relationship with conduct disorder and self-reported delinquency". In: *International Journal of Law and Psychiatry* 58, pp. 48–53.
- Song, Hayeon et al. (2014). "Does Facebook make you lonely?: A meta analysis". In: *Computers in Human Behavior* 36, pp. 446–452.
- Stark, Philip B. (2018). "Before reproducibility must come preproducibility". In: *Nature* 557, pp. 613–613.
- Steege, Sara et al. (2016). "Increasing Transparency Through a Multiverse Analysis". In: *Perspectives on Psychological Science* 11.5, pp. 702–712.
- Steers, Mai-Ly N. (2016). "'It's complicated': Facebook's relationship with the need to belong and depression". In: *Current Opinion in Psychology* 9, pp. 22–26.

- Steinberg, Laurence (2004). "Risk Taking in Adolescence: What Changes, and Why?" In: *Annals of the New York Academy of Sciences* 1021.1, pp. 51–58.
- Steptoe, Andrew, Angus Deaton, and Arthur A. Stone (2015). "Subjective wellbeing, health, and ageing". In: *The Lancet* 385.9968, pp. 640–648.
- Sternheimer, Karen (2003). *It's Not the Media: The Truth About Pop Culture's Influence On Children*. Boulder, Colorado: Westview Press, p. 242.
- Strasburger, Victor C. (1989). "Children, Adolescents, and Television—1989: II. The Role of Pediatricians". In: *Pediatrics* 83.3.
- The Children's Society (2018). *The Good Childhood Report 2018*. Tech. rep. URL: https://www.childrenssociety.org.uk/sites/default/files/thegood%7B%5C_%7Dchildhood%7B%5C_%7Dreport%7B%5C_%7D2018%7B%5C_%7D0.pdf.
- The Children's Society and Barnardo's (2018). *Factors affecting children's mental health over time*. Tech. rep. URL: <https://www.childrenssociety.org.uk/sites/default/files/usoc-mental-health-briefing-2018.pdf>.
- Thomson, Elizabeth, Thomas L. Hanson, and Sara S. McLanahan (1994). "Family Structure and Child Well-Being: Economic Resources vs. Parental Behaviors". In: *Social Forces* 73.1, pp. 221–242.
- Thurstone, Louis L. (1929). "Theory of attitude measurement." In: *Psychological Review* 36.3, pp. 222–241.
- Tilley, Carol L. (2012). "Seducing the Innocent: Fredric Wertham and the Falsifications That Helped Condemn Comics". In: *Information & Culture: A Journal of History* 47.4, pp. 383–413.
- Tremblay, Mark S. et al. (2011). "Systematic review of sedentary behaviour and health indicators in school-aged children and youth". In: *International Journal of Behavioral Nutrition and Physical Activity* 8.1, p. 98.
- Tromholt, Morten (2016). "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-Being". In: *Cyberpsychology, Behavior, and Social Networking* 19.11, pp. 661–666.
- Trout, J. D. (1994). "Austere Realism and the Worldly Assumptions of Inferential Statistics". In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1, pp. 190–199.
- Twenge, Jean M. (2018). *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy-and Completely Unprepared for Adulthood-and What That Means for the Rest of Us*. Atria Books, p. 352.
- (2019). "Stop Debating Whether Too Much Smartphone Time Can Hurt Teens, and Start Protecting Them". In: *TIME Magazine*.
- Twenge, Jean M., Thomas E. Joiner, et al. (2017). "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time". In: *Clinical Psychological Science* 6.1, pp. 3–17.
- Twenge, Jean M., Gabrielle N. Martin, and W. Keith Campbell (2018). "Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology." In: *Emotion* 18.6, pp. 765–780.
- Twenge, Jean M., Brian H. Spitzberg, and W. Keith Campbell (2019). "Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness". In: *Journal of Social and Personal Relationships* 36.6, pp. 1892–1913.

