

The Self-Regulation of Virtue: Reactions to Moral Exemplars

Antonio Fabio Bella



Lancaster University

Faculty of Science and Technology – Department of Psychology

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy
July 2021

Declaration of Authorship

The research described in this thesis is the result of my own work. Mentioned work of other authors has been appropriately referenced and any contribution made to the research by others has been explicitly acknowledged.

The material contained in this thesis has not been previously published by another person or submitted, in whole or in part, for any other academic degree at another university or institution.

Abstract

Extant research has investigated the response to moral exemplars primarily from an emotion perspective, with a focus on either positive or negative reactions. By contrast, the present project, articulated across four studies (N=1,814) in the US and UK, captured simultaneously the positive and negative response to others' moral goodness adopting an integrative self-regulation approach that examined how the self negotiates its standards and standing *vis-à-vis* virtuous people and their actions. Participants viewed and rated a set of real-life moral scenarios portraying agents performing virtuous actions (Study 1), and two suitable vignettes were identified for further investigation. Through EFA (Study 2) and CFA (Study 3), a novel instrument to measure the self-regulation of virtue was assessed and improved. This moral self-regulation inventory consists of a broadening scale measuring the extent that individuals praise the agents, feel uplifted and inspired to better themselves (moral self-improvement), and a defensive scale measuring the extent that individuals experience resentment and even disparage the agents and their actions (moral self-defence). Path modelling (Study 2) and SEM (Study 3) determined that moral comparisons based on opinion and ability (upward/downward) were at the root of these reactions, and motivational dispositions (approach/avoidance and promotion/prevention focus) were associated with them; prosociality (helping behaviour) was linked with moral self-improvement activated by both excellent and lesser good deeds (Study 4). Participants were also clustered in independent latent profiles and groups at various stages of the model (motivation, comparison, self-regulation), and the associations between the profiles/groups across stages reproduced the relational patterns observed through SEM, corroborating robustness of the results. By integrating the literatures on social comparison,

motivation, and moral emotions within a self-regulation framework, these findings advance theory in moral psychology, with practical implications on how to maximise the social upsides of moral goodness while containing its possible drawbacks.

Keywords: virtue, self-regulation, moral self-improvement, moral self-defence, social comparison, regulatory focus, approach/avoidance, structural equation modelling.

Acknowledgements and Dedication

The present work is the result of countless hours of reading, designing, planning, reflecting, all spiced up with testing, analysing, writing, revising, and then starting over again, cycle after cycle. Although this iterative process might sound painful, it actually was exhilarating. Indeed, never in my life have I had the opportunity to manage in full autonomy such a large research project on a subject that I personally chose and nurtured across all its phases, with a sense of full ownership and accountability. I am immensely grateful to my supervisors, Dr. Jared Piazza and Dr. Neil McLatchie, for entrusting me with such freedom and supporting me beyond the call of duty. Their understanding and generosity have kept me highly motivated throughout these years. It is my sincere hope that the quality of this output and the growth of my skills as a researcher might be equally inspiring to them.

The many colleagues that make up the vibrant postgraduate research community at Lancaster University contributed to making my doctorate something very different from a solitary endeavour. Consultations, exchange of ideas, and mutual encouragement have been essential components of the process that materialised in this work. There would be no better reward than knowing that in these interactions I have been able to offer at least a fraction of what I have received from my colleagues.

I am also deeply thankful to my wife, Janice, and my dog, Islay, for their unwavering trust in me and their remarkable flexibility to adjust to the ever-changing circumstances of our lives. Without their moral support, optimism, and sense of humour, this achievement would not have been possible. At the end of the day, this labour of love could not be dedicated but to them.

List of Figures

Theoretical Framework and Conceptual Model

Figure I:	28
Figure II:	47
Figure III:	58
Figure IV:	66
Figure V:	67

Study 1: Stimuli and Model Development

Figure A1:	90
Figure A2:	91
Figure A3:	92
Figure A4:	93
Figure A5:	94
Figure A6:	96
Figure A7:	99
Figure A8:	99
Figure A9:	103
Figure A10:	103
Figures A11a-b:	109
Figures A12a-b:	112
Figures A13a-c:	121
Figure A14:	130
Figure A15:	135

Study 2: Model Assessment

Figure B1:	149
Figure B2:	165
Figure B3:	166
Figure B4:	170
Figure B5:	171
Figure B6:	173
Figure B7:	174
Figures B8a-b:	177
Figure B9:	185
Figure B10:	189
Figure B11:	190
Figures B12a-b:	194

Study 3: Model Improvement

Figure C1:	220
Figure C2:	221
Figure C3:	222
Figure C4:	222
Figure C5:	223
Figure C6:	223
Figure C7:	224
Figure C8:	225
Figure C9:	227
Figure C10:	228
Figure C11:	229
Figure C12:	229
Figure C13:	230
Figure C14:	230
Figure C15:	237
Figure C16:	237
Figure C17:	238
Figure C18:	240
Figure C19:	241
Figure C20:	241
Figures C21a-b:	243
Figures C22a-b:	245
Figures C23a-b:	248
Figures C24a-b:	251
Figures C25:	254
Figure C26:	267
Figure C27:	269
Figure C28:	274

Study 4: Model Extension

Figure D1:	288
Figures D2a-c:	291
Figure D3:	316
Figure D4:	317
Figure D5:	317
Figure D6:	318
Figure D7:	319
Figure D8:	320
Figure D9:	320
Figure D10:	322
Figure D11:	322
Figure D12:	326
Figure D13:	328
Figure D14:	329

Figure D15:.....	330
Figure D16:.....	334
Figures D17a-b:.....	337

List of Tables

Study 1: Stimuli and Model Development

Table A1:	79
Table A2:	83
Table A3:	83
Table A4:	85
Table A5:	85
Table A6:	89
Table A7:	98
Table A8:	102
Tables A9a-b:.....	106
Table A10:	108
Table A11:	110
Table A12:	111
Table A13:	113
Tables A14a-b:	117
Tables A15a-d:	118
Table A16:	123

Study 2: Model Assessment

Table B1:	154
Table B2:	155
Table B3:	156
Table B4:	157
Table B5:	159
Table B6:	159
Table B7:	161
Table B8:	161
Table B9:	163
Table B10:	168
Table B11:	172
Table B12:	175
Table B13:	178
Table B14:	180
Table B15:	182
Tables B16a-b:.....	184
Table B17:	187
Table B18:	191
Table B19:	192
Table B20:	193
Table B21:	193

Study 3: Model Improvement

Table C1:	218
Table C2:	218
Table C3:	231
Table C4:	231
Tables C5a-b:.....	232
Table C6:	233
Table C7:	233
Tables C8a-d:.....	234
Table C9:	235
Table C10:	235
Tables C11a-b:.....	236
Table C12:	239
Table C13:	239
Tables C14a-b:.....	240
Table C15:	244
Table C16:	247
Table C17:	250
Table C18:	253
Table C19:	253
Table C20:	257
Tables C21a-b:.....	258
Tables C22a-b:.....	258
Tables C23a-b:.....	259
Table C24:	265
Table C25:	268
Table C26:	272
Table C27:	278

Study 4: Model Extension

Table D1:	321
Table D2:	323
Table D3:	326
Table D4:	328
Table D5:	329
Table D6:	331
Table D7:	332
Table D8:	333
Table D9:	334
Table D10:	335
Table D11:	336

Contents

THE SELF-REGULATION OF VIRTUE: REACTIONS TO MORAL EXEMPLARS

Declaration of Authorship	2
Abstract	3
Acknowledgements and Dedication	5
List of Figures	6
List of Tables	9
Contents	11
Theoretical Framework and Conceptual Model	16
A Historical Account of Virtue and Moral Beauty	18
Overarching Goal of the Research	25
Objectives and Research Framework	26
Moral Elevation Through the Lens of the “Broaden-and-Build” Theory	28
Resentment and “Do-Gooder” Derogation	30
The Mechanisms of the Moral Response to Virtue	31
Moral Comparisons	31
Moral Self-Regulation	35
Moral Self-Regulation and the Componential Process Theory of Emotion	44
Other Factors Affecting Moral Comparisons and Self-Regulation	47
Individual Differences	48
Situational Factors and the Vignettes	54
Effects on Social Behaviour	56
Summary of Key Hypotheses and the Conceptual Model	57
Adaptiveness of Moral Self-Regulation Modes	60
Participant Clusters	61
A Path Toward Causality	62
Research Roadmap	64
Study 1: Stimuli and Model Development	68
Introduction	68
The Moral Vignettes	68
Measurement Instruments	71
Moral Self-Regulation Inventory	71
Moral Affect Scales	74
Moral Judgments and Moral Comparisons	74
Association Patterns Between Variables	75
Summary of Key Objectives	75

Methods	77
Participants	77
Research Design and Materials	78
Measures	80
Moral Appraisals	80
Moral Self-Regulation	81
Moral Affect	84
Open-Ended Questions	85
Procedure	85
Analytic Approach	86
Results and Preliminary Reflections	88
Multivariate Outliers	88
Moral Appraisals	89
Moral Self-Regulation	96
Moral Affect	100
Analysis of Distribution Bias and Assumptions	104
Relationships Between Key Variables	106
Moral Comparison and Self-Regulation	106
Moral Self-Regulation and Affect	110
Other Correlation Patterns	114
Multiple Linear Regression	114
Development of Preliminary Path Models	119
Findings from the Open-Ended Questions	122
Discussion	124
Stimuli	124
Moral Self-Regulation and Affect	126
Moral Self-Regulation Inventory	127
Moral Affect Scales	127
Moral Self-Evaluation	128
Other Measures of Moral Appraisal and Comparison	129
Relationships Between the Main Variables	133
Correlations	133
Regression and Path Models	134
Summary of Key Findings and Limitations	135
 Study 2: Model Assessment	 140
Introduction	140
Measurement Model	141
Moral Self-Regulation and Affect	141
Moral Appraisals	142
Characteristic Adaptations	142
Structural Model	148
Methods	150
Research Design	150
Participants	150
Sample Size Determination	150
Data Exclusions	151
Multivariate Outliers	152
Final Sample Composition	152
Materials and Procedure	153
Measures	153

Socio-Demographic Questions	153
Characteristic Adaptations	154
Moral Appraisals	157
Moral Self-Regulation	158
Moral Affect	160
Analytic Approach	161
Results and Preliminary Reflections	163
Descriptive Statistics	163
Characteristic Adaptations	163
Moral Appraisals	164
Moral Self-Regulation: Item and Scale Analysis	166
Moral Affect: Item and Scale Analysis	171
Correlations Between the Main Variables	174
Moral Self-Regulation Scale Dimensionality: Exploratory Factor Analysis	178
Multiple Linear Regression	181
The Moderation Hypothesis	185
The Mediation Hypothesis and Path Models	187
Path Model Fitted to the Total Sample	189
Structural Invariance Between Vignettes	191
Path Models Fitted by Vignette	193
Discussion	196
Measurement Model	196
Item Selection	196
Merging of Cognitive/Conative/Affective Items	196
Dimensional Structure	198
Structural Model	200
Moral Comparison and Self-Regulation	200
Structural Invariance	200
Characteristic Adaptations	200
Strengths and Limitations	207
Study 3: Model Improvement	211
Introduction	211
Methods	214
Research Design	214
Participants	214
Sample Size Determination	214
Data Exclusions	215
Multivariate Outliers	215
Final Sample Composition	216
Materials and Procedure	216
Measures	217
Analytic Approach	219
Results and Preliminary Reflections	221
Descriptive Statistics	221
Primary Analyses	225
Testing the Measurement Model: Confirmatory Factor Analysis	226
Testing the Structural Model: Full Structural Equation Modelling	242
Secondary Analyses	254
Measurement Invariance of the Endogenous Latent Variables	254
Structural Invariance Between Vignettes	259

A Person-Centred Analysis: Latent Profile Modelling	261
Discussion	275
Measurement Model	275
Endogenous Latent Variables	275
Exogenous Latent Variables	279
Structural Models	279
Complementary Insights	281
Limitations	283
Beyond Self-Regulation: Measuring Social Behaviour	285
Study 4: Model Extension	287
Introduction	287
Objective 1: Behavioural Effects	287
Objective 2: Mechanisms of the Behavioural Effects	288
Objective 3: Further Exploration	292
Regulatory Mode	292
Social Comparison Orientation	295
Social Desirability	296
Internal Locus of Control	296
Narcissism and Modesty/Humility	297
Self-Growth Constructs	298
Methods	301
Data Collection	301
Experimental Design and Materials	301
Questionnaire Structure and Procedure	302
Participants	303
Sample Size Determination	303
Data Collection and Cleaning (Step 1)	304
Data Collection and Cleaning (Step 2)	305
Multivariate Outliers	305
Final Sample Composition	305
Measures	306
Moral Self-Regulation	306
New Measures	308
Analytic Approach	314
Results and Preliminary Reflections	316
Descriptive Statistics	316
Manipulation Check	323
Test of H1 – Differential Prosocial Effects of the Manipulation	324
Relationship and Mechanism: Mediation Tests	326
Test of H2 – Simple Mediation Model	327
Test of H3 – Serial Mediation Model	328
Test of H4 – Moderated Mediation Model	330
Regulatory Mode: Correlation Patterns	331
A Wider Behavioural Model of Moral Self-Improvement	333
Exploratory Analyses (Post-Task Constructs)	338
Social Desirability	338
Social Comparison Orientation	339
Discussion	340
Impact of the COVID-19 Pandemic	340
Significance of the Main Findings	341

Limitations	345
General Discussion and Conclusive Remarks	347
Methods	347
Hypotheses and Conceptual Model	348
Main Findings	349
Theoretical Contributions	351
Methodological Advances	353
Stimuli	353
Moral Self-Regulation Inventory	354
Independence of Moral Comparison and Self-Regulation Measures	355
Causal Inferences	356
Directions for Future Research	357
Personality Traits	357
Moral Scenarios	357
Trait Measures of Moral Self-Regulation	358
Cross-Cultural Validation	359
Manipulation of Motivational Variables	359
Antisocial Behaviour	360
Neurophysiological Measures	361
Recommended Priorities	364
Practical Implications	364
Amplifying the Upsides of Moral Self-Improvement	365
Containing the Downsides of Moral Self-Defence	366
A Closing Note	367
References	369
Appendices	418

"Few things are harder to put up with than the annoyance of a good example."

Mark Twain (1894). *Pudd'nhead Wilson*

Theoretical Framework and Conceptual Model

In recent years, the field of moral psychology has experienced revived interest – a “multidisciplinary renaissance”, to say it after Haidt and Kesebir (2010, p. 797) – with the proliferation of empirical research and theorising, and the rise of a promising cross-disciplinary convergence of approaches derived from fields such as social psychology, sociobiology, cognitive science, moral and experimental philosophy. Much of the output has gravitated toward the “dark” side of morality, investigating predominantly violations of ethical standards (Cornwell & Higgins, 2015b), that is, *immoral* (transgressions) rather than *moral* behaviour (good deeds). It could be considered one of the effects of the “negativity bias”, according to which humans show a tendency to “give greater weight to negative entities” (Rozin & Royzman, 2001, p. 296), hence the “bad” ends up being stronger than the “good” (Baumeister et al., 2001). Moreover, moral psychologists and philosophers have been considerably devoted to the examination of rare, unrealistic, and highly hypothetical moral dilemmas, such as the trolley problems (Bauman et al., 2014; Kahane, 2015; Kahane et al., 2015), neglecting more common real-life moral situations which would be worthy of equal, if not more, attention.

If psychology as a science was born in the nineteenth century (see e.g., Baumeister et al., 2007), for many decades its development focused fundamentally on “understanding, treating and preventing psychological disorders” (Peterson & Seligman, 2004, p. 3), thus perpetuating the negativity bias and addressing less common problems. It is only in the 1950s that the landscape started to change, with the advent of humanistic psychology first, and of positive psychology at the turn of the century¹. These movements sought to emphasise the psychology of human growth, virtue, and flourishing, bringing about the “bright” side of human life and morality, thus moving beyond the study of “repairing damage within a disease model of human functioning” by complementing it with an understanding of how “normal people flourish under more benign conditions” (Seligman & Csikszentmihalyi, 2000, p. 5). These developments brought to the fore concepts such as self-actualisation, that is, individuals’ motives to realise their full potential (Maslow, 1954; Rogers, 1956), and character strengths, or in other words, positive traits whose development makes life worth living (Seligman & Csikszentmihalyi, 2000). The acknowledgment that virtue, positive moral character, and moral praise have been largely understudied (Bartels et al., 2015) has led to the call to bring them back centre stage (Pizarro & Tannenbaum, 2012), redressing the balance between the study of morality and immorality.

Moral psychology was also characterised by another tendency that Cornwell and Higgins (2015b) call the “ought premise”: the assumption that the focus of psychological inquiry in morality ought to be the moral imperative of the fulfilment of *duties* and *obligations*. This ethical orientation was inherited from centuries of philosophical speculation, which cemented in Kant’s deontology, and in Bentham’s and Mill’s

¹ For the interested reader, a discussion on the thematic convergence as well as the supposed independence of the schools of humanistic and positive psychology can be found in Robbins (2008).

utilitarianism; it still profoundly permeates moral thinking in our present society. However, again in the 1950s, the publication of Anscombe's provocative article "Modern Moral Philosophy" (1958) challenged the moralistic fallacy and legalistic assumption that – in the absence of a lawgiver, such as God – morality can be reduced to a set of rules. Anscombe essentially claimed that a rigorous approach to ethics should necessarily include the notion of *virtue*, independent of *obligation*, as part of human flourishing (Crisp & Slote, 1997). Anscombe's arguments were echoed by the publication of MacIntyre's (1981) "After Virtue" a couple of decades later.

The present research attempts to address these issues and integrate perspectives, setting out to study virtue and positive moral behaviour (good deeds) performed by real people. Some of these deeds are more ordinary moral actions, while others are uncommon and remarkable; together, they contribute to making the present investigation useful to a deeper understanding of the response to real-life moral behaviour. As will become apparent, the study of virtue does not provide a partial, overly optimistic perspective on morality, but also opens a window into less obvious maladaptive responses that are deserving of careful consideration.

A Historical Account of Virtue and Moral Beauty

Following the Aristotelian tradition, the term *virtue* refers to positive "states of character" (von Wright, 1963) or positive "traits of character" (Adams, 2006)². Virtues manifest themselves through specific *character strengths*, which can be viewed as the "distinguishable routes" to displaying virtues, or the "psychological ingredients –

² The distinction between "states" and "traits" of character hinges on respectively the more *transient* or *permanent* nature of these moral attributes, and mirrors other non-moral personal attributes (for definitions, see e.g., Baumert et al., 2017). The debate on whether rather stable character traits do or do not exist was as vigorous in moral psychology and philosophy as it had been a few years earlier for personality traits in personality and social psychology. For an account of the two opposing perspectives, respectively for and against the existence of traits, see e.g., Jayawickreme et al. (2014) and Harman (2000); for a "Hegelian synthesis" of the two, see e.g., Fleeson & Nofle (2008).

processes or mechanisms – that define the virtues” (Peterson & Seligman, 2004, p. 13). Character strengths “contribute to fulfilments that comprise the good life, for oneself and for others” and are “morally valued trait-like personality characteristics” (Ruch et al., 2019, p. 1). They differ from positive aspects of temperament and personality in that they indicate specific *moral* attributes. Over time, departing from the Aristotelian tradition, they have also come to differ from other *non-moral* excellences, for instance, intellectual capabilities such as intelligence (Dent, 1984), but also sensory-motor or artistic skills such as athleticism and creativity (for a slightly different perspective on artistic and aesthetic values, see Adams, 2006). Echoing a definition of personality traits, Cohen and Morse (2014) conceptualise moral character as “an individual’s disposition to think, feel, and behave in an ethical versus unethical manner, or as the subset of individual differences relevant to morality” (p. 45).

The perimeter of what can be considered within or outside the moral domain has been – and still is – the subject of long-standing controversies. For a long time, philosophers understood morality as a *code of conduct* (norms and duties) endorsed by individuals and groups (Luco, 2014). However, Rawls (1975) found a conception of morality as a set of duties to be too restrictive and extended its perimeter to include not only the realm of the *right* (duty), but also of the *good* (virtue) and *worth* (utility), cardinal concepts that he considered capable of identifying the key axes through which a wide variety of moral properties could be analysed. In moral psychology, some researchers described morality as a “collection of biological and cultural solutions to the problems of *cooperation* recurrent in human social life”³ (Curry et al., 2019, p. 106;

³ For a collection of definitions of morality centred around *cooperation*, see Curry and colleagues (2019), who cite Haidt and Kesebir (2010), Rai and Fiske (2011), Tomasello and Vaish (2013), Greene (2015), and Sterelny and Fraser (2016). Although much of morality involves cooperation and the restraint of selfishness, the problem with these definitions is that they are too narrow and do not adequately encompass all its

emphasis added). Other psychologists retained a similar monistic approach, but instead of focusing on cooperation, tried to condense the essence of morality in *care/harm* (Gray et al., 2012), in yet another reductionist attempt to distil the core of the concept in one single element. By contrast, other theorists proposed a pluralistic view, identifying multiple *moral foundations* (Graham et al., 2011) that represent the basic template at the root of moral judgment: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation, and the recently added liberty/oppression (more could be discovered, according to the authors).

Whichever definition is considered more appropriate, recent research has revealed that moral character has a fundamental importance in attributions of person perception and evaluation (Goodwin et al., 2014). These attributions share a few critical characteristics. A prominent one is the fact that they are based on the detection in the performer of a good deed of *ideal* motives, that is, pure motivations to pursue what is fine and noble, either for its own sake (Darwall, 2003) or for the altruistic desire to improve their lives, with no consideration for personal recognition and public praise. Indeed, concern for others has been identified as one of the central ideal motives that pervade positive moral character traits; this means that virtue entails high consideration of what others need and want, and how one's behaviour affects others' needs and wants, as claimed by Cohen and Morse (2014). Attributions of moral character based on others' perceived intentions are ubiquitous and pervasive (Cushman, 2015; Reeder, 2009; Uhlmann et al., 2014), and occur spontaneously, not just when experimental measures draw attention to them (Critcher et al., 2020). Because others' intentions are not directly observable, to make these attributions people rely on inferences based on

manifold facets, particularly the foundation of liberty/oppression and, to a lesser degree, sanctity/degradation.

contextual factors that are filtered through the lens of their ingrained dispositions, which account for what is available (stored in memory), applicable (fitting), accessible (ready to be activated), and salient (subject to selective attention) to them from the surrounding environment (Higgins, 1996).

A second prominent feature of moral character traits is the fact that they manifest *abilities* (Cohen & Morse, 2014); according to Aristotle, character strengths – like other abilities – are acquired through practice (London, 2001). Individuals can develop stable dispositions to act for the good if they consistently exercise them through noble actions. For example, it is only by learning how to temper fear that individuals become capable to dominate it and act courageously. Additionally, it could be argued that virtues could also be learned by observation of others (Bandura, 1986): as people witness others perform praiseworthy actions that reveal their noble intentions toward others, they apprehend what good character looks like. Consequently, they might desire to emulate them (see e.g., Algoe & Haidt, 2009) and eventually develop those character traits.

A further distinctive aspect is the fact that when consideration of others' needs becomes ingrained in one's motives and consistently manifests itself in real-life behaviour, virtues tend to become integrated within that person's *identity* (Cohen & Morse, 2014): when that happens, the person's self-concept prominently features moral values and concerns. Aquino and Reed's (2002) notion of "moral identity internalisation" represents the extent to which those moral values are central to the individual's self-concept.

In a nutshell, virtues are positive character traits which reveal other-orientated motives that are central to the person's sense of self and are developed by consistently performing good deeds. Anscombe (1958) insisted on conceptually distinguishing between moral virtues and moral obligations because the former characterise the moral

character of the *person*, whereas the latter refer to specific requirements of moral *behaviour*. In this sense, she emphasised that ethical systems essentially based on *moral conduct* (a set of behavioural duties and prohibitions) are insufficient and must be complemented by an understanding of *character strengths* that are conducive to virtue.

The centrality of virtue has been acknowledged by philosophers all around the world. In ancient China, the concept of *de* or *mei de* (i.e., “inner character” or “virtue”) is found in the Daodejing, attributed to Lao-Tzu (for an account of virtue in Daoism, see Cline, 2004) and later substantiated in the “Five Constants” (*wu ch’ang*), Confucius’s five essential virtues of benevolence, justice, propriety, wisdom, and integrity (Runes, 1983). In India, the concept of *kama muta* (i.e., “being moved by love” in Sanskrit) has been used for centuries to designate the experience of being touched and inspired by moral beauty in situations of intense communal relationships; recent research has shown that the elicitation of *kama muta* through acts of virtue increases interpersonal closeness and sharing of communal values that extend beyond in-groups, leading to a heightened humanisation of out-groups (Blomster Lyshol et al., 2020). In the Buddhist traditions, the ten *paramitas* (i.e., “perfections” in Pali) of generosity, discipline, renunciation, wisdom, persistence, patience, honesty, determination, loving-kindness, and equanimity (Van Horn, 2017) have been at the centre of moral teaching for centuries. In ancient Greece, the concept of *areté* broadly corresponds to “excellence” (and specifically “virtue” in the moral domain) and represents the *acme* of Aristotle’s ethics (London, 2001). The Aristotelian tradition was later adopted and reorganised by Thomas Aquinas to serve the theological purposes of Medieval Christianity and culminated in the doctrine of the four cardinal virtues (prudence, temperance, justice, and fortitude) and the three theological virtues (faith, hope, and charity) (O’Meara, 1997). As the Middle Ages drew to a close and the scientific method opened up new speculative perspectives, beyond theological

justifications, the idea of virtue slowly faded away to make room for the rule of reason, which culminated in the concept of “universal reason” brought about by the philosophers of the age of enlightenment. In ethics though, the move from premises about human virtue to conclusions about the rational authority of moral rules implied a move from *is* to *ought*; this is often referred to as “Hume’s guillotine”, the impossibility to derive values and norms from facts (see e.g., Saariluoma, 2020), which involves the blurring of the boundaries (and subsequent contamination) between *descriptions* of “nature” (*is*) and *prescriptions* of “duty” (*ought*)⁴; and this very move reveals the above-mentioned moralistic fallacy that Anscombe (1958) critiqued.

Virtue is highly prominent in the work of novelists and poets across cultures. Children’s narratives – from “Pinocchio” and “The Little Mermaid” through to the “Chronicles of Narnia” – vividly depict heroes and villains in their quest of virtue and goodness as opposed to power and fame (for an account of children’s literature on virtue, see Bennett, 2010). Similarly, adult fiction and poetry have exemplified moral character through the traits and values embodied by the protagonists of the stories told by renowned authors such as Homer and Aesop, Dante and Shakespeare, Tolstoy and Dostoevsky, Mann and Hesse, and many others.

Beyond aesthetic appreciation, arguments about the social and moral effects of literature date back to the Greek philosophers Plato and Aristotle (Cain, 2005; Mendelson-Maoz, 2007), and continue until the Middle Ages, but then stalled for several centuries. The debate was reignited in the last fifty years by the work of philosophers such as Palmer, Eldridge, and Nussbaum (see Cates, 1998). Recently, psychologists have also contributed theorising and research on the reader’s response to

⁴ Further reading on the is/ought fallacy is available in Nelson (2019).

fiction and poetry. First, the psychological literature shows a robust effect of fiction reading on social cognition, that is, the ability to perceive, interpret, and respond to social information (Fiske & Taylor, 2013): a recent meta-analysis found a significant positive (albeit small) impact in terms of empathy, theory of mind, and prosociality (Dodell-Feder & Tamir, 2018), explained by the phenomenon of “narrative transportation”, according to which readers mentally enter the world that a story evokes (Van Laer et al., 2014). This phenomenon has also moral implications: Zbikowski and Collins (1994) hinted at viewing literature as the “laboratory of moral life building” and a few years later, Hakemulder (2000) borrowed the same idea, describing literature as a “moral laboratory” and providing a theoretical/empirical framework of the manifold effects of reading literature, including those in the moral domain. In this respect, Koopman and Hakemulder (2015) as well as Carr (2005) emphasised the role of the interplay between the cognitive sphere (understanding of values and moral knowledge) and the affective sphere (intuitive motive apprehension and empathy development). Both spheres are at the root of the processes of identification with literary characters described by Oatley (1994) through the Aristotelian concept of “mimesis”, often translated as imitation, but more precisely corresponding to the notion of *simulation*. Empirical research on children’s moral education was also carried out by Hickman (1981) and Hart et al. (2019), who provided further evidence of the role of literature as an important contributor to moral character building⁵.

⁵ A very different school of thought moved the focus from the reader’s response to the literary text to its meaning, emphasising the role of language as a mere set of signs to be deciphered rather than a force acting on the world (see Tompkins, 1980). This formalist view, which flourished at the start of the twentieth century, denied any educational effect of literature, purported the exclusion of any external interests (ethics included) in the process of reading and evaluation of the literary text (Mendelson-Maoz, 2007), and hotly debated the location of meaning in the text itself or in the reader’s mind (Tompkins, 1980).

In sum, philosophers and psychologists are now converging in the rediscovery of virtue and positive moral exemplars as vehicles to elevate the human conscience and guide it towards “the good life” (Higgins et al., 2014).

Overarching Goal of the Research

If virtue can have the remarkable effect of leading toward the good life, what contribution can moral psychology make at present through the study of moral goodness? Nowadays, mass media and social media often report narratives of good deeds performed by unknown heroes and celebrities alike⁶. Also, contemporary authors write ever new stories of courage and generosity, justice and compassion that continue to inspire readers all over the world thanks to their ability to strike a chord, elicit awe, win their hearts over, and ultimately motivate them to desire to perform similar actions to the benefit of others. They show that not only real, but also fictitious stories of moral beauty can infuse positivity and produce benefits beyond the direct advantage experienced by the recipients of the good deeds.

At the beginning of the new millennium, Haidt and his research team started a psychological inquiry into the positive emotion that follows “witnessing acts of human moral beauty or virtue” (Haidt, 2000, p. 1). Interestingly, in one of these studies (Algoe & Haidt, 2009) the authors recorded an occurrence of 18% of cases whereby participants reported having primarily *negative* feelings (or did not follow the instructions) in response to acts of goodness; as these instances were extraneous to the main research question, those participants were excluded from the analysis. Instead of examining both sides of the coin of the moral response to good deeds (positive *and* negative reactions), Algoe and Haidt’s approach deliberately ignored one side and focused exclusively on the

⁶ See, for example, <https://www.goodnewsnetwork.org>.

other. The present research, by contrast, set out to look at both sides of the coin, investigating why some people experience positive reactions when exposed to good deeds and others deviate from this prediction. Indeed, this two-sided account of morality is necessary because people make widely different judgments about the same moral act. For instance, some people consider sacrificing one's life to save a stranger an ultimate act of unselfishness, while others argue that it imposes an unnecessary life-long cost on family members, whose kinship needs ought to be prioritised; some people consider organ donation a supreme act of generosity, while others view it as a violation of religious beliefs. The present research attempted to shed light *simultaneously* on the two sides of the reactions to virtue: the more positive side of moral praise in association with uplifting feelings of moral admiration, and the less positive side of moral condemnation in association with denial and resented feelings of discomfort.

A notable literature on the positive and negative reactions to acts of virtue has been published in moral psychology during the last couple of decades, for instance the frameworks of "moral elevation" and "do-gooder derogation" that will be discussed in more detail in the following sections. However, these still appear to be independent, self-contained endeavours; one of the key theoretical contributions of the present research is their conceptual and empirical integration under the overarching theme of the *self-regulation of virtue*. If virtue can be double-edged and can potentially generate a wider spectrum of cognitions, motivations, affects, and behaviours, primarily positive, but also negative, then it is important to explore and understand their underpinnings and explicate the underlying psychological processes. By doing so, the investigation could hint a range of strategies to maximise the positive effects and minimise the potential negative impact of the response to virtue.

Objectives and Research Framework

The objective of the present research programme is to investigate reactions to descriptions of virtuous acts by moral exemplars. A set of moral stimuli in the form of vignettes depicting good deeds was developed and presented to participants in a series of studies to record their reactions on various scales, along with relevant individual difference measures. The research adopted a framework wherein virtuous moral agents performed morally motivated good deeds (the moral actions) in favour of third parties (the beneficiaries). Research participants were considered in the position of external observers, as if they were witnesses of the deeds⁷. The beneficiaries were other individuals (human or not, but never the participant) or collective entities (a group, an organisation, a nation, humankind). The phenomena analysed in the research were framed and interpreted within four fundamental theoretical domains:

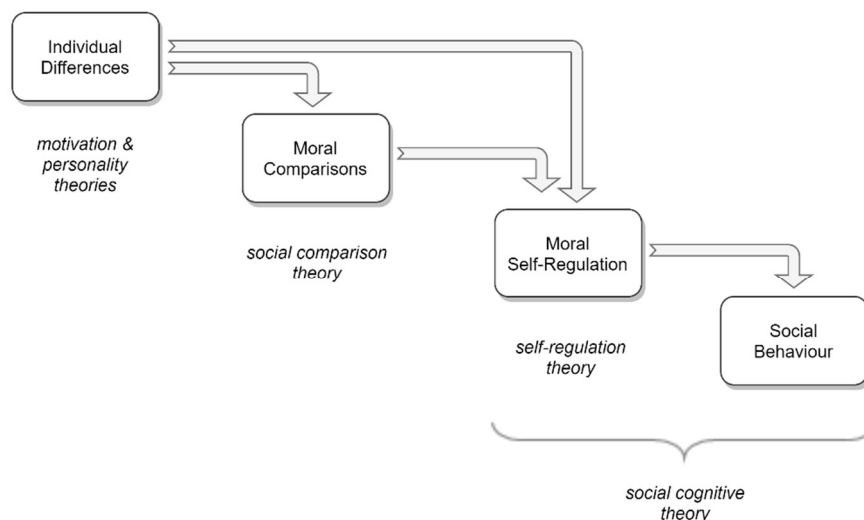
1. motivation and personality theories (see e.g., DeYoung, 2015; Elliot & Thrash, 2002; Higgins, 1997);
2. social comparison theory (see e.g., Festinger, 1954);
3. self-regulation theory (see e.g., Forgas et al., 2009);
4. social cognitive theory (see e.g., Bandura, 1991).

The response to the good deeds depicted in the vignettes was measured in terms of moral comparative and self-regulatory processes, capturing the mechanisms through which people managed their self-views in the face of the agents' virtuous actions and their own moral standards. Indeed, self-regulatory mechanisms were found to be dependent on the moral comparisons underlying the judgments that people make about the moral agents/acts in relation to their moral self as well as ethical norms and beliefs.

⁷ In the present research, the real-life stories depicted in the vignettes presented to the participants sought to create an experimental environment that simulated (in a broad sense) their "witnessing" of the deeds; hence, the recurring use in this thesis of the word "witness".

At the same time, certain dispositional traits were also found to be associated with specific patterns of moral comparison and self-regulation. Lastly, proximate behavioural outcomes were measured to analyse their relationship with moral self-regulation and compare the level of prosocial behaviour elicited by different degrees of positive moral deeds and self-regulation. All these processes and their theoretical frameworks are sketched out in Figure I and will be described in greater detail in the following chapters.

Figure I:
Relationships across the main construct categories and their underlying theoretical frameworks



Moral Elevation Through the Lens of the “Broaden-and-Build” Theory

In moral psychology, the study of the positive emotions experienced by people who witness displays of virtue took a considerable step forward when Haidt (2003) rediscovered what Thomas Jefferson wrote back in 1771 to a friend who was asking for his recommendation on the best books for his library:

[E]very thing is useful which contributes to fix us in the principles and practice of virtue. When any... act of charity or of gratitude, for

instance, is presented either to our sight or imagination, we are deeply impressed with its beauty and feel a strong desire in ourselves of doing charitable and grateful acts also. (p. 350)

Haidt named the emotion described by Jefferson *moral elevation* and initiated a line of research which defined it, distinguished it from other positive emotions (e.g., joy, but especially gratitude, awe, admiration), and delineated its main antecedents and consequences. Thanks to that research, we now know that elevation is elicited by acts of virtue or moral beauty, induces a warm feeling in the chest (sometimes even moves to tears), and motivates to aspire to do something good for others (Haidt, 2000). Further research also found a direct link between elevation and actual prosocial behaviour, beyond the mere intention (Schnall et al., 2010; Schnall & Roper, 2012).

The findings on the moral emotion of elevation seem to harmonise with Fredrickson's (2001) "broaden-and-build" theory of positive emotions. This theory posits that positive emotions, such as joy, interest, pride, and so forth, in the short term "broaden" individuals' thought-action repertoires and in the long term "build" their physical, intellectual, and social resources. This is different from the mechanism of negative emotions, which – according to Fredrickson and Branigan (2005) – typically do not develop enduring capabilities for the future, but narrow down individuals' thought-action repertoire, so that individuals can act quickly when required by the circumstances, for example averting an imminent threat.

The example of the positive emotion of joy can be revealing: regarding the first function, joy *broadens* by creating the urge to play, push the limits, and be creative, while in terms of the second function, the joy experienced during play *builds* by making individuals share amusement and develop lasting bonds for their future lives (Fredrickson, 2001). According to Haidt (2003), elevation fits well with the broaden-and-

build model, in that it opens people's sensitivity to the needs of others (broaden) and makes them cultivate social skills and relationships that will help them in the long run (build).

Resentment and “Do-Gooder” Derogation

A separate line of research has more recently examined what appears to be an opposite phenomenon to moral elevation: the *resentment* experienced toward a moral agent who performs a good deed and the manifestation of “do-gooder” derogation or even “antisocial punishment”. The term “do-gooder” has been defined as “individuals or groups who deviate from the majority on moral grounds, offering morality as the justification for their nonnormative behavior” (Minson & Monin, 2012, p. 206). The term “antisocial punishment” has been used to refer to the sanctioning, under certain circumstances, of highly cooperative people (Pleasant & Barclay, 2018). The premise of this kind of psychological inquiry is founded on evidence that being moral is not always well received by others. Research has identified specific circumstances that help explain this counterintuitive phenomenon. One of these is the fact that the disparagement of moral agents often occurs when people perceive that others' virtuous behaviour endangers the positivity of their self-image (Monin et al., 2008) or their reputation, particularly in competitive contexts (Pleasant & Barclay, 2018). In other cases, the unfavourable evaluation of the good deeds and their protagonists is due to the perception of deviations from typical behaviour, that is, the social norm (Herrmann et al., 2008; Kawamura & Kusumi, 2020).

Do-gooder derogation and antisocial punishment have been studied not only among adults, but also children (see e.g., Tasimi et al., 2015), and in several life domains and contexts. Well-known cases are the denigration of moral advocates who publicly criticise immoral behaviours such as political or environmental wrongdoing, racial or

sexual discrimination (Bashir et al., 2013; Hornsey, 2005); in organisational contexts, people are sometimes vilified for “working too hard” (Pleasant & Barclay, 2018); scorn often accompanies moral vegans and vegetarians (Minson & Monin, 2012). However, there may be a difference between the derogation of a minority of moral “activists” and “rebels” (Monin et al., 2008) or vegans and vegetarians (MacInnis & Hodson, 2017) who deviate from normative or usual behaviour and the disparagement of virtuous people who save a life or help someone in need. Does the phenomenon of moral detraction extend to all kinds of good deeds or is it confined to non-normative goodness? Is it dependent on personal inclinations or beliefs of the witnesses?

The bittersweet contradictory co-existence of elevation/praise and resentment/derogation of virtuous exemplars is a puzzling paradox of morality. Centuries of philosophical reflection and countless inspiring narratives provide an account of the adaptiveness and social utility of virtue. Yet, the phenomenon of do-gooder derogation brings to the surface a parallel reality of denigration and ridicule that is far from being socially desirable. So, all in all, is moral goodness double-edged?

The Mechanisms of the Moral Response to Virtue

Analysis of the literature suggests that the answer might come from the analysis of two psychological mechanisms that appear to be related to the response to other people’s virtue: moral comparisons and moral self-regulation.

Moral Comparisons

The social reality in which humans are immersed creates multiple opportunities for acknowledging similarities and differences between individuals. People care deeply about who they are and are motivated to seek information from the environment to evaluate themselves, acquiring knowledge about their relative strengths and weaknesses (Gregg & Sedikides, 2018). The need for social comparison has been

identified in several species, and has been claimed to be phylogenetically old, biologically powerful, and important for adaptation and survival (Buunk & Mussweiler, 2001). When they evaluate themselves, people do it with reference to certain standards: these can be internal, for instance individuals' own goals or values, but often they are also anchored to external entities, such as social norms or other people. When people think about information concerning other individuals in relation to the self, whether carefully and consciously or not, they engage in *social comparison* processes (Wood, 1996). More broadly, these comparative processes comprise "acquiring, thinking about, and reacting to social information" (Wood, 1996, p. 521).

The phenomenon of social comparison was first described by Festinger (1954), who specified that people make comparisons with others in terms of *opinions* and *abilities*, reflecting questions such as, respectively, "how correct is my opinion?" and "how smart or skilled am I?" (see Gerber, 2018). When applied to the moral domain, the former are comparisons focused on the "rightfulness" of the *act* (how right or appropriate something is), whereas the latter are comparisons focused on the "skills" of the *person* (how capable someone is).

Kelley (1971) echoed this distinction when he proposed that people perform evaluations through a "reality system", which distinguishes *correct* from *incorrect*, and an "achievement system", which distinguishes *success* from *failure*. In the moral domain, the reality system deals with evaluations of *right* or *wrong*, referred to behaviours that are deemed to be appropriate or inappropriate based on certain moral standards; the achievement system deals with evaluations of *good* or *bad*, referred to virtuous achievements or temptation failures experienced by people. In support to the analogy between Festinger's and Kelley's claims, Monin maintained that "the reality

system is undoubtedly of the domain of opinions, while the achievement system is of the domain of abilities” (Monin, 2007, p. 55).

Moral comparisons based on opinion and ability represent the sort of information that individuals seek to constantly refine moral self-evaluations that help them establish who they are relative to others, assess how close they are to what they want to be, and how to behave to do so. They are extremely pervasive and occur at an explicit but also an implicit level, spontaneously and effortlessly (Alicke et al., 2012; Dunning, 2000; Wood, 1996). Based on Monin’s (2007) above-cited claim, when people witness a good deed performed by a moral agent in favour of a third party, it is plausible to anticipate that opinion-based and ability-based comparison processes will be elicited.

Research on social comparison initially focused on the choice of comparison targets (Gerber, 2018) through the so-called “rank-order paradigm”: participants perform a test and are subsequently given a bogus score (e.g., 410) and rank (e.g., fourth out of seven). At that point, they are asked to select the rank of a participant whose score they wish to view; rank choices from first to third are indicative of *upward* comparisons (target better than self), rank choices from fifth to seventh imply *downward* comparisons (self better than target), and rank choice fourth denotes *lateral* comparison (target and self at parity).

Two studies published in a supplement of the Journal of Experimental Social Psychology in 1966 that used this paradigm are now considered the first classic social comparison experiments (Wheeler, 1991): the first was informative of upward comparison driven by pressures toward uniformity (Wheeler, 1966) and the second provided evidence of downward comparison driven by experimentally induced self-threat (Hankmiller, 1966).

Importantly for the present research, about a decade later Brickman and Bulman (1977) argued that upward comparisons are encouraged by adaptive forces to use social information for the purpose of self-growth, while downward comparisons are preferred when hedonic motives are prevalent, especially in the presence of insecurity, shame, or guilt. At the same time, Berger (1977) added that comparisons with similar others offer performance models that represent the best available standards to judge one's abilities, hence providing inspiration for imitative processes. Taken together, these findings contribute insights that can help identify mechanisms potentially at play when people witness others' good actions.

First, it is plausible to assume that witnesses of good deeds could make quick judgments on whether the action performed by the moral agent is appropriate; this evaluation is anchored to certain moral standards held by the witnesses, results in an appraisal of *goodness* and *propriety* versus those standards, and implies opinion-based comparisons. The standards that are part of the comparison may have different degrees of perceived objectivity and social acceptability (Goodwin & Darley, 2012), such that – as mentioned earlier – different people could evaluate what appears to be a good deed along a wide spectrum of degrees of positivity.

Additionally, it is plausible to assume that witnesses of good deeds could also draw moral inferences, using perceptions of circumstantial evidence and person cues to make trait attributions (Reeder, 2009) and establish how morally *virtuous* the moral agent is compared to themselves, implying ability-based comparisons. Witnesses could explicitly or implicitly wonder whether they would have the same strength of moral character as the moral agents; they could ask themselves whether they would be morally so good as to perform the same action, should they find themselves in similar situations. These

comparative processes could result in upward, downward, or lateral moral comparisons (Wills, 1991).

In summary, judgments about the *deeds* (goodness and propriety) bear upon opinion-based comparisons, reflecting questions such as “do we share the same moral views?” or “do we agree that it is a good deed?”; judgments about the *persons* (moral agent and self) are related to ability-based comparisons, implying questions such as “am I as morally capable as you to perform that action?”. Because the focus of the present research is on the processes through which people manage their self-views when witnessing others’ virtuous actions, both action and person judgments are relevant and represent ways to indirectly capture opinion- and ability-based comparisons referred to the reality and the achievement systems mentioned above. Managing the self-concept following the outcome of these kinds of comparisons constitutes one of the facets of the psychological processes that social psychologists call *self-regulation*.

Moral Self-Regulation

Self-regulation has to do with all the adaptive transformations that the self performs to conform to some standard in its ongoing relationship with the environment (Forgas et al., 2009). This standard broadly refers to “concepts of how something ideally should be” (Baumeister, 2007, p. 843) and can comprise aspirations, ideals, rules, or norms that individuals hold at an explicit or implicit level. Self-regulation is, therefore, the ongoing iterative psychological function of the self that underpins contextualised goal-directed behaviour, encompassing a blend of *cognitive*, *affective*, and *conative* components that play out in the social arena.

As a function of the self, some self-regulation theories regard the self as a decision maker, a doer, an agent that governs and models people’s behaviour (Leary & Tangney, 2012). However, the self is more than that. Over the centuries, philosophers and

psychologists have described its many facets, converging on the idea that it represents a central autonomous organising construct in the social and behavioural sciences (Dunning, 2007; Leary & Tangney, 2012), but still wrangle about the details of its definitional content and perimeter. For that reason, in 2004 Leary wrote an editorial to the journal *Self and Identity*, pleading clarity around the concept and its use.

In the present research, the word “self” is used with multiple connotations, as a stand-alone term as well as a prefix to other defining notions, such as concept, esteem, regulation, improvement, defence, enhancement, protection, and others. Depending on the context, it can be viewed as: a) a set of core beliefs about oneself (self-as-known), b) an experiencing subject (self-as-knower), c) an executive agent regulating individuals’ psychological functioning (see Leary & Tangney, 2012)⁸. The concept of self-regulation, around which this research project revolves, certainly involves the governing function, but also implies the contents regulated by the function and the subject who regulates those functions, proving the intricate interdependencies that are at the roots of the long-standing controversies about the meaning of the term “self”.

Based on the above-mentioned definition, self-regulation can be viewed within a *cybernetic* framework. A cybernetic system is a dynamic apparatus with control mechanisms based on feedback loops: in a *discrepancy-reducing* loop, an input is compared to a standard and adjustments are made to shift it to a state that is *closer* to the standard, whereas in a *discrepancy-enlarging* loop, an input is shifted to a state that is *farther away* from the standard or at least *not closer* to it (Carver & Scheier, 2002).

In this perspective, self-regulation constitutes a much broader psychological function compared to the narrower conception to which some social psychologists have

⁸ For a more philosophically orientated description of eight different connotations of the term “self”, see Olson (1998).

confined it. It encompasses more than the mere deliberative effortful processes of self-control, willpower, and other executive functions (Carver & Scheier, 2016; Forgas et al., 2009); rather, it also includes other dynamic processes that occur automatically, partly inaccessible to conscious control (Forgas et al., 2009; Papies & Aarts, 2016), through the workings of implicit goals that lie at various levels of the individual's goal hierarchy (Kruglanski et al., 2002; Shah & Kruglanski, 2000).

Moral self-regulation refers to the complex set of cognitive, affective, and conative interactive patterns related to the *moral* domain. Moral judgment and moral behaviour are inextricably related to self-regulatory processes. According to Pyszczynski and colleagues (2012), the two most fundamental self-regulatory functions are those that evolved for the *preservation* and the *expansion* of the self, because they enable it to *defend* against threats and *improve* its capabilities. One of the greatest challenges that the self must face is how to harmonise motives to remain open to new experiential input while at the same time minimising the impact of information that could destabilise cherished self-views (Alicke et al., 2012).

The broadening and shielding functions of the self operate simultaneously in multiple life domains, including morality. The monitoring subsystem of the self-regulation function constantly scans the environment in search of “cues that signal advantageous or disadvantageous circumstances” (Leary & Guadagno, 2013, p. 340). When individuals witness an act of moral goodness, this search for person and situation cues is in action and forms the grounds, as mentioned earlier, for appraisals that subsume moral comparisons. The outcome of these comparisons can be construed in two different ways: opportunities or threats (Lockwood & Matthews, 2007).

If they are construed as *opportunities*, the witnesses of the good deeds view the agents as exemplars to identify with and imitate, embrace the outcome of the

comparison as an incentive to become better persons, and set aspirational goals to develop and expand the self (see e.g., Buunk & Ybema, 2003); this broadening, advancing regulatory strategy is defined here *moral self-improvement*⁹ (see Kurman, 2006; Sedikides & Hepper, 2009; Taylor et al., 1995). It can be considered a particular instantiation of a broader category of self-growth concepts that became popular in psychology with Maslow (1954), whose notion of “self-actualisation” represents its culmination. Self-growth has been also studied in the context of academic learning (Meece et al., 2006), career development (Bartley & Robitschek, 2000), and achievement goals (Elliot & Thrash, 2002). Only more recently it has been investigated in the context of morality and human flourishing, particularly in research on personal narratives, generative goals, the redemptive self, and eudaimonia (see e.g., Bauer & McAdams, 2010; Bauer et al., 2015; McAdams, 2008).

The outcome of moral comparisons can also result in the perception of *threats* if the witnesses of the good deeds perceive them to be aversive to the integrity of the self, for instance if they push self-views below the tolerance level (Alicke & Sedikides, 2009). People are motivated to protect the positivity and stability of their self; that is, they seek to maintain a phenomenal experience of the self that is adequate, capable, strong, unitary, and coherent (Steele, 1988). Any perceived threat to the self, especially in a domain that is important to the individual, could potentially elicit conscious or non-conscious responses aimed at defusing them and re-establish perceptions of self-worth. Individuals can reach a new state of equilibrium by modelling aspects of the self-concept through various alternative defensive mechanisms, among which two have become prominent in the social psychology literature:

⁹ In this thesis, moral self-improvement and self-improvement will be used interchangeably, but will always refer to the *moral* domain.