- Twomey, Conal and Gary O'Reilly (2017). "Associations of Self-Presentation on Facebook with Mental Health and Personality Variables: A Systematic Review". In: *Cyberpsychology, Behavior, and Social Networking* 20.10, pp. 587–595.
- Understanding Society (2018). *New Funding for Understanding Society*. URL: <https://www.understandingsociety.ac.uk/2018/07/23/new-funding-for-understanding-society>.
- University of Essex, Institute for Social and Economic Research (2018). *Understanding Society: Waves 1-7, 2009-2016 and Harmonised BHPS: Waves 1-18, 1991-2009*.
- University of London Institute for Education, Centre for Longitudinal Studies (2017). *Millennium Cohort Study: Sixth Survey, 2015*.
- University of Michigan. Survey Research Center (2018). *Panel Study of Income Dynamics (PSID): Main Interview, 1968-2015*. Ann Arbor, MI.
- U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau (2014). *Child Health USA 2014*. Tech. rep. Rockville, Maryland: U.S. Department of Health and Human Services. URL: <https://mchb.hrsa.gov/chusa14/special-features/flourishing-resilience.html>.
- Utz, Sonja and Johannes Breuer (2016). "Informational benefits from social media use for professional purposes: Results from a longitudinal study". In: *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 10.4.
- Vahedi, Zahra and Lesley Zannella (2019). "The association between self-reported depressive symptoms and the use of social networking sites (SNS): A meta-analysis". In: *Current Psychology*, pp. 1–16.
- Valkenburg, Patti M. and Jessica Taylor Piotrowski (2017). *Plugged in: how media attract and affect youth*. Yale University Press, p. 328.
- Van Harmelen, Anna-Laura et al. (2017). "Adolescent friendships predict later resilient functioning across psychosocial domains in a healthy community cohort." In: *Psychological Medicine* 47.13, pp. 2312–2322.
- Vanden Abeele, Mariek, Kathleen Beullens, and Keith Roe (2013). "Measuring mobile phone use: Gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use". In: *Mobile Media & Communication* 1.2, pp. 213–236.
- Vanman, Eric J., Rosemary Baker, and Stephanie J. Tobin (2018). "The burden of online friends: The effects of giving up Facebook on stress and well-being". In: *The Journal of Social Psychology* 158.4, pp. 496–508.
- van't Veer, Anna Elisabeth and Roger Giner-Sorolla (2016). "Pre-registration in social psychology—A discussion and suggested template". In: *Journal of Experimental Social Psychology* 67, pp. 2–12.
- Vazire, Simine (2018). "Implications of the Credibility Revolution for Productivity, Creativity, and Progress". In: *Perspectives on Psychological Science* 13.4, pp. 411–417.
- Verduyn, Philippe, David Seungjae Lee, et al. (2015). "Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence." In: *Journal of Experimental Psychology: General* 144.2, pp. 480–488.
- Verduyn, Philippe, Oscar Ybarra, et al. (2017). "Do Social Network Sites Enhance or Undermine Subjective Well-Being? A Critical Review". In: *Social Issues and Policy Review* 11.1, pp. 274–302.
- Viner, Russell, Max Davie, and Alison Firth (2019). *The health impacts of screen time: a guide for clinicians and parents*. Tech. rep. Royal Society for Paediatrics and Child

- Health. URL: https://www.rcpch.ac.uk/sites/default/files/2018-12/rcpch%7B%5C_%7Dscreen%7B%5C_%7Dtime%7B%5C_%7Dguide%7B%5C_%7D-%7B%5C_%7Dfinal.pdf.
- Wagenmakers, Eric-Jan, Gilles Dutilh, and Alexandra Sarafoglou (2018). “The Creativity-Verification Cycle in Psychological Science: New Methods to Combat Old Idols”. In: *Perspectives on Psychological Science* 13.4, pp. 418–427.
- Wagenmakers, Eric-Jan, Ruud Wetzels, et al. (2012). “An Agenda for Purely Confirmatory Research”. In: *Perspectives on Psychological Science* 7.6, pp. 632–638.
- Wait until 8th (2018). *Wait until 8th*. URL: <https://www.waituntil8th.org> (visited on 06/24/2019).
- Wang, Bing-qian et al. (2017). “The association between attention deficit/hyperactivity disorder and internet addiction: a systematic review and meta-analysis”. In: *BMC Psychiatry* 17.1, p. 260.
- Wang, Jin-Liang et al. (2018). “The Reciprocal Relationship Between Passive Social Networking Site (SNS) Usage and Users’ Subjective Well-Being”. In: *Social Science Computer Review* 36.5, pp. 511–522.
- Wartella, Ellen and Byron Reeves (1985). “Historical Trends in Research on Children and the Media: 1900–1960”. In: *Journal of Communication* 35.2, pp. 118–133.
- Wellman, Barry (2004). “The Three Ages of Internet Studies: Ten, Five and Zero Years Ago”. In: *New Media & Society* 6.1, pp. 123–129.
- Wenninger, Helena, Hanna Krasnova, and Peter Buxmann (2019). “Understanding the role of social networking sites in the subjective well-being of users: a diary study”. In: *European Journal of Information Systems* 28.2, pp. 126–148.
- Wertham, Fredric (1954). *Seduction of the Innocent*. Rinehart & Company, p. 397.
- Wicherts, Jelte M., Marjan Bakker, and Dylan Molenaar (2011). “Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results”. In: *PLoS ONE* 6.11. Ed. by Rochelle E. Tractenberg, e26828.
- Wilcockson, Thomas D. W., David A. Ellis, and Heather Shaw (2018). “Determining Typical Smartphone Usage: What Data Do We Need?” In: *Cyberpsychology, Behavior, and Social Networking* 21.6, cyber.2017.0652.
- Williams, James S. et al. (2009). *Growing Up in Ireland National Longitudinal Study of Children: The lives of 9-year-olds*. Dublin.
- Wonneberger, Anke and Mariana Irazoqui (2017). “Explaining Response Errors of Self-Reported Frequency and Duration of TV Exposure Through Individual and Contextual Factors”. In: *Journalism & Mass Communication Quarterly* 94.1, pp. 259–281.
- World Health Organisation (2019). *Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age*. Tech. rep. World Health Organisation.
- Wright Mills, C. (1954). *Nothing to Laugh At*.
- Yoon, Sunkyoung et al. (2019). “Is social network site usage related to depression? A meta-analysis of Facebook–depression relations”. In: *Journal of Affective Disorders* 248, pp. 65–72.