- maximising positive self-views by “distorting” the contents of the comparison, manipulating them so that the outcome is favourable to the self (Wood & Taylor, 1991), uplifting their positive traits and abilities beyond what objective facts would warrant (Sedikides & Alicke, 2012), for example convincing themselves that in many ways they have done greater deeds in other aspects of their lives;
- minimising negative self-views by shielding themselves from unmanageable self-criticism, for example convincing themselves that, after all, the deed was not so praiseworthy or that the agents had egotistic ulterior motives undeserving of any merit (Reeder et al., 2005).

These two defensive self-regulatory strategies are generally referred to respectively as *self-enhancement* and *self-protection* (Alicke & Sedikides, 2009). In the moral domain, by exaggerating their moral traits and achievements (moral self-enhancement), individuals psychologically level or even exceed the moral stature of the virtuous agent, re-establishing a desirable degree of positivity of the self; by trivialising the value of the moral act or downgrading the virtuousness of the moral agent (moral self-protection), they indirectly attenuate negative self-views, restoring a tolerable level of self-worth. These are both guarding, defensive strategies of the self that overall form what hereon will be referred to as *moral self-defence*¹⁰.

It must be noted that these defensive mechanisms are distinct from accurate, objective self-assessment (Gregg & Sedikides, 2018), in that they involve manipulation of the information so that the outcome becomes favourable to the self (Wood & Taylor, 1991). Therefore, they represent essentially self-serving *deceiving* processes. The idea that the human mind can deceive itself dates back to the third century B.C.E. and is

¹⁰ Moral self-enhancement and self-enhancement, moral self-protection and self-protection, moral self-defence and self-defence will be used interchangeably in this thesis, always referring to the *moral* domain.

attributed to the Greek orator Demosthenes (384-322 B.C.E.), but it was Sigmund Freud who introduced it in the field of psychology (Cramer, 1998). In his early publications (Freud, 1894, 1896), defence mechanisms were defined as psychological strategies utilised by individuals, often non-consciously, to ward off sources of psychological threat and protect them from unpleasant or unacceptable thoughts (e.g., harm, death, inadequacy, inferiority) or feelings (e.g., anxiety, guilt, fear). By resorting to them, individuals would maintain or restore psychological equilibrium and preserve positive self-regard. Anna Freud (1937), his daughter, undertook a systematic categorisation and integration of the defence mechanisms in a unified coherent framework.

Despite originating in Sigmund Freud's work, defence mechanisms do not belong only to the realm of psychopathology, but broadly affect normal psychological functioning (Cramer, 1998). Their explanation in contemporary social psychology does not refer to any specific psychoanalytic content (sexual or aggressive impulses) or mechanism (conflicts between Ego, Superego, and Id), but still retains much of the framework based on the shunting of distressful psychological contents (Paulhus et al., 1997). Furthermore, there is no need today to anchor defence mechanisms deep into the unconscious; indeed, there is growing acceptance of the idea that they are *pre-conscious* processes, "available to discovery, but not persistently within awareness" (Paulhus et al., 1997, p. 551).

The two defensive strategies considered in the present research (self-enhancement and self-protection), along with the broadening strategy of self-improvement, are not the only possible self-regulatory processes elicited by witnessing others' good deeds. People might also engage in moral "self-assessment" and "self-verification", respectively seeking *accuracy* in the evaluation of their moral stature, or *consistency* with long-standing self-beliefs about their morality (Sedikides, 2012). Indeed, self-assessment

(Gregg & Sedikides, 2018) and self-verification (Swann & Read, 1981) theories have been proposed to explain some of the phenomena related to how the self negotiates the social reality. For example, individuals with low self-regard might be motivated to self-verify instead of self-defending when confronted with a virtuous moral agent, preferring to stick to existing negative beliefs about themselves rather than trying to enhance them in order to avoid feeling bad about themselves; although not specific to the moral domain, research on depressed individuals has lent credit to this mechanism (Giesler et al., 1996). However, consistent with Pyszczynski and colleagues' (2012) claim about the more fundamental nature of expansion and preservation motives, self-improvement and self-defence were deemed to be an appropriate starting point for the present research, without implying that other mechanisms should be ruled out.

Moral comparisons, and in general all social comparisons, are in the service of broader regulatory processes (Wood, 1996). When people “look up” in a moral comparison (upward comparison), evaluating themselves as morally inferior to a moral agent, they could potentially either self-improve or self-enhance/self-protect, depending on whether they see the moral agent/action as an ideal exemplar to learn from and imitate (opportunity) or a competitor who exposes the limitations of the moral self and therefore represents a menace to deal with (threat). The rationale for the association between upward comparison and self-improvement comes from studies that identified two inter-related motives: the desire to get better by learning from others who are more skilled (Berger, 1977) and the inspiration to emulate aspirational exemplars (Brickman & Bulman, 1977; see also research and theorising on observational learning, e.g. Bandura, 1986). Both motives appear to be fostered in cooperative interactions. However, in competitive contexts, the literature suggests that upward comparison induces painful affective states, which are associated with self-defence (Wheeler, 1991).

When people “look down” in a moral comparison (downward comparison), evaluating themselves as morally superior to a moral agent, in theory, they could either self-improve or self-enhance/self-protect, depending on perceptions of opportunity or threat to the self. Evidence of the link between downward comparison and self-defence was offered by Brickman and Bulman’s (1977) studies mentioned earlier; in principle, downward comparison could also be associated with self-improvement, although empirical evidence is scarce (Wood & Taylor, 1991). Therefore, at the outset of the present research there were no *a priori* hypotheses as to whether either mode of self-regulation would be prevalent when participants engaged in each of the two types of moral comparisons, and it was a matter of empirical discovery the extent to which they were likely to occur in each condition.

Extant literature also suggests a link between motivation and moral self-regulation. Broadening or advancing moral self-regulation (self-improvement) is usually associated with strong future-orientated growth or achievement goals (Wood & Taylor, 1991); that is, when people have a strong drive to get better at what they do, they use information from their environment to learn about how they can improve. Under these conditions, if individuals witness a good deed, it seems plausible to assume that they would tend to make favourable judgments about it, experience positive feelings and prosocial action tendencies, for example the desire to emulate the moral agent and do themselves something good for others. However, in the absence of strong growth motivations, people who engage in upward comparisons could construe them as threatening and potentially self-defend (see Sedikides, 2012).

On the other hand, extant literature also indicates that defensive self-regulation is usually underpinned by psychological well-being concerns (Wood & Taylor, 1991) rooted in the present; that is, people who deeply care about “feeling good” about themselves

are typically prone to self-regulation that allows them to maintain or promptly restore flattering views of the self whenever life events signal a threat. Those individuals could potentially display a variety of responses. Self-enhancement usually results in reinstating positive self-views, and therefore could be characterised by moderately positive feelings (Alicke & Sedikides, 2009), with more ambivalent moral judgments and action tendencies, ranging from slightly positive to neutral or indifferent. Self-protection is usually the result of more turbulent regulatory processes (Alicke & Sedikides, 2009), which realign negative self-views to the tolerance level, and thus could leave the witness of the good deed with a sense of inner struggle and stronger negative affect characterised by feelings of annoyance or resentment; judgments of the moral agent could be more negative than in self-enhancement and action tendencies in a grey area or sometimes even antisocial. More infrequent, yet possible, should be the case of self-improvement resulting from downward comparison (whereby people already think they are superior to the comparison target). Based on these assumptions from the literature, the present research sought to shed light on all these phenomena and measure the likelihood of their occurrence.

The interdependencies between moral comparison and self-regulation processes is evident in the social comparison literature, especially the body of research based on the rank-order paradigm mentioned earlier (Gerber, 2018). However, it must be noted that this experimental design tends to conflate comparison and self-regulatory processes: the choice of the comparison target is interpreted as a regulatory mechanism to improve or enhance/protect the self as a result of growth or well-being motives. One of the key methodological contributions of the present research is an attempt to clearly distinguish these two sets of phenomena by measuring comparison and self-regulation

mechanisms through distinct constructs, related but separable. The details will be explicated in the next chapters.

The psychological literature in the field also posits that self-regulatory processes can be conceptualised either as situational *states*, or dispositional and relatively stable *traits*¹¹. In this research, moral self-regulation was conceptualised and operationalised as a situational process (*state*); that is, a temporary “online” dynamic response to a specific eliciting stimulus (exposure to an act of moral goodness). At the same time though, the research also investigated how certain motivational *traits* affect the way individuals experience others’ good deeds and self-regulate, as outlined in the following sections.

Moral Self-Regulation and the Componential Process Theory of Emotion

One of the important aspects implicated in the self-regulation of virtue is emotion. Various phenomena of emotional activation have already been mentioned in this introduction, for instance when describing the emotion of moral elevation that accompanies moral self-improvement, or the emotion of resentment associated with moral self-protection and do-gooder derogation. Emotion and self-regulation are indeed highly intertwined instantiations of cognitive, affective, and conative processes that are immanent in all aspects of social life.

Although universal consensus on a definition of emotion has not been reached yet (Izard, 2007), the componential process theory of emotion (see e.g., Scherer, 2005) is gaining traction. According to this theory, emotions are dynamic episodes made up of distinct *components*: cognitive appraisals activated by meaningful elicitors, neurophysiological and motor changes in the body, action tendencies, and subjective feelings. Thus, emotions are reactions instigated by events that are highly relevant to

¹¹ For an account of self-regulation as state or trait, see e.g., Baumeister et al., 2006; Matthews et al., 2005.

the needs and goals of an individual (Scherer, 2009) and are experienced in the form of feelings, which “derive from sensory processes that tell the organism what is happening” (Izard, 2007, pp. 262-263), enabling self-regulatory adjustment.

Critical to the componential view of emotion is the role of appraisals (Frijda, 1986; Ortony et al., 1988; Smith & Ellsworth, 1985): these are *cognitive* evaluations of the significance of the environment for the *well-being* of the individual (Moors et al., 2013). The pivotal role of cognitive appraisals in emotion processes has a long tradition that begins with Aristotle, who described anger, for instance, as the consequence of specific thoughts, which in turn can motivate aggressive behaviour (Lazarus, 1999).

Appraisals revolve around specific dimensions, which have been categorised in different ways by psychologists. Russell (1980) introduced the systematic use of multivariate analysis techniques in emotion research, and from various sorting tasks he was able to recover through multidimensional scaling two main dimensions of affect, valence and arousal, around which some of the main affect terms could be arranged in a circumplex.

The existence of these two dimensions was consistently replicated in several later studies, but a severe limitation of this research concerned the narrow set of affect labels utilised; the quality and number of the dimensions recovered from multivariate analysis is strongly dependent on the scope and variety of the initial input. By expanding the range of the affect terms used as input with participants, Smith and Ellsworth (1985) were able to identify a wider set of appraisal dimensions through principal component analysis: pleasantness, effort, certainty, attention, control, responsibility.

A few years later, Ortony and colleagues (1988) reduced the appraisal categories to three, theorising that they are valenced evaluations along the axes of desirability, praiseworthiness, and appealingness, respectively when the focus of the appraisal is

events, agents, and objects. This means that an emotion elicitor provokes distinctly valenced reactions that are dependent on how the elicitor itself is construed: for example, a desirable event will likely trigger a positive emotional response, while a blameworthy agent and a repulsive object will likely instigate a negative emotional response. Importantly, these cognitive appraisals are often posited to take place at the very early stages of the chain of events that constitute an emotion episode (Frijda, 1986; Plutchik, 2001). Scherer recently reaffirmed the fundamental role of multilevel appraisals as the triggers of motor expressions, physiological changes, and action tendencies, whose complex interactions are integrated in central representations and categorisation labels that constitute feelings (Scherer, 2019). From this perspective, while cognitive appraisals are located at the source of emotion episodes, subjective awareness of their occurrence emerges downstream in the form of feelings, after bodily changes activate action tendencies.

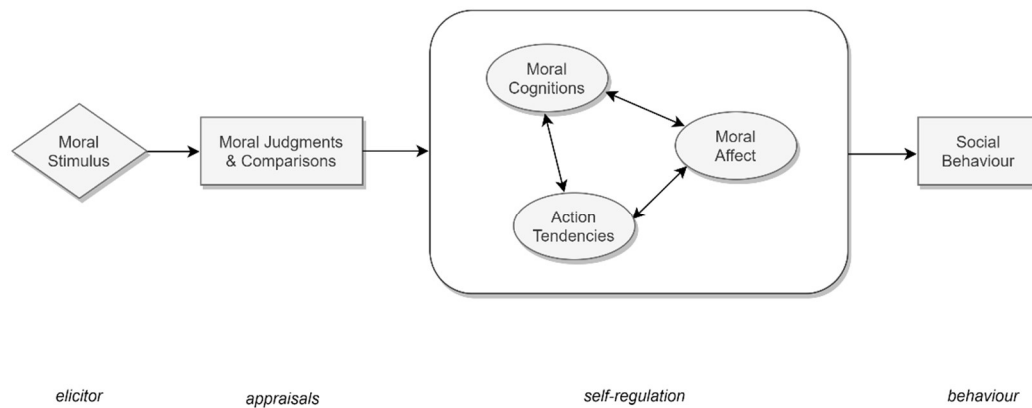
The moral experience of virtue can be studied through this “componential” lens. An act of moral goodness can be viewed as the elicitor, the event that gives rise to an emotion episode. This elicitor, depending on its relevance and meaningfulness to the individuals’ moral concerns, could trigger early cognitive appraisals and moral comparisons, which in turn could instigate corresponding self-regulatory processes, characterised by more complex sets of cognitions, action tendencies, and subjective feelings capable of energising specific social behaviours. These components are precisely the constituents of an emotion episode as described by the componential process theory. The only component omitted here is neurophysiological and motor changes, outside of the scope of this research.

When this componential perspective of emotion is widened even further, explicitly acknowledging and including the *self*, it is possible to analyse the same phenomena

under a more comprehensive nomological network, a *self-regulatory* framework, which is the perspective adopted in the present research, as illustrated in Figure II.

Figure II:

A streamlined nomological network of the moral experience of virtue observed through the lens of moral self-regulation



Other Factors Affecting Moral Comparisons and Self-Regulation

If moral comparison and self-regulation are indeed implicated in the wide spectrum of the processes elicited by displays of virtue, then answering questions about the mechanisms that govern this variety of responses is essential. For instance, what determines whether an act of goodness triggers upward or downward moral comparisons? In the presence of upward comparisons, what instigates self-improvement as opposed to self-enhancement or self-protection?

To answer these questions beyond what already discussed, it is necessary to consider two classes of phenomena: those that pertain to the person (individual differences, e.g. personality traits) and those that pertain to the situation (contextual factors, e.g. features of the virtuous acts/agents). A more detailed description of these phenomena will follow in the next sections, examining individual differences first, and then situational factors.

Individual Differences

The analysis of individual differences in the present research attempted to establish the effect of specific dispositions on participants' judgments of the moral scenarios depicted in the stimuli and the ensuing regulatory processes. These dispositions fall into three categories: personality traits, motivational orientations, and self-beliefs.

Personality Traits. Personality traits are defined as relatively stable patterns of thought, affect, motivation, and behaviour that represent human universals (McCrae & Costa, 1997). Two of these broad personality traits potentially relevant to the present investigation are humility and narcissism.

Humility forms with honesty one of the six factors of the HEXACO model (Lee & Ashton, 2004). It does not explicitly appear in the five-factor model (Costa & McCrae, 1992), but one of the facets of agreeableness is modesty, which is also one of the facets of honesty-humility in HEXACO. Humility and modesty have not been clearly distinguished in the psychological literature until recently, and the fact that laypeople and research participants tend to conflate them has not helped empirical research (Exline & Geyer, 2004). In the last decade, research has offered evidence that general humility is a broader construct that includes modesty content (Davis & Hook, 2014), consistent with the factorial structure of the HEXACO model. Its most central feature is hypo-egoic non-entitlement, or in other words, "the belief that, no matter how extraordinary one's accomplishments or characteristics may be, one is not entitled to be treated special because of them" (Banker & Leary, 2019, p. 1). Research has specified that general humility can apply to a variety of life situations or domains, giving rise to more specific constructs such as intellectual humility (unassuming self-restrained approach to the negotiation of one's ideas with those of others: see McElroy et al., 2014) and cultural humility (other-orientated interpersonal stance characterised by lack

of assumptions about the superiority of one's background in multicultural contexts: see Hook et al., 2013). The underlying thread that unifies these aspects of humility is partly intrapersonal (moderate and accurate views of one's strengths and weaknesses) and partly interpersonal (behaviours that mitigate attention to the self, facilitating cooperation, reducing envy or jealousy in groups, handling conflict, power struggles, cultural differences, and disagreement in a respectful and unassuming manner) (Davis et al., 2016). The intrapersonal dimension of humility could be particularly relevant in the context of the present research: indeed, if individuals tend to hold moderate views of their attributes, including moral attributes, they should be more likely to engage in upward comparisons in response to someone's remarkable acts of moral goodness.

The opposite could be true for narcissism, a personality trait characterised by entitled self-importance (Krizan & Herlacher, 2018). Narcissism has been found to be associated with heightened social comparison processes, particularly downward comparisons, given the highly flattering self-views held by narcissists (Krizan & Bushman, 2011). Krizan and Herlacher (2018) recently developed an integrated "narcissism spectrum model" whose unifying feature is the tendency of narcissists to view their own needs and goals as more significant than those of others, which makes them exhibit an inflated sense of deservingness. These authors have factor analysed a wide array of narcissism items from some of the most widely used measurement instruments in the field. Their results confirmed earlier findings – acknowledged in the literature since the early 1990s (Wink, 1991) – about the existence of two distinct manifestations, both related to a common narcissistic phenotype based on entitled self-importance. These two manifestations are *grandiose* and *vulnerable* narcissism¹². Krizan and Herlacher

¹² For a review of the manifold labels of these main narcissistic categories, see Dickinson & Pincus, 2003.

(2018) showed that the expression of entitled self-importance can be either bold, assertive, exploitative, and exhibitionist (grandiose narcissism: see also Crowe et al., 2016) or reactive, hypersensitive, volatile, and vindictive (vulnerable narcissism: see also Hendin & Cheek, 1997). According to them, both expressions carry a strong antagonistic potential, but while the grandiose type is expressed in a callous and manipulative way, the vulnerable type is often revealed through anger and hostility. Additional evidence exists of the associations between grandiose narcissism with high self-esteem, and vulnerable narcissism with low-self-esteem (Rohmann et al., 2012). Following downward comparison, individuals scoring high on self-esteem and narcissism (grandiose) were found to self-enhance and experience more positive affect (Campbell et al., 2000), while individuals low in self-esteem and high in narcissism (vulnerable) tended to self-protect and experience more negative affect and hostile intentions (Hart et al., 2018).

Beyond Personality Traits: Characteristic Adaptations. More recent theorising has broadened the concept of personality structure beyond the notion of traits, to encompass other dimensions, particularly what McAdams and Pals (2006) first, and then DeYoung (2015), defined *characteristic adaptations*. These are “relatively stable goals, interpretations, and strategies, specified in relation to an individual’s particular life circumstances” (DeYoung, 2015, p. 38). By capturing both traits and characteristic adaptations, personality psychology can offer a more holistic framework, capable of understanding the person “as a whole” (McAdams & Pals, 2006) and providing not only an account of how individuals differ from each other, but also an explanation of why they do so (DeYoung, 2015). This is possible because the strategies, goals, and interpretations that shape individuals’ characteristic adaptations inform their goal-

directed self-regulated behaviour in all life domains (including morality) in a cyclical configuration typical of cybernetic systems (DeYoung, 2015)¹³.

Among the fundamental *strategies* that form part of characteristic adaptations, the present research considered a set of three motivational dispositions that individuals consolidate starting from the crucial early years of their development, as they learn from their caretakers how to socialise in their environment. These motivational dispositions are:

- *hedonic orientations*: chronic energisations of behaviour either toward positive, appetitive, and rewarding stimuli (approach) or away from negative, aversive, and punishing stimuli (avoidance) (Elliot & Thrash, 2002; Elliot, 2008);
- *regulatory focus* (Higgins, 1997; Higgins, 2014): motivational tendencies to either achieve growth and realise ideal hopes and aspirations (promotion focus) or to ensure security, and fulfil duties and obligations (prevention focus);
- *regulatory mode* (Kruglanski et al., 2000): motivational propensities to either emphasise movement, action, and state shift (locomotion mode) or appraisal and accurate evaluation (assessment mode).

Further insights into the nature and utility of these motivational constructs will be discussed in the next chapters, when they are introduced in the models (Studies 2-4). Although not much research exists on the relationships between them and moral comparison and self-regulation, the definitions of approach and promotion focus are consistent with strivings toward self-improvement, while the definition of avoidance aligns with tendencies toward self-defensive regulation.

¹³ Another theoretical approach that expands the realm of personality beyond the traditional traits has been recently advanced by Fleeson and Jayawickreme (2015) with their “whole trait theory”, which bears some similarities, although in a different framework, with McAdams’s and DeYoung’s integration of traits with characteristic adaptations.

More nebulous appears the relationship between prevention focus and moral self-regulation. Despite their conceptual independence, theorising on the associations between regulatory focus and hedonic orientation has pointed out the stronger conceptual links of promotion combined with approach and prevention combined with avoidance, as opposed to the more tenuous conceptual links of promotion combined with avoidance and prevention combined with approach (see Cornwell & Higgins, 2015b). For this reason, prevention focus could potentially function similarly to avoidance in the prediction of self-defensive regulation (positive association).

Further, although research on regulatory mode is less developed, assessment and locomotion mode underpin two distinct motivational orientations in the goal pursuit process (Cornwell & Higgins, 2014): assessment entails epistemic concerns (need for truth) that appear consistent with strong comparison tendencies and iterative appraisals/reappraisals cycles that could trigger and amplify self-defensive regulation, whereas locomotion is governed by needs for control and psychological motion that could predispose to action, state-shift, hence self-improvement.

Among the fundamental *interpretations* that form part of characteristic adaptations, the present research investigated self-esteem. Self-esteem is related to global beliefs of worth, that is, “the feeling that one is good enough” (Rosenberg, 1965); it encompasses both positive and negative overall views about the self. The relationship between self-esteem and moral comparison has been controversial, due to the conflicting results that emerged in empirical research. Studies on direct comparison choice based on personality attributes suggested that people low in self-esteem are more likely to engage in downward comparison than people high in self-esteem (Friend & Gilbert, 1973); the same pattern was found in studies based on ability (Smith & Insko, 1987). This body of research seems to indicate that the choice of downward comparison

targets works as a *defensive* mechanism for people low in psychological resources when facing potential self-threats.

However, subsequent research in interpersonal relationships unveiled a more faceted reality. In those studies (see e.g., Crocker et al., 1987), participants were not asked to choose a comparison target, but provided separate evaluations for self and others on various measures of personality and ability. With this research paradigm, upward or downward comparisons can be inferred through the difference between self-other scores on the target measures. Results revealed that people higher in self-esteem rated themselves better than others on those target measures. Additionally, in studies where participants were divided in depressed and non-depressed groups, the non-depressed group showed significantly higher self-enhancement than the depressed group (Campbell, 1986). This second body of evidence seems to suggest that downward comparison functions as a self-serving mechanism for people high in self-esteem who strive to *maintain* flattering self-views. Further evidence (Tice & Masicampo, 2008) replicated the finding that high trait self-esteem is linked with self-enhancement, and low trait self-esteem with self-protection.

In the present research, self-esteem was conceptualised as a characteristic adaptation, like regulatory focus and hedonic orientation. Research suggests that trait self-esteem is positively associated with promotion focus and negatively associated with prevention focus (McGregor et al., 2007), and also positively associated with approach and negatively associated with avoidance (Heimpel et al., 2006).

Beyond these assumptions based on the literature, the present research also hypothesised that the characteristic adaptations could define specific conditional processes, interacting with moral comparisons, thus functioning as *moderators* of moral self-regulation. This is because of the expectation that characteristic adaptations,

depending on their intensity, could lead to different outcomes in terms of modes and levels of moral self-regulation given certain types and strengths of moral comparisons. For example, depending on the level of approach motivation, higher or lower degrees of self-improvement could be expected in the presence of certain levels of upward comparisons.

Situational Factors and the Vignettes

Although the individual differences described above constitute personal signatures that are relatively stable across situations (Fournier et al., 2015; McAdams & Pals, 2006), contextual factors significantly influence their functioning. A situation is defined as “a set of circumstances outside the person consisting of objectively quantifiable properties (often including other people) that may be perceived and interpreted by a person” (Baumert et al., 2017, p. 528). Given the huge number of these situational variables (and their possible combinations), in the present research they were not *systematically* manipulated, but critical sets of features of the situation (relative to the moral act and agent) were purposely assembled in each moral vignette, and these vignettes were pitted against each other¹⁴.

The moral acts were chosen to be relatively substantial accomplishments in the domain of moral goodness and not just trivial everyday expressions of courtesy or kindness. However, the degree of goodness was carefully dosed across the vignettes to depict a wide variety of deeds, from more accessible acts up to outstanding displays of virtue, each with different gradients of risk/cost to the moral agents and advantage to the beneficiaries. The stimuli also varied in the degree of “obligation” and “supererogation” of the deeds; these categories subsume respectively the deontological

¹⁴ A more classical manipulation against a control condition was conducted in the last study.

normativity of the actions (obligation, i.e., rules that ought to be followed) or their nature of being optional, beyond the call of duty (supererogation, i.e., permissible acts in terms of both performance/commission or omission: see Archer, 2018). Furthermore, the stimuli were created so that they showcased two fundamental categories of moral exemplars: the brave and the caring (Walker & Frimer, 2007). These reflect two basic domains of moral goodness, courage and care, which map respectively onto the broader social dimensions of agency and communion (Wiggins, 1991; Trapnell & Paulhus, 2012), dominance and nurturance (Wiggins, 1979), competence and warmth (Fiske et al., 2007). Given the emphasis in more recent literature on an additional aspect of moral goodness, the stimuli also portrayed a third fundamental moral domain, that is, justice (see e.g., Piazza et al., 2019; Walker & Hennig, 2004). Lastly, the vignettes featured various instantiations of other situational variables referred to the moral agents, such as their gender, ethnicity, profession (for a more complete account of situational variables in moral psychology, see e.g. Christensen & Gomila, 2012). All these features defined a set of boundary conditions that demarcate the range of circumstances to which the model refers and applies. As previously mentioned, the many variables that make up the contextual factors were not systematically manipulated in the present research, and a selection of them (those specifically referred to crucial characteristics of the deeds and the agents) were measured to guide selection of the vignettes: a) goodness of the deed; b) propriety of the deed; c) level of care/courage/justice¹⁵.

The use of vignettes in moral psychology has become popular because they enable the collection of data about participants' thoughts, feelings, attitudes, judgments, and behaviours that might otherwise be difficult to investigate in naturalistic settings. More

¹⁵ For further details on the selection criteria of the vignettes, see next chapter on Study 1.

details about the moral vignettes will be discussed in the next chapter. For now, suffice to say that for the present research twelve vignettes were initially created and tested. Of these, two were progressed to the next stages of the research, as they represented emblematic moral scenarios that best summarised a suitable assortment of critical contextual features, allowing to test to what extent the conceptual model could apply to different templates of virtue.

Effects on Social Behaviour

Social psychologists agree that social and moral comparisons are in the service of self-regulatory processes (Wood, 1996); there is also wide consensus that, in turn, self-regulation is in the service of behaviour, representing one of the important predictors of success in life (Baumeister, 2007). In a social cognitive perspective, behaviour is an inherent aspect of self-regulation. Following Bandura's (1991) social cognitive theory, Zimmerman (2005) refers to "behavioural regulation" as one of the regulatory systems that operates in reciprocal causation with the other regulatory components, "person" and "environment". In the context of this investigation, the behavioural component of wider self-regulatory systems was considered distinct from self-regulation at the intrapersonal level, which Zimmerman (2005) refers to as "covert self-regulation"¹⁶. Nevertheless, they are highly interconnected and therefore it was paramount to determine to what extent intrapersonal moral self-regulatory mechanisms affect overt behaviour.

Because of the expected prevalence of moral self-improvement responses to high moral exemplars (relative to self-defence), the most pertinent kind of behaviour to examine here was *prosocial* behaviour. In her seminal article, Wispé (1972) defined it as

¹⁶ To avoid any confusion, in the present research the term self-regulation strictly refers only to "covert self-regulation" within the person component of social cognitive self-regulation theory (see Zimmerman, 2005).

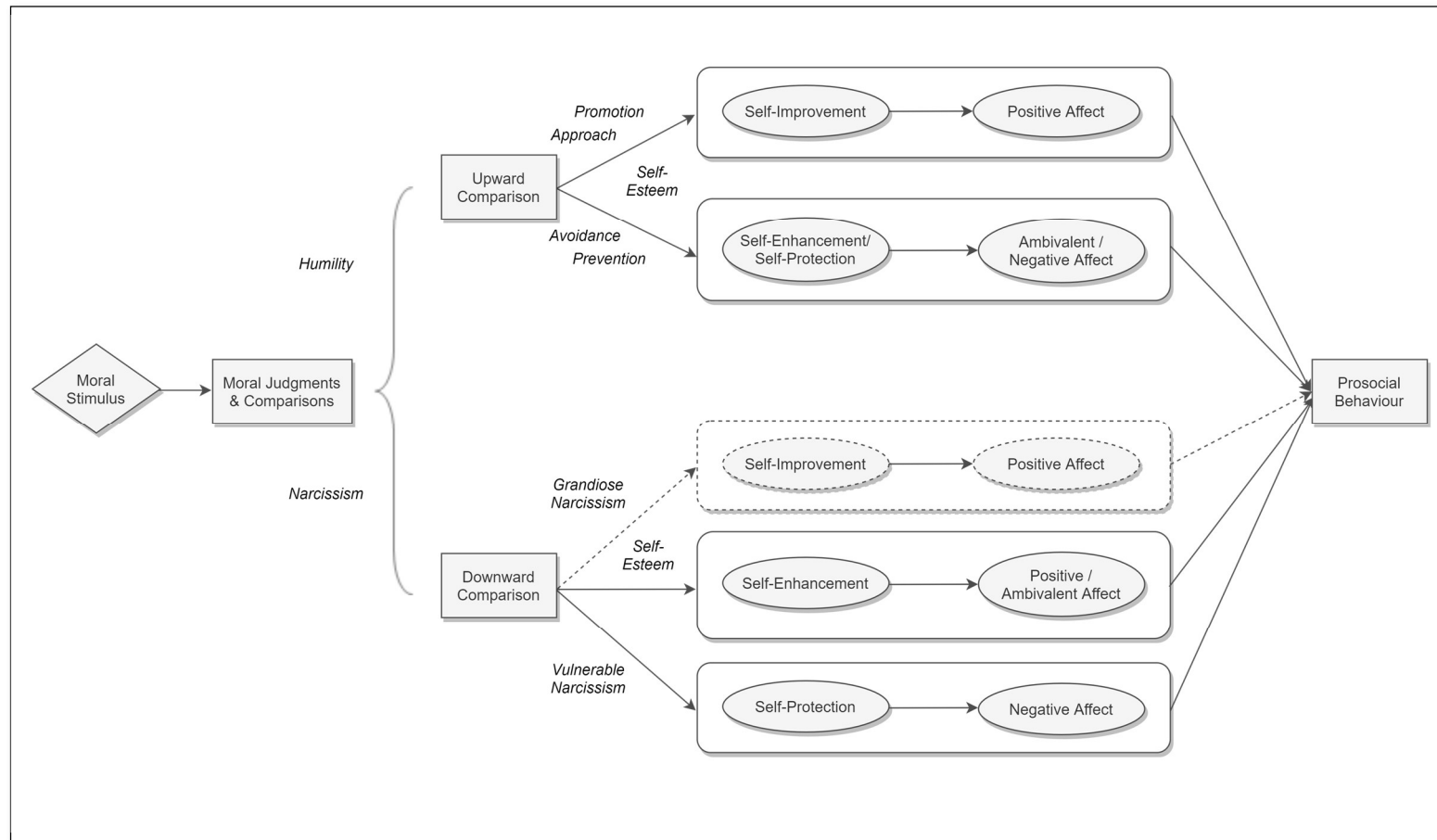
behaviour that has positive social consequences and consists of acts valued by society that improve the life and well-being of other persons. Existing research suggests that others' good deeds eliciting moral elevation promote *helping* behaviours (Schnall et al., 2010; Schnall & Roper, 2012), a particular manifestation of prosocial behaviour characterised by voluntarily aiding or donating to others with the selfless aim of enhancing their welfare (Wispé, 1972). Similarly, recent evidence suggests that others' good deeds can induce the emotion of *kama muta* (Blomster Lyshol et al., 2020), which can promote other positive social outcomes, such as increased humanisation of out-groups. The same kind of prosocial effects should be observed when studying the experience of moral exemplars from a self-regulation rather than an emotion perspective: virtuous actions eliciting moral self-improvement should lead to significantly higher levels of helping behaviour relative to morally neutral or mildly positive moral actions. To test this hypothesis, the present research examined one of the vignettes against a control scenario, measuring the difference in the helping behaviours that they induced; the control scenario was created so that it matched the key components of the experimental vignette, keeping constant the critical contextual elements except for the manipulating factor (the level of goodness of the deed).

Summary of Key Hypotheses and the Conceptual Model

Grounded in the psychological research and theorising available to date in the field, the considerations discussed thus far elucidate some of the key aspects related to the moral response to virtue, highlighting fundamental hypotheses about their mechanisms, antecedents, and proximate consequences. These are summarised in Figure III, which graphically portrays the nomological network and the conceptual model at the outset of the research.

Figure III:

The initial conceptual model of the moral experience of virtue observed through the lens of moral self-regulation, inclusive of a selection of key moderators (in italics)



The model illustrates graphically key initial assumptions and hypotheses of the research, all based on the literature reviewed in this introductory chapter. Translating them into a narrative form, these hypotheses predict that, following the presentation of moral stimuli (vignettes) portraying moral exemplars, participants will make initial judgments, about the goodness of the acts and of the moral agents, that imply specific moral comparisons. Of these comparisons, those based on ability will identify two critical participant groups: those engaging in upward and those engaging in downward comparisons.

For participants in upward comparison, the perceived positive self-discrepancy in moral character (self worse than the agents in the vignettes) will be stronger as their own level of self-assessed humility increases. If this discrepancy is experienced as an opportunity, self-improvement and positive affect will be elicited, which in turn will be likely to induce prosocial behaviours; if it is experienced as a threat, self-defence processes will be triggered (enhancing or protective self-regulation) and these, in turn, will likely hamper the enactment of prosocial behaviours. The strength of self-improvement effects will be positively moderated by promotion focus, approach, and self-esteem, while the strength of self-defensive effects will be positively moderated by prevention focus and avoidance, and negatively moderated by self-esteem.

For participants in downward comparison, the perceived negative self-discrepancy in moral character (self better than the agents in the vignettes) will be stronger as their own level of self-reported narcissism increases. It is theoretically possible that downward comparison experienced as an opportunity could induce self-improvement and positive affect, but most likely it will instigate defensive mechanisms, either self-enhancing or self-protective, to preserve their perceived superiority. The strength of self-enhancement effects will be positively moderated by self-esteem, leading to

positive or ambivalent affect, whereas the strength of self-protection effects will be negatively moderated by self-esteem, leading to negative affect.

Moral self-improvement will inspire degrees of prosociality directly proportional to its level (the higher self-improvement, the higher the degree of prosociality) and greater deeds will induce higher levels of prosociality than smaller deeds; by contrast, self-defence will proportionally inhibit prosocial behaviour, such that the higher self-defence, the lower the degree of prosociality.

Adaptiveness of Moral Self-Regulation Modes

The psychological literature on the self proposes a variety of interpretations about the degree of adaptiveness or maladaptiveness of the different types of self-regulatory processes.

Self-improvement is usually depicted as conducive to positive outcomes. In studies by Karney and Frye (2002) as well as Hui, Bond, and Molden (2012), participants reported better relations in their marriages thanks to self-improvement strivings; Ryff (1991) found that participants' self-improvement endeavours were gratified by benefits in various dimensions of well-being, such as self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life, personal growth. Cross-cultural research in the US (Tsai, Chiang, & Lau, 2016) also revealed that self-improvement benefits were more pronounced among Asian Americans than European Americans.

More complex are the outcomes of the processes of self-enhancement and self-protection. The case of self-enhancement is particularly thought-provoking, due to the contradictory effects that make it "a mixed blessing" (Paulhus, 1998, p. 1207). Among the empirical studies that have investigated the consequences of self-enhancement, those by Robins and Beer (2001) are highly informative: they assessed short-term effects

through a group task in the laboratory and long-term effects in a longitudinal study in a real-world context. In the first experiment, they found that after completion of the task, self-enhancers who convinced themselves that they had done well tended to be narcissistic and ego-involved in the task, and reported increased levels of positive affect versus baseline. However, in the second study, the authors followed undergraduate students during the course of their studies (four years), and repeated measures of critical outcome variables revealed a downward trajectory for self-esteem, satisfaction with university, and well-being; importantly, self-enhancement did not correlate with actual academic performance measured through GPA and likelihood to graduate.

Further research found that self-enhancement in the short term was beneficial in terms of intrapersonal adjustment, in that it helped individuals restore self-worth while dealing with various kinds of self-threats, thus serving a stress-buffering function (Alicke & Sedikides, 2009). However, and perhaps more importantly, in the long run, it often ended up being dysfunctional, particularly in terms of interpersonal adjustment, since it was linked to the progressive deterioration of the individual's ability to use social feedback as a means of personal growth (Sedikides, 2009) and to the worsening of the quality of social relations; these effects were found to be more marked among individuals who, while self-regulating, tended to arrogantly exaggerate *agentive* traits (Dufner et al., 2018), inviting dislike and derision (Alicke & Sedikides, 2009), hampering the effectiveness of their social integration, and damaging others' experience of inclusiveness (Sedikides & Luke, 2007). For this reason, Crocker and Park (2004) suggested that self-enhancement in the service of establishing domination over others or in the pursuit of immediate emotional rewards, such as feeling good about oneself, tends to be so detrimental that it can subvert its greater goals.

Participant Clusters

Beyond conducting the traditional variable-centred analysis of the conceptual model described above, a complementary objective of the present research was to assess the existence of independent latent profiles of participants with meaningfully different characteristics. This *person-centred* analysis was carried out on multiple levels of the model (motivation, comparison, self-regulation) expecting to generate specific participant clusters for each of them. For instance, the analysis could potentially yield a group of individuals with a strong propensity to avoidance motivation, a group who engaged in downward comparison, and a group whose response pattern was characterised by strong self-defence; specific relations between these clusters could also be assessed, seeking to confirm the significance of theoretically driven association patterns deriving from the variable-centred analysis.

A Path Toward Causality

Through the methods designed to test the conceptual model, psychological traits and states in the nomological network were measured primarily as *quantities* and the analysis attempted to delineate associations across these quantities, offering initial probabilistic *causal* explanations about these relationships.

Rohrer (2018) recently pointed out that researchers often avoid making causal claims, especially with observational data, because they are afraid of being unable to justify their claims. They are trained that “correlation does not imply causation”. This is undoubtedly a valid principle, but progress toward a causal understanding of psychological phenomena *can* be made in the presence of specific experimental designs, assumptions, and analytic methods. Grosz and colleagues (2020) speculate that the taboo against causal inference does not prevent researchers and readers from drawing causal conclusions even when they are not overtly formulated. They claim that the inference simply remains opaquely at an implicit level; after all, “as humans, we cannot

avoid thinking in terms of causality” (Asendorpf, 2012, p. 391). These authors are not alone, as similar arguments have been put forward for example by Hernán (2018). Pearl and MacKenzie (2018) even claimed that in recent times the so-called “causal revolution” has provided the philosophical framework, the mathematical language, and the analytical tools to harmonise the realm of statistics (in particular, of probability) with that of causality. From their viewpoint, these advances have allowed to dissipate the fog that has enwrapped causal inference for a long time, giving birth to what they call the “new science of cause and effect” (Pearl & MacKenzie, 2018).

Leveraging these recent developments, the present research project aimed to generate and refine a new model of the self-regulation of virtue. While attempting to unearth some of the mechanisms hidden underneath the observable surface, causal inferences were drawn from “noisy observations” (Pearl, 2009). Based on the claims on causality advanced by Pearl, Rohrer, Hernán, and others, this thesis did not shy away from using explicit causal language. It adopted a *probabilistic* approach to causality (whereby certain causes are *likely* to generate specific effects under certain probability assumptions), thus excluding a deterministic view of cause and effect as *necessities* (whereby events are determined completely and necessarily by previously existing causes).

In this context, a *cause* was intended as a probabilistic determinant of an effect, which “listens” to its cause and determines its value based on what it “hears” from the cause itself (Pearl & MacKenzie, 2018). Along the same lines, a *causal model* was viewed as an inference engine that – based on prior knowledge – uses assumptions, queries, and empirical data to produce estimates (with a certain level of uncertainty) and ultimately provides answers to scientific questions of interest (Pearl & MacKenzie, 2018). All these elements are applied and described in more detail in the following

chapters, where descriptions of putative correlations are complemented by initial causal inferences, made possible partly thanks to experimental designs and partly through specific analytic approaches (independent of design) such as exploratory/confirmatory factor analysis (EFA/CFA), path modelling, structural equation modelling (SEM), and other latent variable modelling methods, such as latent profile analysis (LPA). With Pearl's "causal revolution", experimentation and observation, manipulation and description should no longer be considered incompatible pairs, but complementary components that contribute to causal and counterfactual expressions of the reality under investigation (Bollen & Pearl, 2013).

Research Roadmap

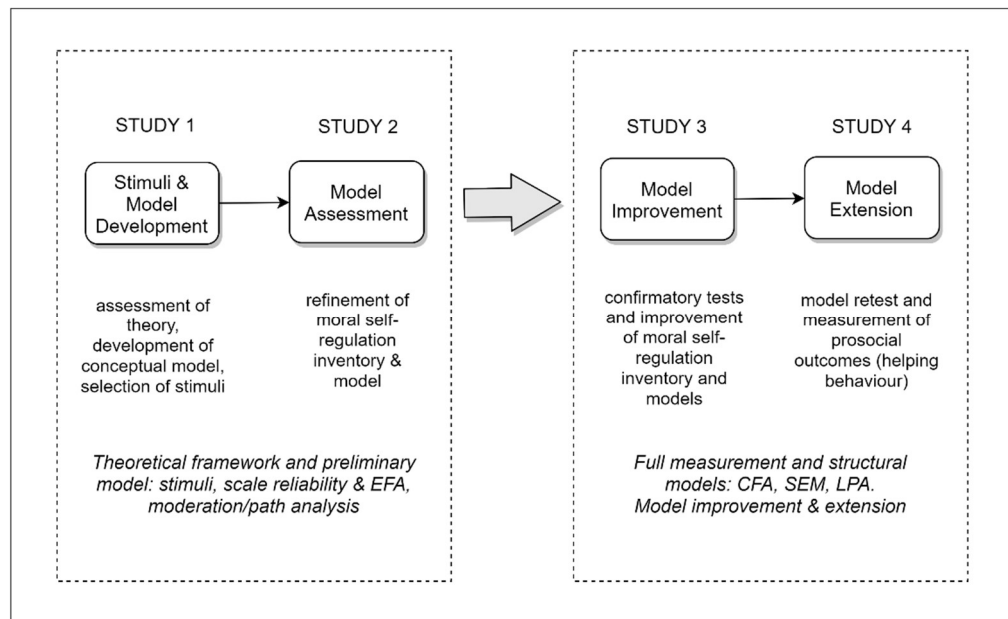
The process put in place before data collection, analysis, and reporting of this research project was aligned with the traditional steps recommended by Dubin (1969): a) a review of the literature allowed to establish a provisional nomological network, highlighting key constructs expected to be implicated in the phenomenon under investigation; b) appropriate measurement instruments were identified (if available) or created (if unavailable) for those constructs; c) tentative associations between the constructs were specified, alongside conditional processes and boundary conditions¹⁷; d) hypotheses and predictions were formulated and preregistered. Following this theory-building groundwork, data collection and analysis enabled to refine the measures and their relationships, until the theoretical model was tested through confirmatory tests and further improved and expanded to encompass a wider range of critical consequences.

¹⁷ Conditional processes are contextual or individual difference factors that define existence, magnitude, and direction of certain effects (see Hayes & Preacher, 2013) and boundary conditions are the temporal/conceptual limits that set the range of a theory (see Busse et al., 2017).

In terms of empirical research, the plan included a sequence of four core studies across generative/exploratory and confirmatory/integrative stages (Fig. IV):

- *Study 1 – Stimuli and model development*: creation of a set of new stimuli in the form of vignettes and choice of the most suitable ones for the project; selection of critical variables and construction/assessment of new scales to operationalise the constructs of moral comparisons and moral self-regulation; preliminary analysis of mutual relations across the key variables;
- *Study 2 – Model assessment*: analysis of the reliability and dimensionality of the moral self-regulation/affect constructs, selection of critical antecedents, analysis of their mutual relationships, and development of provisional path models;
- *Study 3 – Model improvement*: confirmatory tests and improvement of the measurement and structural models, and identification of motivation, comparison, and self-regulation typologies, with measurement of their associations;
- *Study 4 – Model extension*: retest of the model with integration to include social behavioural outcomes (helping behaviour) elicited by self-improvement states, through the adaptation of an existing experimental paradigm.

Figure IV:
Overview of the research plan and analytic framework



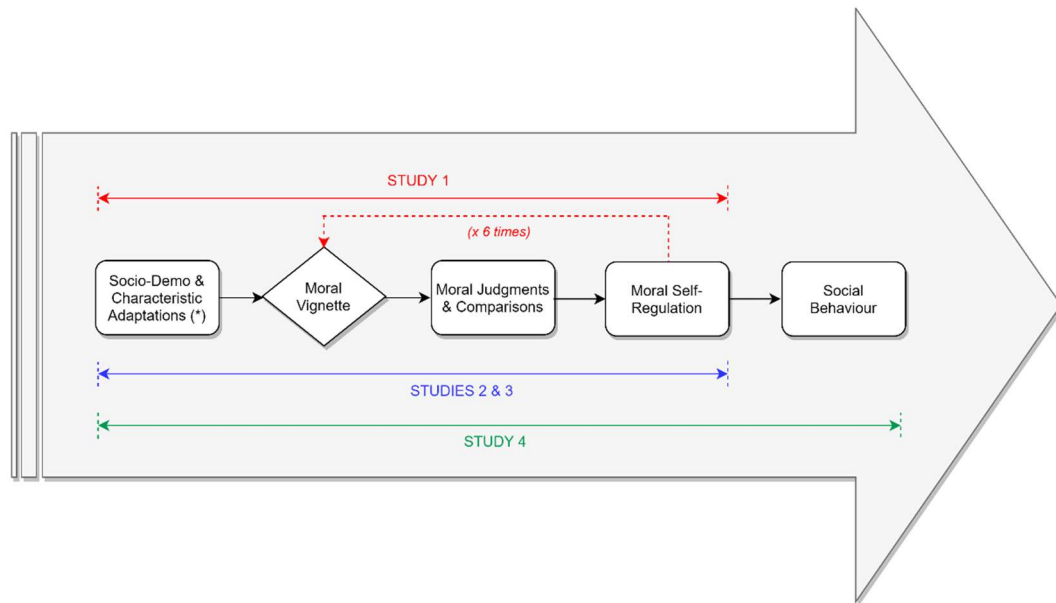
All the studies across the four phases of the research were preregistered¹⁸ on the Open Science Framework (OSF) website (<https://osf.io>).

The four studies had a similar structure and design, and the questionnaire followed a consistent flow across all of them, as illustrated in Figure V. The main differences concerned the presentation to each participant of multiple stimuli in Study 1 (mixed design) versus a single stimulus in Studies 2-4 (between-subjects design), and the integration of a behavioural task in Study 4.

¹⁸ Despite being essentially exploratory in nature, even the first two studies were preregistered. Some deviations from the preregistered plans are flagged out where appropriate throughout this thesis.

Figure V:

Graphical illustration of the questionnaire flow across the four empirical studies



(*) Note: characteristic adaptations did not appear in Study 1.

Study 1: Stimuli and Model Development

Introduction

The first study of the research plan outlined in the previous chapter was designed to provide learning on three key aspects: the stimuli, the measurement instruments, and the relationships across some of the main variables in the conceptual model. Its structure and implementation reflected the need to use the available resources as efficiently as possible, knowing that the following phases would be highly resource intensive.

The Moral Vignettes

Due to the prevailing interest in moral violations, empirical research in moral psychology does not offer a wealth of stimuli depicting good deeds. Most of them can be found in studies focused on positive moral emotions. An example is represented by the early work on the “other-praising” emotions of elevation, gratitude, and admiration by Algoe and Haidt (2009): in the elevation condition, the stimulus for one of the studies was a video featuring a young man who, as a boy, had established a shelter for homeless people in Philadelphia. Later, a series of studies on moral elevation by Erickson and colleagues (2017) used several videos depicting virtuous actions, for instance, a father pulling his paralysed son in a marathon, a teenager helping others escape gangs, and athletes with disabilities displaying courage and grit. In another study on moral elevation, Silvers and Haidt (2008) used a video from an episode of “The Oprah Winfrey Show”, where a musician told the story of how he was saved from a life of gang violence by his music teacher. This video was later used in other studies on moral elevation, among which those by Schnall et al. (2010), Lai et al. (2014), and Piper et al. (2015).

Instead of video clips, written stories administered on a computer screen were used in a set of studies by Freeman et al. (2009), and Aquino et al. (2011): one of these stories was about the members of an Amish community in the US who showed remarkable forgiveness and support to the family of a young man who massacred five young Amish girls in a schoolhouse and subsequently took his life. Other written stories were used in research by Thomson and Siegel (2013), who provided evidence that moral elevation was effectively induced among participants presented with them.

Stimuli in written format were also utilised in several neuroimaging studies on moral dilemmas. These are hypothetical scenarios involving a moral conflict, where people are forced to make difficult decisions about what they would do (or not do) if they were in those situations. The choices are usually between moral imperatives of which neither is objectively preferable or perfectly acceptable. Although the first to use moral dilemmas was Kohlberg (1964), it was Foot (1967) who introduced the one that became the most popular: the trolley dilemma. Many others were proposed in the following years, such as the footbridge, the lifeboat, the hostage, and so forth. The variety of moral dilemmas that proliferated in the literature became so wide that “the obtained evidence is neither necessarily comparable nor replicable across studies” (Christensen & Gomila, 2012, pp. 1249-1250). For this reason, researchers have tried to offer guidance on how to standardise the moral stimuli (even beyond dilemmas), so that more rigorous control can be achieved (see e.g., Christensen & Gomila, 2012; Clifford et al., 2015). The guidelines concern both formal and substantive elements: presentation format, expression style, word framing, word count, order of presentation, type of question, participant perspective, situational antecedents, intentionality, certainty, normality, justifications, factors related to the moral agents, type of moral action, type of outcome (Christensen & Gomila, 2012).

Moral dilemmas are not the kind of stimuli that are useful to study the response to others' good deeds, but the issue of standardisation was very relevant for the present research. The videos used in previously cited work differ remarkably from each other in length and format, and prove the relevance of the issues highlighted by Christensen and Gomila, especially when comparing experimental with control conditions in the same study. Written stimuli are more flexible than videos because they can be more easily created *ex-novo* and more easily modified after pre-testing to address emerging improvement needs; they can also be easily created in larger numbers, potentially manipulating each relevant factor. Therefore, a decision was made for the present research to generate a set of new stimuli in the form of vignettes with text and images; they were designed to cover a wide range of good deeds, display differential degrees of virtue and normativity, a variety of moral exemplars, while sharing essentially the same format and length.

While all the vignettes were intended to depict good deeds, they also contained ambiguous cues subject to diverging interpretations¹⁹; this is because the social reality of real-life situations is invariably steeped in complex textures of subtle “signifiers” that can take on more than just one “signified” (de Saussure, 1959). Depending on specific histories and accretion processes (Fleeson & Jayawickreme, 2015) that crystallise certain propensities, individuals select and deselect, take in and filter out different signs (*reinforcers* in behaviourist theories) from the surrounding environment and form mental representations, evaluating and making sense of them by attributing meaning. Thus, it was expected that the narratives in the vignettes, which describe real-life

¹⁹ For example, in the vignette named after him, Cory rushed into a burning building after fighting off members of the security staff who tried to restrain him. This action may be interpreted in a positive light, putting the emphasis on the favourable outcome, or a negative light, putting the emphasis on the transgression of the security staff orders. For further details, check the Methods section in this chapter.

situations and facts as reported by the media in the public domain (see Appendix 2), could be subject to diverse interpretations and meanings (Monni et al., 2020) as a function of participants' personality structures (traits and characteristic adaptations discussed earlier). It is for this reason, among others, that in Study 2 measures of individual differences between participants were introduced (for instance, regulatory focus), as they could potentially account for differential propensities to decode and interpret the ambiguous signs in the vignettes, contributing to giving rise to dissimilar moral judgments and self-regulatory processes.

One of the key objectives of Study 1 was to enable decisions about which moral vignettes were apt be progressed to the following stages of the research. The main criteria for their evaluation included their ability to:

- depict morally motivated good deeds in different moral domains, such as courage, care, and justice, and with different degrees of obligation and supererogation;
- portray incremental gradients of exemplarity of moral character through a diverse range of deeds, from more ordinary acts of goodness to truly uncommon altruistic actions;
- trigger moral judgments subsuming direct or indirect moral comparisons and elicit functional and dysfunctional moral self-regulatory processes through the written narrative format.

Measurement Instruments

Moral Self-Regulation Inventory

Besides new stimuli, appropriate measurement instruments were required for the research, particularly for the construct of moral self-regulation. A widely accepted state measure of the self-regulation of virtue was nowhere to be found in the literature; thus, one of the goals of this research was to develop a satisfactory operationalisation of the

concept, capable of capturing the nuances of the distinct mechanisms through which the self negotiates other people's virtuous acts. The psychological literature provided useful starting points.

First, existing work on moral elevation investigated specific thematic areas in common with adaptive broadening moral self-regulation. Scholars have used in disparate combinations various items known to tap into the different components of moral elevation, and often analysed them individually, correlating each of them with the outcome measure under investigation (Schnall et al., 2010), or creating a scale and computing the composite mean of the items or groups of items (Aquino et al., 2011). Schnall and colleagues (2010) attempted data reduction through principal component analysis and retrieved one principal component capable to summarise moral elevation items. However, observing the response to acts of moral goodness from the perspective of self-regulation (instead of emotion) meant slightly shifting the point of view; while certain elements overlap with an emotion-focused analysis, new aspects related to the self had to be integrated into a measure of adaptive moral self-regulation, for example the level of identification between the self and the moral agent.

Second, a completely new set of items had to be developed to measure defensive moral self-regulation. The literature does not provide a consolidated state instrument to measure these processes when activated by an eliciting positive stimulus. What is available though, is an inventory of items that measures individual differences in the propensity to adopt self-enhancing and self-protective strategies (Hepper et al., 2010). While the content was sometimes relevant, the wording had to be changed from the average frequency in the deployment of those strategies to the situational occurrence of a set of transitory thoughts, action tendencies, and feelings elicited by the presentation of the vignettes.

Despite this critical difference, the existing trait scales provided helpful hints. For instance, Hepper and colleagues' item "When you do poorly at something, thinking it was due to the situation, not your ability" confronts participants with the strategy of deflecting responsibility for lack of ability to circumstantial factors not depending on themselves; this item inspired reactions of motivated avoidance to accept the possibility of lack of moral character to perform the same moral deed as the agent in the vignette, which came together in items such as SD5 ("These extreme behaviours should not be considered the standard we live by") and others tested in Study 2. Another example is the item "Thinking of yourself as generally possessing positive traits or abilities to a greater extent than most people", which presents aggrandising self-construals that were reconfigured in item SD2 ("In many ways, I have done greater deeds than Francia").

Another source of inspiration was provided by studies on do-gooder derogation. For instance, Minson and Monin (2012) used the semantic differential and asked their participants to rate vegetarians across several attributes, among which was "humble-conceited"; this gave rise to the idea of assessing to what extent participants reckoned the moral agent in the vignette felt superior, which inspired item SD6 ("Francia probably thinks she's better than everyone else"). Also, Monin (2007) described trivialisation as one of the critical strategies to deflect self-threats from unflattering moral comparisons, and this notion was distilled into item SD1 ("It's not such an extraordinary action"), as well as others tested in Study 2. Collectively, these sources provided input to the generation of a set of new self-growth and self-shielding indicators; these tentatively formed the first version of the newly developed moral self-regulation inventory, which was pre-tested in a streamlined variant in the present study.

Only a selection of six self-improvement and six self-defence items were tested in Study 1 because of the limitations imposed by the methodology, primarily concerning

participants' fatigue related to the length of the questionnaire (see the Methods section in this chapter). Consequently, Study 1 did not investigate whether the self-defence items formed one or two distinct subscales (self-enhancement and self-protection), deferring investigation of this aspect to Study 2.

Moral Affect Scales

The moral affect scales were intended to measure feelings or subjective experiences of pleasure/displeasure self-reported by participants. They were measured after participants answered the moral self-regulation questions. Several scales exist in the literature in this area, for example, the Affect Circumplex (Russell, 1980), the AIM - Affect Intensity Measure (Larsen et al., 1986), the PANAS - Positive and Negative Affect Schedule (Watson et al., 1988), the SAM - Self-Assessment Manikin (Bradley & Lang, 1994), and others. However, not all of them would have been wholly relevant to the measure of the response to good deeds, and they would have been too long and time-consuming to administer in their entirety. Therefore, two brief scales were developed to focus on the discrete affect terms underlying the positive and negative feelings more directly implicated in the response to others' good deeds. For example, item PA1 ("I felt uplifted") came from the emotion word "uplifted" used, among others, by Aquino and colleagues (2011) in their seminal study on moral elevation. Item PA2 ("The story was inspiring") was borrowed from the emotion word "inspired" used by Algoe and Haidt (2009) in their study on moral elevation. The feeling of envy that features in item NA4 ("To be honest, I felt envious") was suggested by a study by Smith and colleagues (1994), which links this negative emotion to the sense of inferiority evoked by a position of advantage enjoyed by the envied person.

Moral Judgments and Moral Comparisons

A further need in the initial phase of the current research programme concerned the measurement of specific judgments underlying moral comparisons. Especially important was to develop a measure capable of capturing the participants' perception of the differential in terms of moral stature and capability between the self and the moral agents portrayed in the stimuli. To this end, a *direct* and an *indirect* measure of moral character discrepancy (or simply moral discrepancy) were constructed and tested to choose the most suitable one (further details in the Methods section).

Association Patterns Between Variables

Lastly, besides filling the gaps in terms of stimuli and measurement instruments, Study 1 was designed to explore patterns of mutual relationships between critical variables in the nomological network illustrated in the previous chapter (Fig. III). No specific hypotheses were made about the prevalence of either mode of self-regulation (self-improvement or self-defence) in upward or downward comparison; self-improvement and self-defence were simply assumed to be theoretically possible in association with each of the two comparison types. Empirically assessing this assumption was one of the objectives of the present study.

The relations between moral self-regulation and affect needed to be explored, too. In this case, although Study 1 was exploratory, it seemed plausible to anticipate that moral self-improvement would be associated with a prevalence of positive affect, whereas moral self-defence would link more closely with negative affect, consistent with the conceptual model. These patterns were scrutinised based on the empirical data collected in this first study to inform hypotheses for the following studies.

Summary of Key Objectives

In summary, Study 1 was designed under the general framework of the conceptual model described in the previous chapter and was carried out as a preliminary empirical

exploration of the moral landscape of the response to moral exemplars. Its main objectives were to:

- test a set of new moral scenarios in the form of vignettes, in order to verify their suitability to the research and select the optimal stimuli to progress to the following studies;
- test a selection of items from the newly developed moral self-regulation inventory, providing an early assessment of the quality of their wording and their psychometric properties, and identifying possible improvements;
- test and select one of two measures of moral discrepancy between the self and the moral agent underlying comparative processes in terms of moral character;
- provide initial insights into the nature and strength of the relationships between moral discrepancy, moral self-regulation, and affective states elicited by exposure to virtuous acts of moral goodness.

Methods

Participants

Study 1 was conducted between August and September 2018 using a structured questionnaire hosted by Qualtrics. Recruitment was completed through a variety of methods, including face-to-face invitations by the experimenter at the Lancaster University campus, posts on social media, email requests to postgraduate research groups at Lancaster University and other universities in the UK. Data collection took place on campus using university tablets, or remotely.

The sample size reflected the exploratory nature of the study, the number of research stimuli, the number of items for each of the new scales, the length of the online questionnaire, and accessibility of voluntary participants in the allocated timeframe. Power analysis with G*Power ver. 3.1 (Faul et al., 2007) in a streamlined scenario indicated that, to detect a medium-sized effect in a multiple linear regression with two to three predictors with $\alpha = .05$ and power set at .80, a sample of 68 to 77 participants per vignette would be necessary (see Supplemental Materials SMA1²⁰). Based on the research design discussed in the next section, the intended overall sample size for analysis was determined to be 150 participants.

Among the 212 subjects who started the questionnaire, 66 dropped out before finishing it; the data from these 66 individuals were therefore deleted. The final sample consisted of 146 participants, of which 115 were female (78.8%) and 31 male (21.2%). Their age ranged from 19 to 81 years, with median of 51 and mean of 49 years ($SD = 15$). Participants were primarily residents of the UK (67.8%) and the USA (10.3%); the rest

²⁰ Hereon, the Supplemental Materials will be abbreviated and indicated with three letters and numbers, e.g. SMA1-1, SMB1-1, etc. The first two letters SM indicate Supplemental Material; the third letter A indicates Study 1, the third letter B indicates Study 2, etc.; the numbers indicates the progressive count.

resided in various countries in Europe (17.4%) and other continents (4.1%); one participant did not disclose the country of residence. The participants' main nationality largely reflected the country of residence, with minor variations. Most participants were employed (54.1%), one was a student (0.7%), and the remainder had other occupations (6.2%) or did not reveal their employment status (39%). The median completion time was 34 minutes. Further details about the socio-demographical variables, including political orientation and level of religiosity/spirituality, can be found in SMA3.

Research Design and Materials

The online questionnaire consisted of two parts: a) a brief introductory section, with general socio-demographic questions and a moral self-evaluation question; b) the experimental section, with the administration and rating of the moral vignettes.

The structure of the study was a mixed design. Each participant was presented with a random selection of six out of twelve moral vignettes, and therefore each vignette was evaluated by a range of 68 to 76 participants. The independent variable was the vignette; the dependent variables consisted of measures of moral appraisal, moral self-regulation, and moral affect (plus an open-ended question for general comments), which were collected for all the vignettes presented to each participant.

The twelve vignettes depicted various good deeds performed by moral agents in favour of a third party. The vignettes were all based on true stories in the public domain gathered by the experimenter. They covered a variety of aspects of morality: several focused primarily on care, while others were more centred on courage or justice. The format of the vignettes comprised text (from 134 to 183 words) and visuals (2 to 5 pictures) taken from Internet websites that reported the stories. Each vignette was

labelled after the name of the protagonist of the story: Cory, Wesley, Arnaud, Francia, Matthew, Sarah, Joey, Markus, Ruxandra, Alvaro, Sunita, and Nicholas²¹.

The twelve vignettes can be viewed in Appendix 1 and the questionnaire in SMA2. A summary of the content of the stories is reported in Table A1:

Table A1:
Summary of the content of the stories in the 12 vignettes

Short label	Full name of moral agent	Story content
Cor	Cory	a man who saved a woman from the fire that broke out in her building
Wes	Wesley	a man who jumped down the tracks of the New York subway to save someone who fell off the platform
Arn	Arnaud	a policeman who swapped places with a hostage during a terrorist attack in France and ended up being killed while saving the woman
Fra	Francia	a woman who saved a friend's life by donating her a kidney
Mat	Matthew (abbr. Matt)	a runner of the London marathon who compromised his finishing time by helping someone who was about to collapse toward the end of the race
Sar	Sarah	a woman who, despite being insulted by an online troll, showed him compassion and paid surgery for the back condition that he could not afford to treat
Joe	Joey	a man who offered his shirt and hat to a shirtless person shivering on a freezing train in winter in the New York subway
Mar	Markus	a man who left his successful career behind and flew to South America to volunteer and help children in need
Alv	Alvaro	a former bullfighter in Colombia who repented and became a campaigner against the brutality of bullfighting
Rux	Ruxandra (abbr. Ruxa)	a woman who became vegetarian and blogs to raise awareness on animal cruelty in the food industry
Sun	Sunita	a woman who campaigns so that Western countries, the largest contributors to global warming, provide financial aid to developing countries for the protection of the environment
Nic	Nicholas	a lawyer who challenged the Anti-Homosexuality Act in Uganda and pushed the Constitutional Court to declare it illegal on the basis of discrimination

²¹ Hereon, in several tables and graphs (particularly in the Supplemental Materials) the vignettes will be abbreviated with the short labels shown in Table A1.

Measures

Moral Appraisals

Before the presentation of the vignettes, participants rated the perceived level of their own morality (hereon, moral self-evaluation, or simply self-evaluation) through the question: “To what extent do you see yourself as a morally good person?”. Participants were asked to answer using a slider bar, with a unipolar scale ranging from 0 (*not at all*) to 100 (*very much*)²².

After the presentation of each vignette, a few closed-ended questions asked participants to evaluate the moral character of the agents (hereon, moral agent evaluation, or simply agent evaluation) and the perceived level of moral similarity of the agents to themselves (hereon, similarity). These moral evaluations were measured with unipolar scales ranging from 0 to 100, using slider bars. For example, the following item was used to assess Francia’s moral character: “To what extent do you consider Francia to be a morally good person?” (0 = *not at all*, 100 = *very much*).

A further set of moral evaluations revolved around the quality and nature of the moral action performed by the protagonists of the stories: participants rated the perceived level of goodness of the deeds (hereon, goodness), the extent to which they believed the moral actions were “the right thing to do” (hereon, propriety), the level of relevance of the actions to the circumstances of their own life (hereon, relevance), and the extent to which they would have done the same thing in that specific situation (hereon, in-shoes). Again, these moral evaluations were measured with unipolar scales ranging from 0 to 100 (0 = *not at all*, 100 = *very much*) using slider bars.

²² It must be noted that in all the four studies of this research, for all the scales with slider bars (whether unipolar or bipolar) participants saw the labels of the anchor points, but not the scale numbers.

Participants were also asked to make a *direct* comparison between their own perceived moral character and that of the moral agent, again using a scale 0-100 with slider bars. In the instance of Francia's vignette, the anchor points were: 0 = *Francia is much less moral than I am*, 50 = *Francia and I are equally moral*, 100 = *Francia is much more moral than I am*; participants were also allowed to bypass this question, by choosing *I don't compare myself to Francia*.

A variable measuring an *indirect* form of moral character comparisons was computed by subtracting the self-evaluation score from the agent evaluation score. This variable, called moral discrepancy, could potentially vary from -100 to +100. Positive values indicated an upward comparison (agent evaluated more positively than the participant), whereas negative values indicated a downward comparison (participant evaluated more positively than the agent); zero indicated parity in this indirect form of moral comparison (lateral comparison).

Lastly, participants were asked to rate the degree of courage, care, and justice of the moral actions performed by the agents; these three independent evaluations were measured again with slider bars using unipolar scales varying from 0 = *not at all* to 100 = *very much*. The wording of the question, for example for the item courage, was the following: "To what extent does the story depict an act of courage?".

Moral Self-Regulation

Following the moral appraisals, participants were asked to rate the extent to which they agreed with a set of items aimed at measuring the moral self-regulatory response elicited by the presentation of the vignettes. These items comprised a variety of cognitions and action tendencies through which participants elaborated on their initial moral judgments while negotiating their self-concept; therefore, these items underpin processes wherein participants: a) reconsidered how the good deed performed by the

moral agents bore on their own moral self; and b) regulated the impact of the agents' actions on their self to orientate their own behaviour.

The items were divided into two distinct sets of cognitions/action tendencies: *positive* and *negative*. The positive cognitions/action tendencies constituted the new scale of moral self-improvement, and the negative cognitions/action tendencies the new scale of moral self-defence; the latter contained a mix of self-enhancement and self-protection items. Together, the moral self-improvement and moral self-defence scales formed the moral self-regulation inventory.

The list of moral self-regulation items tested in Study 1 is shown in Tables A2-A3. All these indicators were measured with 0-100 agreement scales using slider bars (0 = *strongly disagree*, 50 = *neither agree nor disagree*, 100 = *strongly agree*). They were presented in randomised order, and all of them were worded in the direction of high self-enhancement or self-protection (no reverse-coding), following Hepper and colleagues' (2010) trait scales.

The moral self-regulatory items included in Study 1 represent a narrow selection of those developed in the early stage of the project. Because of the research design and the length of the questionnaire, a choice was made to prioritise and test only six items for the self-improvement scale and six items for the self-defence scale.

Table A2:*Moral self-improvement items tested in Study 1 (example of Francia's vignette)*

Item code	Label	Item wording
SI1	Admirable	Actions like this are truly admirable
SI2	Awakened	When I read these stories, I feel awakened to the good in the world
SI3	Humanity	This story strengthens my faith in humanity
SI4	Values	Francia and I share the same values
SI5	BeBetter	Francia has shown me how to be a better person
SI6	ForOthers	I feel like I want to do something good for others

Table A3:*Moral self-defence items tested in Study 1 (example of Francia's vignette)*

Item code	Label	Item wording
SD1	Ordinary	It's not such an extraordinary action
SD2	MeGreater	In many ways, I have done greater deeds than Francia
SD3	Praise	Francia's actions may be good, but I bet she is seeking the praise of others
SD4	Untrue	This story is too good to be true
SD5	Extreme	These extreme behaviours should not be considered the standard we live by
SD6	Superior	Francia probably thinks she's better than everyone else

The item generation process involved a combination of deductive *theory-driven* approaches and inductive *data-driven* methods (DeVellis, 2003). As previously noted, some of the self-improvement items were inspired by measures of moral elevation utilised in previous research; for example, Aquino and colleagues (2011) used the item “The person/people in the story have shown me how to be a better person” (corresponding to item SI5: “Francia has shown me how to be a better person”), Schnall and colleagues (2010) used the item “optimistic about humanity” (corresponding to item SI3: “This story strengthens my faith in humanity”). The items of the self-defence scale were generated mostly within the theoretical framework of self-enhancement and self-protection. Some of the items refer to self-serving construal processes (Sedikides, 2012), for example item SD4 (“This story is too good to be true”), or dismissive reasoning, for example item SD5 (“These extreme behaviours should not be considered the standard we live by”). Other items refer to the trivialisation of the deed, for example item SD1 (“It's not such an extraordinary action”) or the vilification of the agent

(Sedikides, 2012), for example item SD6 (“Francia probably thinks she’s better than everyone else”). Other items refer to self-aggrandising processes, for example item SD2 (“In many ways, I have done greater deeds than Francia”).

Following protocols recommended in the literature (see e.g., Bastos et al., 2010), the new experimenter-generated items were pre-tested qualitatively in a few brief interviews with members of the target population to assess content validity, as well as clarity and comprehension of their verbal expression. Then a selection of them was included in the present study, where they were presented in randomised order.

Moral Affect

Following the moral self-regulation measures, participants were asked to rate the extent to which they agreed with a set of items aimed at measuring their perceived affective reactions to the vignettes. These items assessed how participants felt after reading the moral stories.

The full list of moral affect items tested in Study 1 is shown in Tables A4-A5. Two items were designed to capture specifically the perception of opportunity (PA3: “I felt challenged in a positive way”) and threat (NA2: “I felt as if I was threatened by something”) that the literature considers central to self-regulatory mechanisms. Item NA1 (“I had a mix of conflicting feelings”) was included to capture ambivalent feelings that could be associated with self-enhancement (see conceptual model in Fig. III). The other items assessed further affective states relevant to the response to good deeds, for example inspiration, envy, guilt, resentment.

Like with the moral self-regulation scales, all the affective items were measured on 0-100 agreement scales using slider bars (0 = *strongly disagree*, 50 = *neither agree nor disagree*, 100 = *strongly agree*). Again, the affect items included in this study were a narrow selection of those developed in the early stage of the project, and a choice was

made to prioritise and test only three items for the positive affect scale and five for the negative affect scale. All the items were presented in randomised order.

Table A4:
Positive moral affect items tested in Study 1

Item code	Label	Item wording
PA1	Uplifted	I felt uplifted
PA2	Inspired	The story was inspiring
PA3	Challenged	I felt challenged in a positive way

Table A5:
Negative moral affect items tested in Study 1

Item code	Label	Item wording
NA1	Conflicted	I had a mix of conflicting feelings
NA2	Threatened	I felt as if I was threatened by something
NA3	Guilty	It made me feel guilty
NA4	Envious	To be honest, I felt envious
NA5	Resentful	I felt resentful

Open-Ended Questions

At the end of the closed-ended questions, participants had the opportunity to answer an open-ended question for each vignette. Specifically, the question asked: “Would you like to make any comments regarding this story or questionnaire? Please, feel free to share any relevant thoughts or feelings”. Participants’ answers were analysed to provide additional insight into the general content of the moral stories, as well as the wording of the questions and the items of the new scales.

Procedure

The study had received prior ethical approval by the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University (UK). Participants were invited to follow a link to an online questionnaire. They read information about the study, including that they could withdraw at any time without giving a reason, and gave their consent to participate. After the introductory socio-

demographic and the moral self-evaluation questions, participants viewed a random selection of six vignettes, providing answers to the moral appraisal, the moral self-regulation, the moral affect, and the open-ended questions. Following this, they were debriefed and thanked for their participation. They also learned that, if they wanted, they could take part in a prize draw to win an Amazon voucher. The draw took place on 27th October 2018 and on the same day the vouchers were emailed to the winners.

Analytic Approach

The dataset from Study 1 was analysed using IBM SPSS Statistics ver. 25-26, as well as R ver. 3.6 (R Core Team, 2020) and RStudio ver. 1.2 (RStudio Team, 2020)²³, employing a variety of descriptive and exploratory techniques. First, multivariate outlier analysis was conducted to detect extreme data points. Next, analysis of the distributions and key assumptions for the main variables was conducted. Regression and correlation analyses were carried out to single out critical variables and investigate key relationships between them. Path modelling was used for the most promising vignettes to identify preliminary models that could shed light on the pathways between those critical variables. Lastly, thematic analysis of participants' answers to the open-ended questions was carried out to obtain information used to improve content and wording of the items and the questionnaire. Overall, the analysis yielded a detailed understanding of strengths and weaknesses, as well as commonalities and differences, across the twelve vignettes, and provisional measurement and path models reflective of the conceptual model.

During the analysis, the conceptual model was implicitly referenced as the theoretical backdrop, but no systematic hypothesis testing was carried out between

²³ Specific R packages used for analysis and visualisation are referenced in text/notes in the next sections.

vignettes. Methods originally developed in a confirmatory framework, such as path modelling, were used here in a model *generation* rather than validation approach. The use of path modelling (and SEM in general) as a preliminary exploratory procedure has recently become more common as a result of the issues related to the use of these techniques in the absence of strong hypotheses for the specification of a theoretical model (Marsh et al., 2014). In the past, tentative models would often yield poor fit, leading to extended sequences of model modification using the same datasets (Kaplan, 1995, 2009). Nowadays, to avoid such practice, researchers often conduct preliminary path modelling (or SEM) to generate initial models; these provisional models are then fully tested and validated in subsequent research, often among larger samples, through the conventional confirmatory procedures typical of these methods. By clearly separating the exploratory and the confirmatory stages, the number of iterations of the model modification process in the validation stage is considerably reduced, thanks to the stronger bases on which the models subject to validation are specified.

The critical evaluation of the combined findings from all the analyses of Study 1 contributed to making informed decisions for the following stages of the research, such as choices about the key vignettes, items, and hypotheses.

Results and Preliminary Reflections

The final dataset from Study 1 was virtually complete, as only one participant did not answer one socio-demographic question (level of religiosity/spirituality). Otherwise, all the other variables had no missing data.

Multivariate Outliers

Following the preregistered analysis plan, multivariate outlier analysis was carried out. The analysis involved the calculation of centred leverage values, Mahalanobis distance, and Cook's distance. Centred leverage values and Mahalanobis distance are informative about the distance of the data points from the centroid of the predictors' space. Following Belsley et al. (1980), the cutoff points for centred leverage values were identified through the formula:

$$2(1+k)/n$$

where n represents the sample size and k the number of predictors.

The cutoff points for the Mahalanobis distance were identified through the critical chi-square at the chosen probability level (here $\alpha = .001$) with degrees of freedom equal to the number of predictors (Stevens, 1984).

Cook's distance provided complementary information about the overall influence of the data points on the regression equation. The cutoff points were identified through the formula (Fox, 2015):

$$4/(n-k-1)$$

where n represents the sample size and k the number of predictors.

As a conservative strategy, participants were considered extreme multivariate outliers if they lay beyond at least two out of three of the above-mentioned cutoff points. The analysis was conducted for the two outcome variables in the conceptual

model: positive and negative moral affect. A total of 18 participants were found to be multivariate outliers; the complete list by vignette is shown in Table A6.

Table A6:
Number of outliers and participant ID for each vignette

Vignette	Number of outliers	Participant ID
Cory	2	59, 146
Wesley	3	11, 37, 70
Arnaud	2	67, 70
Francia	3	16, 77, 85
Matthew	1	70
Sarah	2	19, 122
Joey	2	134, 146
Markus	2	46, 54
Ruxandra	1	70
Alvaro	1	70
Sunita	2	65, 143
Nicholas	3	70, 131, 137

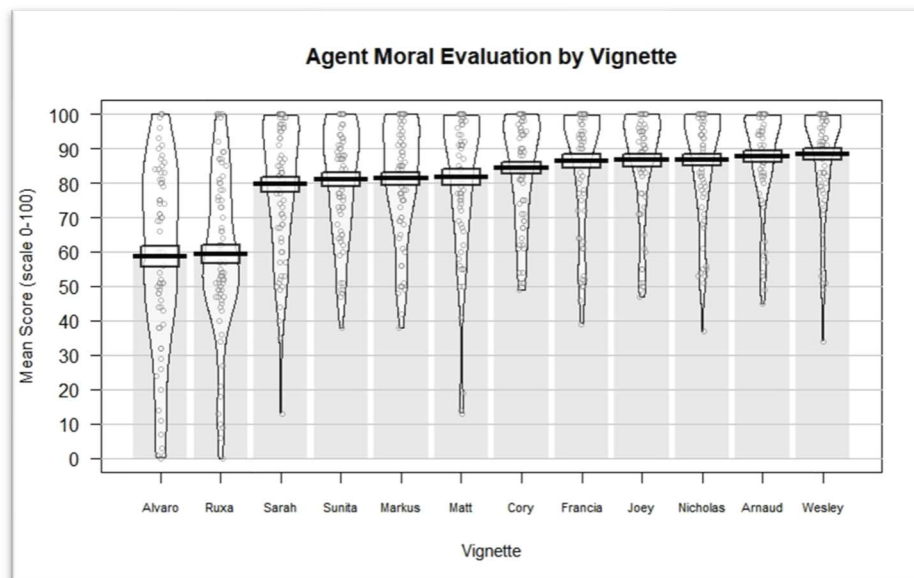
It can be noticed that outliers recurring across multiple vignettes were rare. Participant 70 was an outlier in six vignettes, participant 146 in two vignettes, and each of the other outliers was detected in only one vignette. Crucially, even when the distance from the predictors' centroid was large, the influence on the regression equation for both positive and negative affect was noticeable only in two vignettes for participant 70 (Ruxandra and Matthew) and no influence was noticeable for participant 146. Because the impact caused by the outliers was deemed trivial, the full sample of 146 participants was retained.

Moral Appraisals

The details of the descriptive statistics for the moral appraisal variables can be found in SMA4. Regarding the moral self-evaluation, the mean score was very high ($M = 82.86$, $SD = 12.85$) and the distribution was asymmetric. The mean scores of the moral agent evaluation for ten of the twelve vignettes were quite close to each other and

varied between 79.53 and 88.32, as can be seen in Figure A1²⁴. Wesley received the highest evaluation of moral character ($M = 88.32$, $SD = 15.22$). The other top ratings were for Arnaud ($M = 87.86$, $SD = 14.87$), Nicholas ($M = 86.75$, $SD = 15.75$), Joey ($M = 86.57$, $SD = 15.41$), and Francia ($M = 86.40$, $SD = 17.15$). The two vignettes with the lowest scores received considerably lower ratings, with greater variability around the means: Ruxandra ($M = 59.28$, $SD = 23.53$) and Alvaro ($M = 58.70$, $SD = 27.22$).

Figure A1:
Moral agent evaluation (moral character): bean plots with mean scores for the 12 vignettes

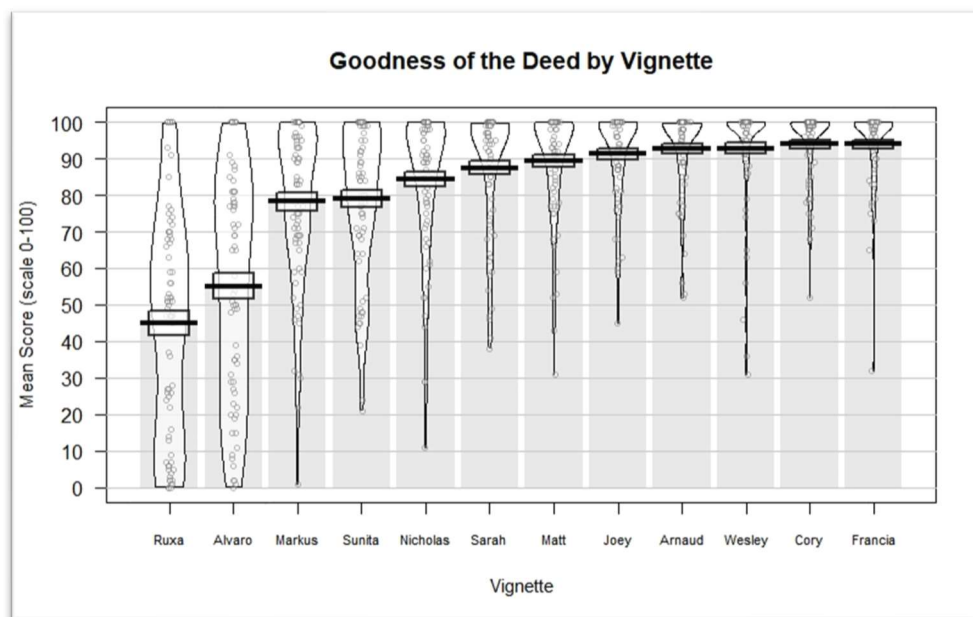


²⁴ This and the following bean plots were created with the R package yarr (Phillips, 2017).

The mean scores of the level of goodness of the deeds were on average even higher and varied between 78.23 and 93.90 for the same top ten vignettes (Fig. A2). Francia's deed ($M = 93.90$, $SD = 10.90$) and Cory's deed ($M = 93.90$, $SD = 10.21$) received the highest evaluation, followed closely by the deeds performed by Wesley ($M = 92.87$, $SD = 14.58$) and Arnaud ($M = 92.64$, $SD = 11.87$). Again, the two moral actions with the lowest scores obtained substantially lower ratings, with greater variability around the means: Alvaro ($M = 55.07$, $SD = 30.97$) and Ruxandra ($M = 45.01$, $SD = 31.10$).

Figure A2:

Goodness of the deed: bean plots with mean scores for the 12 vignettes

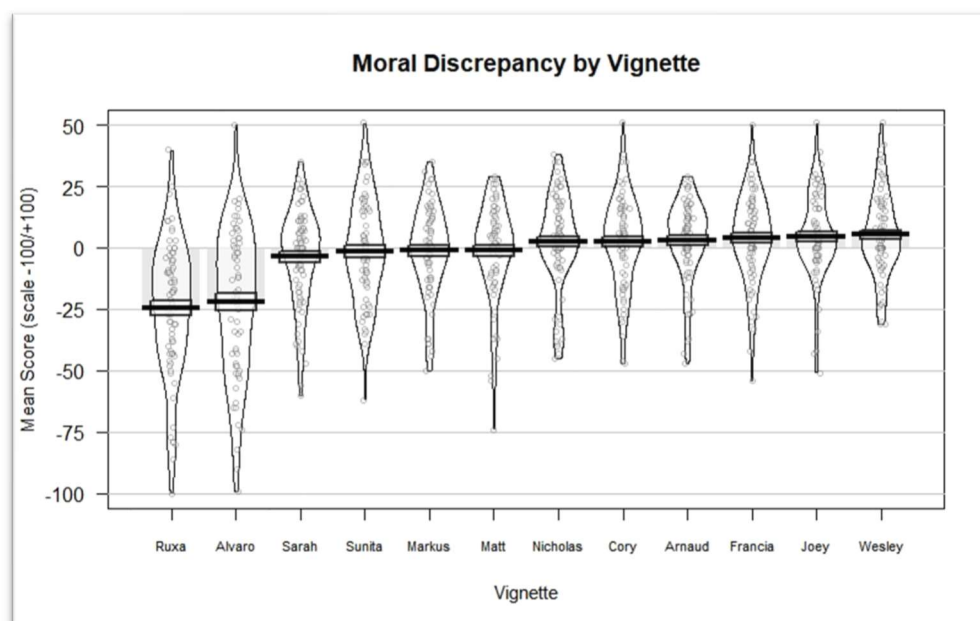


With regard to the measure of *indirect* moral comparison (moral discrepancy, i.e. the difference between agent evaluation and self-evaluation), when aggregating all the evaluations across the twelve vignettes, the mean score was negative ($M = -2.63$, $SD = 22.79$), indicating a slight tendency toward downward comparisons. Figure A3 displays the mean scores for each of the twelve vignettes. Half of the vignettes exhibited an

overall upward moral comparison, although the values were not far from zero; this phenomenon is not surprising, considering the uncommon goodness of many of the moral acts performed in the stories. Among the remaining six vignettes, which were on average in downward comparison, two of them (Alvaro and Ruxandra) showed moral discrepancy values that were much lower than the others; this was mostly a reflection of the lower rating of the moral character of the protagonists of those vignettes and explains why the overall moral discrepancy across the twelve vignettes was negative.

Figure A3:

Moral discrepancy (indirect moral character comparison): bean plots with mean scores for the 12 vignettes

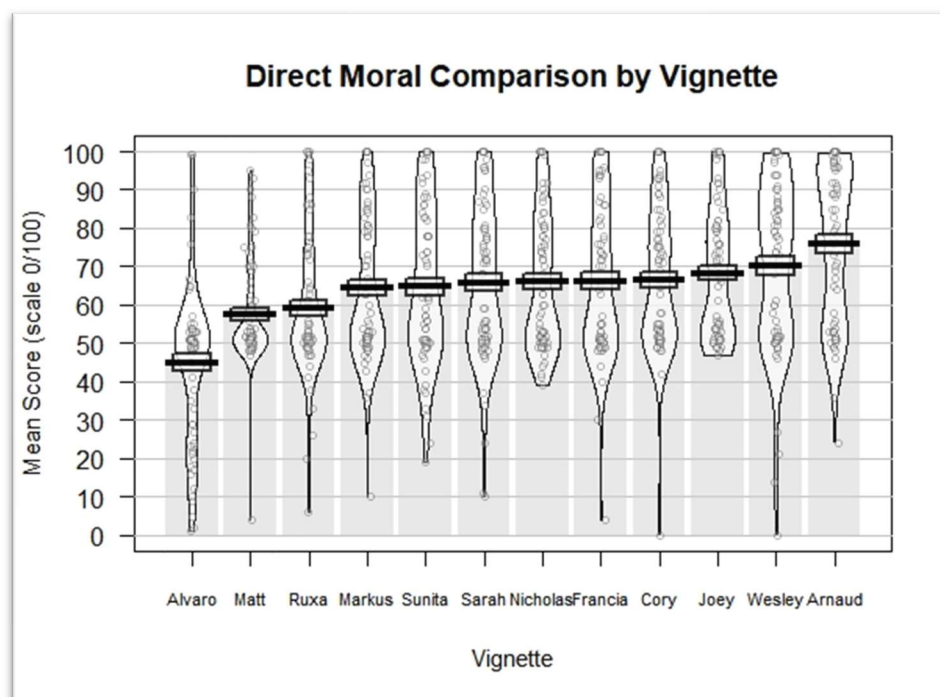


The *direct* moral comparison measure depicted a slightly different scenario (Fig. A4). In aggregate, the mean was well above the scale mid-point (i.e., 50) indicating an overall upward comparison: $M = 64.22$, $SD = 20.54$. Thus, when participants were forced to compare themselves directly with the protagonists of the vignettes, they offered a

less flattering image of their own relative moral standing. Eleven vignettes out of twelve showed an average upward comparison and only Alvaro's vignette displayed an average downward comparison. Interestingly, with this direct measure, Ruxandra showed an average upward comparison, while the previous indirect measure revealed an average downward comparison. Because the self-evaluation question was asked at the beginning of the questionnaire, well before showing the vignettes and rating the moral agents, its score reflected a chronic assessment of the moral self, independent of the protagonists of the vignettes; therefore, it could be argued that the indirect measure of moral comparison was indicative of a trait-like comparison of the self with the moral agent. By contrast, the direct measure – obtained immediately after viewing the vignettes – could be considered more indicative of a state-like comparison, the self-evaluation having been made just as salient as the agent evaluation at the time of the measurement.

Figure A4:

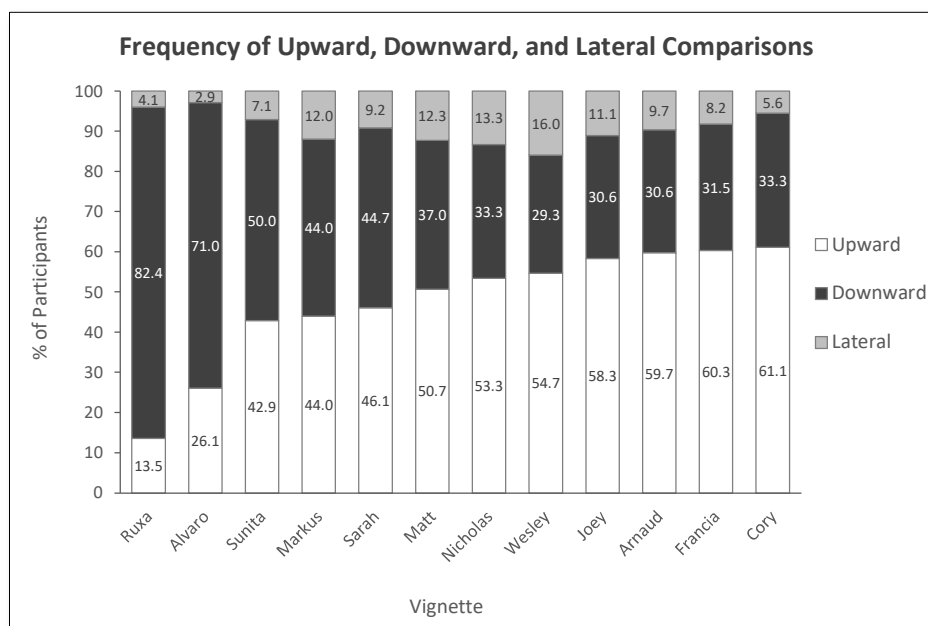
Direct moral character comparison: bean plots with mean scores for the 12 vignettes



Another way to look at the indirect moral comparison measure is to examine the proportion of upward, downward, and lateral comparisons for each vignette; this means comparing the percentage of participants who showed a negative, positive, or null value of moral discrepancy for each vignette (see Fig. A5).

Figure A5:

Types of moral comparisons for the 12 vignettes: clustered bar chart of frequency distributions (%)



The moral actions in the vignettes were meant to showcase different moral domains (Fig. A6): care, courage, and justice (sometimes partly overlapping). In terms of care, Francia's deed was perceived as the most caring ($M = 95.60$, $SD = 7.20$), but many others received high ratings, too (Joey, Wesley, Matt, Cory, and Arnaud). Once again, the deeds performed by Ruxandra ($M = 59.73$, $SD = 32.11$) and Alvaro ($M = 58.32$, $SD = 35.78$) appeared at the bottom of the rank by a considerable margin.

In terms of courage, Arnaud's ($M = 97.46$, $SD = 7.04$), Cory's ($M = 96.76$, $SD = 6.83$), and Wesley's deeds ($M = 96.41$, $SD = 8.80$) were perceived to be the bravest, followed

by Francia's ($M = 93.89$, $SD = 9.74$) and Nicholas's ($M = 89.35$, $SD = 16.47$). Below the 50 mark (mid-point of the scale) were the acts performed by Alvaro, Matthew, Sarah, and Ruxandra (the last with a very low mean of 28.62 and a large standard deviation of 30.09).

In terms of justice, Nicholas's deed ($M = 92.55$, $SD = 12.30$) was perceived to be by far the most just, followed at a distance by Sunita's deed ($M = 73.63$, $SD = 26.62$). For once, the repairing nature of Alvaro's deed endowed it a relatively high score ($M = 66.36$), although with noticeable variability ($SD = 33.52$), while Francia's deed, the most caring of them all, received the lowest rating ($M = 17.82$, $SD = 23.48$).

A comparison of the perceived nature of the deeds across all the twelve vignettes showed that the moral domain of justice is the most discriminating: it was perceived to be highly pertinent to a more limited set of vignettes, particularly Nicholas (social justice for minority groups) and not at all relevant to other vignettes, for instance Francia and Cory. On the other hand, a substantial overlap was observed between courage and care, especially in Francia, but also in Arnaud, Cory, and Wesley (all brave actions aimed at taking care of someone in need).

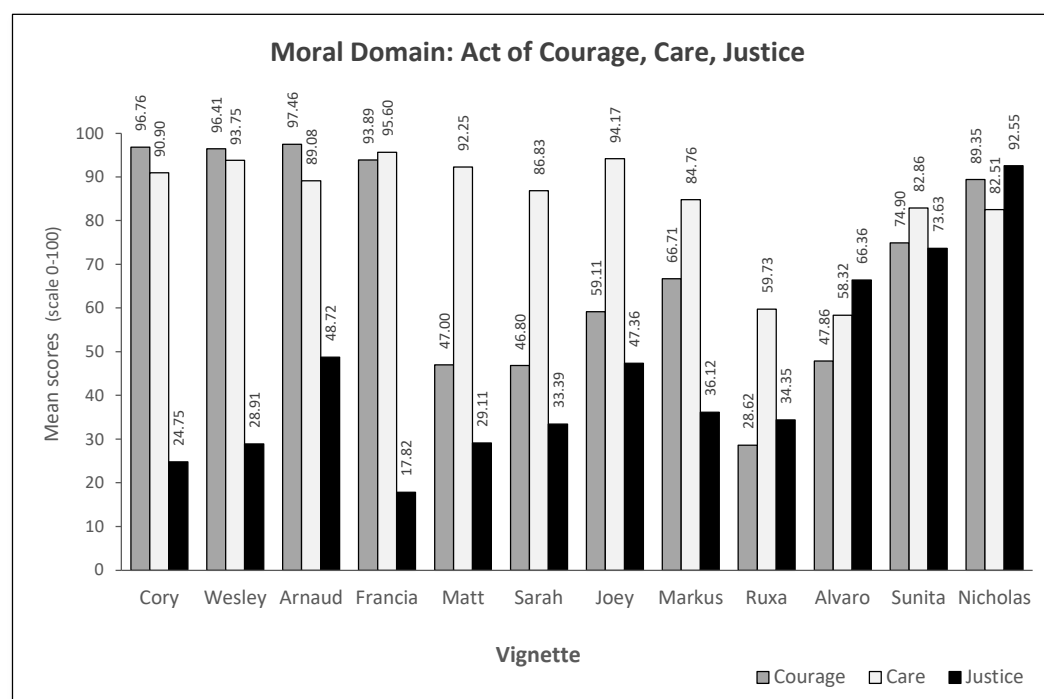
These findings seemed to suggest that Nicholas and Francia could be the two most interesting vignettes to further explore in the next stages of the research: the former focussing on justice, the latter other on care (and courage); the former with more mixed judgments, the latter positively received by almost all participants. In the rest of the analysis of Study 1, these two vignettes were scrutinised with particular attention to confirm the initial impression that they were good candidates for selection.

A third vignette stood out from the rest for being very different: Ruxandra (the vegetarian). It emerged as arguably the least positive scenario, with the lowest means for goodness and propriety of the deed, the second lowest mean for the moral character

of the protagonist, the lowest moral discrepancy, and the highest proportion of downward comparisons. This could be a third promising vignette for further investigation, particularly to explore how less positive judgments affect moral self-regulation and prosocial behaviour.

Figure A6:

Moral domain: perception about the nature of the deed (courage, care, justice) for the 12 vignettes (mean scores on the 0-100 scale)



Moral Self-Regulation

Whereas the detail of the descriptive statistics for the items of the moral self-regulation inventory can be found in SMA5, a summary of the reliability analysis for the self-improvement and self-defence scales (six items each) is illustrated in Table A7 and discussed below. At this stage, the analysis aimed at determining if any items were decidedly inadequate, with a view to retaining the same item list for all the vignettes, rather than optimising the scale for each vignette and ending up with different scales

across vignettes. In addition, consideration was given to rewording the items if clear directions for improvement were identified.

Overall, the self-improvement scale showed satisfactory internal consistency, with Cronbach's alpha always above .80. Across all the vignettes, Cronbach's alpha ranged between .813 (Arnaud) and .875 (Matthew). Sometimes the original 6-item scale could be marginally improved by deleting one item in certain vignettes, but that small improvement would cause a drop in Cronbach's alpha in the other vignettes.

The self-defence scale overall showed greater variability and lower internal consistency than the self-improvement scale. In three vignettes Cronbach's alpha was equal or higher than .700 (Alvaro, Markus and Ruxandra), in two vignettes it was comprised between .642 and .660 (Matthew, Sarah and Sunita), and in all the remaining vignettes it fell in the region of .5 or even .4. This lower level of internal consistency was due to the fact that some of the six self-defence items did not correlate strongly with the scale. Reliability of the scale would improve with the removal, for example, of item SD5/Extreme (which recurred four times as a candidate for deletion or rewording), and item SD4/Untrue (twice).

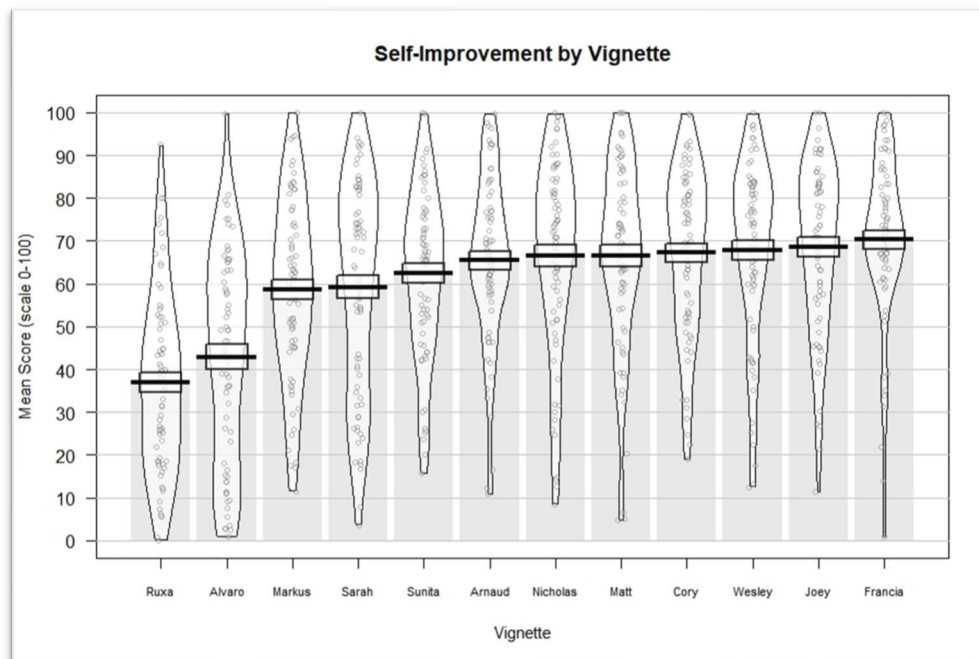
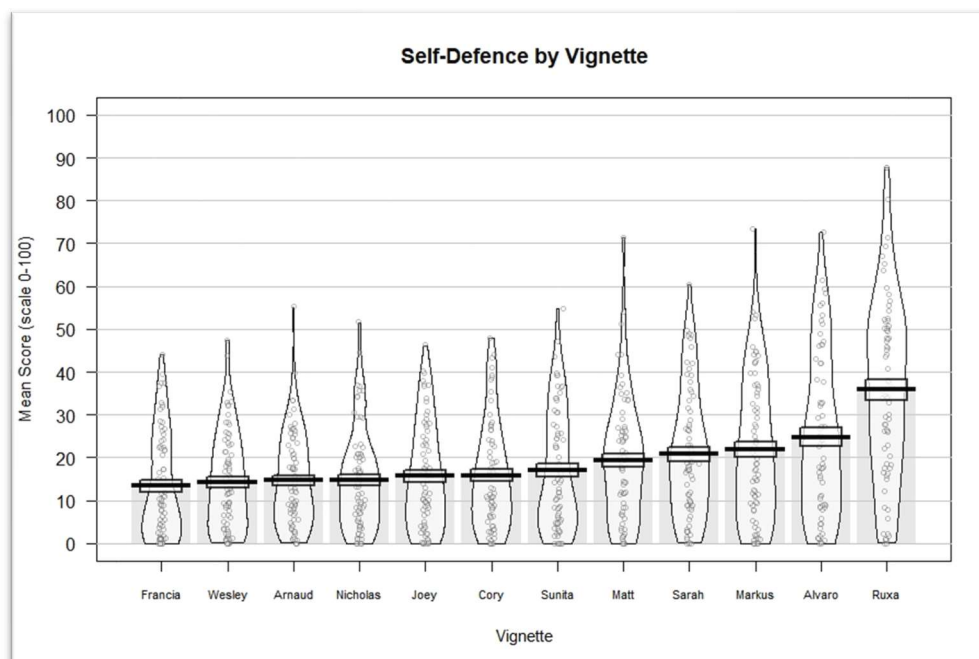
Table A7 shows Cronbach's alpha for both the self-improvement and the self-defence scales for all the twelve vignettes, highlighting the items that could be deleted to improve internal consistency. More detailed analyses of reliability for the moral self-regulation scales are presented in SMA7.

Table A7:
Cronbach's alpha for the self-improvement and self-defence scales for the 12 vignettes

Vignette	Moral self-regulation scales			
	Self-improvement		Self-defence	
	original scale with 6 items	reduced scale if any item deleted	original scale with 6 items	reduced scale if any item deleted
Cory	.814	.820 (SI5/BeBetter)	.536	.537 (SD5/Extreme)
Wesley	.826	.831 (SI1/Admirable)	.507	.519 (SD5/Extreme, SD1/Ordinary)
Arnaud	.813	-	.468	.577 (SD5/Extreme)
Francia	.860	.864 (SI5/BeBetter)	.596	.599 (SD4/Untrue)
Matthew	.875	-	.642	.668 (SD5/Extreme)
Sarah	.871	.877 (SI4/Values)	.656	.700 (SD5/Extreme, SD6/Superior)
Joey	.821	-	.539	.564 (SD2/MeGreater)
Markus	.853	-	.749	-
Ruxandra	.836	.853 (SI6/ForOthers)	.790	-
Alvaro	.857	-	.710	.713 (SD4/Untrue)
Sunita	.815	.823 (SI4/Values)	.660	.676 (SD1Ordinary)
Nicholas	.862	-	.524	.604 (SD5/Extreme)

Note: in parenthesis the items that, if removed, would improve Cronbach's alpha.

The composite means for the self-improvement and self-defence scales are shown in Figures A7-A8.

Figure A7:*Moral self-improvement scale: bean plots with mean scores for the 12 vignettes***Figure A8:***Moral self-defence scale: bean plots with mean scores for the 12 vignettes*

Overall, participants showed higher levels of self-improvement ($M = 61.12$, $SD = 23.65$) than self-defence ($M = 19.05$, $SD = 17.74$). Francia's deed generated the highest self-improvement mean ($M = 70.34$, $SD = 19.79$), followed closely by Joey's ($M = 68.67$, $SD = 20.08$), Wesley's ($M = 67.88$, $SD = 21.32$), Cory's ($M = 67.22$, $SD = 19.72$), Matthew's ($M = 66.53$, $SD = 22.23$), and Nicholas's ($M = 66.51$, $SD = 22.80$); relative to these, Alvaro's and Ruxandra's deeds induced less improvement regulation (respectively, $M = 42.95$, $SD = 24.75$, and $M = 36.91$, $SD = 21.02$).

In terms of self-defence, Ruxandra's vignette produced the highest mean score ($M = 35.86$, $SD = 21.77$), followed by Alvaro's ($M = 24.83$, $SD = 18.78$). Francia's vignette generated the lowest score ($M = 13.42$, $SD = 12.20$).

Moral Affect

Whereas the detail of the descriptive statistics for the items of the moral affect scales can be found in SMA6, a summary of the reliability analysis for the positive and negative affect scales (three and five items respectively) is illustrated in Table A8 and discussed below. As for moral self-regulation, at this stage the analysis aimed at determining if any items were clearly inadequate across all the vignettes, with the aim of having a set of items with balanced performance across all of them, rather than optimal for any single vignette. In addition, if necessary, consideration was given to rewording poorer items.

Overall, the original three-item positive affect scale showed an acceptable degree of internal consistency, with Cronbach's alpha varying from .690 to .875; however, for most vignettes, Cronbach's alpha could be improved by removing item PA3/Challenged, which often inadequately correlated with the scale. The modified two-item positive affect scale showed an improved level of reliability, ranging between .778 (Francia) and .911 (Alvaro), which could be considered respectively acceptable and fully satisfactory.

The exception was Joey's vignette, for which overall the positive affect scale did not seem to be particularly efficient: in this case, the best candidate for deletion was item PA2/Inspired (Cronbach's alpha would improve from .690 to .699), while removing item PA3/Challenged would cause Cronbach's alpha to drop from .690 to .625, which is much lower than all other vignettes.

The original five-item negative affect scale showed greater variability across vignettes, and overall exhibited lower internal consistency than the positive affect scale. This pattern mirrors the pattern observed for defensive self-regulatory processes. In six vignettes, the original five-item scale was the optimal one, with Cronbach's Alpha varying from .764 (Ruxandra) to .514 (Alvaro). In the remaining six vignettes at least one item could be removed with an improvement of Cronbach's alpha, particularly item NA1/Conflicted, which correlated less markedly with the scale, especially in the vignettes where the moral action was more highly regarded. However, the elimination of that item would cause a drop in the internal consistency of the negative affect scale of the other vignettes. Joey's vignette appeared to have the lowest internal consistency also in the negative affect scale: with the original five items, Cronbach's alpha was only .430, and improved to .574 (which is still low) if three items were deleted.

Table A8 displays Cronbach's alpha for both the positive and negative affect scales for all the twelve vignettes, showing the items that could be deleted to improve reliability. More detailed reliability analyses for the moral affect scales can be found in SMA8.

Table A8:
Cronbach's alpha for the positive and negative affect scales across the 12 vignettes

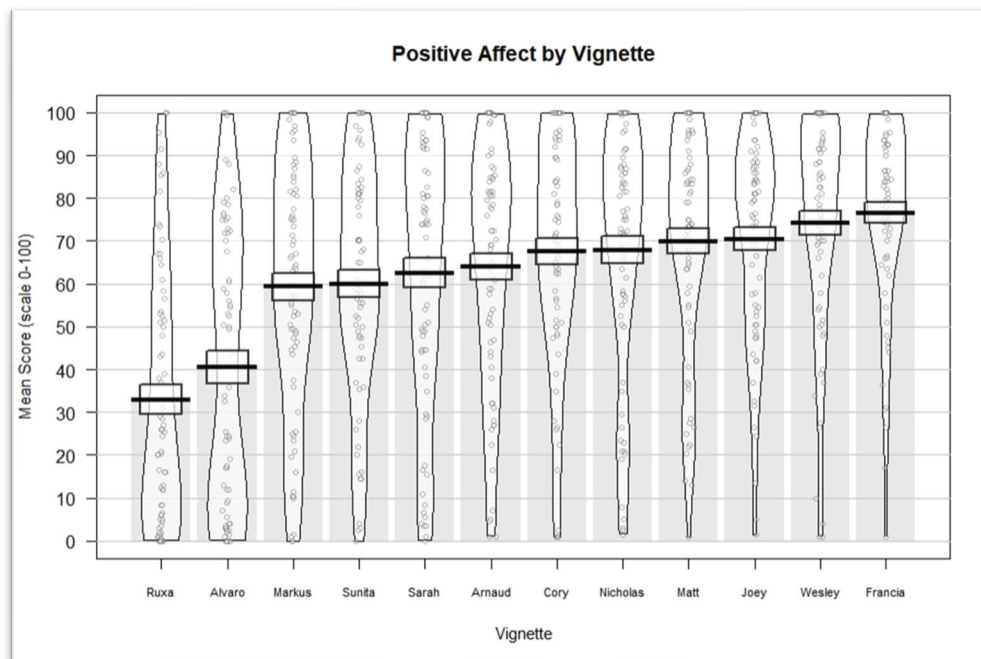
Vignette	Moral affect scales			
	Positive affect		Negative affect	
	original scale with 3 items	reduced scale if any item deleted	original scale with 5 items	reduced scale if any item deleted
Cory	.751	.782 (PA3/Challenged)	.580	.736 (NA1/Conflicted)
Wesley	.783	.809 (PA3/Challenged)	.551	.665 (NA1/Conflicted)
Arnaud	.816	.857 (PA3/Challenged)	.640	-
Francia	.695	.778 (PA3/Challenged)	.746	-
Matthew	.819	.861 (PA3/Challenged)	.613	-
Sarah	.875	-	.664	.704 (NA1/Conflicted, NA5/Resentful, NA2/Threatened)
Joey	.690	.699 (PA2/Inspired) [.625 if PA3/Challenged removed]	.430	.574 (NA1/Conflicted, NA5/Resentful, NA2/Threatened)
Markus	.834	-	.642	.700 (NA4/Envious, NA3/Guilty, NA1/Conflicted)
Ruxandra	.869	.887 (PA3/Challenged)	.764	-
Alvaro	.845	.911 (PA3/Challenged)	.514	-
Sunita	.808	-	.756	-
Nicholas	.733	.815 (PA3/Challenged)	.651	.763 (NA1/Conflicted)

Note: in parenthesis the items that, if removed, would improve Cronbach's alpha.

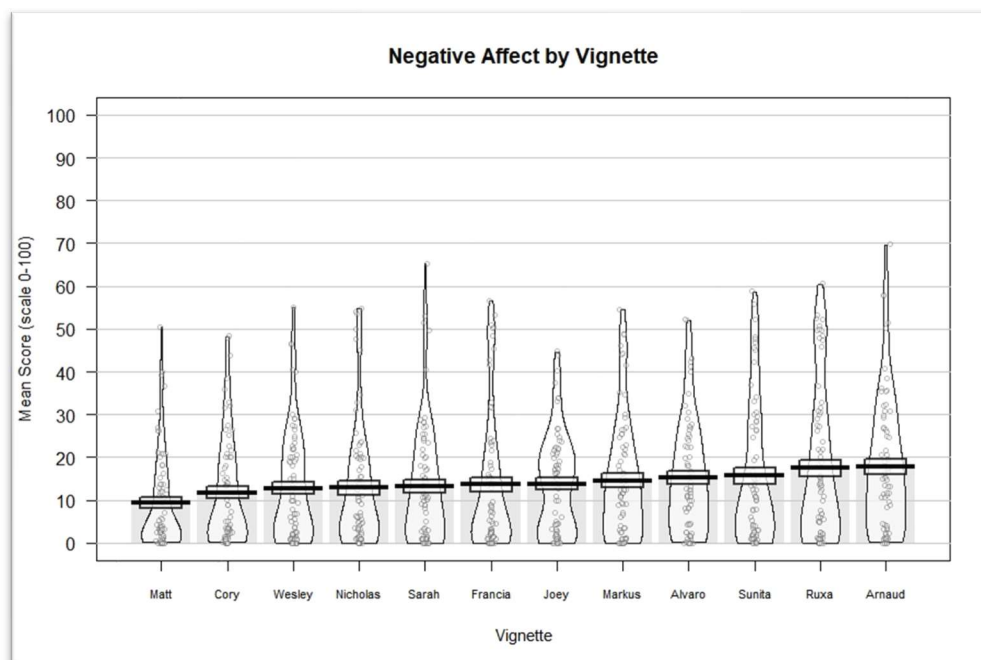
The composite means for the positive affect and negative affect scales are shown respectively in Figures A9-A10.

Figure A9:

Positive moral affect scale: bean plots with mean scores for the 12 vignettes

**Figure A10:**

Negative moral affect scale: bean plots with mean scores for the 12 vignettes



Overall, participants reported experiencing more positive than negative affect (respectively, $M = 62.26$, $SD = 30.52$, and $M = 14.03$, $SD = 14.47$). The results for positive affect by vignette mirrored the self-improvement scores: Francia's action elicited the highest levels of positive affect ($M = 76.62$, $SD = 22.11$), followed closely by Wesley's ($M = 74.23$, $SD = 26.11$). High levels of positive affect were reported also for the deeds performed by Joey ($M = 70.44$, $SD = 24.40$), Matthew ($M = 69.95$, $SD = 26.05$), Nicholas ($M = 67.85$, $SD = 28.88$), and Cory ($M = 67.56$, $SD = 27.16$). Alvaro's and Ruxandra's deeds elicited considerably lower levels of positive affect (respectively, $M = 40.54$, $SD = 32.89$, and $M = 32.99$, $SD = 30.23$).

The highest negative affect value was observed in Arnaud's vignette ($M = 17.91$, $SD = 15.98$); strongly negative feelings probably arose because of the tragic ending of the story (i.e., the death of the protagonist). Ruxandra's deed followed closely behind ($M = 17.47$, $SD = 17.86$); in this case, some degree of negative affect arose most likely due to the polarising or controversial nature of her moral action (being vegetarian and campaigning for it). Matthew's deed elicited the lowest levels of negative affect ($M = 9.31$, $SD = 11.30$).

Overall, analysis of the moral self-regulation and affect scales confirmed that Nicholas and Francia could be considered good candidates to be progressed to the next phases of the research, with Ruxandra as a strong runner-up.

Analysis of Distribution Bias and Assumptions

All the main moral appraisal, self-regulation, and affect variables in the dataset were measured through interval²⁵ scales varying from 0 to 100, making them 101-point

²⁵ It could be argued that the scales are actually *ratio*, and not just interval, depending on whether or not the zero score on the 0-100 scale is regarded as a *true* zero point. However, Stevens's (1946) traditional theory of measurement, which distinguishes between nominal, ordinal, interval, and ratio levels of measurement, has been repeatedly challenged (see e.g., Velleman & Wilkinson, 1993; Williams, 2020), in that the distinction between these categories is often arbitrary.

scales. Following Jöreskog and Sörbom's (1996) criterion reported by Schumacker and Lomax (2004), quantitative variables with 15 or more points in the scale (like those used here) can be referenced as *continuous*; continuous variables show better psychometric properties than traditional *ordinal* Likert-type scales with three, four, or five levels.

However, because the vignettes depict stories of moral goodness (some truly *uncommon* levels of goodness), violations of the assumption of univariate normality were expected in the distribution of several variables. Effectively, the moral appraisal variables often exhibited negative skew; in some cases, the mode was the highest point in the scale (100). High levels of kurtosis could also be observed in the distributions of several moral appraisal variables. Regarding moral self-regulation, several self-improvement items and the composite mean showed long lower tails, and several self-defence items and the composite mean exhibited long upper tails. Similarly, several positive affect items and the composite mean showed long lower tails, and several negative affect items and the composite mean long upper tails. The Kolmogorov-Smirnov and the Shapiro-Wilk normality tests (Tab. A9a) highlighted violation of normality for all these variables. However, even more important than violations of univariate normality are violations of *multivariate* normality. Mardia's (1970) tests indicated violation of multivariate normality for kurtosis, but not for skewness (Tab. A9b). As noted by Byrne (2016), violations caused by skewness represent an issue for the application of parametric testing (based on means), while violations caused by kurtosis are particularly critical for the use of asymptotic estimation methods in factor analysis and structural equation modelling (based on variance/covariance).

Attempts to eliminate the distribution bias through transformations could not be pursued because of the opposite direction of the skew for self-improvement/positive affect and self-defence/negative affect. Since this kind of bias was likely to persist in the

ensuing studies, to reduce its impact a variety of analytic steps were undertaken, for example bootstrapping, non-parametric testing, robust and non-asymptotic estimation techniques (details to follow where appropriate).

Tables A9a-b:

Tests of univariate and multivariate normality for key moral appraisal, self-regulation, and affect variables (all vignettes aggregated: N=876 observations)

Tab. A9a	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Moral discrepancy	.116	876	<.001	.950	876	<.001
Goodness of the deed	.232	876	<.001	.755	876	<.001
Propriety of the deed	.189	876	<.001	.835	876	<.001
SI1/Admirable	.181	876	<.001	.837	876	<.001
SI2/Awakened	.116	876	<.001	.899	876	<.001
SI3/Humanity	.139	876	<.001	.881	876	<.001
SI4/Values	.094	876	<.001	.944	876	<.001
SI5/BeBetter	.129	876	<.001	.912	876	<.001
SI6/ForOthers	.106	876	<.001	.917	876	<.001
SD1/Ordinary	.247	876	<.001	.764	876	<.001
SD2/MeGreater	.230	876	<.001	.800	876	<.001
SD3/Praise	.269	876	<.001	.687	876	<.001
SD4/Untrue	.285	876	<.001	.640	876	<.001
SD5/Extreme	.205	876	<.001	.842	876	<.001
SD6/Superior	.274	876	<.001	.660	876	<.001
Self-improvement scale	.081	876	<.001	.961	876	<.001
Self-defence scale	.118	876	<.001	.916	876	<.001
PA1/Uplifted	.109	876	<.001	.893	876	<.001
PA2/Inspired	.142	876	<.001	.876	876	<.001
PA3/Challenged	.147	876	<.001	.893	876	<.001
NA1/Conflicted	.246	876	<.001	.786	876	<.001
NA2/Threatened	.305	876	<.001	.542	876	<.001
NA3/Guilty	.288	876	<.001	.677	876	<.001
NA4/Envious	.298	876	<.001	.637	876	<.001
NA5/Resentful	.313	876	<.001	.515	876	<.001
Positive affect scale	.109	876	<.001	.916	876	<.001
Negative affect scale	.166	876	<.001	.860	876	<.001

Note: a. Lilliefors significance correction

Tab. A9b	Statistic	Sig.	Result (multivariate normality)
Mardia's skewness	-41594.404	.999	YES
Mardia's kurtosis	83.394	<.001	NO
Overall multivariate normality			NO

Relationships Between Key Variables

Moral Comparison and Self-Regulation

Correlation analysis was conducted to investigate the zero-order associations between moral comparison and moral self-regulation, which unveiled asymmetric probabilities of experiencing certain modes of self-regulation in association with different moral comparison types. Although these results were preliminary and the analysis was purely correlational, a clear pattern seemed to emerge: when aggregating all the evaluations made by participants across all the vignettes that were presented to them ($N = 876^{26}$), moral discrepancy correlated positively with self-improvement ($r = .601, p < .001$) and negatively with self-defence ($r = -.395, p < .001$), as documented in Table A10 (with bootstrap confidence intervals).

Here (and in the ensuing studies) the Pearson's r correlation coefficient was used (with the related significance test) even if violations of normality in the dataset would, in theory, suggest the use of non-parametric coefficients and tests. In such instances, traditional statistics textbooks would recommend the use of Spearman's rho, which measures the relationship between two variables based on rank-order. However, empirical research based on Monte Carlo simulations (see e.g., Havlicek & Peterson, 1976) has provided evidence of the robustness of Pearson's correlation coefficient (and the relative significance test) to violations of various parameter assumptions, such as skewness and kurtosis²⁷.

²⁶ It must be noted that the aggregation of all evaluations across vignettes violates the assumption of independence of the observations, as participants viewed and rated six vignettes randomly selected from a pool of twelve. However, the following analysis by vignette showed in most cases the same pattern.

²⁷ To ensure robustness of Pearson's r coefficient vs Spearman's rho coefficient (and the associated significance tests), a comparison between the results from the use of the two measures was carried out (for further details, see SMA9). As the results corroborated Havlicek and Peterson's claim reported above, in the present study Pearson's r correlation coefficient was used. Because the ensuing studies were conducted among larger samples, Pearson's r remained the elective measure of bivariate correlation. In addition, bootstrapping was used wherever possible to provide robust confidence intervals.

Table A10:

Zero-order correlation between moral discrepancy and moral self-regulation measures with bootstrap confidence intervals (all vignettes aggregated: N=876 observations)

		Self-improvement	Self-defence
Moral discrepancy	Pearson Correlation	.601***	-.395***
	Sig. (2-tailed)	<.001	<.001
	N	876	876
	Bias	<.001	.001
	Std. Error	.023	.034
	Bootstrap ^b		
	BCa 95% C.I. Lower	.552	-.464
	Upper	.646	-.327

Note: *** $p < .001$ (2-tailed)

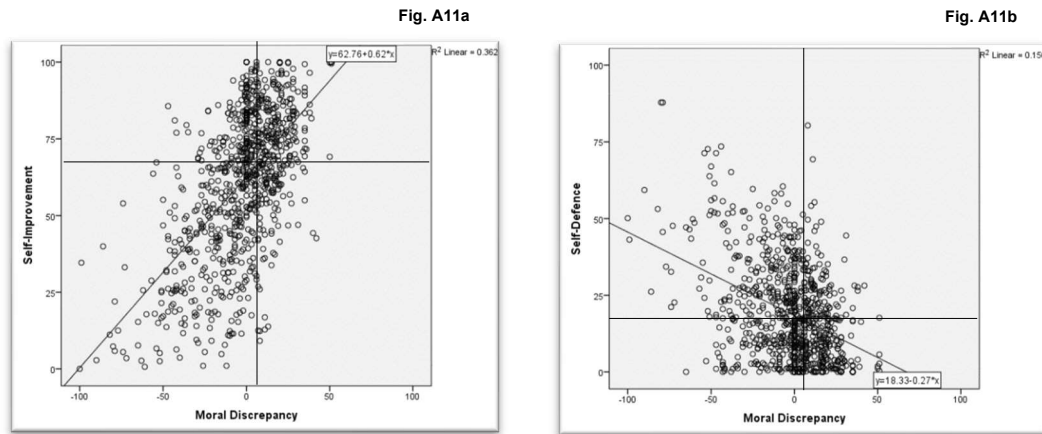
b. Bootstrap results are based on 1000 bootstrap samples.

The scatter plots in Figures A11a-b illustrate graphically the distributions of these moral evaluations and the corresponding self-regulatory modes (all vignettes aggregated). The scatter plot in panel “a” clearly shows a strong prevalence of data points in the upper/right-hand quadrant (upward comparison and higher levels of self-improvement) and the lower/left-hand quadrant (downward comparison and lower levels of self-improvement), indicating a strong positive correlation between moral discrepancy and self-improvement. Similarly, panel “b” clearly shows a prevalence of data points in the lower/right-hand quadrant (upward comparison and lower levels of self-defence) and the upper/left-hand quadrant (downward comparison and higher levels of self-defence), indicating a lower, but still large²⁸ negative correlation between moral discrepancy and self-defence.

²⁸ Following Hemphill’s (2003) comments on how to interpret correlation coefficients in social psychological studies, Gignac & Szodorai (2016) carried out meta-analyses to provide improved guidelines to discern the magnitude of correlations, particularly for research on individual differences. They concluded that the typical correlation (in absolute value) should be expected to be around .20, and therefore correlations above .30 should be considered large, and below .10 small. These guidelines for large correlations considerably depart from Cohen’s (1988, 1992) initial recommendations of Pearson’s r values of .50 to demarcate large effects.

Figures A11a-b:

Relations between moral discrepancy and the two modes of moral self-regulation: scatter plots (with medians and regression lines) aggregating mean scores across the 12 vignettes (N=876)



For the most part, the results of correlation analysis with all vignettes aggregated held up when analysing the data at the level of the individual vignettes (assuming their independence): in seven out of twelve vignettes the correlation pattern remained the same, while in the remaining five vignettes only the positive correlation between moral discrepancy and self-improvement consistently emerged as significant (Tab. A11). The significant correlation between moral discrepancy and self-regulation suggested the possibility to modify the conceptual model (Fig. III), enabling the use of the variable moral discrepancy and the analysis at total sample instead of by subsample of upward and downward “comparers”, as previously hypothesised.

Table A11:

Zero-order correlation between moral discrepancy and moral self-regulation measures by vignette

	Vignette		Self-improvement	Self-defence
Moral discrepancy	Cory	Pearson Correlation	.443***	-.125
	n=72	<i>Sig. (2-tailed)</i>	<.001	.294
	Wesley	Pearson Correlation	.441***	-.079
	n=75	<i>Sig. (2-tailed)</i>	<.001	.501
	Arnaud	Pearson Correlation	.500***	.023
	n=72	<i>Sig. (2-tailed)</i>	<.001	.848
	Francia	Pearson Correlation	.429***	-.304**
	n=73	<i>Sig. (2-tailed)</i>	<.001	.009
	Matthew	Pearson Correlation	.578***	-.340**
	n=73	<i>Sig. (2-tailed)</i>	<.001	.003
	Sarah	Pearson Correlation	.552***	-.361**
	n=76	<i>Sig. (2-tailed)</i>	<.001	.001
	Joey	Pearson Correlation	.500***	-.084
	n=72	<i>Sig. (2-tailed)</i>	<.001	.483
	Markus	Pearson Correlation	.431***	-.455***
	n=75	<i>Sig. (2-tailed)</i>	<.001	<.001
	Ruxandra	Pearson Correlation	.534***	-.428***
	n=74	<i>Sig. (2-tailed)</i>	<.001	<.001
	Alvaro	Pearson Correlation	.610***	-.372**
	n=69	<i>Sig. (2-tailed)</i>	<.001	.002
	Sunita	Pearson Correlation	.562***	-.483***
	n=70	<i>Sig. (2-tailed)</i>	<.001	<.001
	Nicholas	Pearson Correlation	.608***	-.150
	n=75	<i>Sig. (2-tailed)</i>	<.001	.199

Note: ** $p < .01$ (2-tailed)

*** $p < .001$ (2-tailed)

n denotes the sample size by vignette

Moral Self-Regulation and Affect

The other important set of relationships in the conceptual model was the association between moral self-regulation and affect. In line with the hypotheses, the empirical data from Study 1, when aggregating all the vignettes, showed that self-improvement strongly positively correlated with positive affect ($r = .840, p < .001$), and self-defence moderately correlated positively with negative affect ($r = .285, p < .001$). Negative affect also negatively correlated with positive affect ($r = -.355, p < .001$). The results with bootstrap confidence intervals can be seen in Table A12.

Table A12:

Zero-order correlation between moral self-regulation and moral affect measures with bootstrap confidence intervals (all vignettes aggregated: N=876 observations)

		Positive affect	Negative affect
Self-improvement	Pearson Correlation	.840***	.050
	<i>Sig. (2-tailed)</i>	<.001	.135
	N	876	876
	Bias	-.001	-.002
	Std. Error	.012	.034
	Bootstrap ^c	.816	-.014
		.860	.116
Self-defence	Pearson Correlation	-.355***	.285***
	<i>Sig. (2-tailed)</i>	<.001	<.001
	N	876	876
	Bias	<.001	<.001
	Std. Error	.032	.034
	Bootstrap ^c	-.417	.213
		-.291	.350

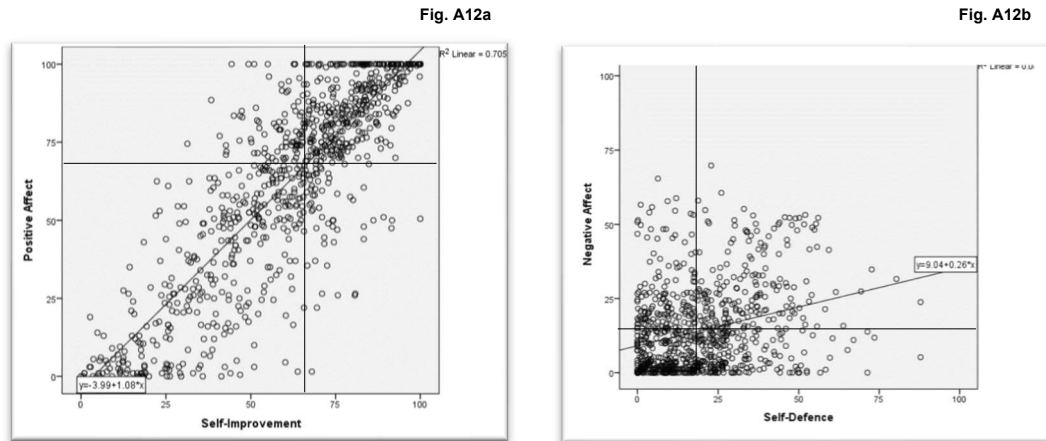
Note: *** $p < .001$ (2-tailed).

c. Bootstrap results are based on 1000 bootstrap samples.

The scatter plots in Figures A15a-b illustrate graphically the distributions of the self-regulatory modes and affective outcomes (all vignettes aggregated). The scatter plot in panel “a” clearly shows a strong prevalence of data points in the upper/right-hand quadrant (high levels of self-improvement and positive affect) and the lower/left-hand quadrant (low levels of self-improvement and positive affect), indicating a strong positive correlation between self-improvement and positive affect. Similarly, panel “b” shows a prevalence of data points in the lower/left-hand quadrant (low levels of self-defence and negative affect) and the upper/right-hand quadrant (high levels of self-defence and negative affect), indicating a positive but less strong correlation between self-defence and negative affect.

Figures A12a-b:

Relations between the two modes of moral self-regulation and moral affect: scatter plots (with medians and regression lines) aggregating mean scores across the 12 vignettes (N=876)



Correlation analysis at the level of the individual vignettes (assuming their independence) showed a replication of the results observed for all vignettes aggregated in eight out of twelve vignettes; in the remaining four vignettes only the positive correlation between self-improvement and positive affect consistently emerged as significant (Tab. A13).

Table A13:*Zero-order correlations between moral self-regulation and affect measures by vignette*

Vignette			Positive affect	Negative affect
Cory n=72	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.728*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.347** .003
Wesley n=75	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.784*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.332** .004
Arnaud n=72	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.765*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.438*** <.001
Francia n=73	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.790*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.365** .002
Matthew n=73	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.837*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.428*** <.001
Sarah n=76	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.883*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.040 .730
Joey n=72	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.761*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.114 .342
Markus n=75	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.834*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.294* .010
Ruxandra n=74	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.812*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.114 .334
Alvaro n=69	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.872*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.396** .001
Sunita n=70	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.806*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.209 .082
Nicholas n=75	Self-improvement	Pearson Correlation <i>Sig. (2-tailed)</i>	.803*** <.001	
	Self-defence	Pearson Correlation <i>Sig. (2-tailed)</i>		.300** .009

Note: * $p < .05$ (2-tailed) ** $p < .01$ (2-tailed) *** $p < .001$ (2-tailed)
n denotes the sample size by vignette.

Other Correlation Patterns

Analysis of zero-order correlations between other critical variables showed that, with all vignettes aggregated, the moral self-evaluation was not significantly correlated with either broadening ($r = .011, p = .756$) or defensive self-regulation ($r = .001, p = .972$), and either positive ($r = .001, p = .980$) or negative affect ($r = -.058, p = .084$); it also did not correlate with the appraisals of the goodness ($r = .062, p = .067$) and the propriety of the deed ($r = .020, p = .562$).

By contrast, the goodness of the deed correlated strongly with self-regulatory and affective states: positively with self-improvement ($r = .647, p < .001$) and positive affect ($r = .607, p < .001$), negatively with self-defence ($r = -.492, p < .001$) and negative affect ($r = -.077, p = .022$). The same pattern was found for the propriety of the deed: positive correlation with self-improvement ($r = .502, p < .001$) and positive affect ($r = .430, p < .001$), negative correlation with self-defence ($r = -.415, p < .001$) and negative affect ($r = -.168, p < .001$). Details of these correlation patterns can be found in SMA9.

Multiple Linear Regression

After obtaining a general overview through the analysis of the moral appraisals, the self-regulatory, and the affective variables, including their zero-order correlations, the next step was to regress separately the two outcome variables (positive and negative affect) on all their main predictors, measuring the variance explained by the models and the strength of these relationships. The predictors entered in the forward selection procedure were the following: age, religiosity, political orientation²⁹, moral self-evaluation, agent evaluation, moral discrepancy, goodness of the deed, propriety of the

²⁹ Given the near perfect correlation between the variables of political orientation on economic matters and political orientation on social matters ($r = .915, p < .001$), the two variables were averaged in a new “political orientation” variable.

deed, relevance, similarity, courage, care, justice, in-shoes, self-improvement, and self-defence. A further set of multiple linear regressions was also conducted regressing separately self-improvement and self-defence on all their main predictors. The forward selection process usually has the advantage to identify fewer significant predictors than other procedures, such as the backward elimination process (Field et al., 2012)³⁰. In addition, it is a simpler stepwise approach that in this exploratory context was preferred to more complex regularised methods.

Multiple linear regression assumes independence of the observations and normal distribution of the residuals of the dependent variable. As previously discussed, these assumptions were not met by the data. However, at this stage these analyses were purely exploratory and were not associated with any specific decision criteria. They aimed to provide some initial evidence of the strength of the main predictors of self-regulatory and affective outcomes in relation to the conceptual model, and encompassed only continuous variables, without analysing interactions. Given the purely exploratory purpose, a few sociodemographic variables were added to the analysis, even if not included in the conceptual model, to ensure they had no major impact on the outcome variables. Indeed, past research has shown that holders of traditional religious beliefs and conservative views are more prone to rejection and defensive reactions toward deeds that conflict with their morals (with regard to homosexuality, see e.g., Janssen & Scheepers, 2019; with regard to vegetarianism, see e.g., Hoffarth et al., 2019).

³⁰ In fact, forced entry and backward elimination did produce a larger number of predictors, many of weak effect size (e.g., both forced entry and backward elimination in Francia's vignette found seven significant predictors of positive and negative affect, instead of four with forward selection).

When aggregating all the vignettes, the variance explained in the model with positive affect as dependent variable was much greater (adjusted $R^2 = .715$) than the variance explained in the model with negative affect as dependent variable (adjusted $R^2 = .177$). In line with the general assumptions of the conceptual model in Figure III, the best predictor of positive affect was self-improvement ($\beta = .754$) and the best predictor of negative affect was self-defence ($\beta = .298$). The variance explained in the model with self-improvement as dependent variable was greater (adjusted $R^2 = .577$) than the variance explained in the model with self-defence as dependent variable (adjusted $R^2 = .293$). The best predictors of self-improvement were agent evaluation ($\beta = .328$) and goodness of the deed ($\beta = .244$), and the best predictors of self-defence were goodness of the deed ($\beta = -.267$) and propriety of the deed ($\beta = -.196$). Moral discrepancy was a significant predictor of self-defence.

Bearing in mind these initial results, a further set of regressions with the forward selection procedure was conducted independently for the top vignettes of Francia, Nicholas, and Ruxandra. The total variance explained by the models, reflected in the adjusted R^2 , was again considerably higher for the prediction of positive affect and self-improvement than for negative affect and self-defence (Tab. A14a-b).

Tables A14a-b:

Multiple linear regressions for the vignettes with Francia, Nicholas, and Ruxandra (forward selection): model summaries. Dependent variables: positive/negative affect, self-improvement/self-defence

Tab A14a: Dependent variables - positive and negative affect

Vignette	Dependent variable: positive affect				Dependent variable: negative affect			
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R	R Square	Adjusted R Square	Std. Error of the Estimate
Francia	.790	.624	.619	13.743	.525	.276	.244	13.946
Nicholas	.820	.672	.663	16.817	.506	.256	.224	12.750
Ruxandra	.812	.659	.655	17.763	.510	.260	.229	15.68701

Tab A14b: Dependent variables – self-improvement and self-defence

Vignette	Dependent variable: self-improvement				Dependent variable: self-defence			
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R	R Square	Adjusted R Square	Std. Error of the Estimate
Francia	.690	.477	.454	14.734	.435	.189	.178	12.244
Nicholas	.806	.649	.629	13.965	.244	.060	.046	11.982
Ruxandra	.758	.575	.550	14.102	.607	.368	.350	18.898

The significant predictors and the regression coefficients for Francia, Nicholas, and Ruxandra are exhibited in Tables A15a-d. One thing worth noting is that the variable religiosity/spirituality appears as a positive predictor of both self-improvement and self-defence/negative affect; this phenomenon could be explained by the fact that the question was inadvertently double-barrelled: conflating in a single construct two variables (religiosity and spirituality) that could be non-significantly correlated could have caused this inconsistent behaviour. From Study 2 onward, two distinct questions were asked.

Importantly, multicollinearity was never an issue in these analyses, since the variance inflation factor (VIF) was never greater than 2.4 (values above 10 are considered problematic; see Myers, 1990; full regression output available in SMA10). Also, the best regression models (with predictors) for the top three vignettes were

better than using the mean of the dependent variable (intercept-only models, see SMA10).

Tables A15a-d:

Multiple linear regressions for the three main vignettes (forward selection): significant coefficients. Dependent variables: positive/negative affect, self-improvement/self-defence

Tab. A15a: Regression coefficients - Dependent variable: positive affect

Vignette	Significant Predictors	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	Sig.
		B	Std. Error	Beta		
Francia	(Constant)	14.497	5.981		2.424	.018
	Self-improvement	.883	.082	.790	10.786	<.001
Nicholas	(Constant)	-23.484	11.184		-2.100	.039
	Self-improvement	.812	.117	.642	6.924	<.001
	Agent evaluation	.428	.170	.234	2.525	.014
Ruxandra	(Constant)	-10.101	4.194		-2.409	.019
	Self-improvement	1.167	.099	.812	11.806	<.001

Tab. A15b: Regression coefficients - Dependent variable: negative affect

Vignette	Significant Predictors	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	Sig.
		B	Std. Error	Beta		
Francia	(Constant)	-10.521	6.745		-1.560	.123
	Self-defence	.486	.125	.409	3.879	<.001
	Self-improvement	.208	.088	.259	2.376	.020
	Relevance	.121	.059	.220	2.056	.044
Nicholas	(Constant)	24.244	8.014		3.025	.003
	Propriety	-.247	.077	-.332	-3.192	.002
	Self-defence	.368	.126	.312	2.914	.005
	Religiosity/Spirituality	.106	.046	.244	2.301	.024
Ruxandra	(Constant)	5.058	8.410		.601	.550
	Self-improvement	.427	.095	.503	4.500	<.001
	Self-defence	.231	.085	.303	2.719	.008
	Age	-.250	.121	-.212	-2.056	.044

Tab. A15c: Regression coefficients - Dependent variable: self-improvement

Vignette	Significant Predictors	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
Francia	(Constant)	9.425	7.966		1.183	.241
	Propriety	.524	.091	.529	5.753	<.001
	Similarity	.207	.076	.249	2.710	.009
	Religiosity/Spirituality	.111	.052	.189	2.144	.036
Nicholas	(Constant)	-.476	9.550		-.050	.960
	Goodness	.533	.108	.459	4.946	<.001
	Similarity	.231	.073	.242	3.175	.002
	Religiosity/Spirituality	.148	.050	.215	2.962	.004
	Moral discrepancy	.275	.108	.235	2.539	.013
Ruxandra	(Constant)	7.275	3.681		1.976	.052
	Goodness	.223	.066	.330	3.360	.001
	Courage	.251	.058	.360	4.314	<.001
	Similarity	.160	.068	.210	2.362	.021
	In-Shoes	.122	.053	.209	2.312	.024

Tab. A15d: Regression coefficients - Dependent variable: self-defence

Vignette	Significant Predictors	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
Francia	(Constant)	37.975	6.088		6.238	<.001
	Propriety	-.292	.072	-.435	-4.045	<.001
Nicholas	(Constant)	28.780	6.146		4.683	<.001
	Goodness	-.152	.071	-.244	-2.135	.036
Ruxandra	(Constant)	54.761	4.380		12.504	<.001
	Goodness	-.464	.073	-.615	-6.376	<.001
	Religiosity/Spirituality	.148	.069	.208	2.152	.035

Development of Preliminary Path Models

Path analysis was carried out using the R package lavaan ver. 0.6-5 (Rosseel, 2012) assuming independence of the observations. Estimation used robust maximum likelihood (MLR), and pathway coefficients were tested 2-tailed at 95% c.i. Composite mean scores of the scales were used, assuming perfect reliability (i.e., no measurement error).

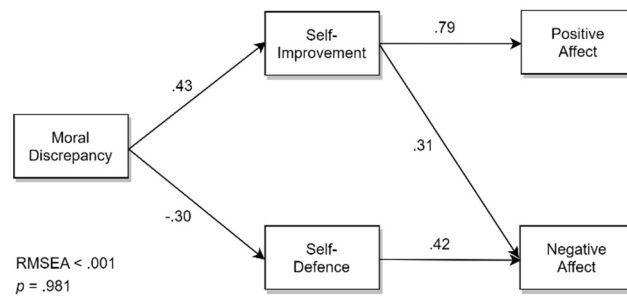
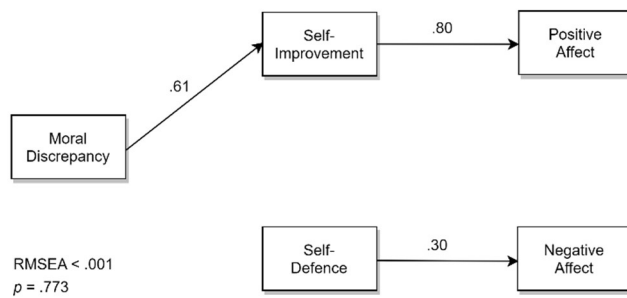
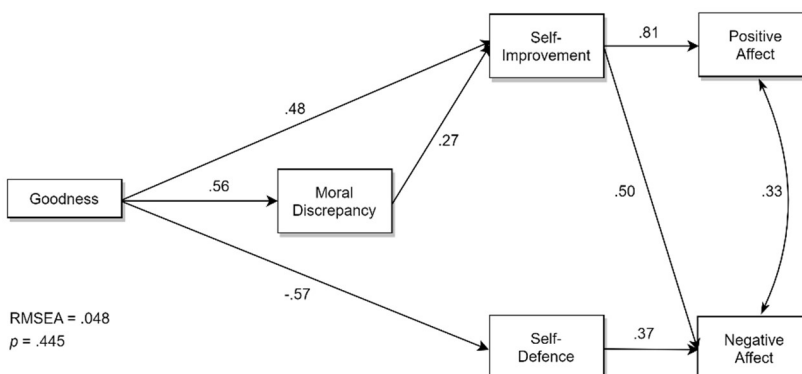
To fulfil parsimony objectives, the analysis was conducted in two phases. In the first phase, the models for the three vignettes of Francia, Nicholas, and Ruxandra were fitted using the variables in the conceptual model available from this study (i.e., moral discrepancy, self-improvement, self-defence, positive affect, and negative affect). The model fit was assessed with the root mean square error of approximation (RMSEA), a widely used fit index, highly diagnostic, as it returns a parsimony-adjusted value: were the fit satisfactory (RMSEA less than .080) or excellent (less than .050) and the p -value non-significant, those models were accepted; were it not satisfactory (RMSEA greater than .080) and the p -value significant, a second model was fitted with the addition the most impactful predictors (using the results from the multiple regression and the modification indices) until the model fit became satisfactory³¹.

The diagrams in Figures A13a-c illustrate the path models with satisfactory fit for the three above-mentioned vignettes, indicating standardised regression weights and RMSEA/ p -value. The other main fit indices and the full results are reported in SMA11. Only Ruxandra's model required the additional variable of the goodness of the deed to achieve non-significant and satisfactory RMSEA, presumably because of the wide range of evaluations of the protagonist's controversial moral action.

³¹ It must be emphasised again that path modelling was explicitly used here in a model *generative* approach, deferring confirmatory approaches to the following phases of the research.

Figures A13a-c:

Path models for the vignettes of Francia, Nicholas, and Ruxandra. Estimator: robust maximum likelihood (MLR). Standardised regression weights. Error terms not displayed to facilitate legibility

Fig. A13a: Francia**Fig. A13b: Nicholas****Fig. A13c: Ruxandra**

Findings from the Open-Ended Questions

Lastly, thematic analysis was conducted for each vignette on the answers to the open-ended questions. The aim was to identify additional insight to support the decision about which vignettes to progress to the following studies and identify improvements in the text or visuals that would make the vignettes suitable to be progressed.

Six key themes were identified across the twelve vignettes. In a few cases, especially the vignettes portraying popular public figures, prior knowledge of the protagonist of the story biased the judgment (Cory and Sarah³²), for good or bad. Some scenarios depicted stories that were considered too heroic and extreme (especially Arnaud and Wesley) and others quite ordinary (Matthew and Joey). Other scenarios were thought to be lifestyle choices rather than morally motivated good deeds (Markus and Ruxandra). In a few cases, there was a perception of obligation to carry out the specific deed performed by the protagonist, especially in the vignette portraying Arnaud (a policeman). Lastly, Alvaro's vignette contained multiple actions (a negative one, bullfighting, and a positive one, campaigning against bullfighting) and the halo effect of the former on the latter biased the judgment of the good deed (i.e., participants declared that in their mind Alvaro's repentance could not wash away the killing of many innocent bulls³³).

All these insights seemed to confirm the preliminary findings, suggesting that Francia's and Nicholas's vignettes could be the best candidates to be progressed to the

³² Although a celebrity too, Francia was not mentioned as a public figure whose knowledge consciously affected the results.

³³ This "halo effect" could be ascribed to the phenomenon of *identity continuity*, according to which transformative events in people's lives can lead them to become "different" persons, but not "new" persons. According to this view, the case of the overall poor judgment of Alvaro's character could be explained by participants' seamless judgment of his identity as a unitary psychological entity made of the sum total of all his identities from past to present, with no discontinuity. For a discussion on identity continuity, see for example Gomez-Lavin et al., 2020.

next phases of the research, with Ruxandra a third possible option. A summary of the key themes is reported in Table A16 and a more detailed analysis with anonymised quotes can be found in SMA12.

Table A16:

Summary of the key themes from the answers to the open-ended questions

Theme number	Theme label	Vignettes
Theme 1	Knowledge of the protagonist	Cory, Sarah, Sunita
Theme 2	Extreme situation	Wesley, Arnaud, Francia
Theme 3	Everyday kindness	Matthew, Joey
Theme 4	Obligation	Arnaud, Matthew, Nicholas
Theme 5	Lifestyles and personal choices	Markus, Ruxandra
Theme 6	Multiple actions and halo effects	Alvaro

Discussion

The preliminary exploratory investigation conducted in Study 1 was designed to meet three key objectives: a) provide insights on the newly developed stimuli, assessing their suitability and identifying the best candidates to progress to the next phases of the research; b) provide feedback on the preliminary measurement instruments; c) provide an initial assessment of the nature and strength of the relationships between some of the main variables in the conceptual model. The research questions were framed in the context of the conceptual model and nomological network previously illustrated (Fig. III) and the results offered answers that facilitated decisions for the next stages of the research.

Stimuli

The combined analysis of the findings from the closed- and the open-ended questions enabled a clear understanding of the strengths and weaknesses of the twelve vignettes assessed in the study.

Overall, the written narrative format of the vignettes proved to be adequate for the elicitation of comparative and self-regulatory processes. The existing literature in moral psychology, experimental philosophy, and neuroscience offered a limited range of ready-to-use materials depicting good deeds, for the most part in video format. Generally speaking, video clips deliver a richer set of cues than written vignettes (e.g., the flow of images in movement, the protagonists' looks and voices, sometimes even background music). Therefore, although examples of written stimuli about good deeds do exist in the literature, there was a legitimate question about the suitability of the newly developed vignettes to fulfil their purpose in the present research. To make them realistic and engaging, the twelve vignettes included both text and visuals.

The overall response from participants indicated that the vignettes:

- were perceived to portray people performing morally motivated deeds with varying gradients of goodness, some more universally regarded as positive (e.g., Francia and Cory) and others more polarising (especially Arnaud and Ruxandra);
- depicted moral actions characterised by diverse combinations of courage, care, and justice. High levels of courage and care often overlapped (e.g., Francia, Wesley, Cory, Arnaud); care was prevalent in the vignettes with Joey, Matthew, Sarah, and Markus; justice was more discriminant and comparatively stronger in Nicholas, Alvaro, and Sunita;
- elicited ability-based moral comparison processes, with most vignettes showing a prevalence of upward comparison (e.g., Cory, Francia, Arnaud), and a few indicating a prevalence of downward comparison (Ruxandra and Alvaro);
- activated both modes of moral self-regulation, mainly self-improvement (especially Francia), but also self-defence (especially Ruxandra), which correlated respectively with positive and negative affect.

Overall, these results suggested that, while several vignettes met the objectives, those portraying Francia (organ donor) and Nicholas (lawyer fighting for social justice) showed the mix of suitable characteristics that made them the best candidates to be progressed to the following phases of the research. Indeed, Francia focused more on care/courage and Nicholas on justice; Francia garnered near-universal positive judgments and Nicholas a more diverse range of judgments; both induced participants to engage in upward but also downward comparisons; and both elicited a mix of broadening and defensive self-regulatory processes, leading to corresponding positive and negative affective states. This blend of similarities and differences was considered ideal to test a conceptual model across a common core (good deeds performed by moral exemplars) and specific features (in terms of moral domain, gradient of perceived

goodness of the deed, proportion of adaptive/maladaptive self-regulatory modes and affective states).

A third vignette, Ruxandra (vegetarian), revealed a highly distinctive profile of moral judgments, richer in low scores on goodness and propriety, more downward comparisons, stronger self-defensive processes and negative affect (relative to the other vignettes). This result is consistent with the moral psychology literature that describes denigration of vegans and vegetarians. Minson and Monin (2012) found that people expect vegetarians to feel morally superior to them, and the more they anticipate superiority, the more they tend to negatively judge them; the anticipated reproach constitutes a moral threat that is aversive to the self, instigating the disparagement of vegetarians. It could be argued that the same mechanisms could have been at play in response to Ruxandra's vignette, who not only is vegetarian, but also campaigns against animal cruelty, which amplifies the moral significance and potentially might have magnified the perception of threat. Thanks to these characteristics, after the two top vignettes of Francia and Nicholas, Ruxandra would represent an interesting case to further explore in the future.

Moral Self-Regulation and Affect

Given the absence in the psychological literature of a measurement instrument capable of capturing the nuances of the broadening and defensive processes elicited by moral exemplars, a new moral self-regulation and affect inventory was developed, so that it could apply to a wide array of positive moral scenarios. Due to the restrictions imposed by the length of the questionnaire and the accessibility of voluntary participants, Study 1 assessed a limited set of items from the full inventory. The key objective was to use the results to identify any items that needed to be modified or dropped, as they did not correlate strongly with the others and the related scales.

Overall, the answers to the items showed a wide range of values, but also non-normal distributions, with both skewness and kurtosis. Because of the opposite direction of the skew for the self-improvement/positive affect and the self-defence/negative affect items, it was not possible to identify one single transformation that could reduce the bias for all of them. The solutions adopted hereon to account for violations of normality included, depending on the case, bootstrap samples, non-parametric tests, robust and non-asymptotic estimation techniques that make fewer or no specific demands on the distributions.

Moral Self-Regulation Inventory

The six self-improvement items worked well as a scale, which exhibited satisfactory internal consistency for most vignettes, particularly for the top three (Francia, Nicholas, and Ruxandra). None of the items seemed to require modifications and all were progressed to Study 2.

The self-defence scale showed lower internal consistency. Item SD4 (“This story is too good to be true”) did not correlate well with the scale and its removal would improve the scale reliability for several vignettes; it was decided to drop it from the scale. Item SD5 (“These extreme behaviours should not be considered the standard we live by”) was also problematic, but it was reworded to emphasise its uncommon rather than extreme nature (“Actions uncommon as this one should not be considered the standard we live by”) and brought forward to the following study for further assessment. Item SD3 (“Francia’s actions may be good, but I bet she is seeking the praise of others”) performed well but was slightly reworded too, so that it focused on one specific action (“Francia may have done a good deed, but I bet she is seeking the praise of others”).

Moral Affect Scales

Despite the existence in the literature of several scales to measure affect, they were deemed too long and partly inadequate to the present research. Thus, new scales tailored to measure the affective response to others' good deeds were developed. A subset of them was assessed in Study 1 (three and five items respectively for positive and negative affect), with the aim of dropping or modifying those that were unsatisfactory. The scales exhibited adequate internal consistency for the top three vignettes. In the positive affect scale, item PA3 ("I felt challenged in a positive way") was the weakest, but it was decided to retain it and test it again in Study 2, when a wider set of the scale items would be assessed. Item PA2 ("The story was inspiring") performed well but was slightly reworded ("I was inspired by the story") to put the emphasis on the participants' subjective feeling rather than the story. No modifications were made to the negative affect items and all were retained.

Moral Self-Evaluation

An important moral appraisal measure was the evaluation of the self, that is, the strength of the moral character that participants attributed to themselves. Unsurprisingly, on average participants rated their own morality very highly. Although measured differently, this phenomenon is in line with previous research on the "better-than-average effect", according to which people overwhelmingly rate themselves better than an average peer on various attributes (for a review, see e.g., Alicke & Govorun, 2005). This effect is thought to stem – among other factors – from the need to maintain favourable self-views (Alicke & Sedikides, 2009) and relies on various cognitive processes of generalisation and abstraction that make people vulnerable to gaps between the judgment of their own actions and objective reality (Alicke et al., 2012). The better-than-average effect was found to be even more striking when people compared themselves

with others on moral attributes (relative to competence attributes), giving rise to the so-called “Muhammad Ali effect” (Allison et al., 1989).

The moral self-evaluation was also found not to be significantly correlated with either self-regulatory mode and either affective state; it also did not correlate with the appraisals of the goodness and the propriety of the deed. These results seemed to depend on the fact that the self-evaluation scores were quite high for nearly all participants, while the appraisals of the deeds and the downstream psychological mechanisms of self-regulation and affect exhibited a wider range of values.

It must be noted that the measure of self-evaluation was collected in this research in the first part of the questionnaire (alongside the socio-demographic variables) before the presentation of the vignettes. Therefore, it could be argued that it represents a sort of trait measure of moral self-evaluation, capturing chronic self-perceptions of moral stature crystallised in people’s self-concept and relatively stable over time, although modifiable (see Monin & Jordan, 2009). This aspect will be discussed next in more detail.

Other Measures of Moral Appraisal and Comparison

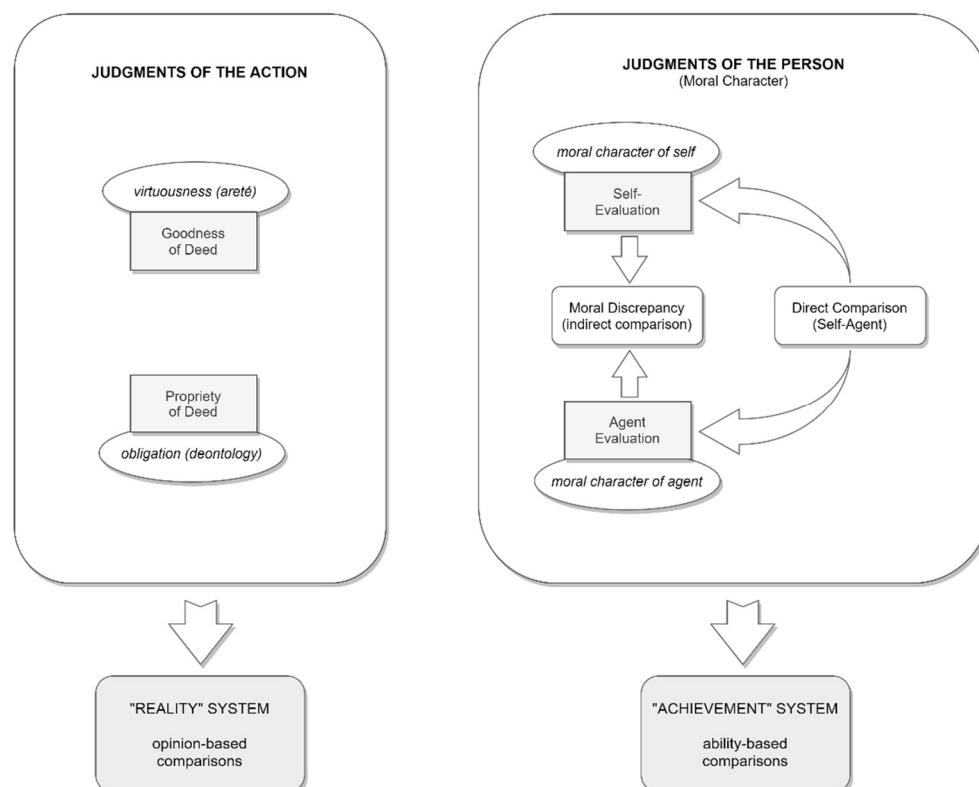
The moral appraisal variables measured judgments pertaining to two distinct evaluation targets: the moral *actions* described in the vignettes and the moral *character* ascribed to the protagonists of the stories. At a deeper level, these evaluations subsume the two different kinds of epistemic assessments mentioned in the introductory chapter: one based on a reality system and the other based on an achievement system (Kelley, 1971). The items measuring the level of goodness and the level of propriety of the deeds belong to the first kind of epistemic assessment: they are judgments that distinguish *behaviour* in terms of what is *right* and what is *wrong* based on specific moral standards held by participants. Thus, they provide an insight into the moral comparison processes that Festinger (1954) defined “opinion-based”. Specifically, the goodness refers to the

level of virtuousness of the deed (its “aretaic” aspect), while the propriety refers to the level of obligation of the deed (its “deontic” aspect). The items that measure the moral character of the agent and the self belong to the second kind of epistemic assessment: they are judgments that imply an evaluation of the moral stature of the *person* and allow to gauge the *capability* to perform certain moral actions. Thus, they provide an insight into the moral comparison processes that Festinger defined “ability-based”.

Figure A14 displays a conceptual illustration of the two classes of moral judgments (action- and person-based) and the corresponding types of moral comparisons (opinion- and ability-based) that they imply.

Figure A14:

A conceptual illustration of the two classes of moral judgments and the underlying types of moral comparisons based on opinion and ability



Analysis of the direct and the indirect measures of moral comparison revealed a dissociation between the two variables, with different patterns emerging from them:

- forcing a conscious, explicit, *direct* comparison with the protagonists of the stories seemed to make participants surrender a less flattering image of their own relative moral standing;
- inferring implicit comparison processes *indirectly* from the difference between self-evaluation and agent evaluation (evaluated at different stages of the questionnaire), without necessarily making the comparison salient to conscious thought at the time of measurement, allowed to uncover a more psychologically meaningful perception of the self/other character gap, more strongly correlated with downstream self-regulatory processes.

It could be argued that the direct and indirect comparison measures capture two distinct types of evaluations, the former more dependent on the temporary state induced by the vignette (which yielded more upward comparisons), the latter anchored to a more stable trait-like evaluation of the moral self-concept (which yielded more downward comparisons). The fact that the indirect measure exhibited stronger correlations with downstream psychological processes might be explained by this anchoring to an enduring ideal reference point based on their *best* moral performance; it could be speculated that this is more likely to drive actual behaviour than a less self-flattering direct comparison surrendered when a state-like assessment of the moral self is made salient in relation to highly praiseworthy moral exemplars.

From a methodological standpoint, moral discrepancy also provides a measure of the output of moral comparisons which is clearly distinct from self-regulation processes. This approach marks an important methodological difference from classical social comparison research based on the rank-order paradigm, which ends up confounding

measures of comparison and self-regulation. For all these reasons, the indirect measure of moral discrepancy was considered the more suitable measure of moral comparisons and used in the following studies.

It is also important to mention here that, in an attempt to achieve parsimony, the conceptual model developed at the start of the research restricted the perimeter of interest to ability-based comparisons, forgoing opinion-based comparisons. This was due to the assumption that the moral actions depicted in the vignettes would be widely perceived in a positive light, with little variation around high mean scores. If the goodness of the deeds were highly regarded by most participants, its contribution to generating variance in the model would be limited. The empirical findings from Study 1 revealed a more intriguing phenomenon. In some vignettes, especially Ruxandra and Alvaro, the deeds garnered widely diverging judgments, with means below the scale mid-point; in other vignettes, for instance Nicholas and Sunita, the mean judgments were moderately positive, and in others overwhelmingly positive (Francia, Cory, and Wesley). Therefore, at this stage, this finding seemed to suggest a differential role played by moral comparisons based on opinion and ability as a function of the moral scenarios:

- for uncontested deeds, widely regarded as highly praiseworthy, moral judgments underpinning ability-based comparisons could represent the more critical response driver, whereas opinion-based comparisons in a parsimonious model of the self-regulation of virtue could be regarded as redundant (explaining marginal incremental variance);
- for more controversial deeds, with mixed appraisals due to antagonistic moral standards, ability-based comparisons could not suffice to fully explain self-regulatory mechanisms, in particular defensive, which could be activated to a

significant extent by opinion-based comparisons implying less positive views about the actions and the agents.

This insight required further empirical scrutiny and was investigated more carefully with the availability of the data from the ensuing studies among larger samples.

Relationships Between the Main Variables

Correlations

One of the important preliminary findings from Study 1 was the pattern of correlations between moral discrepancy and self-regulation. The existing social psychological literature supported the notion that upward and downward comparisons could be associated with both self-improvement and self-defence, depending on whether they are experienced as opportunities or threats (Sedikides, 2012) and whether they are associated with growth or well-being motives (Wood & Taylor, 1991). Consequently, the conceptual model adopted the initial assumption of relative independence of moral comparisons and self-regulatory modes, deferring further analysis once empirical data would become available. Study 1 provided initial evidence of the existence of a specific probabilistic correlation pattern, whereby when witnessing virtuous acts, at an aggregated level across vignettes, self-improvement appeared more likely to occur following upward comparison, and self-defence more likely following downward comparison. The strength of these correlations was strong, and the result held up for most of the vignettes analysed separately. Participants engaging in upward comparisons experienced the moral stories primarily as inspiring opportunities to better themselves, whereas those engaging in downward comparisons experienced them primarily as threats and tended to self-defend in an effort to safeguard and validate their supposed moral superiority. The association between downward comparison and self-defence is consistent with that part of the literature which claims that people tend

to disproportionately protect moral self-beliefs compared to other kinds of beliefs (Ellemers et al., 2019; Jordan & Monin, 2008; Mazar et al., 2008): if individuals hold a certain self-evaluation about their morality, they are strongly motivated to maintain it and defend it from external threats (Pagliaro et al., 2016), adjusting their moral reasoning to justify and affirm their self-views (Haidt, 2001).

Regression and Path Models

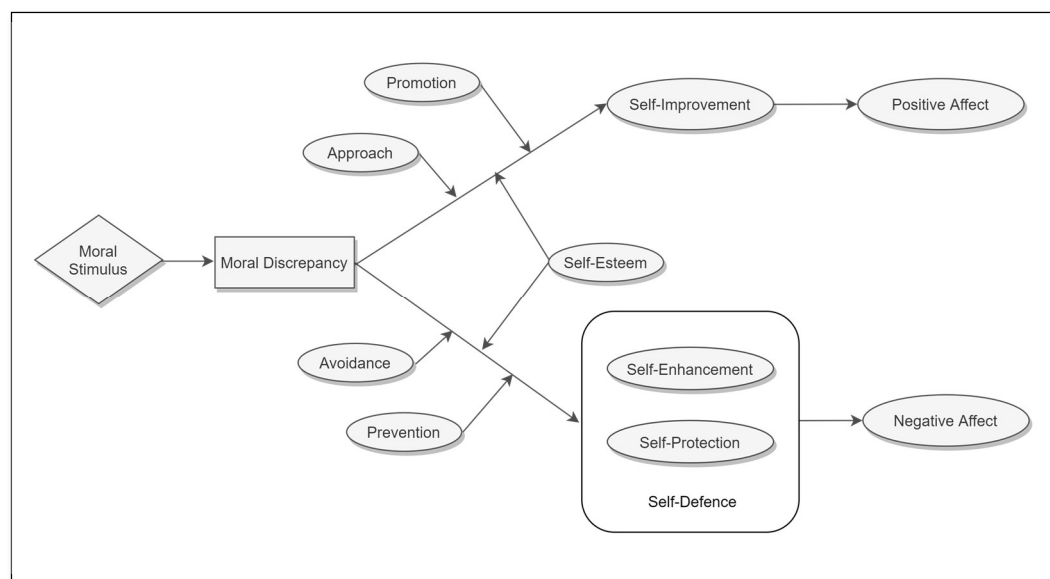
Multiple regression highlighted that self-improvement and self-defence were the best predictors of respectively positive and negative affect, and goodness, propriety, and discrepancy were among the strongest predictors of the moral self-regulation constructs. This analysis could only specify *direct* associations between predictors and a single outcome at a time (Hoyle, 1995), but not *indirect* pathways or other conditional processes between all the variables simultaneously. More specifically, they could not feature the mediated paths through which the variables exert their mutual influence. To investigate those pathways, preliminary path modelling was carried out for exploratory purposes for the top three vignettes (Francia, Nicholas, Ruxandra). While holding the conceptual model in the background, the technique of path modelling was used in a “model generative” way rather than a strictly confirmatory setting, allowing the identification of initial structural models that could be tested again in subsequent studies.

Francia’s and Nicholas’s models resulted in satisfactory fit using only moral discrepancy and the two modes of self-regulation to predict moral affect. Ruxandra’s vignette required the addition of the goodness of the deed to achieve satisfactory fit, corroborating the notion that more controversial moral actions could only be explained by a wider range of predictors, particularly those moral appraisals that subsume opinion-based judgments. These results provided clear initial evidence across the three

vignettes of the plausibility of the theoretical framework and seemed convincing enough to suggest a first revision of the conceptual model (Fig. A15), hence informing new and more specific hypotheses for Study 2.

Figure A15:

Graphical illustration of the revised and simplified conceptual model of the moral experience of virtue, observed through the lens of moral self-regulation, inclusive of a narrower selection of moderators



For the following Studies 2 and 3, the robustness of the self-regulatory framework suggested to prioritise the integration of the model with motivational predictors rather than personality traits, and so regulatory focus and approach/avoidance (along with self-esteem) were preferred over humility and narcissism, whose exploration was conducted in Study 4.

Summary of Key Findings and Limitations

In summary, Study 1 met its three main objectives and offered a wealth of insights about: a) the most suitable vignettes for an investigation of the self-regulation of virtue; b) the reliability of the preliminary version of the self-regulation and affect inventory; c)

the importance of the measures of act- and person-based judgments to infer moral comparison processes; d) the significant associations of upward comparisons with self-improvement and downward comparisons with self-defence.

A synthesis of these initial findings is graphically illustrated in the preliminary path models (Fig. A13a-c), which enabled to revise and simplify the initial conceptual model (Fig. A15), focusing on fewer critical variables. A side-by-side comparison between the initial and the revised models can elucidate the significance of the learning distilled from this initial study.

It must be acknowledged though, that these results must be taken with caution. The sample was not random or representative; due to resource constraints, it was necessary to rely on a convenience sample of voluntary participants recruited during the summer. In general terms, the total sample size of 146 participants might look satisfactory, but each vignette was allocated to a random selection of about 70 participants. While being a relatively modest sample size, it is was determined to be sufficient to detect a medium-sized effect based on a simplified *a priori* power analysis.

Although the research design was mixed and the analytic procedures also encompassed more sophisticated techniques such as path modelling, power analysis was conducted within the framework of multiple linear regression, assuming independence of the observations, normality of the residuals of the dependent variables, homoscedasticity of the residuals, and absence of indirect pathways between predictors. The data violated these assumptions, but the results can be considered robust enough for an early exploratory study. The standardised coefficients from multiple regression and the standardised regression weights from path analysis were moderate to large (usually above .30 and in several cases above .50) and the pattern recurred across vignettes. The magnitude of these regularities was such that, even if the

point estimates might be slightly overestimated due to the above-mentioned methodological weaknesses, they should not be too far from the true population values.

The actual analysis slightly differed from the preregistered plan. No systematic frequentist significance testing of the mean differences across vignettes was conducted and no Bayes factors were computed. Due to the exploratory nature of the study and for the sake of conciseness, the details of a multilevel model analysis were omitted from this chapter³⁴.

Participants provided their answers to nearly all the questions on 0-100 scales using slider bars. This format makes the scales virtually continuous, but by no means the distributions were automatically assumed to be normal. It was known upfront that the nature of the measurement of the phenomenon at hand would likely generate skewed distributions, which could have been assimilated to zero-one-inflated beta distributions (ZOIB; Liu & Kong 2015), for which more complex Bayesian ordinal regression methods could have returned more accurate estimates (see e.g., Bürkner & Vuorre, 2019). The decision to simplify the analysis and use traditional ordinary least squares (OLS) estimates was justified by the exploratory nature of this stage of the research. Other important analyses, such as path modelling, relied on robust estimators that did not assume normal distributions. In path modelling, all the variables were considered observed, using the composite means for each of the scales, thus ignoring the existence of latent variables and measurement error. Again, this simplification was deemed to be acceptable in the context of an early exploratory study with a relatively small sample.

³⁴ Results from the multilevel model analysis for positive affect indicated a significant random intercept effect at both participant and vignette level, and no significant random slope effects, and for negative affect only a significant general random intercept effect at both participant and vignette level (see SMA13).

Like most empirical research on virtue and good deeds available to date, the present research programme, including Study 1, was designed around self-reports. In particular, the new moral self-regulation inventory used, at least partly, the so-called “indirect” self-reports (especially the self-defence scale) which attempt to obscure the constructs being measured through *subtle* items (Paulhus & Vazire, 2007). This strategy often allows to reduce demand characteristics. Yet, self-reports remain partly vulnerable to various forms of bias. On the one hand, people might want to portray themselves in certain ways. When responding to the good deeds described in the vignettes, participants might try to present themselves in ways that are consistent with their *ideal-self* or their *ought-self* (Higgins et al., 1986), that is, consistent with the way that they would want to see themselves in an ideal world or with the way that they believe was socially desirable, aligned to certain social norms. By doing so, they would either consciously suppress their true negative thoughts and feelings, or alternatively they would show admiration and approval beyond their authentic thoughts and feelings. On the other hand, participants might simply be vulnerable to non-conscious self-deception processes: they might have no deliberate intention to lie, and yet fail to reveal their true thoughts and feelings, again either suppressing or over-emphasising their true internal states. These processes of impression management (the former) and self-deception (the latter) have been well documented in the social psychology literature and fall under the umbrella of social desirability or socially desirable responding (see e.g., Crowne & Marlowe, 1960; Paulhus, 1994). In impression management individuals consciously dissemble and dissimulate (they knowingly lie, fake, exaggerate), in self-deception they are convinced of the genuineness of their responses, and yet they self-favour, self-enhance, self-defend, deny (Paulhus, 1984; Paulhus & Vazire, 2007). Both processes are thought to originate from approval needs and Paulhus (1984) recommends controlling

for them (especially impression management). Because social desirability was not measured in Study 1, it could have slightly inflated self-improvement and deflated self-defence. The exploration of social desirability phenomena in the next studies was considered a potential option to reduce the bias of a methodology relying purely on self-reports.

A third way in which self-report data could be potentially biased is the presence within the sample of participants who respond carelessly or apply insufficient effort in their responses (for a review, see Curran, 2016). Whereas in socially desirable responding answers to survey questions are not truthful due to consciously or non-consciously motivated reasons to generate a desired outcome, in careless or insufficient effort responding answers are not truthful because of the lack of interest or attention, or the desire to minimise effort or time (Curran, 2016). In the past, it was thought that careless or insufficient effort responding weakened the relationships among constructs, but recent evidence has shown that the effect can be unpredictable, depending on the type of careless or insufficient effort responding (Kam & Chan, 2018). In Study 1 no specific countermeasures were adopted before data collection to identify and isolate this phenomenon; potentially, outlier analysis could have identified careless responders (at least partly), but none were removed, providing partial reassurance that strong effects caused by careless or insufficient effort responding should not have occurred. As an improvement measure, for the following studies it was decided to implement specific actions to enhance data quality.

The preliminary results from Study 1 were put to the test in Study 2, a larger exploratory investigation designed to consolidate and deepen understanding of the conceptual model before the confirmatory stage (Study 3).

Study 2: Model Assessment

Introduction

The results from Study 1 provided insights that were instrumental to the choice of appropriate stimuli, the selection and improvement of the measurement instruments, the choice of a narrower number of variables of interest, and the revision of the conceptual model. These findings still needed to be complemented by additional information before a model of the self-regulation of virtue could be tested in a confirmatory study, specifically: a) validity and reliability of the *full* moral self-regulation and affect inventory, inclusive of all the items generated in the early stage of the research; b) evidence of the moderating role of the characteristic adaptations and the significance of these variables as additional factors for the understanding of the response to moral exemplars. The provision of this information was the objective of the present study, which was designed to contribute a more thorough assessment of:

- the *measurement model*, or in other words, the measures for the operationalisation of the key variables, including the moral comparison variables and the moral self-regulation and affect inventory;
- the *structural model*, that is, the web of mutual relationships across all the key variables and the role that each of them plays to explain the phenomenon under investigation.

As discussed in the previous chapter, the two vignettes of Francia and Nicholas emerged from Study 1 as the best candidates to be progressed to the next stages, each of them sharing similarities and differences that made them suitable to develop a relatively comprehensive model of the self-regulation of virtue under two different

scenarios. A third promising vignette (Ruxandra) was identified for potential follow-up studies.

Further investigation of participants' response to the two main moral scenarios was expected to facilitate discernment in two important areas:

1. the possibility to develop a *single* measurement model of the self-regulation of virtue, capable of capturing equally well the latent constructs across the two moral scenarios;
2. the possibility to develop either a *single* structural model of the self-regulation of virtue, capable of showing strong fit with the data for two different moral scenarios, or alternatively *two* separate models, one for each scenario.

The construction of a single measurement model was an important objective of this phase of the research, whereas falling back to the alternative of two slightly different structural models (one for each vignette) with a "common core" and specific features did not represent a major conceptual impediment to the development of substantive theory. Following the plan laid out at the beginning of the project, Study 2 was therefore designed to complete the development of the measurement model and advance the understanding of the structural model before Study 3.

Measurement Model

Moral Self-Regulation and Affect

In Study 1 only a limited number of items of the moral self-regulation and affect inventory could be tested; most of them were retained with no modifications, one was dropped and three were slightly reworded based on the findings. Study 2 re-tested them and explored the suitability of all the other items that had been previously generated.

At this stage of the research plan, Study 2 aimed to examine the reliability of the moral self-regulation and affect scales, but also to determine the dimensionality of the

moral self-defence scale. As previously discussed, beneath the self-defence items could potentially lie *one* latent factor (self-defence) or *two* conceptually distinct latent factors (self-enhancement and self-protection). The other scales of self-improvement, positive affect, and negative affect were hypothesised to be unidimensional, and one latent factor was expected to be retrieved from exploratory factor analysis (EFA) for each of them.

Moral Appraisals

Analysis of the regression and path models of three top vignettes in Study 1 helped select a subset of key moral appraisal variables considered critical to the understanding of underlying moral comparison processes: the level of goodness and propriety of the deed, and evaluations of the moral character of the agent and the self. The difference between agent and self-evaluations constituted the variable of moral discrepancy, an indirect measure of moral comparison that was preferred to the direct one.

Other moral appraisal variables used in Study 1 (e.g., the judgments of the level of courage, care, and justice, the relevance of the deed, the perception of similarity to the moral agent) were abandoned for the sake of parsimony. However, two novel moral appraisal variables were considered worthy of investigation and added, both measuring the perception of the normativity of the moral action (further details in the Methods section of this chapter).

Characteristic Adaptations

Study 2 also began the analysis of the role of self-esteem and two motivational dispositions (regulatory focus and hedonic orientation) in determining moral comparison and self-regulatory mechanisms experienced by participants exposed to the moral exemplars in the vignettes. These variables, as discussed in the first chapter, are part of what McAdams and Pals (2006) and DeYoung (2015) define *characteristic*

adaptations, that is, relatively stable goals, interpretations, and strategies that concur with other dispositions to form the structure of personality. Self-esteem could be viewed as a particular kind of interpretation of one's self-worth, while hedonic orientation and regulatory focus represent goal pursuit motives and strategies.

Hedonic Orientation and Regulatory Focus. Hedonic orientation and regulatory focus operate according to the same cybernetic principles as the moral self-regulatory functions at the core of this research. As previously clarified, a cybernetic system can function through discrepancy-reducing loops, whereby an input is compared to a standard and adjustments are made to shift it to a state that is closer to the standard, or through discrepancy-enlarging loops, whereby an input is shifted to a state that is farther away from the standard or at least not closer to it (Carver & Scheier, 2002). In a cybernetic framework, the hedonic orientations of approach and avoidance can be considered respectively discrepancy-reducing and discrepancy-enlarging motives or tendencies (Carver, 2006), as they involve energisation toward or away from certain standards, references, or goals. Regulatory focus as well can be observed from a cybernetic viewpoint, in that promotion focus involves growth needs toward the ideal-self (motion from "0" to "+1"), whereas prevention focus underlies the safeguard of the ought-self while maintaining a satisfactory status quo (i.e., preserving "0") or impeding loss of ground (i.e., ensuring "not -1") (Scholer & Higgins, 2012).

Promotion/prevention focus and approach/avoidance were conceptualised in the present research as relatively stable motivational orientations operationalised as trait measures. Many different measurement instruments can be found in the literature for these constructs and no consensus exists on which ones are preferable. Some of them are based on different theoretical paradigms. For instance, approach and avoidance have been conceptualised within biological theories of personality (Corr, 2015), such as

reinforcement sensitivity theory (Gray, 1982). This theory, in its initial formulation, postulated the existence of two conceptual nervous systems (Elliot & Thrash, 2002): the behavioural activation system (BAS) related to approach, and the behavioural inhibition system (BIS) related to avoidance. The BIS/BAS scales developed by Carver and White (1994) have enjoyed widespread adoption over the years as a tool to operationalise reinforcement sensitivity constructs, but have also shown deficiencies in their psychometric properties. When the reinforcement sensitivity theory was substantially revised (Gray & McNaughton, 2000), researchers in the field tried to develop new measurement instruments that also included the fight/flight/freeze system (FFFS), such as the Jackson-5 (Jackson, 2009) and the reinforcement sensitivity theory of personality questionnaire (Corr & Cooper, 2016). Approach and avoidance have also been conceptualised as neurophysiological reflexes (Lang & Bradley, 2010), sensitivities to reward and punishment (Torrubia et al., 2001), or temperaments (Elliot & Thrash, 2002), each with their own measurement tools.

Across these diverse conceptualisations and theoretical traditions, the constructs of approach and avoidance share a common essence: they are at least partly heritable and biologically based, emerge in early childhood, remaining relatively stable across the lifespan, and entail distinct affective tendencies associated with specific motivational systems (Elliot & Thrash, 2002). In the present research, approach and avoidance were conceptualised as basic *temperaments*, the affective core of personality (Rihmer et al., 2010), following the theoretical perspective and operationalisation proposed by Elliot and Thrash (2002, 2010). They are conceptually independent of each other, although in empirical research sometimes they appear to be weakly negatively correlated (Elliot & Thrash, 2002).

The measurement of trait regulatory focus was initially proposed by Higgins and colleagues in 2001, but researchers have since proposed several new measurement instruments, such as the general regulatory focus measure (Lockwood et al., 2002), the regulatory focus scale (RFS: Fellner et al., 2007), the regulatory focus strategies scale (RFSS: Ouschan et al., 2007), the regulatory focus reference-point scales (Summerville & Roese, 2008). One of the reasons that could explain the proliferation of dispositional regulatory focus measures lies in its conceptual complexity. The promotion and prevention foci imply various distinct components: the self-guides (respectively ideal-self versus ought-self), motivational needs or concerns (nurturance/achievement versus security/maintenance), reference points (gains/non-losses versus losses/non-gains). Each measurement instrument tends to rely more strongly on some of these components but fails to capture all of them to the same extent. A review by Haws and colleagues (2010) based on the analysis of correlation, representativeness, internal consistency, stability, and predictive validity (as suggested by Simms and Watson, 2007) provided evidence that the theoretical and empirical overlap between these instruments is limited.

Besides, the concept of regulatory focus is so intertwined with that of hedonic orientation that balancing approach and avoidance in the expression of the statements that form the measurement instruments has proved to be challenging. Higgins himself made it clear that regulatory focus and hedonic orientation are conceptually independent notions (see e.g., Cornwell & Higgins, 2015a; Higgins, 1997) that should not be conflated. It is therefore important that the scale indicators account for both approach and avoidance motives while measuring each of the regulatory foci. However, Lockwood and colleagues' (2002) scale, for instance, was found to only examine

tendencies to attain gains and elude losses, neglecting non-gains and non-losses (Summerville & Roese, 2008).

Haws et al. (2010) concluded that, among the instruments reviewed, the original regulatory focus questionnaire by Higgins and colleagues (2001) best encapsulated the key tenets of the theory and achieved the most reasonable balance between approach and avoidance, making it suitable to general-purpose theory testing. For these reasons and despite a predominant orientation of the items toward the self-guides and subjective past experiences, it is a very widely used measure of dispositional regulatory focus and therefore was adopted in this research. Like approach and avoidance, promotion and prevention focus are conceptually independent; empirical research has generally failed to detect evidence of correlation (e.g., Chung et al., 2014; Higgins et al., 2001; Polman, 2012), although sometimes a weak positive correlation was found (e.g., Cornwell & Higgins, 2015a; Pfattheicher & Keller, 2013).

Self-Esteem. Along with regulatory focus and approach/avoidance, Study 2 also investigated the impact of self-esteem on moral comparison and self-regulation. This concept has a long history in psychology. There is widespread evidence that healthy levels of self-esteem correlate strongly with goal achievement, quality of interpersonal relationships, general well-being, happiness, and life satisfaction, but the causal relationships are complex and intricate. Baumeister and colleagues (2003), for example, claim that high self-esteem is at least partly the result of good school performance and occupational success rather than the reverse, and that high self-esteem has not been found to cause better quality of interpersonal relationships, although people who hold high self-esteem tend to believe so; on the other hand, the authors also maintain that high self-esteem is an important contributor to happiness because of enhanced personal initiative and pleasant feelings.

As discussed by Baumeister and colleagues (2003), what makes self-esteem problematic to study is the fact that it is a heterogeneous concept, which could comprise unbiased judgments about the self and at the same time overly inflated or deflated self-evaluations that bear little resemblance to objective reality. Consequently, the authors claim that sometimes high self-esteem gets conflated with narcissism and low self-esteem with humility; and for this reason, depending on their nature and categories, high and low self-esteem end up being associated with a diverse range of outcomes, positive and negative.

Tice (1993) claims that folk conceptions about people with high and low self-esteem often include the belief that they exhibit opposite characteristics, but these beliefs may be inaccurate; if individuals high in self-esteem are seen as eager to succeed and be liked, it cannot be said that those low in self-esteem want to fail and be disliked. The real motivational differences between them are to be found in the fact that people with high self-esteem are driven by the desire to succeed, win love and admiration, while those with low self-esteem are more concerned about eluding failure, humiliation, or rejection (Tice, 1993). These underlying motives (succeeding and eluding) link high and low self-esteem quite closely with promotion/approach and prevention/avoidance (without overlapping with them), hence the interest in examining empirically their relationships with the other main variables in the conceptual model.

Unlike regulatory focus and hedonic orientation, the operationalisation of self-esteem has been dominated by one measurement instrument: the Rosenberg self-esteem scale (RSES), developed by Rosenberg in 1965. Although much debate has animated the social psychology literature about its factorial structure (claimed to be unidimensional, bidimensional, or bifactor; see e.g., Greenberger et al., 2003; Hyland et

al., 2014; Salerno et al., 2017), its robust psychometric properties make it the most widely used instrument and therefore it was adopted in the present research.

Structural Model

One of the overarching objectives of Study 2 was to deepen the understanding of the relationships between the main variables in the conceptual model. Some of these relationships had already been explored in Study 1, for instance those between moral discrepancy and the two moral self-regulation modes, as well as those between the self-regulation modes and affect; however, the associations of the characteristic adaptations between themselves and with the other variables still had to be measured. Analysis of the extant literature and the data collected in Study 1 suggested that certain correlation patterns could be expected and therefore some directional hypotheses were preregistered:

- positive correlations between self-improvement and positive affect, and between self-defence and negative affect;
- positive correlation between moral discrepancy and self-improvement, and negative correlation between moral discrepancy and self-defence (or absence of correlation³⁵);
- positive correlations of promotion focus, approach, and self-esteem with self-improvement;
- positive correlations of prevention focus, avoidance, and self-esteem with self-defence;

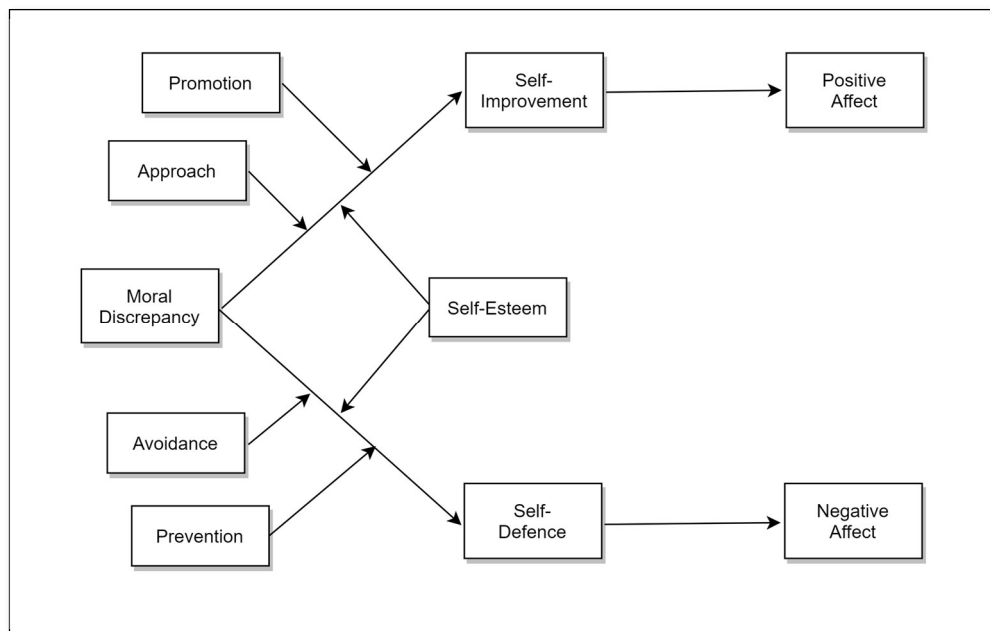
³⁵ Following the results from Study 1, the preregistration mentioned absence of correlation with reference to the most recurring outcome from the separate analysis for each of the top three vignettes, while the negative correlation is referred to the analysis with all twelve vignettes aggregated.

- promotion focus, approach, and self-esteem moderators of the relationship between moral discrepancy and self-improvement;
- prevention focus, avoidance, and self-esteem moderators of the relationship between moral discrepancy and self-defence.

These hypotheses are graphically summarised in the revised conceptual model illustrated in Figure B1.

Figure B1:

Graphical illustration of the revised conceptual model of the self-regulation of virtue, with the characteristic adaptations expected to function as moderators



Methods

Research Design

Study 2 had a between-subjects design: participants were randomly assigned to one of two independent conditions. In each condition, participants were presented with one of two moral vignettes: Francia or Nicholas (tested with no modifications from Study 1).

The study was conducted online, based on a structured questionnaire (see SMB2) hosted by Qualtrics and made available to potential participants during the month of November 2018. Recruitment was conducted through CloudResearch, formerly known as TurkPrime (Litman, Robinson, & Abberbock, 2016), which uses the Amazon Mechanical Turk (MTurk) crowdsourcing platform. Voluntary participants filled out the questionnaire in return for payment.

One of the measures introduced in Study 2 to improve data quality was the implementation of attention checks: after the presentation of the moral vignettes, participants were asked a multiple-choice question about the content of the story that they had just read (see SMB2). If participants did not answer correctly, they were gently reminded that it was important to know the content of the story to answer the following questions, and were subsequently allowed to view the vignette a second time before carrying on with the questionnaire.

Participants

Sample Size Determination

The sample size was determined based on several criteria. Power analysis was run using G*Power ver. 3.1, under the assumption of multiple linear regression with up to fourteen predictors for each outcome variable³⁶. To detect a medium-sized effect with α

³⁶ A list of the fourteen predictors can be found later in this chapter when reporting on multiple linear regression.

= .05 and power set at .80, a minimum sample size of 135 participants per condition would be necessary (full analysis available in SMB1). The sample size was actually set at 250 participants per condition in consideration of the more complex analysis techniques (beyond regression) that were planned, specifically factor analysis and path modelling. For these techniques, no simple formulaic methods to determine the sample size are known. Various rules of thumb have been proposed, for example, ten times the number of free parameters (Blunch, 2013). Since these heuristics were often found to be inadequate (Wolf et al., 2013), Monte Carlo simulations such as those proposed by Muthén and Muthén (2002) are now considered the gold standard, but they can be highly resource-demanding and computationally complex (Barrett, 2007).

A sample of 250 would yield a power of .990 in multiple linear regression (see SMB1) and recent models suggest that correlation coefficient estimates tend to stabilise when the sample size reaches an order of magnitude of about 250 participants (Schönbrodt & Perugini, 2013; Schönbrodt & Perugini, 2018). Therefore, a sample size of 250 participants per vignette was deemed adequate for the present study.

At the end of data collection, the data file comprised a total of 539 subjects, split approximately in half between the two conditions. Among these subjects, 40 did not provide consent or dropped out before completing the questionnaire. After their removal, the resulting sample consisted of 499 participants.

Data Exclusions

Of these 499 participants, 23 answered the check questions incorrectly; they were removed from the sample³⁷. Next, an analysis of the actual duration of the questionnaire was conducted. Pre-testing suggested an average length of about fifteen minutes.

³⁷ Although the check questions were implemented as a screening tool to improve data quality, the removal from the sample of the participants who did not answer those questions correctly was not preregistered.

However, several participants completed it in a much shorter time. Further testing showed that the minimum time necessary to quickly read the questions and answer them was about eight minutes; 51 participants appeared to have completed the survey in less than eight minutes, a time that was deemed insufficient to adequately process the information and provide reliable answers to the questions. Therefore, those participants were removed from the sample³⁸.

With these 74 exclusions, the remaining sample size was 425.

Multivariate Outliers

At that point, following the preregistered analysis plan, outlier analysis was conducted, applying the same methodology as in Study 1, which used the combined analysis of centred leverage values, Mahalanobis distance, and Cook's distance. Analysis of the first two values provided information about the data points farthest from the centroid of the predictors' multivariate space, while analysis of the third value complemented it with information about the data points with the greatest influence on the prediction of the outcome variables. The data points that exceeded two out of three of the cutoff points determined by the centred leverage values, Mahalanobis distance, and Cook's distance were considered extreme multivariate outliers. This analysis enabled the detection of 13 outliers (for the most part highly influential), which were subsequently deleted from the sample.

Final Sample Composition

The final sample was composed of 412 participants: 207 respondents allocated to Francia's vignette and 205 to Nicholas's vignette. It comprised 211 females (51.2%), 200

³⁸ Although the expected average duration of the questionnaire was preregistered with the intention to remove participants who completed it in too short a time, the preregistration did not explicitly mention the planned exclusion of these participants, focusing only on outliers.

males (48.5%), and one participant who self-reported “other” (non-binary) to the gender question (0.2%). Age ranged from 19 to 77 years, with median of 37 and mean of 39 years ($SD = 12$). Participants were nearly entirely US nationals (98.1%). There was no significant difference in the demographical composition of the sample across the two conditions (see SMB3).

Materials and Procedure

The study had received prior ethical approval by the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University (UK). Participants were recruited through the MTurk crowdsourcing platform and invited to follow a link to an online questionnaire. They read the participation sheet and the consent form, learning that they could withdraw at any time without giving a reason. After providing informed consent to voluntarily take part in the study, participants answered the socio-demographic questions, the moral self-evaluation question, the characteristic adaptations scales, and then viewed one vignette randomly selected from a set of two (Francia or Nicholas).

Next, participants in either condition evaluated the moral action and the moral agent in the vignette; lastly, they rated the items of the moral self-regulation and affect inventory, and were given the option to make comments in an open-ended question. Following this, they were debriefed and thanked for their participation, and received payment after the provision of their unique survey code.

Measures

Socio-Demographic Questions

The first part of the questionnaire included a few socio-demographic questions, such as age, gender, nationality, level of religiosity and spirituality, political orientation on social and economic matters. Participants answered through scales ranging from 0 to

100, using slider bars. The questions were asked as in Study 1, with the only exception that religiosity and spirituality were split into two distinct questions following the previous results.

Characteristic Adaptations

Self-Esteem. Participants filled out the Rosenberg self-esteem scale (Rosenberg, 1965), a ten-item measure of global self-worth, with five items relating to positive views and five items to negative views of the self-concept (Tab. B1).

Table B1:
Rosenberg self-esteem items (Rosenberg, 1965)

Item code	Label	Item wording
Es1	Worth	I feel that I am a person of worth, at least on an equal plane with others
Es2	Qualities	I feel that I have a number of good qualities
Es3	Able	I am able to do things as well as most other people
Es4	Positive	I take a positive attitude toward myself
Es5	Satisfied	On the whole, I am satisfied with myself
Es6	Failure	All in all, I am inclined to feel that I am a failure (<i>R</i>)
Es7	NoProud	I feel I do not have much to be proud of (<i>R</i>)
Es8	NoRespect	I wish I could have more respect for myself (<i>R</i>)
Es9	Useless	I certainly feel useless at times (<i>R</i>)
Es10	NoGood	At times I think I am no good at all (<i>R</i>)

Note: (*R*) reverse-coded items

The items were presented in randomised order. Participants answered through scales ranging from 0 to 100, using slider bars. The anchor points were the same as in the original scale: 0 = *strongly disagree*, 33 = *disagree*, 67 = *agree*, 100 = *strongly agree*.

Regulatory Focus. The next set of questions measured trait regulatory focus through the original regulatory focus questionnaire by Higgins and colleagues (2001). This is an eleven-item scale that measures the two constructs of promotion and prevention focus, respectively with six and five items each (Tab. B2).

Table B2:
Regulatory focus items (Higgins, 2001)

Scale	Item code	Label	Item wording
PROMOTION FOCUS	Pm1	Unable	Compared to most people, are you typically unable to get what you want out of life? (<i>R</i>)
	Pm2	Psyched	How often have you accomplished things that got you “psyched” to work even harder?
	Pm3	DoWell	Do you often do well at different things that you try?
	Pm4	NoPerform	When it comes to achieving things that are important to me, I find that I don't perform as well as I ideally would like to do (<i>R</i>)
	Pm5	Progress	I feel like I have made progress toward being successful in my life
	Pm6	NoHobby	I have found very few hobbies or activities in my life that capture my interest or motivate me to put effort into them (<i>R</i>)
PREVENTION FOCUS	Pv1	Cross	Growing up, would you ever “cross the line” by doing things that your parents would not tolerate? (<i>R</i>)
	Pv2	Nerves	Did you get on your parents' nerves often when you were growing up? (<i>R</i>)
	Pv3	Rules	How often did you obey rules and regulations that were established by your parents?
	Pv4	Object	Growing up, did you ever act in ways that your parents thought were objectionable? (<i>R</i>)
	Pv5	Trouble	Not being careful enough has gotten me into trouble at times (<i>R</i>)

Note: (*R*) reverse-coded items

Participants answered using unipolar scales ranging from 0 to 100 (with slider bars), whose anchor points were the same as in the original scales, for example: 0 = *never or seldom*, 25 = *a few times*, 50 = *sometimes*, 75 = *often*, 100 = *very often*. The eleven items were presented in randomised order.

In addition to the original questionnaire, regulatory focus items from other existing scales were used³⁹ (Tab. B3). Four items were taken from the general regulatory focus measure (Lockwood et al., 2002), two items were borrowed from the regulatory focus strategy scale (Ouschan et al., 2007), two items came from the regulatory focus scale (Fellner et al., 2007), and two new regulatory focus items were added with the aim of

³⁹ The extra regulatory focus items were added for exploratory purposes in relation to a parallel project and are not reported in this thesis. They were selected from different existing scales to expand the coverage of dimensions of the construct only partly captured by the original regulatory focus questionnaire (further details in this chapter's Introduction).

capturing concerns about errors of commission versus omission (Crowe & Higgins, 1997).

Table B3:
Additional regulatory focus items and their sources

Scale	Source	Item code	Item wording
PROMOTION FOCUS	General Regulatory Focus Measure (Lockwood et al., 2002)	PmEx1	I see myself as someone who is primarily striving to become the person I ideally want to be, fulfilling my hopes and aspirations
		PmEx2	I frequently imagine how I will accomplish my dreams and ideals
	Regulatory Focus Strategy Scale (Ouschan et al., 2007)	PmEx3	The worst thing you can do when trying to achieve a goal is to worry about making mistakes
	Regulatory Focus Scale (Fellner et al., 2007)	PmEx4	I like trying out lots of different things, and am often successful in doing so
	Newly developed	PmEx5	I would prefer to miss a target rather than do nothing at all
PREVENTION FOCUS	General Regulatory Focus Measure (Lockwood et al., 2002)	PvEx1	I see myself as someone who is primarily striving to become the person I ought to be, fulfilling my duties and obligations
		PvEx2	I frequently think about how I can prevent failures in my life
	Regulatory Focus Strategy Scale (Ouschan et al., 2007)	PvEx3	To avoid failure, it is important to keep in mind all the potential obstacles that might get in your way
	Regulatory Focus Scale (Fellner et al., 2007)	PvEx4	I always try to make my work as accurate and error-free as possible
	Newly developed	PvEx5	I would rather do nothing than make a mistake

To answer these ten additional items, participants used scales ranging from 0 to 100 (with slider bars) with the following anchor points: 0 = *strongly disagree*, 25 = *disagree*, 50 = *neither agree nor disagree*, 75 = *agree*, 100 = *strongly agree*. The ten items were presented in randomised order.

Hedonic Orientation (Approach and Avoidance Temperaments). Next, participants answered the twelve questions from the approach and avoidance temperament questionnaire (Elliot & Thrash, 2010), chosen as the hedonic orientation measure. Each subscale is composed of six items. These were presented in randomised order and

participants answered using scales ranging from 0 to 100 (with slider bars), whose anchor points were the same as in the original scales: 0 = *strongly disagree*, 25 = *disagree*, 50 = *neither agree nor disagree*, 75 = *agree*, 100 = *strongly agree* (Tab B4).

Table B4:

Approach and avoidance temperament items (Elliot & Thrash, 2010)

Scale	Item code	Label	Item wording
APPROACH	Ap1	Energised	Thinking about the things I want really energizes me
	Ap2	Excited	When I see an opportunity for something I like, I immediately get excited
	Ap3	Motivated	It doesn't take a lot to get me excited and motivated
	Ap4	Opportunities	I'm always on the lookout for positive opportunities and experiences
	Ap5	GoodThings	When good things happen to me, it affects me very strongly
	Ap6	Desire	When I want something, I feel a strong desire to go after it
AVOIDANCE	Av1	Nervous	By nature, I am a very nervous person
	Av2	Worry	It doesn't take much to make me worry
	Av3	Anxiety	I feel anxiety and fear very deeply
	Av4	BadExperiences	I react very strongly to bad experiences
	Av5	Escape	When it looks like something bad could happen, I have a strong urge to escape
	Av6	ImagineBad	It is easy for me to imagine bad things that might happen to me

Importantly, all the characteristic adaptation measures were gathered in the initial part of the questionnaire, before presenting participants with the vignettes; this ensured they were unbiased by the moral scenarios and could be modelled as exogenous variables predicting comparative and self-regulatory processes.

Moral Appraisals

The moral self-evaluation question was asked in the first part of the questionnaire, alongside the socio-demographic variables. It was measured exactly as in Study 1.

Next, after viewing one of two vignettes (Francia or Nicholas), participants answered a set of moral appraisal questions, evaluating the story and its protagonist. As in Study 1, they were asked to evaluate the moral character of the agent (agent evaluation) and the perceived level of goodness and propriety of the moral action.

Participants were also asked two new questions about the perceived normativity of the deed through the questions “To what extent do you think most people would consider this act a good deed?” (normative judgment) and “If they had a chance, to what extent do you think most people would do the same?” (normative behaviour). All these moral evaluations were measured through unipolar scales ranging from 0 to 100 (0 = *not at all*, 100 = *very much*) using slider bars.

As in Study 1, moral discrepancy was computed to measure indirect moral comparisons based on ability. Positive values indicated upward comparisons (agent evaluation higher than the participant’s self-evaluation), whereas negative values indicated downward comparisons (agent evaluation lower than the participant’s self-evaluation); zero indicated lateral comparisons (agent and self at parity).

Moral Self-Regulation

Following the moral appraisals, participants were asked to fill out the moral self-regulation inventory, rating how true they believed a set of items were in relation to the vignette that they viewed. In Study 2, the *full* item list was presented in randomised order: seven self-improvement items and thirteen self-defence items. Participants answered using a 0-100 unipolar scale with slider bars, where 0 = *not at all true* and 100 = *very true*. The item lists for moral self-improvement and moral self-defence are shown respectively in Tables B5-B6.

Table B5:*Moral self-improvement items tested in Study 2 (Francia's vignette)*

Code	Label	Item wording
SI1	Admirable	Actions like this are truly admirable
SI2	Awakened	When I read these stories, I feel awakened to the good in the world
SI3	Humanity	This story strengthens my faith in humanity
SI4	BeBetter	Francia has shown me how to be a better person
SI5	Values	Francia and I share the same values
SI6	BeLike	I want to be more like Francia
SI7	ForOthers	I feel like I want to do something good for others

Table B6:*Moral self-defence items tested in Study 2 (Francia's vignette)*

Code	Label	Item wording
SD1	MeGreater	In many ways, I have done greater deeds than Francia
SD2	Remember	Reading about Francia's good deed makes me think about all the good deeds I've done for others
SD3	Devalue	This is not an act that I value all that much
SD4	Ordinary	It's not such an extraordinary action
SD5	People	I know people who have done greater deeds than Francia
SD6	Praise	Francia may have done a good deed, but I bet she is seeking the praise of others
SD7	Ulterior	Francia may have had ulterior motives for doing this
SD8	Uncomfortable	It makes me uncomfortable to dwell on these stories
SD9	Uneasy	I would feel uneasy if I had to interact with Francia
SD10	Uncommon	Actions uncommon as this one should not be considered the standard we live by
SD11	NoPraise	Everyone occasionally does something really good, so Francia isn't more praiseworthy than anybody else
SD12	Superior	Francia probably thinks she's better than everyone else
SD13	Seriously	Francia takes herself too seriously

The items of these two scales could be divided into two broad classes:

- *cognitions* related to the moral action, for example self-improvement item SI1/Admirable and self-defence item SD4/Ordinary, or related to the moral agent, for example self-improvement item SI5/Values and self-defence item SD7/Ulterior;
- *behavioural tendencies* (or action readiness/preparedness), for instance self-improvement item SI7/ForOthers and self-defence item SD9/Uneasy.

Compared to the scales used in Study 1, one new self-improvement item (SI6) and eight new self-defence items (SD2, SD3, SD5, SD7, SD8, SD9, SD11, and SD13) were tested in Study 2. Of the self-defence items already used in Study 1, two were slightly

reworded (SD6 and SD10) based on the findings. From a theoretical standpoint, most of the self-defence items fall under the wider umbrella of self-protection, while for example items SD1, SD2, and SD5 directly or indirectly reflect self-enhancement mechanisms.

Moral Affect

Following the moral self-regulation measures, participants were asked to rate how true they believed a set of affect statements were in relation to how they felt after viewing the vignettes. As in Study 1, the items were divided into the two scales of positive and negative affect. They were presented in randomised order and measured through unipolar 0-100 scales using slider bars, where 0 = *not at all true* and 100 = *very true*.

The range of the subjective feelings measured by the moral affect scales tested in Study 2 was broader than that used in Study 1: it included six positive affect items and eight negative affect items. Compared to the set used in Study 1, the positive affect scale had three new items (PA3, PA4, and PA5) and the negative affect scale had three new items as well (NA2, NA3, and NA8). The wording of item PA2 in the positive affect scale was slightly modified following the results of Study 1. The full list of moral affect items is shown in Tables B7-B8.

Table B7:*Positive moral affect items tested in Study 2 (Francia's vignette)*

Code	Label	Item wording
PA1	Uplifted	I felt uplifted
PA2	Inspired	I was inspired by the story
PA3	Moved	I was moved
PA4	Proud	I felt proud of what Francia did
PA5	Happy	It made me feel happy
PA6	Challenged	I felt challenged in a positive way

Table B8:*Negative moral affect items tested Study 2*

Code	Label	Item wording
NA1	Conflicted	I had a mix of conflicting feelings
NA2	Detached	I felt detached
NA3	Vulnerable	I felt vulnerable
NA4	Threatened	I felt as if I was threatened by something
NA5	Guilty	It made me feel guilty
NA6	Envious	To be honest, I felt envious
NA7	Resentful	I felt resentful
NA8	Irritated	The story irritated me

Analytic Approach

The dataset was analysed using R ver. 3.6, RStudio ver. 1.2⁴⁰, IBM SPSS Statistics ver. 25-26, the PROCESS macro ver. 3.3 (Hayes, 2018), and the Monte Carlo PCA for Parallel Analysis application ver. 2.5 (Watkins, 2006).

For Study 2 various analysis techniques, primarily descriptive and exploratory, were employed: outlier analysis to identify extreme data points, factor analysis to explore the dimensionality of the moral self-regulation and affect inventory, regression and correlation analysis to estimate importance, type of relationship, and effect size of critical predictors (especially the characteristic adaptations), moderation analysis to test the interactions hypothesised in the conceptual model, path modelling to further refine

⁴⁰ Specific R packages used for analysis and visualisation are referenced in text in the following sections.

the models, and structural invariance analysis to determine if one path model was adequate for both vignettes.

Results and Preliminary Reflections

Descriptive Statistics

In Study 2 there were no missing data, except for eleven participants who did not answer the political orientation questions (variables that are not part of the conceptual model). The descriptive statistics of the main socio-demographic variables are reported in Table B9. No significant difference was detected between the two vignettes (SMB3).

Table B9:

Descriptive statistics of the main socio-demographic variables (total sample)

Variable	Mean	Standard Deviation	Bootstrap		
			Standard Error	95% Lower Limit C.I.	95% Upper Limit C.I.
Religiosity	31.68	36.41	1.74	27.96	34.78
Spirituality	47.37	37.04	1.80	43.71	50.85
Political Orientation (Social Matters)	38.16	31.96	1.56	35.06	41.02
Political Orientation (Economic Matters)	45.75	32.70	1.61	42.65	49.12

Characteristic Adaptations

Self-Esteem. The descriptive statistics for the items of the Rosenberg self-esteem scale revealed a wide range of values. The mean scores of the items relative to negative aspects of the self-concept were lower than the mid-point of the scale; conversely, the mean scores of the items relative to positive aspects of the self-concept were higher than the mid-point. As expected, the negative items exhibited positive skewness and the positive items negative skewness; however, these values were not overly problematic. Across all the ten items, the distributions also showed limited kurtosis (not critical).

After reverse-coding the negative items, the distribution of the composite mean of the Rosenberg self-esteem scale showed a long lower tail and was not normally distributed; the mean was 69.96 ($SD = 19.89$), skewness -0.47 and kurtosis -0.44. No

significant differences were observed in the self-esteem scale across the two vignettes (for details, see SMB4).

Regulatory Focus. The descriptive statistics for the dispositional regulatory focus scales revealed satisfactory ranges. After reverse-coding items and taking the average, the distribution of the promotion focus scale was approximately normal, with mean 63.37 ($SD = 16.00$), skewness 0.02 and kurtosis -0.41. The distribution of the prevention focus scale was not normal, with mean 61.97 ($SD = 19.54$), skewness -0.22 and kurtosis -0.49. No significant differences were observed in the promotion and prevention focus scales across the two vignettes (see SMB4 for details).

Hedonic Orientation. The descriptive statistics for the approach and avoidance temperament scales revealed wide ranges for all the items. After averaging the corresponding items, the distribution of the approach scale showed a long lower tail (skewness -0.85) and was moderately pointy (kurtosis 1.16), with mean 64.25 ($SD = 17.30$). The distribution of the avoidance scale was nearly symmetric (skewness 0.05) but platykurtic (kurtosis -0.96), with mean 49.90 ($SD = 24.57$). Both scales were not normally distributed. No significant differences were observed in the approach and avoidance scales across the two vignettes (details in SMB4).

Moral Appraisals

The means and the distributions of the moral appraisal variables exhibited the same patterns as in Study 1. Participants rated their own morality very highly: the distribution of the moral self-evaluation presented a long lower tail, with mean 80.93 ($SD = 15.61$), in line with the value measured in Study 1.

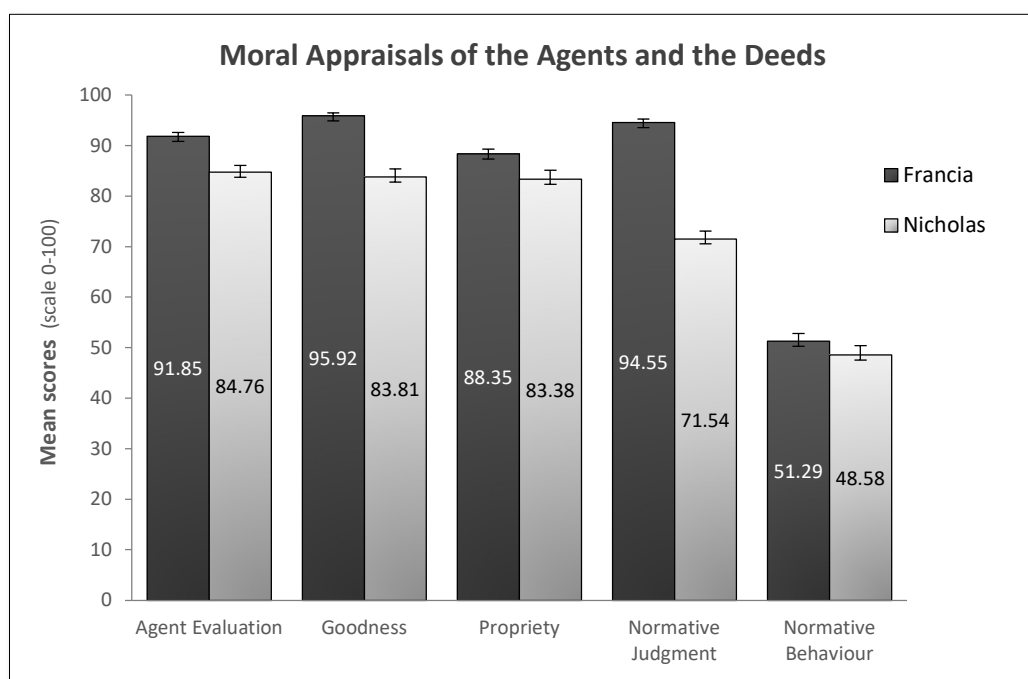
The distributions of the agent evaluation, goodness and propriety of the deed were also asymmetric, with high mean scores and long lower tails. Only the variable normative behaviour had a mean score at approximately the mid-point of the scale for

both vignettes and was more symmetric (and slightly platykurtic). All the moral appraisal variables were not normally distributed.

The mean scores of the moral appraisals of the agents and the deeds for the two vignettes are illustrated in Figure B2 (full details in SMB5).

Figure B2:

Mean scores for the moral appraisals for the two vignettes, with bootstrap standard errors (BCa 95% CI)



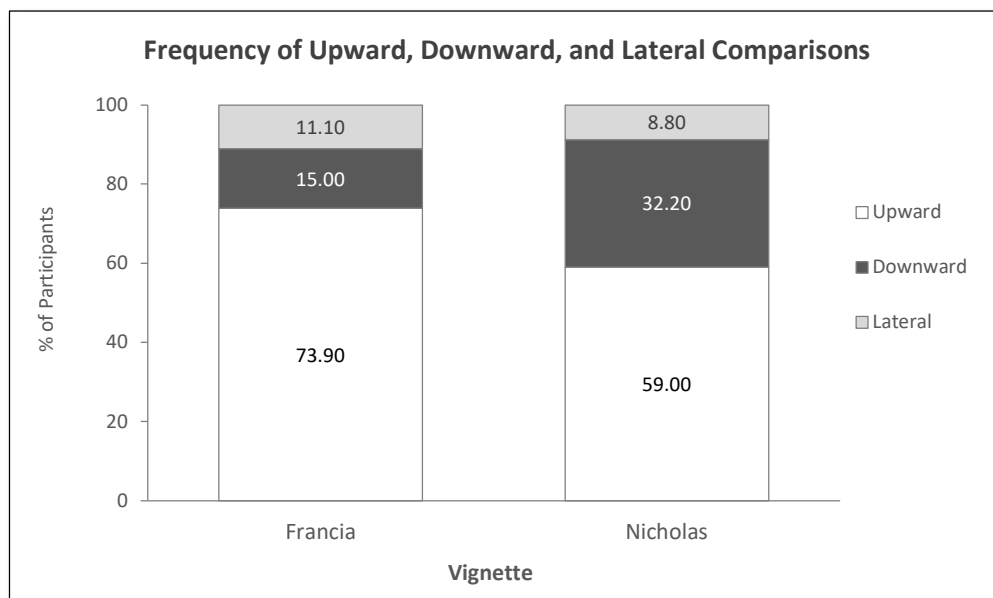
With regard to the moral discrepancy, for both vignettes the values were positive (although not far from zero), indicating an overall upward moral comparison. Francia's mean score was 10.93, whereas Nicholas's was 3.82 (significantly lower, see SMB5).

Another way to look at the moral comparisons is to examine the frequencies of the three types of comparisons for each of the two vignettes, that is, the percentage of participants who engaged in upward, downward, and lateral comparisons. The results of this analysis are displayed in Figure B3: for both vignettes, upward comparison was

predominant, but in Francia's it did so to a greater extent than in Nicholas's, where downward comparison was significantly more marked due to the more mixed appraisal of the moral character of the agent (for detailed results of a chi-square test, including contingency tables with standardised residuals, see SMB5). Again, this pattern was consistent with the findings from Study 1.

Figure B3:

Types of moral comparisons for the two vignettes: clustered bar chart of frequency distributions (%)



Moral Self-Regulation: Item and Scale Analysis

The moral self-regulation items showed for both vignettes negative skewness for the self-improvement indicators and positive skewness for the self-defence indicators. A few items across both scales also exhibited high kurtosis and all of them violated the assumption of normality (details in SMB7).

Overall, the moral self-regulation scales showed satisfactory internal consistency. At total sample level, Cronbach's alpha was above .90 for both the self-improvement

and self-defence scales; within each vignette, Cronbach's alpha was excellent for Nicholas and satisfactory for Francia (Tab. B9).

The seven-item self-improvement scale at total sample level could improve by removing item SI7/ForOthers, which did not correlate well with the scale, especially in Nicholas's vignette (Tab. B9). However, elimination of this item would cause the reliability of Francia's scale, which was already lower, to drop further down, which was not desirable. At this stage, the existing seven-item scale was deemed to be an appropriate solution for the self-improvement scale for both vignettes.

The thirteen-item self-defence scale performed much better than in Study 1, when only six items were tested among a smaller sample. Its internal consistency could even improve for both vignettes by removing item SD2/Remember (Tab. B10), which correlated less strongly with the scale; SD2 was therefore dropped from the self-defence scale. Details of reliability analysis for the self-regulation scales are available in SMB10.

Table B10:

Possible improvements in internal consistency (Cronbach's alpha) for the self-improvement and self-defence scales (total sample and by vignette)

Vignette	Moral self-regulation scales			
	Self-improvement		Self-defence	
	original scale with 7 items	reduced scale if any item deleted	original scale with 13 items	reduced scale if any item deleted
Total sample	.902	.910 (SI7/ForOthers)	.907	.921 (SD2/Remember)
Francia	.878	-	.901	.914 (SD2/Remember)
Nicholas	.915	.933 (SI7/ForOthers)	.911	.925 (SD2/Remember)

Note: in parenthesis the items that, if removed, would improve Cronbach's alpha.

Deletion of item SD2/Remember had implications in terms of the expected dimensionality of the self-defence scale. Item SD2 was one of the items developed to capture self-enhancing mechanisms, alongside SD1/McGreater and possibly others, such as SD5/People. In theory, these items – depending on participants' response – could aggregate to form a *self-enhancement* factor, distinct from a *self-protection* factor that would aggregate the remaining items. Reliability analysis revealed two insights: a) item SD2 did not correlate strongly with the self-defence scale; b) SD1 and SD5 (the other possible self-enhancement items) correlated strongly with the self-defence scale (see SMB10). Consequently, once item SD2 was dropped, a bidimensional structure for the self-defence scale was no longer likely. In fact, *all* the twelve remaining items correlated moderately or strongly with the scale (except for SD10 and a few others, which needed further evaluation), suggesting one single factor. In addition, only two self-enhancement

items would be too few to form a reliable factor (the literature suggests that at least three items should aggregate to form a robust factor⁴¹).

From a theoretical standpoint, Alicke and Sedikides (2009) acknowledge that it is often difficult in practice to clearly distinguish between self-enhancing and self-protecting processes in the absence of a baseline level of functioning indicating levels of aspiration (for self-enhancement) and tolerance (for self-protection). Moreover, they point out that sometimes self-enhancement may serve self-protection interests, as it is often the case with narcissists (Alicke & Sedikides, 2009). To complement these theoretical considerations and the previous empirical findings, the dimensionality of the self-defence scale was tested through exploratory factor analysis, whose results are presented later in this section.

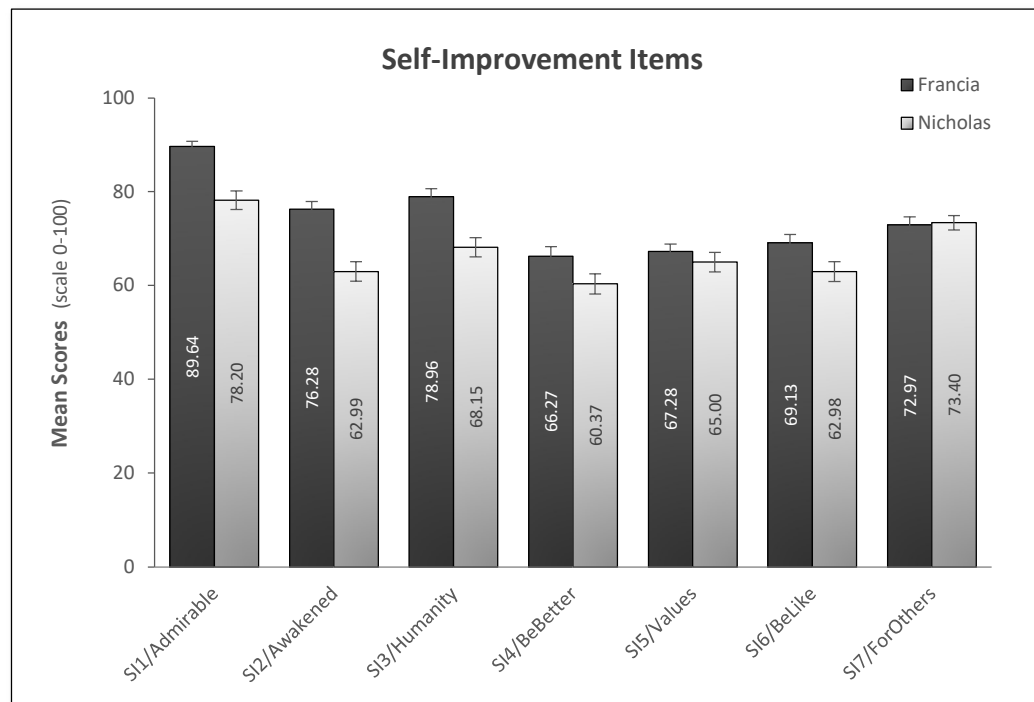
Prior to that, the reconfigured scales (self-improvement with seven items and self-defence with twelve) were analysed in terms of distributions and biases: clear asymmetries and tailedness were observed, and both distributions violated assumptions of normality at total sample as well as at individual vignette level. The descriptive statistics, normality tests, histograms with the full distributions, and probability plots are available in SMB6.

The composite mean scores of all seven self-improvement items for the two vignettes are displayed in Figure B4: all of them were strongly endorsed by participants, particularly in Francia's vignette, which overall generated more self-improvement than Nicholas's, consistent with the pattern that emerged in Study 1.

⁴¹ Costello and Osborne (2005) recommend an even stricter standard, with at least five items per factor, but the more generally accepted criterion is a minimum of three items per factor, provided the sample size is large enough and the loadings are above .400 (Kline, 2016).

Figure B4:

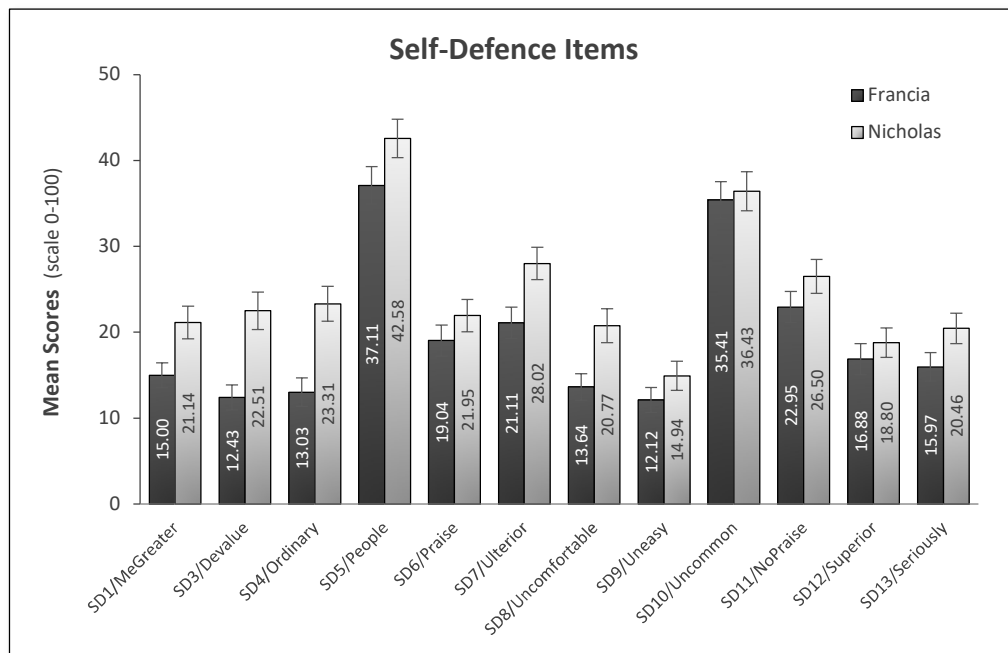
Mean scores of the self-improvement items for the two vignettes, with bootstrap standard errors (BCa 95% CI). All original 7 items retained



The composite mean scores of the twelve retained self-defence items for the two vignettes are displayed in Figure B5: overall, they were less strongly endorsed by participants than the self-improvement items; yet, self-defence processes appeared sizeable, particularly in Nicholas's vignette, which elicited more self-defence than Francia's. More detailed descriptive statistics are available in SMB7.

Figure B5:

Mean scores of the twelve retained self-defence items for the two vignettes, with bootstrap standard errors (BCa 95% CI). Item SD2/Remember removed



Moral Affect: Item and Scale Analysis

Analysis of the individual items of the moral affect scales revealed patterns similar to those observed in Study 1: negative skewness for measures of positive affect, positive skewness for measures of negative affect. Considerable kurtosis recurred across several items. The assumption of normality was violated across all items for both vignettes (details in SMB9).

Reliability analysis of the positive and negative affect scales (six and eight items respectively) was carried out. At total sample level, both scales showed a satisfactory degree of internal consistency (better for Francia than for Nicholas). However, two items did not perform well (see Tab. B11). Cronbach's alpha would improve for the positive affect scale with the removal of item PA6/Challenged, which did not perform well in Study 1 either and was found again here to not correlate strongly with the scale.

Cronbach's alpha would also slightly improve for the negative affect scale with the removal of item NA6/Envious. Therefore, items PA6 and NA6 were dropped from their respective scales. More detailed reliability analyses are presented in SMB11.

Table B11:

Possible improvements in internal consistency (Cronbach's alpha) for the positive and negative affect scales (total sample and by vignette)

Vignette	Moral affect scales			
	Positive affect		Negative affect	
	original scale with 6 items	reduced scale if any item deleted	original scale with 8 items	reduced scale if any item deleted
Total sample	.935	.944 (PA6/Challenged)	.852	.857 (NA6/Envious)
Francia	.913	.931 (PA6/Challenged)	.885	.889 (NA6/Envious)
Nicholas	.945	.949 (PA6/Challenged)	.819	.829 (NA6/Envious)

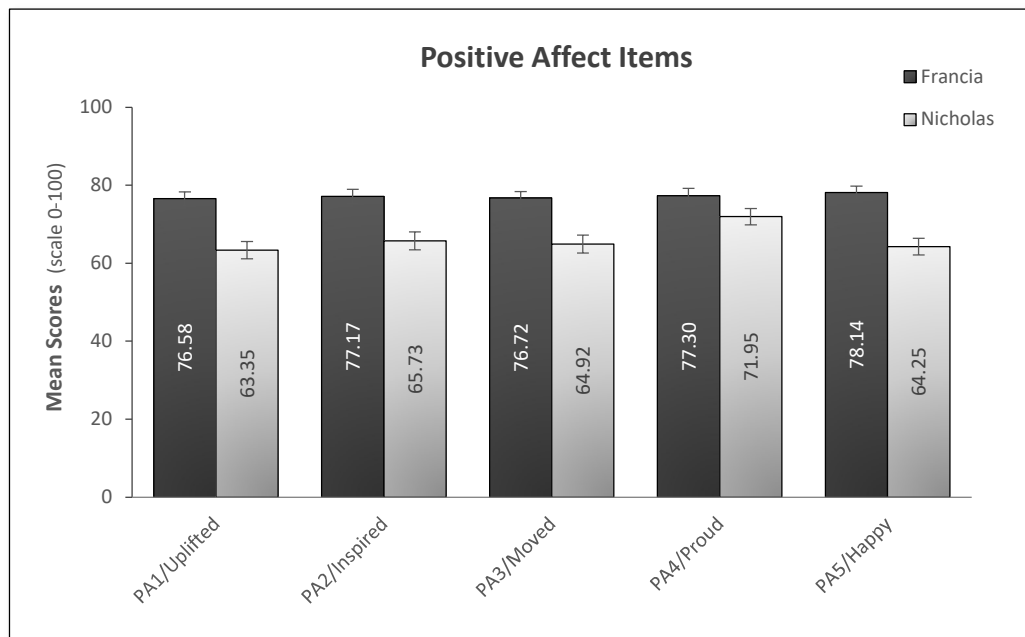
Note: In parenthesis the items that, if removed, would improve Cronbach's alpha.

The reconfigured scales (positive affect with five items and negative affect with seven) were analysed in terms of distributions and biases: clear asymmetries and tailedness were observed, and both distributions violated assumptions of normality at total sample as well as at individual vignette level. The descriptive statistics, normality tests, histograms with the full distributions, and probability plots are available in SMB8.

The composite mean scores of the five retained positive affect items for the two vignettes are displayed in Figure B6: all of them were strongly endorsed by participants, particularly in Francia's vignette, which overall generated more positive affect than Nicholas's, consistent with the pattern from Study 1.

Figure B6:

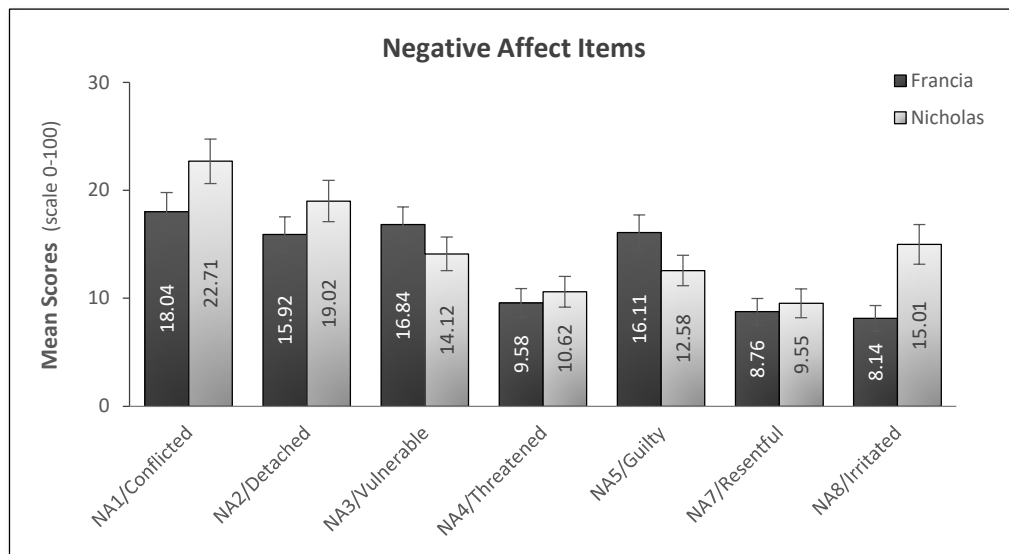
Mean scores of the five retained positive affect items for the two vignettes, with bootstrap standard errors (BCa 95% CI). Item PA6/Challenged removed



The composite mean scores of the seven retained negative affect items for the two vignettes are displayed in Figure B7: overall, they were less strongly endorsed by participants than the positive affect items; yet, negative affective states were substantial, particularly in Nicholas's vignette, which elicited them more strongly than Francia's vignette. When comparing the two vignettes, on average two items scored in the opposite direction relative to all the others: NA3/Vulnerable and NA5/Guilty were endorsed more for Francia's than for Nicholas's vignette. These items were carefully considered in terms of their inter-correlations and communalities. More detailed descriptive statistics are available in SMB9.

Figure B7:

Mean scores of the seven retained negative affect items for the two vignettes, with bootstrap standard errors (BCa 95% CI). Item NA6/Envious removed



Correlations Between the Main Variables

Correlation analysis was carried out, as in Study 1, by computing zero-order Pearson's coefficients across characteristic adaptations, moral discrepancy, moral self-regulation, and moral affect variables (see summaries in Tab. B12-B13, and the full output in SMB12-SMB13).

Among the total sample, Study 2 found evidence of a weak positive correlation between promotion and prevention focus, and a weak negative correlation between approach and avoidance temperaments (Tab. B12). Promotion focus was strongly positively correlated with approach, while prevention focus was weakly negatively correlated with avoidance. Quite striking were also the associations of self-esteem with approach, promotion, and prevention (all positive), and with avoidance (negative).

Both approach and promotion focus correlated positively with self-improvement, whereas self-defence was associated positively with avoidance and negatively with prevention and promotion focus.

These results suggested the possible *redundancy* of self-esteem in the conceptual model: given the above-mentioned strong correlation patterns, dropping the construct could potentially enable a higher degree of parsimony without significant loss of information. Results also suggested the possible *substitutability* of approach and promotion focus for the prediction of moral discrepancy and self-improvement: indeed, approach and promotion correlated strongly with each other, and both predicted moral discrepancy and self-improvement (approach more strongly associated with self-improvement and promotion focus with moral discrepancy). Based on this pattern, it would be plausible to hypothesise a parsimonious model whereby just one of them (e.g., promotion focus) predicted both moral discrepancy and self-improvement.

Table B12:

Zero-order correlations across characteristic adaptations, moral discrepancy, and moral self-regulation (total sample)

		2.	3.	4.	5.	6.	7.	8.
1. Promotion focus	Correlation	.157**	.551***	-.509***	.709***	-.163**	.180***	-.164
	Sig. (2-tailed)	.001	<.001	<.001	<.001	.001	<.001	.001
2. Prevention focus	Correlation	1	.000	-.153**	.205***	-.128**	.015	-.217***
	Sig. (2-tailed)		.995	.002	<.001	.009	.756	<.001
3. Approach	Correlation		1	-.100	.433***	-.140**	.243***	.003
	Sig. (2-tailed)			.043*	<.001	.005	<.001	.954
4. Avoidance	Correlation			1	-.606***	.188***	.032	.100
	Sig. (2-tailed)				<.001	<.001	.524	.042*
5. Self-esteem	Correlation				1	-.240***	.087	-.081
	Sig. (2-tailed)					<.001	.079	.101
6. Moral discrepancy	Correlation					1	.363***	-.441***
	Sig. (2-tailed)						<.001	<.001
7. Self-improvement	Correlation						1	-.488***
	Sig. (2-tailed)							<.001
8. Self-defence	Correlation							1
	Sig. (2-tailed)							

Note: * $p < .05$ (2-tailed) ** $p < .01$ (2-tailed) *** $p < .001$ (2-tailed).

Among the total sample, moral discrepancy correlated negatively with approach, promotion, prevention, and self-esteem, and positively with avoidance. It was also positively associated with self-improvement and negatively associated with self-defence, replicating the findings from Study 1.

When examining the correlations between moral discrepancy and self-regulation within each vignette, Nicholas reproduced the same pattern as the total sample, while Francia was partly inconsistent (no significant correlation was detected between moral discrepancy and self-improvement; see SMB13 for further details). The scatter plots in Figures B8a-b⁴² illustrate graphically the distributions of the moral evaluations and the corresponding self-regulatory modes for Francia's vignette (also fitting the regression lines). Panel "a" of the Figure shows that participants self-improved more or less strongly regardless of whether they engaged in upward or downward comparisons. However, in this sample, the number of participants who engaged in downward comparison was quite small (only 31), so stronger evidence would be desirable. It should also be remembered that these are only zero-order correlations in a bivariate system; path modelling can determine regression weights with higher levels of accuracy within the context of a wider model with multiple variables (see next).

⁴² The plots were created using the R package ggplot2 ver. 3.3.2 (Wickham, 2016).

Figures B8a-b:

Relations between moral discrepancy and the two modes of moral self-regulation in Francia's vignette: scatter plots with linear regression lines



Regarding the relationships between self-regulation and affect, results from the correlation analysis were consistent with those from Study 1: positive correlation between self-improvement and positive affect, as well as between self-defence and negative affect (Tab. B13). Whereas the correlation between self-improvement and positive affect was high already in Study 1, the correlation between self-defence and negative affect was stronger in the present study ($r = .692, p < .001$ at total sample) than in Study 1 ($r = .285, p < .001$ at total sample), when only a few items were tested. The same pattern was consistently observed in both vignettes: Francia $r = .740, p < .001$, Nicholas $r = .663, p < .001$ (see SMB13 for details).

Table B13:*Zero-order correlation between moral self-regulation and moral affect (total sample)*

		Positive affect	Negative affect
Self-improvement	Pearson Correlation	.874***	-.238***
	<i>Sig. (2-tailed)</i>	<.001	<.001
Self-Defence	Pearson Correlation	-.513***	.692***
	<i>Sig. (2-tailed)</i>	<.001	<.001

Note: *** $p < .001$ (2-tailed).

c. Bootstrap results are based on 1000 bootstrap samples.

Because these correlation coefficients were very high (about .7 or higher in absolute value), it was legitimate to carefully consider to what extent these measures assess truly distinct constructs. In light of this, a revised configuration of the model was contemplated, that is, a solution that:

- conflated indicators of positive cognitions/action tendencies and positive affect into a broader moral cognitive/conative/affective self-improvement scale;
- conflated indicators of negative cognitions/action tendencies and negative affect into a broader moral cognitive/conative/affective self-defence scale.

To empirically test this modified configuration of the measures of moral self-regulation, exploratory factor analysis was conducted.

Moral Self-Regulation Scale Dimensionality: Exploratory Factor Analysis

To investigate the dimensional structure of the four scales of the moral self-regulation and affect inventory, EFA was conducted. Principal axis factoring was employed to handle non-normal distributions and a Monte Carlo simulation was run using the Monte Carlo PCA for Parallel Analysis application. At total sample level, after aggregating all the 31 cognitive, conative, and affective moral self-regulation indicators, two or three factors emerged from an initial extraction, a parallel analysis, and an oblimin rotation: a) one factor for the twelve cognitive/conative/affective self-

improvement items; b) one or two factors for the nineteen cognitive/conative/affective self-defence items. The three-factor solution was hardly interpretable, due to multiple cross-loadings of equal magnitude; on the contrary, the two-factor solution was neat, with all the nineteen self-defence items aggregating together (explaining 40.7% of the total variance) and loading separately from the twelve self-improvement items (explaining 14.3% of the total variance). At this stage, the factor loadings could be considered acceptable, although with some weaknesses: about half of them were greater than .700, particularly for the self-improvement factor, and all were greater than .400. Table B14 exhibits the factor loadings of the two-factor solution, while further details about the analysis for the total sample can be found in SMB14.

Table B14:

Exploratory factor analysis of the full 31-item moral self-regulation inventory: factor loadings (total sample)

	Factor 1	Factor 2
	Self-defence	Self-improvement
	(variance explained = 40.7%)	(variance explained = 14.3%)
NA4/Threatened	.826	
NA7/Resentful	.819	
SD9/Uneasy	.768	
SD12/Superior	.764	
SD13/Seriously	.716	
SD6/Praise	.702	
SD8/Uncomfortable	.683	
SD4/Ordinary	.668	
SD1/Greater	.668	
NA8/Irritated	.654	
SD3/Devalue	.640	-.313
SD7/Ulterior	.628	
NA3/Vulnerable	.615	
NA1/Conflicted	.605	
SD11/NoPraise	.599	
NA5/Guilty	.563	
NA2/Detached	.508	
SD10/Uncommon	.411	
SD5/People	.406	
PA2/Inspired		.879
SI3/Humanity		.853
PA1/Uplifted		.837
PA3/Moved		.835
PA5/Happy		.832
SI4/BeBetter		.820
SI6/BeLike		.793
SI2/Awakened		.788
PA4/Proud		.752
SI5/Values		.679
SI1/Admirable	-.309	.643
SI7/ForOthers		.518

Note: Extraction method: Principal Axis Factoring.

Rotation method: oblimin with Kaiser normalisation (rotation converged in 7 iterations).

Factor loadings below .3 are not reported in the table.

Results partly held up when the analysis was carried out at the level of the individual vignettes. For Francia, a two-factor solution emerged, closely replicating the pattern observed for the total sample (see SMB15). For Nicholas, a three-factor solution looked slightly more appropriate. The three factors roughly comprised: 1)

cognitive/conative/affective self-improvement, 2) cognitive/conative self-defence, and 3) negative affect. However, one item did not load on any factor, and several cross-loadings of a similar magnitude emerged (see SMB16 for details). Although at a superficial glance it looked mathematically the most appropriate solution, it was not statistically straightforward and substantively interpretable with ease.

At the present stage of the research, it appeared that the measurement model would fit Francia's vignette more comfortably than Nicholas's. The two-factor solution was retained as the lead option for the measurement model, deferring re-evaluation to Study 3 among a larger sample with confirmatory techniques. Hereon, the broader cognitive/conative/affective constructs were used for moral self-improvement and self-defence.

Multiple Linear Regression

Multiple linear regression was carried out to identify the strongest predictors and their effect sizes. The analysis reproduced the same method used in Study 1, but the two criteria (dependent variables) were now the broader cognitive/conative/affective self-improvement and self-defence scales. The crucial insight expected from the analysis was the contribution of the characteristic adaptations to the explanation of the variance of the outcome variables, but also a replication of the role played by moral comparisons based on the critical moral appraisals.

The predictors entered in the forward selection procedure were the following: age, religiosity, spirituality, political orientation⁴³, self-esteem, promotion focus, prevention

⁴³ As in Study 1, the very strong correlation between the variables of political orientation on economic matters and political orientation on social matters ($r = .834$, $p < .001$) suggested to average the two variables to form an overall "political orientation" variable.

focus, approach, avoidance, self-evaluation, agent evaluation, moral discrepancy, goodness and propriety of the deed.

For the total sample, the models explained just above half of the variance of the dependent variables (adjusted R square); at the level of the individual vignettes, the fit was slightly better for Nicholas than for Francia (Tab. B15).

Table B15:

Multiple linear regressions at total sample and by vignette (forward selection): model summaries. Dependent variables: self-improvement/self-defence

Vignette	Dependent variable: self-improvement				Dependent variable: self-defence			
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R	R Square	Adjusted R Square	Std. Error of the Estimate
Tot. sample	.721	.520	.515	15.57001	.717	.514	.501	12.34800
Francia	.634	.402	.393	14.85618	.715	.511	.491	11.78669
Nicholas	.779	.607	.596	15.65991	.748	.560	.546	12.30510

For the total sample, the characteristic adaptations appeared to be significant predictors of moral self-regulation: approach for self-improvement, prevention focus and avoidance for self-defence. A similar outcome emerged at the level of the individual vignettes except for avoidance, which did not appear among the significant predictors of self-defence for Francia.

The substitutability in the conceptual model of approach and promotion focus emerged again: with both variables entered as independent variables in the regression, approach (but not promotion) was a significant predictor of self-improvement; if approach were not entered in the regression, promotion focus would appear as a significant predictor.

As expected, the sign of the coefficients showed positive relationships between approach and self-improvement, and between avoidance and self-defence (Nicholas). However, the results were not aligned with the hypothesis that prevention focus would

be positively correlated with self-defence; in fact, it turned out to be *negatively* associated with it, both at total sample and by individual vignette. Another interesting pattern was the significance of moral discrepancy as a predictor of self-defence only in Nicholas's vignette; as noted earlier on, one possible explanation of this phenomenon is that the forward selection process usually tends to identify fewer significant predictors than other procedures, such as the backward elimination process (Field et al., 2012). Religiosity (but not spirituality) appeared as one of the significant predictors, but with contradictory results: at total sample, it was positively correlated with both self-improvement ($\beta = .116, p = .002$) and self-defence ($\beta = .112, p = .006$). The correlation with self-improvement was driven by Nicholas's subsample ($\beta = .160, p = .004$), whereas the correlation with self-defence was driven by Francia's subsample ($\beta = .198, p < .001$). It is not easy to interpret these findings, since it could have been anticipated that Nicholas's deed (the safeguard of homosexuals' civil rights) and not Francia's deed (organ donation) would have caused moral defensiveness among holders of traditional religious beliefs (consistent with existing research, see e.g., Janssen & Scheepers, 2019). Effectively, in Study 2 conservative political views did correlate positively with self-defence at total sample ($\beta = .166, p < .001$) as well as within each of the two subsamples. As a general remark, it must be noted that all these standardised regression coefficients were relatively weak. Multicollinearity was never an issue in this analysis, since the VIF was never greater than 2.8 (well below the cutoff point of 10), as can be seen in the full output available in SMB17. Additionally, the best regression models (with predictors) for both vignettes were better than using the mean of the dependent variable (intercept-only models, see SMB17). The main results of the regression analysis are documented in Tables B16a-b.

Tables B16a-b:

Multiple linear regressions at total sample and by vignette (forward selection): significant coefficients. Dependent variables: self-improvement/self-defence

Tab. B16a: Regression coefficients - Dependent variable: self-improvement

	Significant Predictors	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
Tot. sample	(Constant)	-20.851	5.215		-3.998	<.001
	Propriety of Deed	.599	.053	.548	11.267	<.001
	Approach	.190	.046	.147	4.139	<.001
	Goodness of Deed	.283	.061	.222	4.653	<.001
	Religiosity	.072	.023	.116	3.122	.002
Francia	(Constant)	-32.390	9.970		-3.249	.001
	Propriety of Deed	.553	.083	.403	6.687	<.001
	Agent Evaluation	.493	.109	.273	4.537	<.001
	Approach	.197	.065	.173	3.050	.003
Nicholas	(Constant)	-12.592	7.083		-1.778	.077
	Propriety of Deed	.564	.073	.579	7.743	<.001
	Goodness of Deed	.253	.076	.229	3.336	.001
	Approach	.178	.064	.128	2.797	.006
	Religiosity	.108	.037	.160	2.938	.004
	Political Orientation	-.099	.041	-.130	-2.416	.017

Tab. B16b: Regression coefficients - Dependent variable: self-defence

	Significant Predictors	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
Tot. sample	(Constant)	79.359	5.977		13.278	<.001
	Goodness of Deed	-.231	.058	-.232	-3.994	<.001
	Political Orientation	.095	.023	.166	4.125	<.001
	Prevention Focus	-.175	.033	-.195	-5.371	<.001
	Propriety of Deed	-.140	.046	-.163	-3.061	.002
	Avoidance	.059	.030	.083	1.934	.054*
	Agent Evaluation	-.207	.063	-.192	-3.262	.001
	Religiosity	.054	.020	.112	2.764	.006
	Age	-.091	.054	-.062	-1.683	.093
	Promotion Focus	-.179	.057	-.164	-3.147	.002
	Approach	.120	.046	.119	2.584	.010
Francia	(Constant)	126.286	10.794		11.700	<.001
	Goodness of Deed	-.574	.141	-.268	-4.081	<.001
	Prevention Focus	-.197	.043	-.235	-4.548	<.001
	Religiosity	.092	.026	.198	3.552	<.001
	Promotion Focus	-.320	.066	-.322	-4.830	<.001
	Propriety of Deed	-.175	.068	-.147	-2.584	.011
	Agent Evaluation	-.254	.101	-.163	-2.509	.013
	Approach	.171	.066	.174	2.604	.010
	Political Orientation	.070	.032	.122	2.218	.028
Nicholas	(Constant)	55.038	6.376		8.632	<.001
	Goodness of Deed	-.260	.063	-.318	-4.111	<.001
	Political Orientation	.120	.032	.213	3.779	<.001
	Prevention Focus	-.143	.048	-.150	-2.978	.003
	Propriety of Deed	-.167	.057	-.231	-2.944	.004
	Avoidance	.120	.039	.157	3.081	.002
	Moral Discrepancy	-.109	.053	-.145	-2.068	.040

Note: * non-significant at the .05 level (2-tailed).

The Moderation Hypothesis

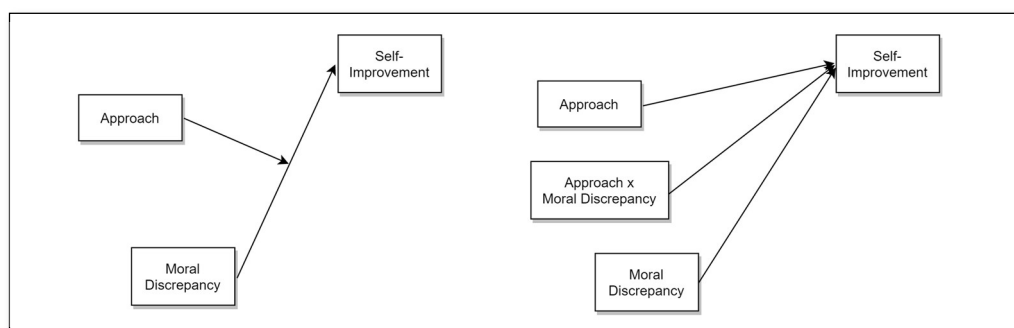
One of the hypotheses of Study 2 was an interaction between moral discrepancy and the variables of regulatory focus, hedonic orientation, and self-esteem. The presence of a significant interaction term between these variables would reveal a moderating role exerted by the characteristic adaptations (an example is illustrated in Fig. B9).

Specifically, the preregistered moderation effects contemplated that:

- moral discrepancy would lead to higher degrees of self-improvement in the presence of high promotion focus, approach, and self-esteem, as opposed to lower degrees of self-improvement in the presence of low promotion focus, approach, and self-esteem;
- moral discrepancy would lead to higher degrees of self-defence in the presence of high prevention focus⁴⁴, avoidance, and self-esteem, as opposed to lower degrees of self-defence in the presence of low prevention focus, avoidance, and self-esteem.

Figure B9:

Two equivalent representations of the conceptual model of moderation: the example of approach (moderator), moral discrepancy (predictor), and self-improvement (outcome)



⁴⁴ The hypothesis of a positive moderating effect of prevention focus on self-defence was based on the assumption of a positive correlation between the two constructs. However, regression analysis uncovered a negative correlation, which would lead to an expectation of a moderation effect with negative sign.

The moderation analysis was conducted with the PROCESS macro ver. 3.3, one by one for each characteristic adaptation interacting with moral discrepancy, each time controlling for the effect of all the other characteristic adaptations. For instance, as in Figure B9, when testing the moderating effect of approach on the relationship between moral discrepancy and self-improvement, the effect was modelled controlling for promotion focus, prevention focus, avoidance, and self-esteem⁴⁵. For exploratory purposes, the analysis was extended to include each characteristic adaptation in the prediction of *both* self-improvement and self-defence (beyond the preregistered hypotheses) and was carried out at the level of the two individual vignettes. All the variables were treated as observed using composite means.

Overall, the analysis showed no systematic evidence of interactions for either vignette; therefore, an overarching moderation effect was *not* supported by the data. The only exception was a negative interaction between approach and moral discrepancy in relation to self-defence in Nicholas's vignette, which was significant (confidence interval not straddling zero); the effect size of the interaction increased as approach tended to higher levels, but it remained weak. A summary of the results of the moderation analysis for the two vignettes is reported in Table B17⁴⁶ (further details in SMB18).

At this point, with the results of correlation and moderation analysis taken together, it seemed reasonable to drop self-esteem and carry on with a more parsimonious model that retained regulatory focus and approach/avoidance.

⁴⁵ The inclusion of self-esteem in the moderation analysis, although redundant in the overall model, is explained by the fact that it was preregistered.

⁴⁶ The moderation coefficients reported in Table B17 are unstandardised. They can be interpreted as follows: when the moderator (e.g., approach) is zero, a 1-unit change in the independent variable (moral discrepancy) corresponds to a change of the dependent variable (e.g., self-improvement) equal to the value of the unstandardised moderation coefficient – all other elements being equal (see Hayes, 2018).

Table B17:

Moderation analysis for the characteristic adaptations for the two vignettes. Dependent variables: self-improvement/self-defence. Predictor: moral discrepancy

Moderators for each dependent variable	Francia			Nicholas		
	R ²	Interaction coefficient	Sig.	R ²	Interaction coefficient	Sig.
Dependent variable: self-improvement						
Promotion focus	.1744	-.0043	.4692	.2986	.0030	.3789
Approach	.1747	-.0041	.4132	.2995	.0031	.3155
Self-esteem	.1722	.0016	.7493	.2972	.0018	.5521
Prevention focus	.1761	.0043	.3281	.2964	.0010	.7759
Avoidance	.1723	-.0014	.7659	.2980	-.0021	.4417
Dependent variable: self-defence						
Prevention focus	.2696	.0029	.4408	.3637	-.0007	.8036
Avoidance	.2779	-.0054	.1009	.3637	-.0006	.7868
Self-esteem	.2749	.0052	.1089	.3654	-.0017	.5094
Promotion focus	.2696	.0037	.3545	.3693	-.0033	.1933
Approach	.2687	-.0028	.4658	.3884	-.0061***	<.001

Note: *** $p < .01$ (2-tailed).

Bootstrap results are based on 500 bootstrap samples.

The Mediation Hypothesis and Path Models

Because based on the empirical evidence the notion of an overarching moderating effect exerted by the characteristic adaptations was no longer tenable, an alternative hypothesis was explored: instead of moderators, regulatory focus and hedonic orientation were considered *predictors* (exogenous variables) of moral comparisons and self-regulatory processes (endogenous variables), with moral comparisons as mediators of self-regulation. To test mediation, unlike with the moderation hypothesis, instead of running a set of separate mediation analyses, a single full path model was tested, first among the total sample, and then (failing structural invariance tests) for each of the two vignettes. The key constructs were modelled as observed variables using the corresponding composite mean scores. The analysis was carried out with the R package lavaan ver. 0.6-6 (Rosseel, 2012), employing the robust maximum likelihood (MLR)

estimator or the weighted least squares adjusted for mean and variance (WLSMV) estimator to handle multivariate non-normality of the distributions (see Savalei, 2018); the free parameters were tested 2-tailed at 95% c.l.

Assessment of the goodness-of-fit of the hypothesised models against the empirical data was based on the joint analysis of the following key fit indices (“scaled” whenever appropriate):

- *absolute fit indices*, e.g. the chi-square statistic (and its p -value), and the GFI (with its “parsimony-adjusted” parent-measure, the AGFI);
- *relative fit indices*, e.g. the CFI and TLI;
- *non-centrality-based fit indices*, e.g. the RMSEA (and its p -value).

Among the above-mentioned indices, the RMSEA (with its p -value) is probably the most widely used and recommended (Schreiber et al., 2006). In the context of this analysis, the chi-square statistic was reported for the sake of completeness rather than for its diagnosticity, given its high sensitivity to factors such as sample size, multivariate non-normality of the distributions, number of parameters to be estimated, size of the correlations, and so forth (Kline, 2016).

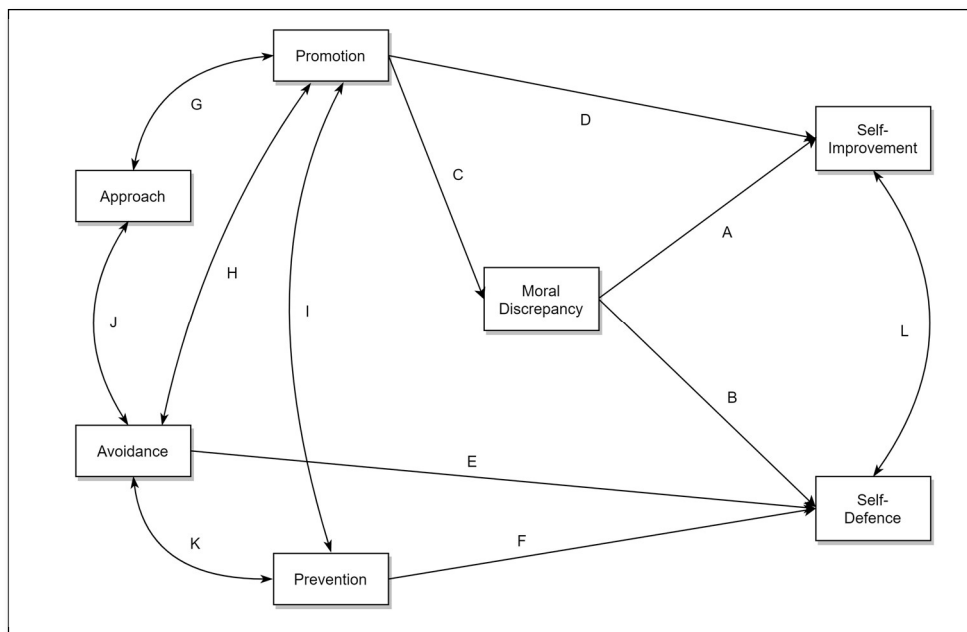
The models were considered to have satisfactory fit if the scaled RMSEA were non-significant and lower than .080 (excellent fit if lower than .050), and if scaled CFI, scaled TLI and GFI were greater than .900 (excellent fit if greater than .950).

The hypothesised model for the analysis included all the pathways that could be assumed to be significant based on available findings. Self-improvement would be predicted by moral discrepancy and promotion focus (or approach); self-defence would be predicted by moral discrepancy, prevention focus and avoidance; moral discrepancy would be predicted by promotion focus (or approach); promotion focus would covary with prevention focus, approach and avoidance, and avoidance with approach and

prevention focus; self-improvement and self-defence would show residual covariances (see Fig. B10).

Figure B10:

Hypothesised path diagram of the revised conceptual model with the key characteristic adaptations as exogenous predictors, moral discrepancy as endogenous mediator, and cognitive/conative/affective self-regulation modes as endogenous outcomes. All variables treated as observed



Note: the letters represent the pathways whose coefficients were hypothesised to be significantly different from zero.

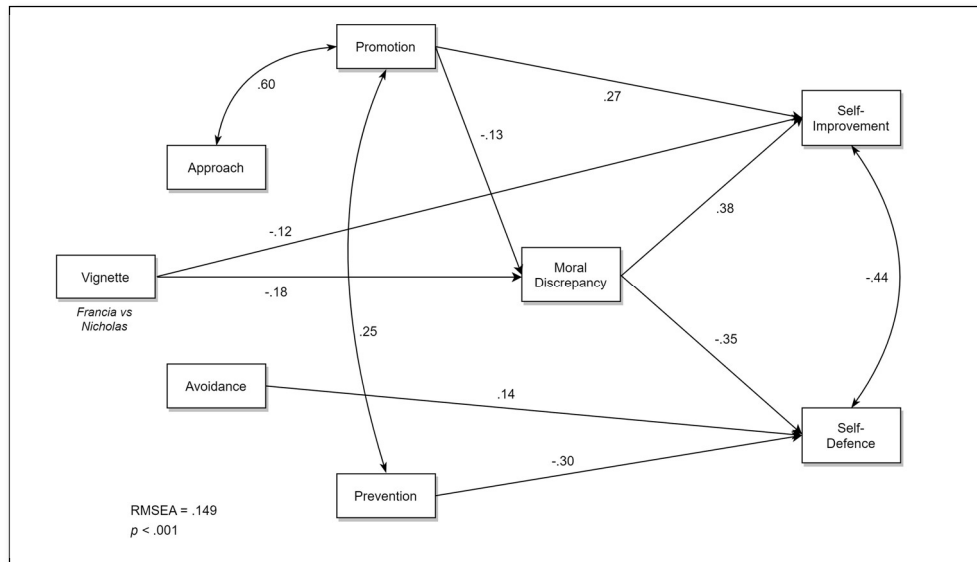
Path Model Fitted to the Total Sample

The path model fitted to the total sample included the variable vignette (the experimental manipulation) as an additional predictor⁴⁷ to test its ability to determine differential ability-based comparisons and moral self-regulation modes. In this case, the WLSMV estimator was used. The optimised path diagram is illustrated in Figure B11, while the complete output of the analysis is available in SMB19.

⁴⁷ Francia was dummy-coded 0 and Nicholas 100, reproducing the metric of the other scales.

Figure B11:

Optimised path diagram for the total sample. All variables treated as observed, using composite mean scores for the scales. Estimator: weighted least squares adjusted for mean and variance (WLSMV). Standardised regression weights. Error terms not displayed to facilitate legibility



The negative sign of the regression weights from vignette to moral discrepancy and self-improvement indicates that Francia's vignette induced significantly higher levels of upward comparison and self-improvement than Nicholas's vignette⁴⁸.

Overall, the goodness-of-fit was inadequate, as shown in Table B18: in particular, the RMSEA of .149 with significant *p*-value indicated poor fit. However, it must be said that the present model was fitted with observed variables, assuming perfect measurement with no error. It was a pragmatic approach suitable to an early exploratory study with a limited sample size, but not accurate. In the next stage of the research (Study 3), with a fresh larger sample drawn from the same population, latent variables were constructed and fitted to the full structural models, allowing a higher

⁴⁸ The values of the regression weights with vignette as predictor (-.12 and -.18) in Figure B11 are standardised. The corresponding unstandardised values are respectively -.056 and -.074 (see SBM19) and are meaningless from a substantive viewpoint, as they depend on the dummy coding of the vignettes. However, what is meaningful is that they are both statistically significant and carry a negative sign.

degree of precision for the point estimates of the regression weights as well as for the model fit.

Table B18:

Main fit indices for the optimised model for the total sample

	Total sample
Chi-Square (scaled)	162.006
df (scaled)	16
<i>p</i> -value (scaled)	<.001
RMSEA (scaled)	.149
<i>p</i> -value (scaled)	<.001
CFI (scaled)	.527
TLI (scaled)	.202
GFI	1.000
AGFI	1.000

Structural Invariance Between Vignettes

To test whether one single path model could be adequate for both vignettes, structural invariance between the two subsamples/vignettes (Francia versus Nicholas) was tested. At a model-wide level, this was achieved by comparing an unconstrained model, whereby all parameters were freely estimated for each vignette, with a fully constrained model, whereby all parameters (regression weights and intercepts) were constrained to be equal between vignettes. If the chi-square difference for the corresponding degree of freedom difference were significant, there would be evidence for lack of invariance, hence distinct models would be needed for each vignette.

Results of structural invariance tests showed that the two vignettes were significantly different (see Tab. B19; full output in SMB20), indicating that two separate models were necessary, one for each vignette subsample.

Table B19:
Structural invariance test by vignette (total sample)

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	18	24659	24868	75.969			
Fully constrained	31	24683	24840	125.481	51.04	13	<.001***

Note: *** $p < .001$ (2-tailed).

Because comparing an unconstrained with a fully constrained model provides a *model-wide* test of invariance, once ascertained that the path models significantly differed by vignette, there was an interest in determining which specific pathways were the drivers of the overall effect. This was done by constraining individual pathways of interest (one at a time) to be equal between vignettes, and then comparing the unconstrained with these single-constrain models. Here, the most interesting paths to test were those linking moral discrepancy with self-improvement and self-defence (paths A and B in Fig. B11), since in Study 1 they behaved slightly differently across the three main vignettes of Francia, Nicholas, and Ruxandra (Fig. A13a-c). Additionally, given the unclear association patterns found in the literature, some of the covariances between the characteristic adaptations were tested for invariance too, for example those between promotion and prevention focus (path I), approach and avoidance (path J), prevention focus and avoidance (path K). The results of these tests are reported in Table B20 (full output in SMB20).

Table B20:

Structural invariance tests by vignette: models with paths A, B, I, J, and K constrained to equality by vignette and tested against the unconstrained model

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	18	24659	24868	75.969			
Path A constrained	19	24670	24876	89.110	15.484	1	<.001***
Path B constrained	19	24664	24869	82.993	6.2492	1	0.012*
Path I constrained	19	24664	24869	82.609	5.8695	1	0.015*
Path J constrained	19	24672	24877	90.843	13.812	1	<.001***
Path K constrained	19	24661	24866	79.969	3.5066	1	0.061

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

Only path K was invariant between vignettes, while all the others were significantly different. Lack of invariance of path A was highly significant at the .001 level. Once again, it must be emphasised that these results are still preliminary, being based on relatively small samples and models with inadequate fit indices obtained with observed instead of latent variables.

Path Models Fitted by Vignette

The results of path analysis by vignette indicated that the models did not achieve satisfactory fit for either Francia's or Nicholas's subsamples (see Tab. B21; full output in SMB21). However, as expected following invariance tests, in both cases the fit was substantially better than for the total sample.

Table B21:

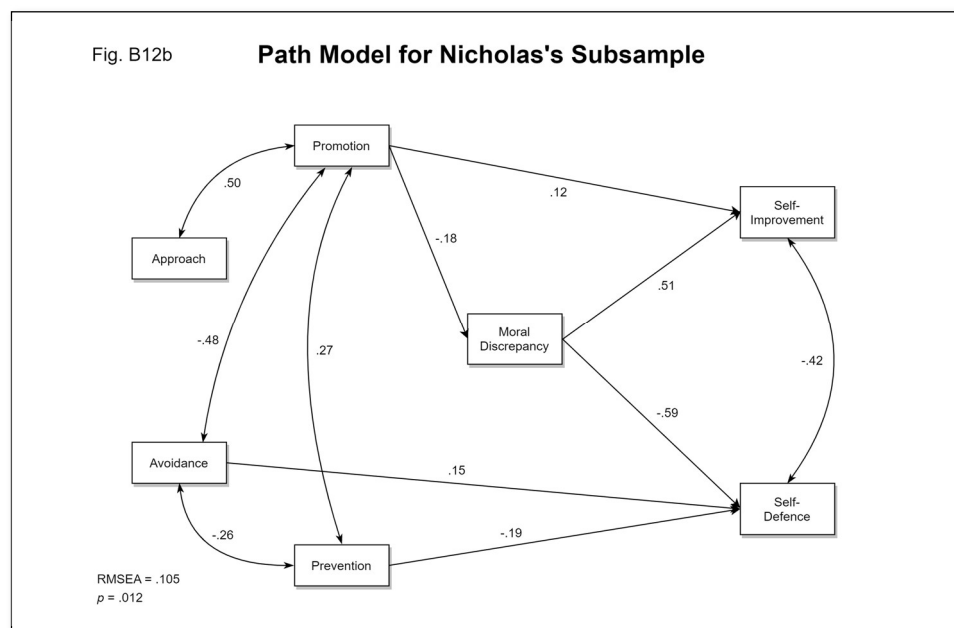
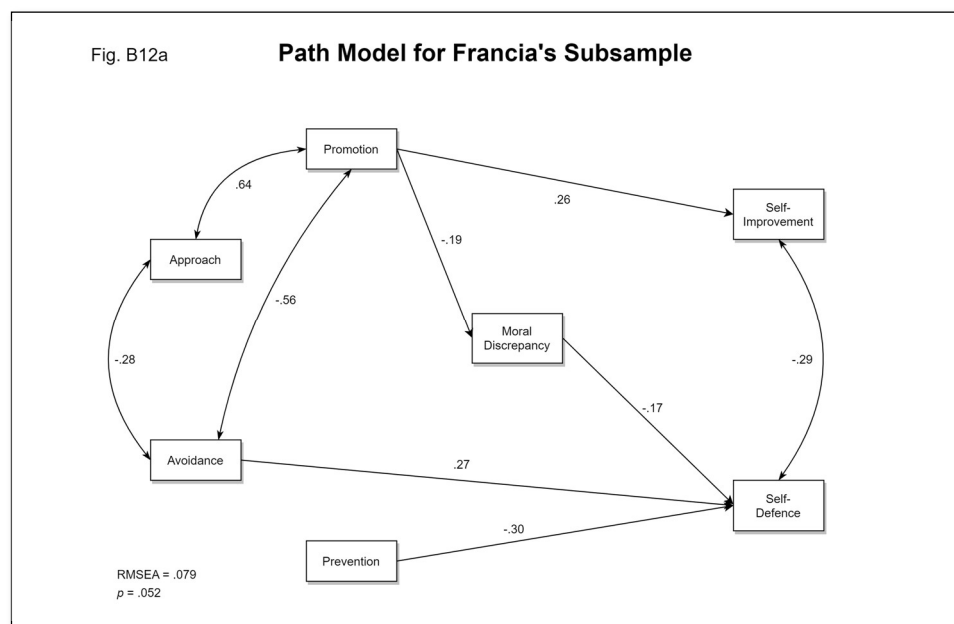
Main fit indices for the optimised models for the two vignettes

	Francia	Nicholas
Chi-Square (scaled)	27.597	32.650
df (scaled)	12	10
p -value (scaled)	.006	<.001
RMSEA (scaled)	.079	.105
p -value (scaled)	.052	.012
CFI (scaled)	.925	.928
TLI (scaled)	.870	.850
GFI	.996	.998
AGFI	.990	.992

The optimised path models for the two vignettes (following the results from invariance tests) are displayed in Figures B12a-b, which also document the significant regression weights and the RMSEA with its p -value.

Figures B12a-b:

Optimised path diagrams for the two vignettes. All variables treated as observed, using mean scores for the scales. Estimator: robust maximum likelihood (MLR). Standardised regression weights. Error terms not displayed to facilitate legibility



As expected, the pathway between moral discrepancy and self-improvement was different for the two vignettes, significant for Nicholas and non-significant for Francia (although this could be an artefact of the methodology). Similarly, the covariance between promotion and prevention focus was significant for Nicholas and non-significant for Francia; the covariance between approach and avoidance was significant for Francia and non-significant for Nicholas. The covariance between prevention focus and avoidance should have been invariant between vignettes, but turned out to be significant for Nicholas and non-significant for Francia.

Although knowingly provisional and inadequate, the path models in Figures B12a-b represent the final and most important output of Study 2. They were used as input for the full structural equation models fitted in Study 3.

Discussion

Study 2 completed the preliminary exploratory and generative stage of the research plan (Fig. IV), consolidating important learning for the development of the measurement and structural models of the self-regulation of virtue.

Measurement Model

Item Selection

Study 2 allowed to remove a few poorly performing indicators from the moral self-regulation and affect inventory: one from the self-defence scale (SD2), one from the positive affect scale (PA6), and one from the negative affect scale (NA6). It also helped identify other potentially weaker items (e.g., SI7, SD5, SD10, NA2, NA5) which showed lower factor loadings onto their corresponding factors. At this stage, they were not dropped to avoid the risk of overfitting, but would be carefully re-evaluated in Study 3 in terms of inter-correlations, factor loadings, internal consistency, goodness-of-fit, and error.

Merging of Cognitive/Conative/Affective Items

Study 2 was instrumental to the fundamental decision about the consolidation of the indicators of the four scales of the preliminary self-regulation and affect inventory into two broader cognitive/conative/affective scales. These were still labelled moral self-improvement and moral self-defence, but must now be intended in their more comprehensive connotation, comprising thoughts/action tendencies and feelings elicited by the moral actions depicted in the vignettes. This data-driven solution is not only empirically, but also theoretically plausible, since affect in this research was measured through self-reports. Self-reported measures of affect are known to be mediated by people's ability to directly access bodily cues necessary for the experience of emotion (Feldman Barrett et al., 2005). As discussed in the introductory chapter, one

of the components of emotion is neurophysiological/motor changes in the body (Scherer, 2005), which are experienced, among other cues, in the form of feelings (Izard, 2007). The ability to access and become aware of these bodily changes is called “interoception”. Recent research and theorising have established that interoception comprises the conversion of sensory information (e.g., visceral and muscular cues) into afferent signals and their transmission to the central nervous system to form mental representations of inner states (Feldman Barrett et al., 2005). In a review published in 2016, Tsakiris and Critchley claimed that interoceptive signals actively interact with cognition, “influencing attention and perception, guiding decision-making and shaping memory and emotion processing” (p. 1). If feelings arising from bodily experiences are cognitively processed to form complex mental representations of inner states, it is likely that self-reports coalesce into undivided cognitive/affective experiences, resulting in strongly correlated measures.

Additionally, the supposed dualism between cognitive and affective aspects of moral judgment has been at the centre of philosophical speculation even before psychology was established as a science, with the opposing traditions of the “rationalistic” and the “sentimentalist” schools (Cushman et al., 2010). In psychology, Kohlberg’s work inherited the dominant Kantian view in Western ethics that for centuries put moral reasoning at the core of moral judgment, proposing a cognitive-developmental approach that was highly influential (see e.g., Kohlberg, 1976). A few decades later, Haidt and colleagues (see e.g., Haidt & Joseph, 2004; Haidt & Kesebir, 2010) offered a socio-intuitionist perspective which, in contrast, highlighted the importance of intuitions as basic semi-innate units that underlie more complex culturally-dependent moral compounds strongly influenced by emotions; in this landscape, which revived Hume’s empiricist conception of the “passions” in moral

judgment (McIntyre, 2012), feelings are the key ethical currency and cognitions are relegated to *post-hoc* rationalisations (Ellemers et al., 2019). Attempts were made during the last few years to reconcile these opposite views of morality: Dedeke (2015) reviewed theorising and empirical research on ethical decision making from the 1980s onward and put forward a cognitive-intuitionist model of moral judgment, which proposed an integration of cognitive/deliberative and affective/intuitive mechanisms. The debate is still open, but the notion of a close interplay between deliberative and intuitive processes in moral judgment is gaining traction (Ellemers et al., 2019), echoing the more recent advances in the field which no longer see controlled deliberative processes and automatic heuristic processes as mutually exclusive and clearly separable (for a review, see e.g., Gawronski & Creighton, 2013). If further empirical evidence is obtained and integrative theories solidify, it could become easier to provide further explanations of why in the present study it appeared difficult to disentangle self-reported cognitive/affective regulatory processes.

Dimensional Structure

Study 2 also provided important insights into the dimensional structure of the moral self-regulation scales. EFA among the total sample indicated that, from a substantive standpoint, a two-factor solution could be more easily explained; analysis of the empirical data revealed that this solution was completely satisfactory for Francia's vignette, but slightly weaker for Nicholas's, for which a three-factor solution statistically showed better fit (albeit hardly interpretable).

As previously discussed, theoretical considerations would make it preferable to develop a *single* valid and reliable measurement instrument of the self-regulation of virtue capable of capturing the same phenomenon across the two (or more) moral scenarios. This balancing act posed a challenge. On the one hand, overall, the moral self-

regulation inventory is supposed to measure the same phenomenon (e.g., experiences of broadening and defensive moral self-regulation) regardless of the stimuli; on the other hand, a slightly different emphasis on specific components of those experiences could be justified by the varying content of the stimuli, while still pertaining to the same category of experience. The issue becomes apparent when items referred to certain components of that experience behave *too differently* across conditions, so that their presence in a single measurement instrument is no longer legitimate. Items NA5/Guilty and NA3/Vulnerable provide two interesting examples: Francia's action of donating a kidney to a friend triggered among participants a sense of guilt and vulnerability that was not induced to the same extent by Nicholas's commitment against social discrimination. These items were still retained after Study 2, but remain candidates for exclusion from the final inventory, with final assessment deferred to confirmatory techniques (CFA and measurement invariance) planned for Study 3 among a larger sample.

In principle, a second round of factor analysis, with a confirmatory procedure, could have been conducted on the data from Study 2 as a complement to the exploratory procedure detailed earlier in this chapter. However, this is not a recommended practice: the risk of conducting EFA (or PCA⁴⁹) and CFA sequentially on the same dataset is that the solution could be too sample-specific, hence with low generalisability (the issue of "overfitting"), particularly if the results of EFA are used to influence decisions about the subsequent CFA (see e.g., Fokkema & Greiff, 2017). Therefore, in Study 2 only exploratory factor analysis was carried out, in line with the nature of the present stage

⁴⁹ PCA is the commonly used acronym for principal component analysis, a similar exploratory technique used when the dimensional structure is not known (for a clear summary of the difference between the two techniques, see e.g., Blunch, 2013).

of the research, deferring confirmatory analysis to the subsequent Study 3 among a new sample.

Structural Model

Study 2 enabled strong progress in the definition of the relationships across key constructs in the nomological network.

Moral Comparison and Self-Regulation

At total sample level, results reproduced the pattern of relationships between moral comparisons and self-regulation already emerged in Study 1: upward comparison was more likely to lead to moral self-improvement, downward comparison to moral self-defence.

Structural Invariance

Structural invariance was rejected, suggesting that specific models were necessary for each vignette. Notably, the magnitude of the association between moral discrepancy and moral self-regulation was much stronger in Nicholas's condition, where the standardised coefficients exceeded .50. Further evidence will be necessary for Francia's vignette, where only the pathway between moral discrepancy and self-defence was significant, and the effect size was moderately low.

Characteristic Adaptations

The study also provided evidence about the importance of the characteristic adaptations as predictors of the moral experience of virtue. Among these, self-esteem – given the very strong correlations with promotion focus, approach, and avoidance – turned out to be redundant and was dropped from the model for the sake of parsimony. Importantly, the role of regulatory focus and hedonic orientation as moderators was not supported by the data. Although theoretical predictions would have suggested a possible interaction between these variables and moral discrepancy, the empirical data

showed no systematic significance. The result was consistent across the two vignettes for all the four motivational variables, except for approach in the prediction of self-defence (with low effect size).

Whereas the moderation hypothesis was not supported, the mediation hypothesis was. In the revised path models, moral discrepancy functioned as a partial mediator of promotion focus (or approach) in the prediction of moral self-improvement (with also a link to moral self-defence), whereas prevention focus and avoidance directly predicted self-defence. These path models did not achieve satisfactory fit, but represent a necessary intermediate step before the specification of full structural equation models fitted among a larger sample using latent variables in Study 3.

Direct Effects on Moral Self-Regulation and Moral Discrepancy. An important finding concerned the direction of the effects of the motivational variables on moral self-regulation. The path models confirmed the hypothesis that promotion focus or approach positively predicted self-improvement, aligning with claims that self-improvement is underpinned by promotion-orientated strategies and approach motives (Sedikides & Hepper, 2009). Indeed, promotion focus and approach represent respectively advancement motives towards ideals (Higgins, 1997) and energisation toward desired goals (Elliot & Thrash, 2010). The prediction that participants high in promotion and approach would tend to score higher (relative to those low in promotion and approach) on self-improvement was met on the basis of a psychological motion toward a desirable moral ideal, that is: a perception of the moral agents as admirable (SI1) moral exemplars, with whom participants shared the same altruistic values (SI5) and sense of humanity (SI3); a perception of their actions as uplifting (PA1) and inspiring (PA2) which made participants feel vicariously proud (PA4), wanting to be more like them (SI6) and to imitate their behaviour for good (SI7).

The hypothesis of the positive association between avoidance and self-defence was also met for both vignettes. Avoidant participants susceptible to thoughts/feelings of anxiety and worry tended to self-defend: the moral exemplars triggered a greater sense of threat (NA4) and more negative feelings of conflict (NA1), resentment (NA7) or even irritation (NA8) (Nicholas's vignette), or alternatively a greater sense of vulnerability (NA3) and guilt (NA5) (Francia's vignette). They were also associated with motivated aggrandising self-serving judgments about themselves (SD1, SD5) or dismissive trivialising judgments about the moral agents (SD6, SD7, SD12, SD13) and their actions (SD3, SD4, SD10).

Particularly interesting is the case of the relationship between prevention focus and self-defence. The two variables were predicted to correlate positively and instead turned out to be negatively associated. The result at total sample was upheld in the analysis by individual vignette and was of moderate effect size. The original hypothesis was based on relationship patterns between the motivational dispositions: Elliot and Thrash (2010) found a strong positive correlation between chronic prevention focus and avoidance temperament ($r = .57, p < .001$), and so in theory both could have correlated positively with self-defence. Additionally, extant research pointed toward the contrasting judgments usually expressed by people high in promotion versus prevention focus in the presence of transgressions of moral standards (the latter tend to blame significantly more harshly than the former; see Cornwell & Higgins, 2015b). Because the actions in both vignettes were initially thought to be supererogatory, it was anticipated that people high in promotion focus would rate them as highly praiseworthy, indicative of immaculate intentions toward an ideal greater good, and that people high in prevention focus would construe them as transgressions of ethical norms to comply with (acts beyond the call of duty, hence violations of required behaviours) and

therefore would judge them more harshly and even blame them. The issue with the latter hypothesis was likely the fact that it would predict moral actions considered *strongly* supererogatory, which perhaps was not the case for the deeds performed by Francia and Nicholas. Although Francia's action was construed as less obligatory than Nicholas's, neither of them was considered so strongly supererogatory to cause people high in prevention focus to consistently blame them in self-defence⁵⁰.

It is interesting to note that prevention focus, while correlating negatively with self-defence, did not correlate positively with self-improvement: the prevention system did not systematically lead participants to feel inspired and uplifted by an admirable moral action (absence of correlation with self-improvement), but simply *inhibited* the instantiation of defensive self-regulation (negative correlation with self-defence), thus functioning as a safeguarding "hygiene factor" when self-threat and advancing drives are modest.

Additionally, in all path models, the regression weight between prevention focus and moral discrepancy was not significant (prevention could be associated with either upward or downward comparison), and at total sample prevention was also unrelated to goodness and propriety (prevention could be associated with more or less positive judgments about the deed). A possible explanation for these patterns is that people high in prevention focus are not motivated by ideal exemplars, such as those in the two vignettes, but rather by *unsuccessful* others from whom they can learn what they should *not* do, as they provide a palette of the kind of failures that they strive to prevent (Lockwood & Matthews, 2007). This interpretation is supported by studies where prevention-primed and chronically prevention-orientated students appeared to be more

⁵⁰ This could have been the case for Arnaud's vignette in Study 1 (the story of the policeman who sacrificed his life during a terrorist attack).

motivated by unsuccessful target students than by successful ones because the former represented the kind of end-states that they aimed to avoid, motivating them to work harder at averting a similar negative outcome (Lockwood et al., 2002). If for individuals high in prevention focus the source of motivation originates more strongly from unsuccessful or negative referential targets, it is easy to see why, after viewing the vignettes, they tended not to consider the moral agents particularly praiseworthy and not to strongly endorse the self-improvement items.

Assuming that a causal interpretation is plausible, overall the findings about the functioning of the four motivational constructs in their *direct* relationships with moral self-regulation can be summarised as follows:

- the promotion and the approach motivational systems have a direct effect on the *activation* of moral self-improvement states (and they could be mutually substitutable in the models);
- the avoidance system has a direct effect on the *activation* of moral self-defence mechanisms;
- the prevention system has a direct effect on the *inhibition* of moral self-defensive processes, but no effect on the activation of moral self-improvement.

Indirect Effects on Moral Self-Regulation Via Moral Comparisons. Promotion and approach orientations also had an indirect effect on moral self-regulation via moral comparative processes. Indeed, they were found to lead to downward comparisons, which in turn fed defensive regulation. This is in line with the mechanism, identified by Scholer and colleagues (2014), according to which individuals tend to inflate their self-evaluations to eagerly sustain promotion and advancement motives toward desired/ideal goals. It must be noted that those findings concern primed states. However, if promotion and approach motives become dispositionally chronicised, then

individuals need to constantly fuel the mechanism of self-inflation to maintain high levels of volition during goal pursuit. This can be effective in an achievement framework, but in the moral domain a systematic tendency to engage in downward comparisons is potentially maladaptive, since it tends to lean individuals toward defensive regulation, as they strive to preserve stability of their inflated self-concept, constantly seeking validation of their perceived moral superiority. In line with this reasoning, in Study 2 participants who scored high, for example, in promotion focus held high self-esteem and high consideration of their own morality, which led to the perception of the moral agents in the vignettes as threats to their moral status (high scores on NA4), triggering self-aggrandising judgments (high scores on SD1) and self-protecting trivialisation or outright dismissal of the virtuous actions (high scores on SD3, SD4, etc.).

Prevalence of Motivational versus Comparative Forces. These findings show that in the model the promotion and approach systems, depending on the mechanism that is elicited, could potentially lead to two opposite outcomes: they could lead *directly* to self-improvement and *indirectly* to self-defence through the mediation of moral discrepancy (in downward comparison). This apparently counterintuitive phenomenon could also be explicated by the interplay between the motivational and the comparative psychological functions. The literature offers evidence that individuals high in promotion and approach can be more focused on the pursuit of their own goals and less interested in comparisons with others, typically because they hold high self-esteem (Taylor et al., 1995) and a more stable self-concept (Campbell, 1990). Therefore, when the drive toward a desired ideal of virtue is the predominant motive and moral comparisons remain in the background, individuals may have a prevailing tendency to self-improve. By contrast, if the juxtaposition and contrast of information about self and agent are more salient than growth and hedonic motives, then individuals high in promotion or

approach (with high self-regard) may tend to perceive the moral agents as “competitors” who endanger their supposed moral superiority; consequently, they may show a tendency to self-defend in order to defuse the threat.

A closer inspection of the models also reveals that for Nicholas’s vignette comparative processes were stronger predictors of moral self-regulation, while for Francia’s vignette motivational tendencies played a more substantial role. This could be linked to the fact that the average value of moral discrepancy was lower in Nicholas’s subsample, that is, fewer participants on average engaged in upward comparison and more participants engaged in downward comparison (relative to Francia). It could be argued that, overall, participants perceived a higher degree of moral character *proximity* to Nicholas than Francia, whose very high stature could have been considered hardly attainable. This phenomenon could have heightened the saliency of moral comparisons in Nicholas’s subsample. The literature on social comparison supports the notion that people often discount comparisons with others who are *too* dissimilar and prefer more realistic comparisons with more similar targets, capable of providing information that is more relevant and actionable for the self (Wheeler, 1991). This is also in line with the framework proposed by Festinger (1954) in his original description of social comparison theory and is particularly true when people are motivated by epistemic needs (i.e., truth).

The idea that people might have stronger or weaker dispositions to compare themselves with others has led researchers to develop a scale to measure individual differences in the orientation to engage in social comparisons. To this end, Gibbons and Buunk (1999) validated the Iowa-Netherlands Comparison Orientation Measure (INCOM), whose relationships with key variables in the conceptual model was explored in Study 4.

Strengths and Limitations

The progress in terms of model generation achieved with Study 2 was substantial. The two best vignettes from Study 1 proved to be successful in eliciting among participants differential moral experiences, which were measured in terms of comparative and self-regulatory processes. A unidimensional structure for each of two moral self-regulation scales was identified, conflating cognitive/conative and affective components of the moral experience of virtue; also, a preliminary choice of items was made, dropping those that showed poor communalities with the corresponding constructs. Parsimonious path models were also fitted to the total sample and by individual vignette, abandoning redundant variables (self-esteem) and determining the role of the motivational dispositions as predictors, not moderators. In a nutshell, the key objectives were met.

At the same time, it is fair to acknowledge that the measurement model still had to fully prove its reliability and validity. Some items, particularly in the self-defence scale, still required further assessment and only a more precise confirmatory technique, such as CFA, would be capable of validating the scales. To accomplish this, a new larger sample size would be necessary. Even more importantly, the structural models generated with a simplified path analysis did not achieve satisfactory fit. Therefore, it remained to be seen whether the models would be capable to adequately explain the phenomenon at hand. A confirmatory technique, such as full structural equation modelling (SEM), would be necessary to provide the evidence. That too would require a new larger sample.

The final sample size of just above four hundred participants for Study 2 was adequate, but it was smaller than the planned five hundred. A certain loss of participants was determined by the need to remove eighty-seven of them due to

incorrect answers to the check questions, completion time incompatible with adequate information processing, and presence of extreme outliers (as defined in the preregistration). All of these were justifiable reasons, yet the bottom line was a certain loss of power. This is always undesirable, but in this case, wise decisions in the planning stage helped contain the possible drawbacks. *A priori* power analysis had indicated that, setting alpha and beta respectively at .05 and .20, a total sample of 218 participants would be necessary for a robust multiple regression analysis capable of detecting a medium-sized effect. The actual total sample size was almost twice as large, so it was adequate not only for regression but also for path analysis.

Once the analysis was completed, it was important to reflect on the data quality from the study, particularly because the sample was drawn from the MTurk crowdsourcing platform. Dropping 87 participants from the initial sample of 499, albeit justifiable, represented a loss of nearly 20% of “information capital”, which is not negligible. There has been debate among academics over the past couple of years about how valid and reliable MTurk samples could be considered for academic research. Whereas in the first years of operation published studies seemed to attest the validity and reliability of the data collected through these platforms (see e.g., Behrend et al., 2011; Buhrmester et al., 2011), more recently various kinds of concerns have been raised:

- the non-naïveté of participants, often unemployed, who take a large number of surveys (Chandler et al., 2014) as their primary source of income (Paolacci et al., 2010);
- the risk that participants take part in the same study multiple times (Woods et al., 2015), affecting independence of the observations;

- the risk that the surveys are taken by “farmers”, individuals who do not have platform membership credentials, but bypass location and other restrictions using server farms (Chmielewski & Kucker, 2020);
- the risk that, instead of human participants, the surveys are filled out by “web-bots”, automated computer programmes (robots) that perform repetitive tasks in a short time (Chmielewski & Kucker, 2020), following random algorithms;
- evidence that a conspicuous share of participants does not pay enough attention and answers carelessly (Aruguete et al., 2019; Curran, 2016) to save time.

Crowdsourcing platforms and researchers have since implemented several practices to improve data quality, for instance the use of captcha technologies (Aruguete et al., 2019), geolocation/IP address verification (Peer et al., 2017), procedures for screening participants who have previously completed the same or related studies (Chandler et al., 2014), attention checks in the form of instructed response questions (Kam & Chan, 2018) or treatments that slow down survey presentation to encourage thoughtful responding (Paolacci et al., 2010), and so forth. Some of these recommended practices were adopted in the present study (IP address verification, duration screening, check questions, etc.) to improve data quality, but the issue of the declining value of certain crowdsourcing platforms remains, especially MTurk.

To be fair, Study 2 was in line with other online studies in terms of loss of participants (for a review, see Thomas & Clifford, 2017) and therefore cannot be considered eccentric in this respect. Yet, careful consideration of MTurk data quality was closely monitored in the following studies, and a switch to other platforms regarded as a possible countermeasure.

In sum, Study 2 achieved to identify one preliminary measurement model and two preliminary structural models of the self-regulation of virtue, completing the initial stage of the research. Study 3 used these models as input into confirmatory methods that offered solid evaluative bases for their empirical and conceptual evaluation.

Study 3: Model Improvement

Introduction

The exploratory stage of Studies 1-2 culminated in the definition of specific measurement instruments for a set of observed variables and preliminary latent variables, and the specification of a network of provisional relationships between them. Exploratory factor analysis returned a unidimensional structure for the latent variables of moral self-improvement and self-defence, and path modelling – applied to an experimental design – provided a grid of associations and probabilistic causal links. At this point, the results needed to be assessed in a confirmatory study among a new sample from the same population.

The goal of Study 3 was to conceptually replicate the findings from Study 2, retesting and improving: a) the measurement model of the moral self-regulation inventory, and b) the structural models of the antecedents and mediators of the self-regulatory response to moral exemplars. To meet this objective, the two vignettes of Francia and Nicholas were tested again within the same experimental survey design as in Study 2, this time among a fresh sample of larger proportions, so that confirmatory analysis techniques could confidently be applied. The fact that the preregistered analysis plan involved confirmatory techniques such as CFA and full SEM did not mean that the study would be limited to the purpose of mere validation of the output of Study 2. Regardless of the goodness of the fitted models, the research data were intended to be used to enhance as much as possible the quality of the findings within the constraints of the available resources; for that reason, after fitting the hypothesised models, theory- and data-driven techniques were planned to help identify the existence of models with stronger fit.

As in the previous studies, CFA was expected to lead to one *single* measurement model, common for both vignettes, for the latent variables of moral self-improvement and self-defence. The fulfilment of this expectation depended on measures of validity and reliability applied to the total sample; in other words, the measurement model had to show strong psychometric properties regardless of the content of the individual vignettes and the characteristics of the sample. For that to be true, the measurement model needed to be *invariant*, particularly in relation to the vignettes, but also to the type of moral comparison (upward, downward, or lateral) and participants' gender (as an example of sample characteristics). As discussed earlier, this objective required striking a balance, avoiding overfitting and underfitting, through the selection of a set of core self-regulatory indicators that satisfied conditions of robustness for both vignettes, even in the presence of two scenarios deliberately selected after Study 1 to be different enough to stretch the generalisability of the models. Because some items of the moral self-regulation inventory retained from Study 2 were not strongly correlated with their respective latent variables, they were scrutinised with particular attention to ensure strong psychometric properties of the final inventory. Specifically, these items were SI7/ForOthers, SD5/People, SD10/Uncommon, NA2/Detached, and NA5/Guilty, all showing poor factor loadings in Study 2, but not dropped from their latent variables to avoid costly decisions when the sample was not large enough and the analysis technique was merely exploratory.

One of the consequences of retaining these supposedly poorer indicators was a larger measurement error and a drop of the goodness-of-fit. Indeed, in Study 2 the results of path modelling did not yield adequate fit indices for either the individual vignettes or the total sample. This was due partly to the use of a streamlined approach with only observed instead of latent variables for the motivational dispositions and the

self-regulatory scales, but also to the larger error that could have been caused by computing scale composite means with weaker items. With Study 3, the use of a confirmatory technique such as CFA should ensure the assessment of the moral self-regulation inventory through a stronger statistical technique, which in turn should result in better model fit thanks to the reduced measurement error.

Study 3 also intended to complement these fundamental variable-centred analyses with *person-centred* analyses. In fact, latent variable modelling can also be applied to identify clusters of participants who share common response patterns to certain variables of interest. The convergence of the findings from these two types of analyses would corroborate the robustness of the results, solidifying the theoretical advances that they underpin. Further details are discussed in the following sections of this chapter.

Study 3 would complete the analysis of the antecedents and mediators of the moral response to virtuous actions. The behavioural consequences included in the initial conceptual model (Fig. III) were intentionally left out of the present inquiry and deferred to a final fourth study. The measurement of behavioural effects of moral self-improvement in Study 4 would complete the current research programme, providing answers to the questions raised in the introductory chapter.

Methods

Research Design

Study 3 had a between-subjects design, whereby participants were randomly assigned to one of two independent conditions. In each condition, participants were presented with one of two moral vignettes: Francia or Nicholas (see AB1).

The study was conducted online and preregistered on the OSF website. A structured questionnaire hosted by Qualtrics (see SMC2) was made available to potential participants during the month of June 2019 through CloudResearch (the same platform used for Study 2, for consistency reasons) in return for payment. Repeat participants from the previous study were automatically excluded via the MTurk worker code.

To improve data quality, and particularly to control for careless responding, instructed response questions were added to the attention checks already implemented in Study 2. As per Kam and Chan's (2018) recommendation, the presence of these questions was introduced and explained in the participation sheet; participants were asked to answer exactly as indicated in two corresponding questions⁵¹. The instructed response questions were located at two different points in the questionnaire: one in the first part, randomised within the regulatory focus items, and the other after the presentation of the vignette, randomised within the moral self-regulation inventory.

Participants

Sample Size Determination

The sample size was determined based on several criteria. First, power analysis using G*Power ver. 3.1, for a multiple linear regression suggested that, to detect a

⁵¹ An example of instructed response question is: "I am competent in panabogy, and please answer 'Not Applicable' to show that you carefully read the question", borrowed from Kam and Chan (2018).

medium-sized effect based on six predictors (those from the path model in Study 2) with $\alpha = .05$ and power set at .80, a minimum sample size of 92 participants per condition would be necessary (see SMC1). This number was substantially augmented in consideration of the complex analyses that were planned (beyond regression), the number of questions and items within the scales (to be factor-analysed), the length of the questionnaire, and the cost of the study. Notably, the sample size had to be adequate to conduct full structural equation modelling separately for each of the two vignettes. For all these reasons, and in the absence of simple formulaic methods to fulfil these criteria, a total sample of 1,100 participants (550 per condition) was deemed appropriate.

Data Exclusions

At the end of data collection, the dataset comprised a total of 1,202 subjects, split approximately in half across the two conditions. Of these, 53 did not provide consent, 52 answered the check questions incorrectly, 61 answered the instructed response questions incorrectly, and 49 dropped out without finishing the questionnaire⁵²; as planned in the preregistration, they were removed from the sample.

As in Study 2, an analysis of the actual duration of the questionnaire was conducted. Pre-testing suggested an average length of about 8 minutes. However, 68 participants completed the survey in less than 3 minutes, a time considered insufficient to adequately process the questions and provide reliable answers. Consistently with the preregistration plan, those participants were also removed from the sample. With all these exclusions, the remaining sample size was 1,024.

Multivariate Outliers

⁵² For some participants, these exclusion criteria overlapped.

At that point, outlier analysis was conducted, replicating the same methodology applied in the two previous studies, using the combined analysis of centred leverage values, Mahalanobis distance, and Cook's distance. The observations that exceeded two out of three of the cutoff points determined by those values were considered extreme multivariate outliers. This analysis enabled the detection of 42 outliers, which were subsequently deleted from the sample, following the preregistered plan.

Final Sample Composition

After the above-mentioned exclusions, the final sample was composed of 982 participants: 490 respondents allocated to Francia's vignette, 492 respondents to Nicholas's vignette. It comprised 460 females (46.8%), 517 males (52.6%), and 5 participants (0.5%) who self-reported "other" (non-binary or transgender) to the gender question. Age ranged from 18 to 75 years, with median of 34 and mean of 36 years ($SD = 11$). All participants were US residents, of which 969 (98.7%) were US nationals and the remaining 13 (1.3%) from other nationalities. The median completion time was 11 minutes. There were no significant differences in the socio-demographical composition of the sample across the two conditions (see SMC3).

Materials and Procedure

The study had received prior ethical approval by the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University (UK).

Participants were recruited through the MTurk crowdsourcing platform and invited to follow a link to an online questionnaire hosted by Qualtrics (see SMC2). They read the participation sheet, learned that they could withdraw at any time without giving a reason, and provided informed consent to voluntarily participate.

Next, participants answered the socio-demographic questions, the moral self-evaluation question, the motivational trait scales, and then viewed one vignette

randomly selected from a set of two (Francia or Nicholas). The two vignettes were tested with no modifications from Study 2. In either condition participants evaluated the moral action and the moral agent in the story through various appraisal questions and the self-regulation items; toward the end, they were given the option to make comments in an open-ended question. Following this, they were debriefed and thanked for their participation, and received payment after the provision of their unique survey code.

Measures

The measures were essentially the same as in Study 2. The only minor differences consisted in the deletion of the self-esteem scale, the additional regulatory focus items, and three moral self-regulation items. Following the exploratory stage results, the cognitive/conative and affective items of the moral self-regulation scales were merged, thus forming the broadened constructs of moral self-improvement and self-defence shown in Tables C1-C2 (with the new labels). These two revised scales constituted the new moral self-regulation inventory that the present study set out to assess.

Table C1:
Moral self-improvement items tested in Study 3 (Francia's vignette)

Code/label	Item wording
SI1/Admirable	Actions like this are truly admirable
SI2/Awakened	When I read these stories, I feel awakened to the good in the world
SI3/Humanity	This story strengthens my faith in humanity
SI4/BeBetter	Francia has shown me how to be a better person
SI5/Values	Francia and I share the same values
SI6/BeLike	I want to be more like Francia
SI7/ForOthers	I feel like I want to do something good for others
SI8/Uplifted	I felt uplifted
SI9/Inspired	I was inspired by the story
SI10/Moved	I was moved
SI11/Proud	I felt proud of what Francia did
SI12/Happy	It made me feel happy

Table C2:
Moral self-defence items tested in Study 3 (Francia's vignette)

Code/label	Item wording
SD1/MeGreater	In many ways, I have done greater deeds than Francia
SD2/Devalue	This is not an act that I value all that much
SD3/Ordinary	It's not such an extraordinary action
SD4/People	I know people who have done greater deeds than Francia
SD5/Praise	Francia may have done a good deed, but I bet she is seeking the praise of others
SD6/Ulterior	Francia may have had ulterior motives for doing this
SD7/Uncomfortable	It makes me uncomfortable to dwell on these stories
SD8/Uneasy	I would feel uneasy if I had to interact with Francia
SD9/Uncommon	Actions uncommon as this one should not be considered the standard we live by
SD10/NoPraise	Everyone occasionally does something really good, so Francia isn't more praiseworthy than anybody else
SD11/Superior	Francia probably thinks she's better than everyone else
SD12/Seriously	Francia takes herself too seriously
SD13/Conflicted	I had a mix of conflicting feelings
SD14/Detached	I felt detached
SD15/Vulnerable	I felt vulnerable
SD16/Threatened	I felt as if I was threatened by something
SD17/Guilty	It made me feel guilty
SD18/Resentful	I felt resentful
SD19/Irritated	The story irritated me

For what concerns the characteristic adaptations, Study 3 used:

- the twelve items of the approach and avoidance temperament scales (Elliot & Thrash, 2010);

- the original eleven items from the regulatory focus questionnaire (Higgins et al., 2001).

The other key measures, for example moral self-evaluation, agent evaluation, goodness and propriety of the deed, and so forth did not change from the previous study.

Analytic Approach

Study 3 represented the initial step of the second stage of the research plan (Fig. IV): the confirmatory/integrative stage. The final dataset was analysed using R ver. 3.6 (R Core Team, 2020), RStudio ver. 1.3 (RStudio Team, 2020)⁵³, and IBM SPSS Statistics ver. 25-26.

Data analysis comprised the usual preliminary inspection of descriptive statistics and assumptions for statistical testing and estimation, and then – following the preregistered plan – primary and secondary analyses were conducted:

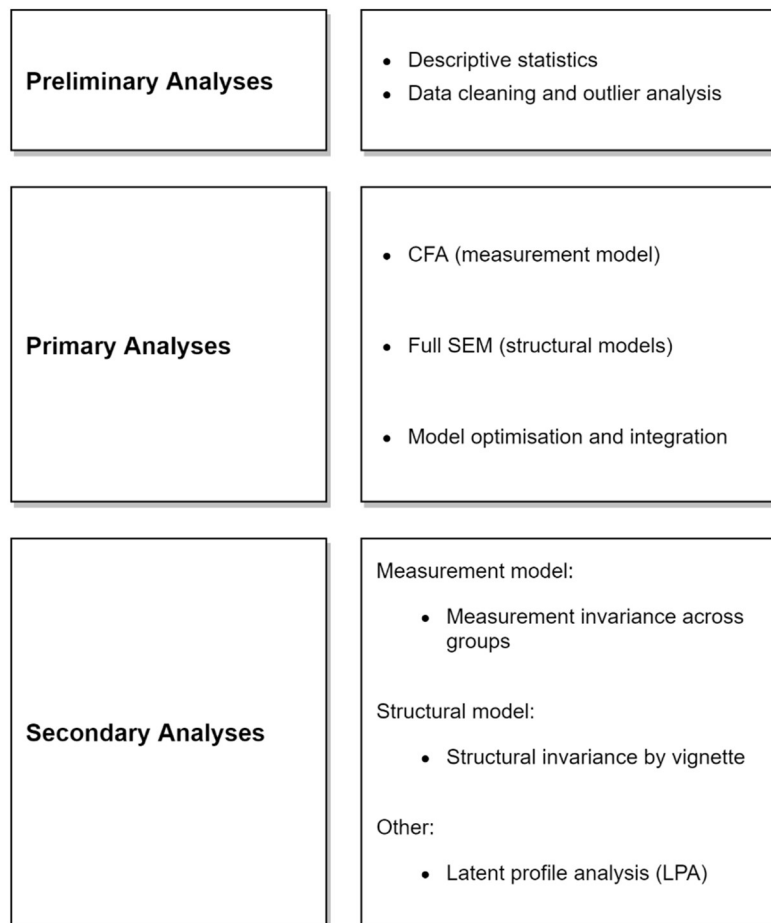
- *primary analyses* encompassed confirmatory factor analysis and full structural equation modelling (testing respectively the hypothesised measurement and structural models through confirmatory methods) as well as strategies for optimising/integrating the models;
- *secondary analyses* comprised further investigation of the measurement invariance across groups (at both measurement and structural level) and latent profile analysis (LPA) to identify clusters of participants with specific characteristics⁵⁴.

Figure C1 summarises the analytic approach, detailing the stages and the tools employed in the present study.

⁵³ Specific R packages used for analysis and visualisation are referenced in text in the following sections.

⁵⁴ LPA was mentioned as a possible secondary analysis in the preregistration of Study 2, but it was considered more appropriate for Study 3, applied to a larger sample and the final measurement model.

Figure C1:
Summary of the analytic approach adopted in Study 3



It must be emphasised again that in the primary analyses, regardless of the goodness of the models initially fitted, *alternative* models (optimised and integrated) were generated and compared to the hypothesised ones. Features of the alternative models include, for instance, removal of weaker items from the latent variables, as well as inclusion or exclusion of specific variables and pathways in the structural models. Regarding the secondary analyses, further methodological details about latent profile modelling appear in the next section.

Results and Preliminary Reflections

The dataset from Study 3 was virtually complete. Only thirty-nine participants did not answer the political orientation questions (not a critical variable in the model).

Descriptive Statistics

The central tendency measures and the distributions of the socio-demographical variables, the moral self-evaluation, the motivational dispositions, and the primary moral appraisals exhibited very similar patterns to those observed in Study 2, providing further evidence of the stability of these measurements. For the sake of brevity, only the mean scores and standard errors for some of the main variables are reported in Figures C2-C6; the full analysis is documented in SMC3-SMC5.

Figure C2:

Mean scores for the socio-demographical variables and the moral self-evaluation (total sample), with bootstrap standard errors (BCa 95% CI)

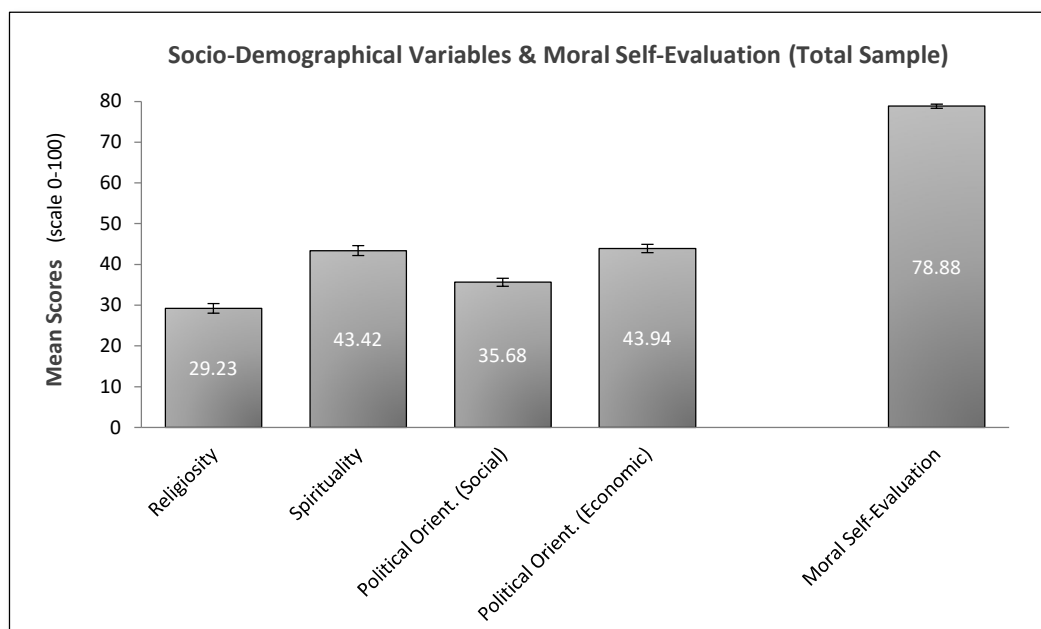
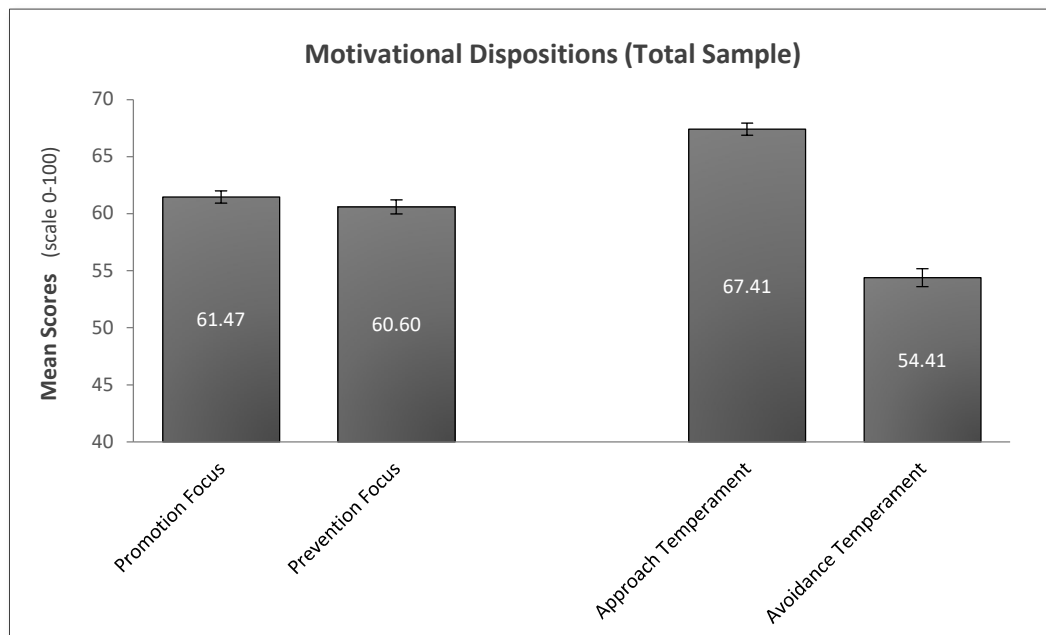


Figure C3:

Mean scores for the motivational disposition scales (total sample), with bootstrap standard errors (BCa 95% CI)

**Figure C4:**

Mean scores for the moral appraisal variables by vignette, with bootstrap standard errors (BCa 95% CI)

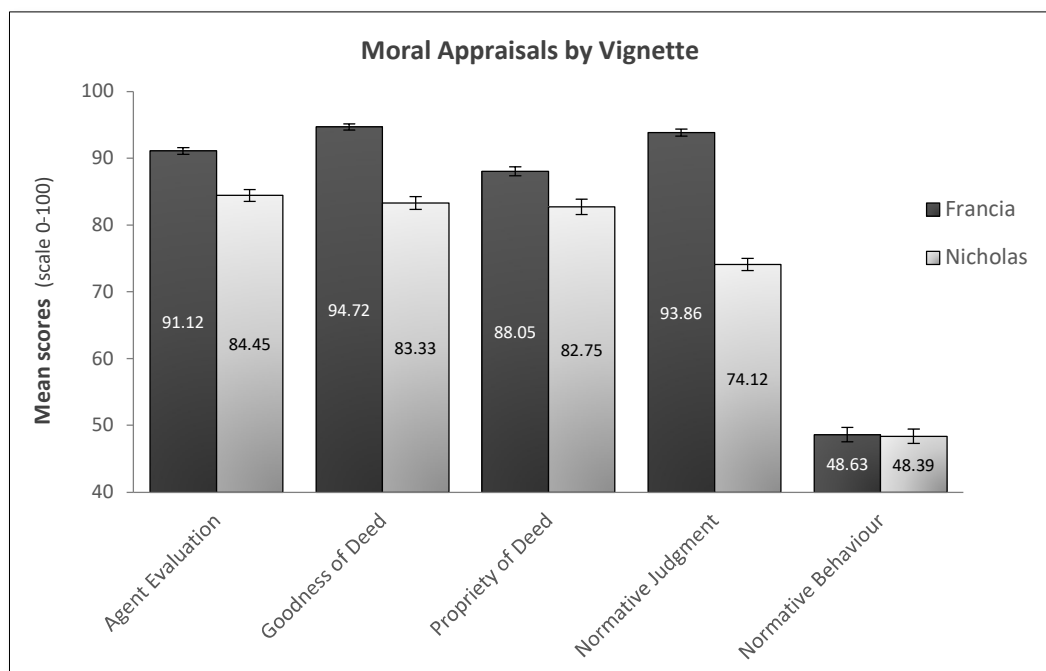
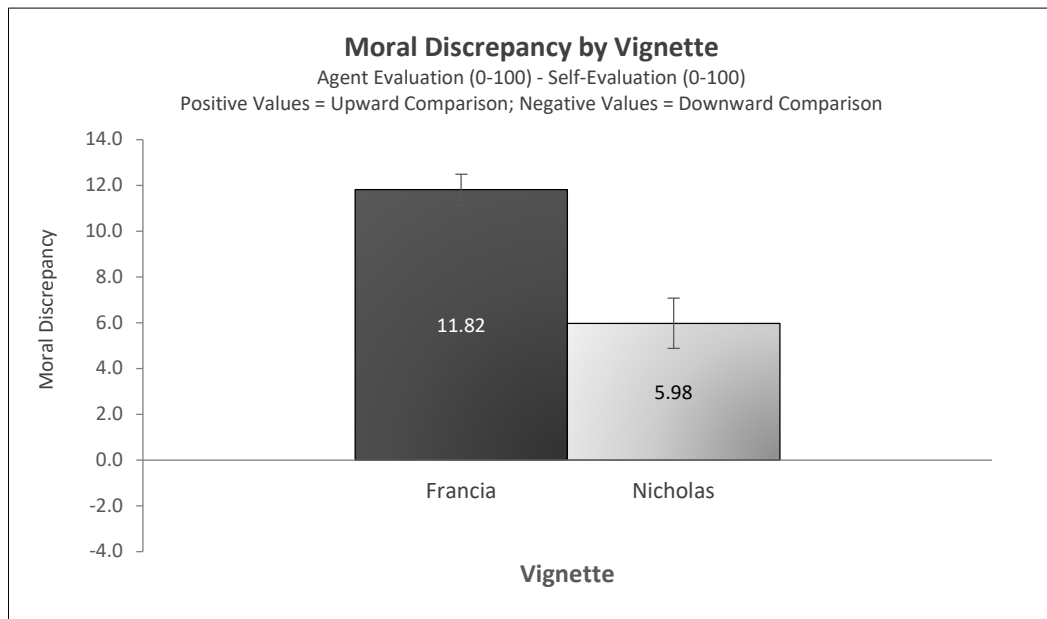
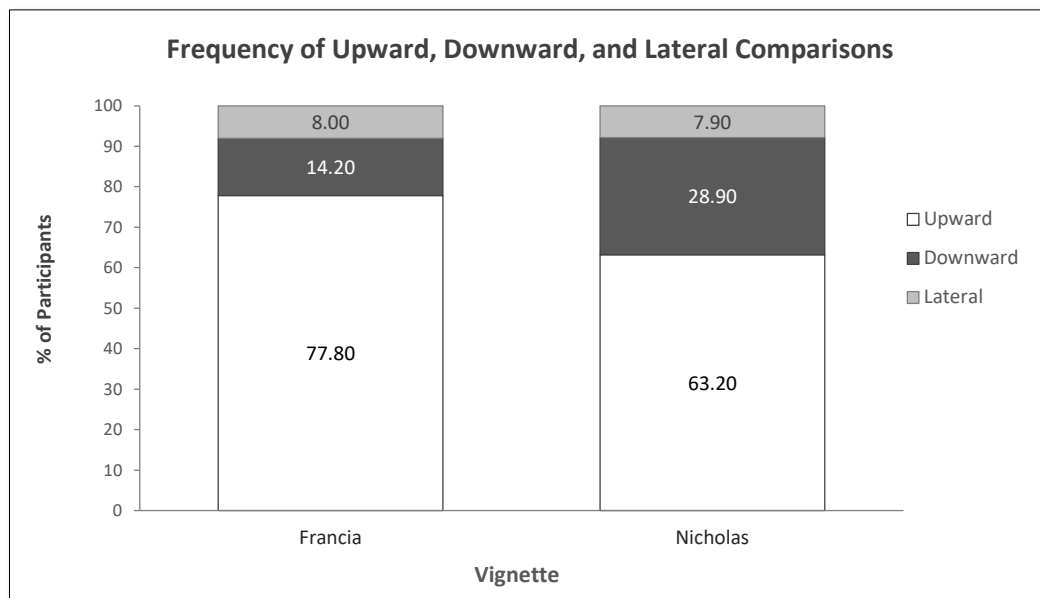


Figure C5:

Mean scores for the moral discrepancy by vignette, with bootstrap standard errors (BCa 95% CI)

**Figure C6:**

Frequency of type of moral comparison by vignette: clustered bar chart of frequency distributions (%)



The composite mean scores of the moral self-improvement and self-defence scales were respectively 69.81 ($SD = 23.60$) and 16.79 ($SD = 18.63$). The moral self-regulation items reproduced similar patterns to those observed in Study 2. The mean scores and the standard errors are illustrated in Figures C7-C8, while the full descriptive statistics are available in SMC6-SMC7.

Figure C7:

Mean scores of the twelve self-improvement items for the two vignettes, with bootstrap standard errors (BCa 95% CI)

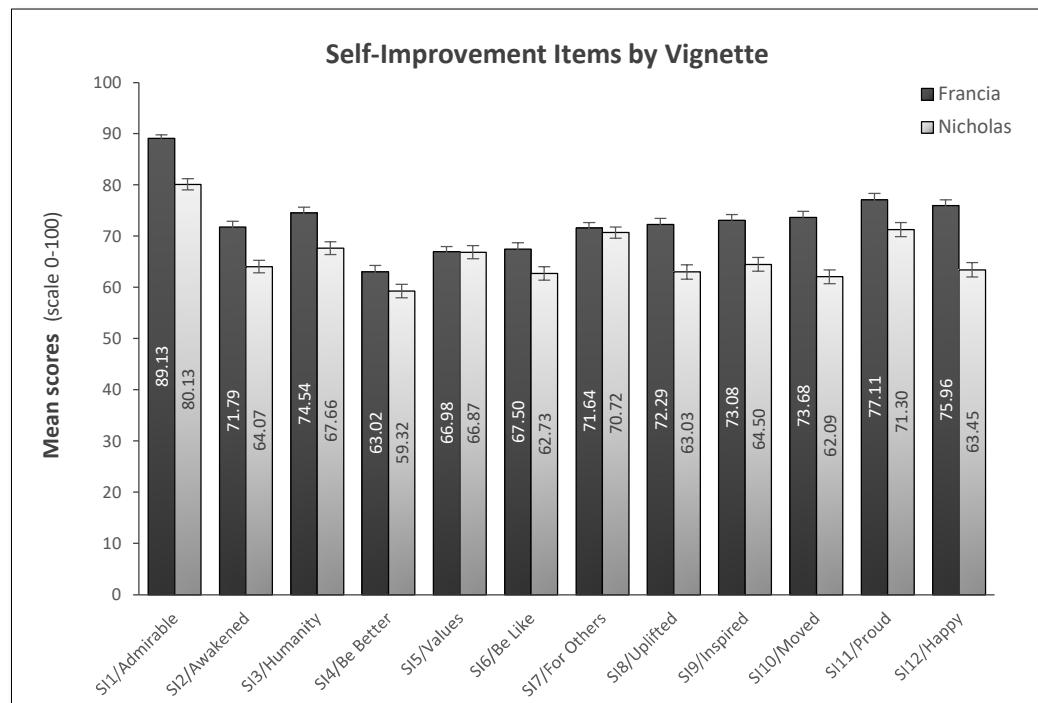
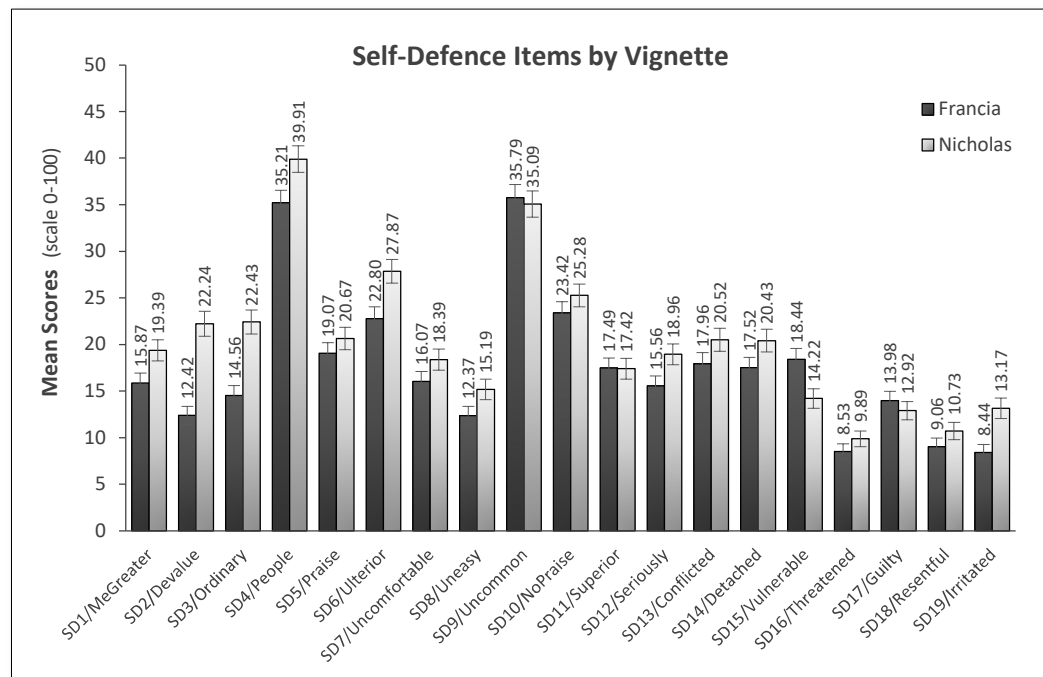


Figure C8:

Mean scores of the nineteen self-defence items for the two vignettes, with bootstrap standard errors (BCa 95% CI)



Primary Analyses

The primary analyses included confirmatory testing of the *measurement* model (the latent variables) as well as the *structural* models (the mutual relationships across the variables). As in Study 2, confirmatory factor analysis (CFA) and full structural equation modelling (SEM) were conducted with the R package lavaan ver. 0.6-6, using the robust maximum likelihood (MLR) estimator (unless otherwise stated), with regression weights tested 2-tailed at 95% c.i. Complementary analyses (e.g., internal consistency of the scales) utilised other R packages, such as psych ver. 1.9.12 (Revelle, 2019).

The assessment of the goodness-of-fit followed the criteria set out in the preregistration. As for Study 2, the fitted models were assessed through the joint analysis of the following key fit indices (“scaled” whenever appropriate, given the use of the MLR estimator): RMSEA and its *p*-value, CFI, TLI, GFI, and AGFI (the chi-square

statistic and its p -value were reported for the sake of completeness). Models were deemed to show *satisfactory* fit if the scaled RMSEA were non-significant and lower than .080 (*excellent* fit if lower than .050⁵⁵) and if CFI, TLI and GFI were greater than .900 (*excellent* fit if greater than .950).

Testing the Measurement Model: Confirmatory Factor Analysis

To test goodness-of-fit of the measurement model, CFA was carried out among the total sample for all the six latent variables of the conceptual model: the two endogenous variables of moral self-improvement and self-defence, and the four exogenous variables of promotion and prevention focus, approach and avoidance. Each of the two self-regulation latent variable models was fitted with the same unidimensional structure that emerged from EFA on the total sample in Study 2. The hypothesised structure of the latent variables is illustrated in Figures C9-C14.

⁵⁵ There was an unintentional oversight in the preregistration, where models were said to be satisfactory if the scaled RMSEA were non-significant and lower than .80 (instead of the correct value of .080), and excellent if lower than .50 (instead of the correct value of .050).

Figure C9:

Hypothesised model of the latent variable self-improvement and its twelve cognitive/conative/affective observed indicators

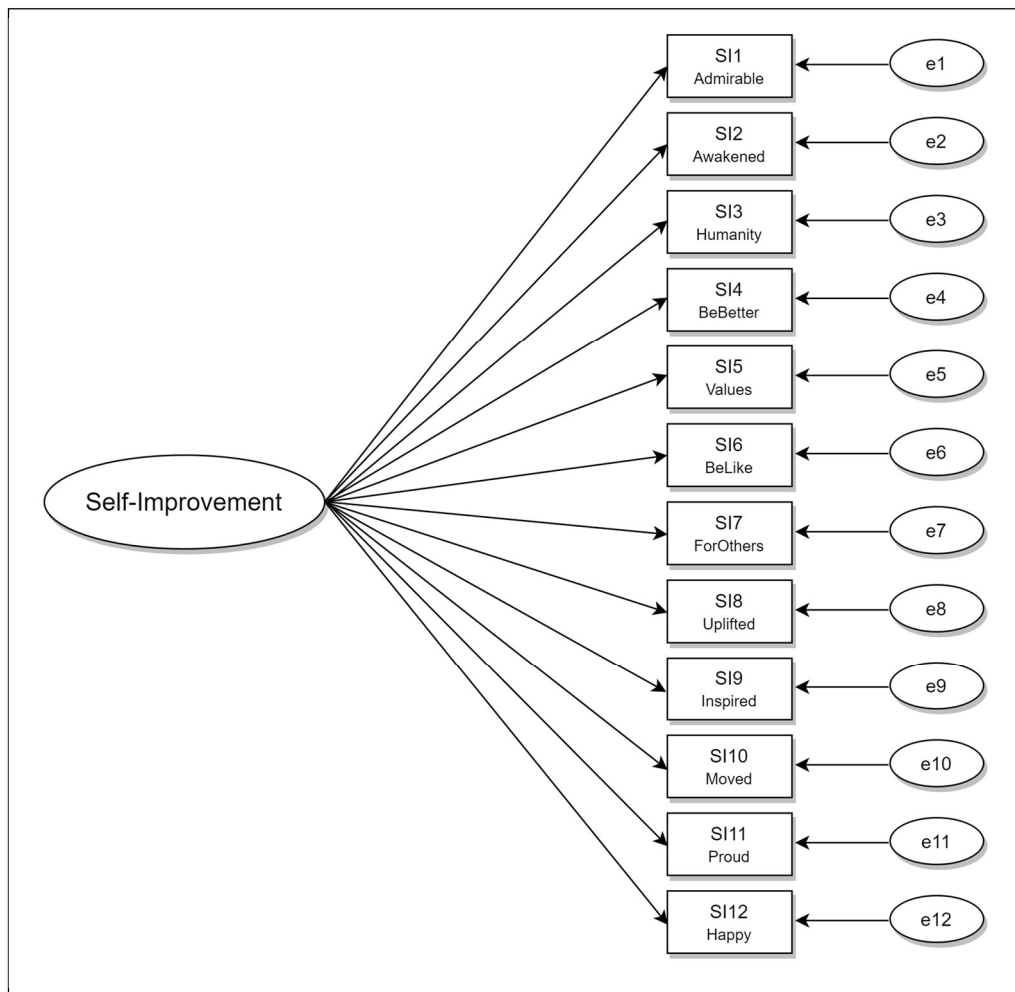


Figure C10:

Hypothesised model of the latent variable self-defence and its nineteen cognitive/conative/affective observed indicators

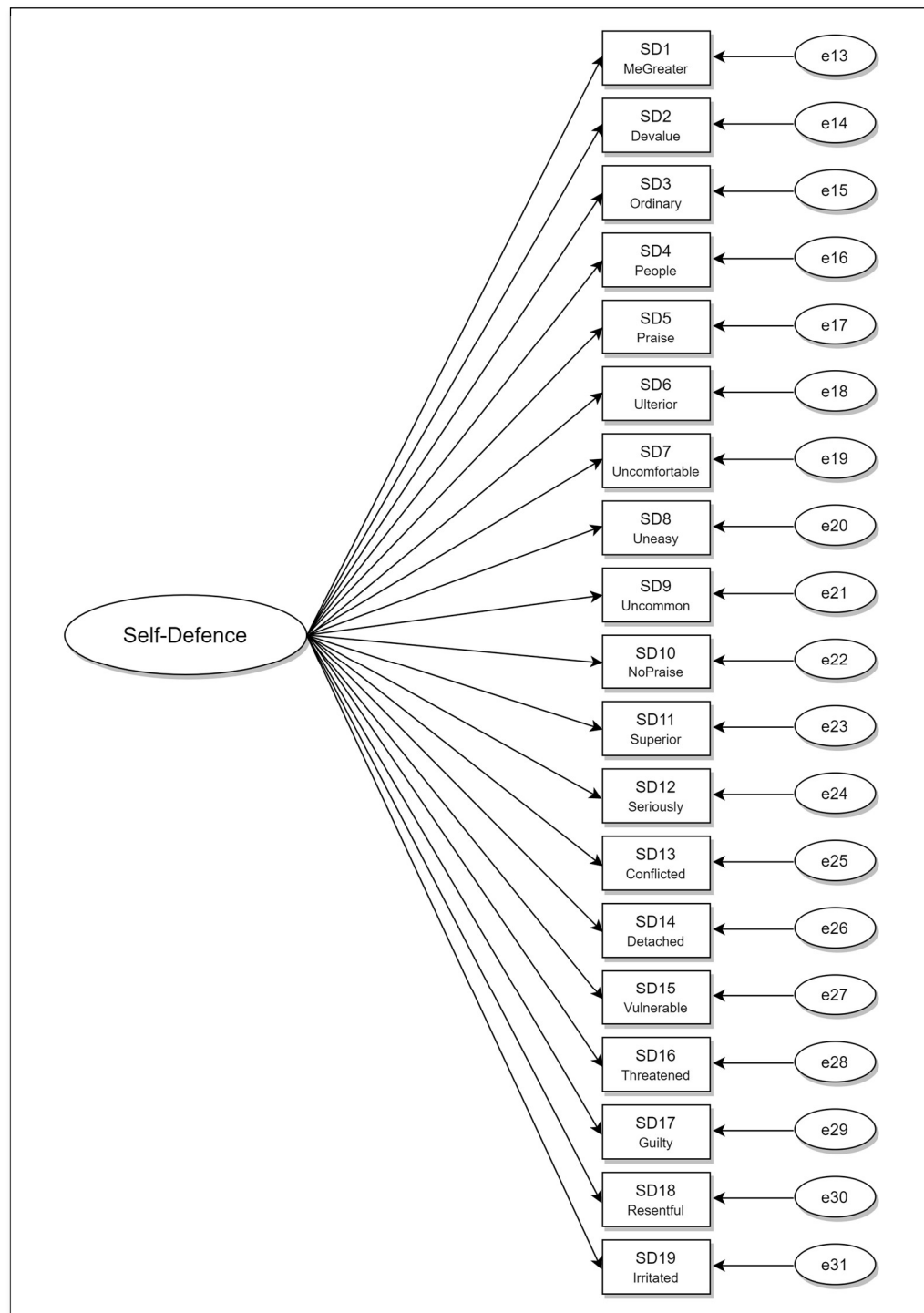
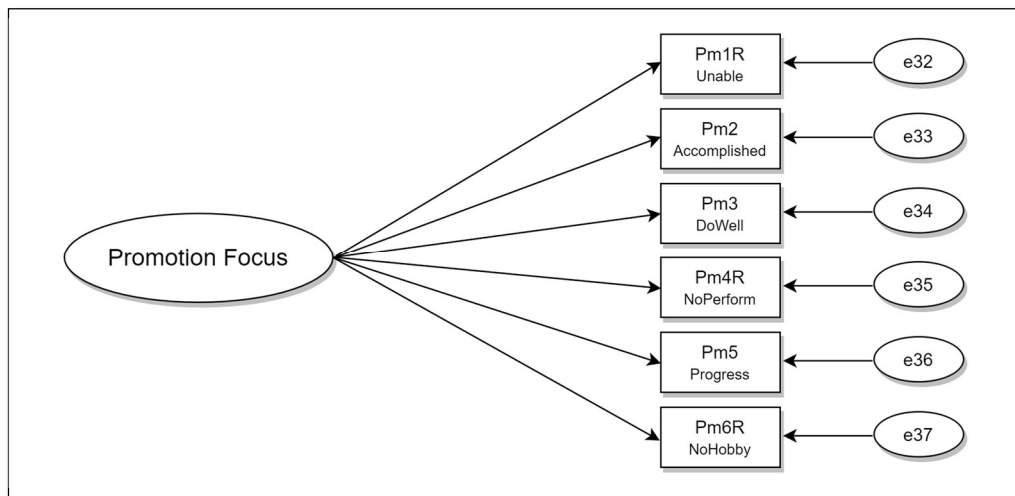


Figure C11:

Hypothesised model of the latent variable promotion focus and its six observed indicators

**Figure C12:**

Hypothesised model of the latent variable prevention focus and its five observed indicators

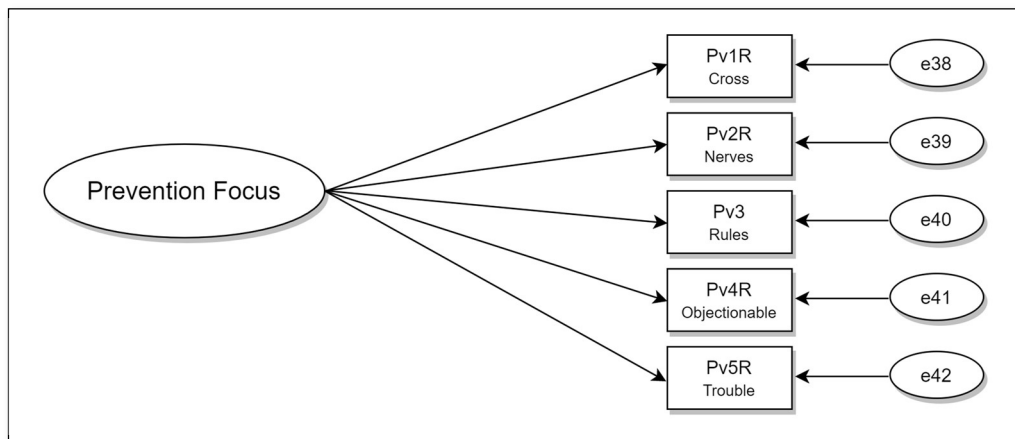
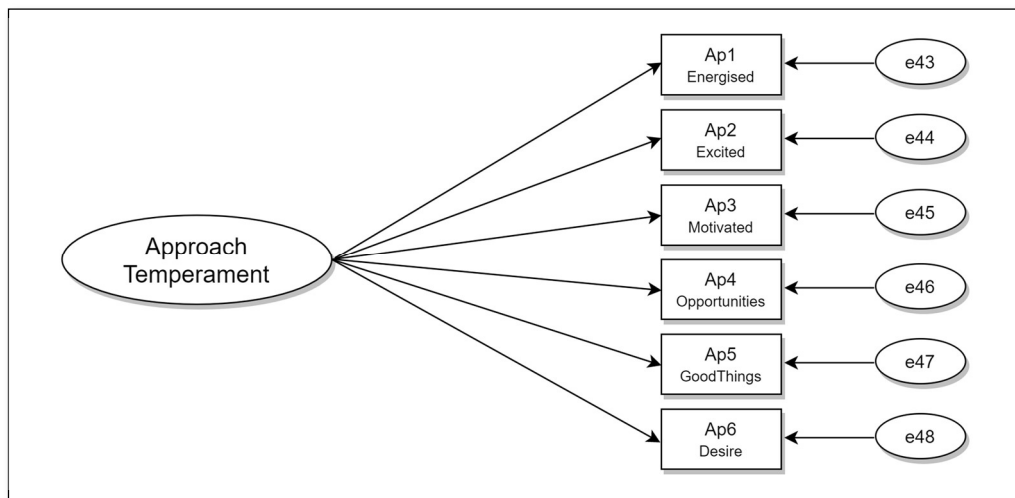
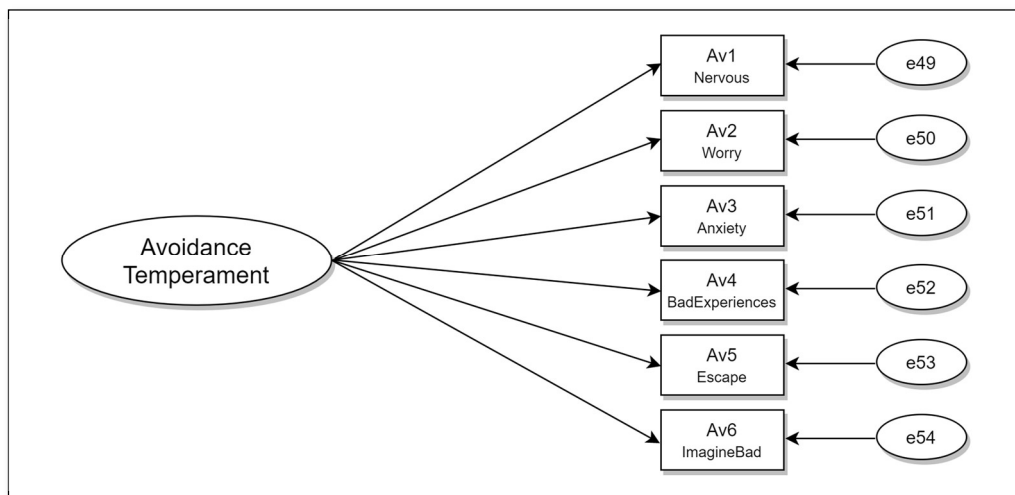


Figure C13:

Hypothesised model of the latent variable approach temperament and its six observed indicators

**Figure C14:**

Hypothesised model of the latent variable avoidance temperament and its six observed indicators



Results from CFA for the endogenous latent variables exhibited satisfactory RMSEA values, below the cutoff point of .080, for both self-improvement and self-defence, but with significant p -values (see Tab. C3); the other fit indices were nearly adequate for self-improvement, but not acceptable for self-defence. These results suggested rejection

of the models, despite the strong internal consistency, as shown by Cronbach's alpha and McDonald's omega greater than .90 for both self-improvement and self-defence (Tab. C4).

Table C3:

Main fit indices from CFA for the endogenous latent variables (total sample)

	Self-improvement	Self-defence
Chi-Square (scaled)	372.714	921.003
df (scaled)	54	152
<i>p</i> -value (scaled)	<.001	<.001
RMSEA (scaled)	.078	.072
<i>p</i> -value (scaled)	<.001	<.001
CFI (scaled)	.945	.861
TLI (scaled)	.933	.843
GFI	.976	.851
AGFI	.961	.619

Table C4:

Internal consistency of the endogenous latent variables (total sample)

	Self-improvement	Self-defence
Cronbach's alpha	.953	.933
McDonald's omega	.956	.934

The factorial structure of the endogenous latent variables is one of the reasons that accounts for the overall insufficient goodness-of-fit. Self-improvement showed strong loadings for nearly all indicators (Tab. C5a), but two were not above the desirable level of .700, namely SI7/ForOthers and SI5/Values. These items were candidates for deletion following Study 2, where they had already factor loadings below .700.

The picture looked less satisfactory for self-defence: nine items out of nineteen showed factor loadings below .700, some of them in the region of .400 and .500, for instance SD15/Vulnerable, SD9/Uncommon, SD13/Conflicted; these indicators had poor factor loadings also in Study 2 and were already considered for deletion.

Tables C5a-b:*Factor loadings from CFA for the endogenous latent variables (total sample)*

Tab. C5a	Self-improvement		Tab. C5b	Self-defence	
	Factor loading	Sig. (2-tailed)		Factor loading	Sig. (2-tailed)
SI1/Admirable	.720	<.001	SD1/MeGreater	.709	<.001
SI2/Awakened	.810	<.001	SD2/Devalue	.749	<.001
SI3/Humanity	.836	<.001	SD3/Ordinary	.693	<.001
SI4/BeBetter	.757	<.001	SD4/People	.524	<.001
SI5/Values	.700	<.001	SD5/Praise	.773	<.001
SI6/BeLike	.772	<.001	SD6/Ulterior	.702	<.001
SI7/ForOthers	.556	<.001	SD7/Uncomfortable	.690	<.001
SI8/Uplifted	.876	<.001	SD8/Uneasy	.772	<.001
SI9/Inspired	.893	<.001	SD9/Uncommon	.418	<.001
SI10/Moved	.873	<.001	SD10/NoPraise	.698	<.001
SI11/Proud	.832	<.001	SD11/Superior	.785	<.001
SI12/Happy	.871	<.001	SD12/Seriously	.765	<.001
			SD13/Conflicted	.543	<.001
			SD14/Detached	.566	<.001
			SD15/Vulnerable	.415	<.001
			SD16/Threatened	.740	<.001
			SD17/Guilty	.487	<.001
			SD18/Resentful	.757	<.001
			SD19/Irritated	.769	<.001

The other reason why the RMSEA p -values were significant became clear when analysing the measurement error within these latent variables. Several error terms covaried significantly, contributing to generating unexplained variance that resulted in the poor fit indices (for the complete CFA output, see SMC8). Potentially, the problems related to low factor loadings and correlated errors could be addressed by simplifying the moral self-regulation inventory and dropping a few unsatisfactory indicators.

CFA for the exogenous latent variables was also carried out. The regulatory focus and hedonic orientation scales exhibited acceptable but not entirely satisfactory psychometric properties. Prevention focus, approach, and avoidance showcased better results, with the RMSEA always below .080 (but with a significant p -value for avoidance).

Promotion focus exhibited less desirable fit indices, with the RMSEA above the cutoff point for acceptance of .080 and significant p -value, suggesting rejection of the model.

Reliability analysis showed similar patterns, with satisfactory values of Cronbach's alpha and McDonald's omega for prevention focus, approach, and avoidance, and weaker values for promotion focus (Tab. C6-C7).

Table C6:

Main fit indices from CFA for the exogenous latent variables (total sample)

	Promotion Focus	Prevention Focus	Approach	Avoidance
Chi-Square (scaled)	109.661	8.275	21.678	48.226
df (scaled)	9	5	9	9
p -value (scaled)	<.001	.142	.010	<.001
RMSEA (scaled)	.107	.026	.038	.067
p -value (scaled)	<.001	.930	.867	.036
CFI (scaled)	.893	.998	.992	.981
TLI (scaled)	.822	.995	.985	.969
GFI	.994	.999	.999	.993
AGFI	.981	.998	.996	.979

Table C7:

Internal consistency of the exogenous latent variables (total sample)

	Promotion Focus	Prevention Focus	Approach	Avoidance
Cronbach's alpha	.724	.838	.850	.898
McDonald's omega	.725	.856	.851	.903

One of the reasons for the lower levels of fit and reliability of promotion focus was the low factor loadings of two items (Tab. C8a): Pm6R/Hobbies (.373) and Pm1R/Unable (.452). Despite being part of a validated scale, widely used in psychological research, these items had to be removed from the promotion focus scale to make the latent variable model satisfactory.

The factor loadings of all the other indicators of the motivational dispositions were acceptable (Tab. C8b-d). Some of them barely exceeded the value of .500 (e.g., Pv5R/Trouble in the prevention focus scale) and certainly contributed to the measurement error. However, these factor loadings were not so small as to justify their removal from such short, validated scales. The modification indices also prompted other possible actions to improve their goodness-of-fit (full output of the initial CFA for the exogenous latent variables available in SMC9).

Tables C8a-d:

Factor loadings from CFA for the exogenous latent variables (total sample)

Tab. C8a	Promotion Focus		Tab. C8b	Prevention Focus	
	Factor loading	Sig. (2-tailed)		Factor loading	Sig. (2-tailed)
Pm1R/Unable	.452	<.001	Pv1R/Cross	.880	<.001
Pm2/Psyched	.602	<.001	Pv2R/Nerves	.730	<.001
Pm3/DoWell	.599	<.001	Pv3/Rules	.558	<.001
Pm4R/NoPerform	.638	<.001	Pv4R/Object	.901	<.001
Pm5/Progress	.736	<.001	Pv5R/Trouble	.509	<.001
Pm6R/NoHobby	.373	<.001			

Tab. C8c	Approach		Tab. C8d	Avoidance	
	Factor loading	Sig. (2-tailed)		Factor loading	Sig. (2-tailed)
Ap1/Energised	.796	<.001	Av1/Nervous	.874	<.001
Ap2/Excited	.800	<.001	Av2/Worry	.769	<.001
Ap3/Motivated	.578	<.001	Av3/Anxiety	.881	<.001
Ap4/Opportunities	.666	<.001	Av4/BadExperiences	.697	<.001
Ap5/GoodThings	.642	<.001	Av5/Escape	.670	<.001
Ap6/Desire	.744	<.001	Av6/ImagineBad	.690	<.001

Overall, the results from CFA revealed satisfactory fit for several latent variables, but not for all, particularly promotion focus and self-defence; the data also indicated that goodness-of-fit and reliability could improve with a few data-driven modifications. It was therefore decided to remove items Pm1R and Pm6R from the promotion focus

scale, items SI5, SI7, SD4, SD7, SD9, SD13, SD14, SD15, and SD17 from the moral self-regulation inventory, and allow the specification of a few error covariances across self-regulation indicators (see next).

Strengthening the Measurement Model. A second run of CFA was carried out on the revised latent variables with a view to enhancing their overall goodness-of-fit and expecting improvements particularly for promotion focus and self-defence. Results from the second run of CFA displayed clear improvements to the goodness-of-fit for both the endogenous and exogenous latent variables. After the second run, both self-improvement and self-defence showcased excellence across virtually all the key fit indices, as shown in Table C9 (full output available in SMC10).

Table C9:

Comparison between the main fit indices from the two runs of CFA for the endogenous latent variables (total sample)

	Self-improvement		Self-defence	
	CFA first run	CFA second run	CFA first run	CFA second run
Chi-Square (scaled)	372.714	124.991	921.003	198.611
df (scaled)	54	33	152	50
<i>p</i> -value (scaled)	<.001	<.001	<.001	<.001
RMSEA (scaled)	.078	.053	.072	.055
<i>p</i> -value (scaled)	<.001	.237	<.001	.067
CFI (scaled)	.945	.981	.861	.953
TLI (scaled)	.933	.974	.843	.938
GFI	.976	.991	.851	.941
AGFI	.961	.982	.619	.894

Table C10:

Comparison between the internal consistency from the two runs of CFA for the endogenous latent variables (total sample)

	Self-improvement		Self-defence	
	CFA first run	CFA second run	CFA first run	CFA second run
Cronbach's alpha	.953	.955	.933	.935
McDonald's omega	.956	.956	.934	.936

The factorial structure of the optimised ten-item self-improvement and twelve-item self-defence scales is documented in Tables C11a-b. Only two self-defence indicators had factor loadings slightly below .700, but overall the scales showed better psychometric properties compared to the first run, including better internal consistency (see Tab. C10).

Tables C11a-b:

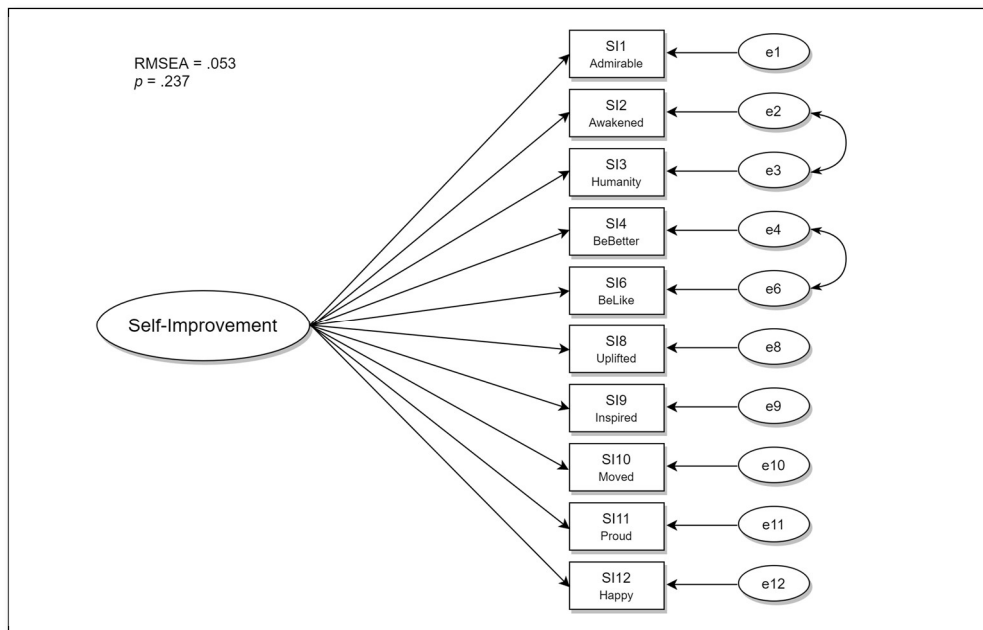
Factor loadings from the second run of CFA for the endogenous latent variables (total sample)

Tab. C11a			Tab. C11b		
Self-improvement			Self-defence		
	Factor loading	Sig. (2-tailed)		Factor loading	Sig. (2-tailed)
SI1/Admirable	.713	<.001	SD1/MeGreater	.724	<.001
SI2/Awakened	.795	<.001	SD2/Devalue	.748	<.001
SI3/Humanity	.820	<.001	SD3/Ordinary	.693	<.001
SI4/BeBetter	.741	<.001	SD5/Praise	.758	<.001
SI6/BeLike	.751	<.001	SD6/Ulterior	.685	<.001
SI8/Uplifted	.884	<.001	SD8/Uneasy	.772	<.001
SI9/Inspired	.898	<.001	SD10/NoPraise	.707	<.001
SI10/Moved	.881	<.001	SD11/Superior	.794	<.001
SI11/Proud	.832	<.001	SD12/Seriously	.775	<.001
SI12/Happy	.879	<.001	SD16/Threatened	.707	<.001
			SD18/Resentful	.735	<.001
			SD19/Irritated	.766	<.001

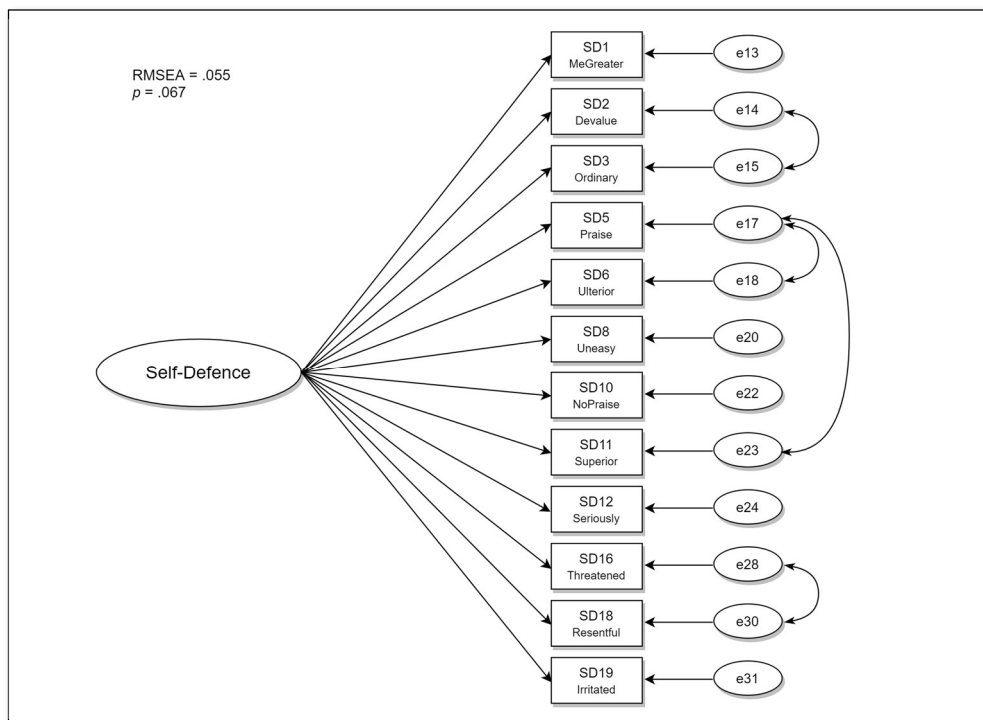
Other than the deletion of the items with poor factor loadings, the latent variables were also modified with the addition of two error covariances between self-improvement indicators and four error covariances between self-defence indicators. The factor structure of the revised latent variables, including the error covariances, are illustrated in Figures C15-C16.

Figure C15:

Unidimensional structure of the revised latent variable of self-improvement and its ten cognitive/conative/affective observed indicators with two error covariances

**Figure C16:**

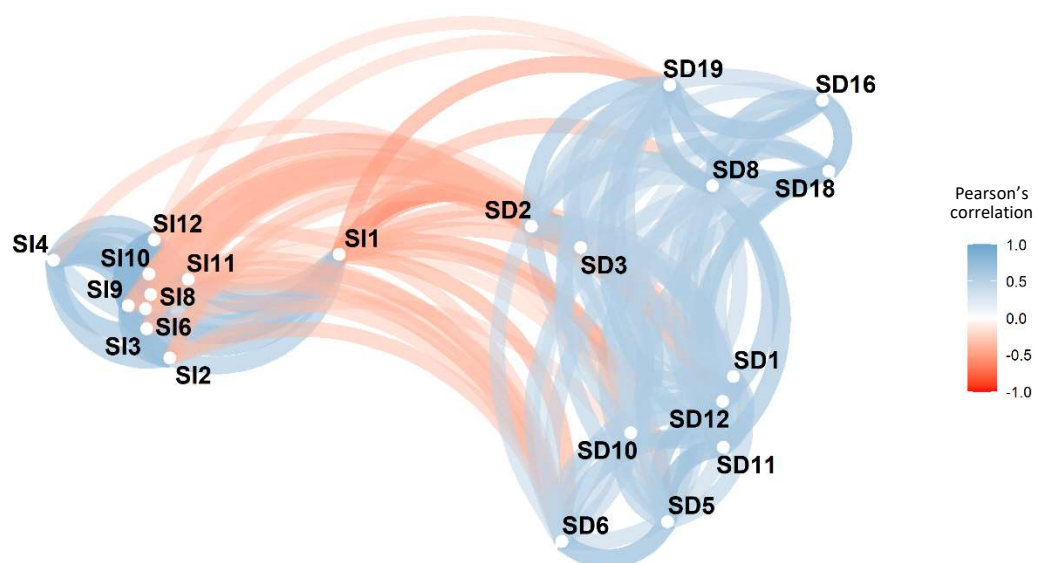
Unidimensional structure of the revised latent variable of self-defence and its twelve cognitive/conative/affective observed indicators with four error covariances



The structure of the correlations across all the twenty-two retained self-regulation indicators at total sample is graphically depicted in Figure C17, realised with the R package *corr* ver. 0.4.2 (Kuhn et al., 2020). The two “constellations” of self-improvement and self-defence can be easily recognised at the two sides of the graph. The item pairs whose error terms significantly covaried are located at the periphery of the two constellations, for instance SI2-SI3, SD5-SD6, and SD16-SD18. The items that most strongly correlated negatively (SI1-SD2) are themselves rather peripheral to their respective self-regulatory constellations. The items that had the strongest factor loadings on the self-improvement construct (SI8, SI9, and SI10) are approximately at the centre of the self-improvement constellation. By contrast, the self-defence constellation is more scattered, and the items with the highest loadings (SD11 and SD12) do not constitute a real barycentre toward which the other items gravitate.

Figure C17:

Graphical illustration of the zero-order correlations across the twenty-two retained moral self-regulation indicators (total sample)



Note: the graph shows Pearson's correlation coefficients greater than .30 in absolute value.

Regarding the exogenous latent variables, the second run of CFA enhanced the goodness-of-fit of both promotion focus and avoidance temperament. For promotion focus, beside the deletion of the two weak items, one error covariance between items Pm3-Pm5 was added, drastically improving the model fit; for avoidance, no items were deleted and the mere addition of one error covariance between items Av1-Av2 brought the RMSEA to non-significance (Tab. C12). These modifications also resulted in an improvement of Cronbach's alpha and McDonald's omega for promotion focus (obviously not for avoidance, which retained the same items: see Tab. C13).

Table C12:

Comparison between the main fit indices from the two runs of CFA for the exogenous latent variables of promotion focus and avoidance (total sample)

	Promotion Focus		Avoidance	
	CFA first run	CFA second run	CFA first run	CFA second run
Chi-Square (scaled)	109.661	.064	48.226	29.192
df (scaled)	9	1	9	8
<i>p</i> -value (scaled)	<.001	.801	<.001	<.001
RMSEA (scaled)	.107	<.001	.067	.052
<i>p</i> -value (scaled)	<.001	.952	.036	.395
CFI (scaled)	.893	1.000	.981	.990
TLI (scaled)	.822	1.008	.969	.981
GFI	.994	1.000	.993	.996
AGFI	.981	1.000	.979	.986

Table C13:

Comparison between the internal consistency from the two runs of CFA for the exogenous latent variables of promotion focus and avoidance (total sample)

	Promotion Focus		Avoidance	
	CFA first run	CFA second run	CFA first run	CFA second run
Cronbach's alpha	.724	.735	.898	.898
McDonald's omega	.725	.745	.903	.903

The factorial structure of the revised scales is documented in Tables C14a-b (full output of the second run of CFA for the exogenous latent variables available in SMC11).

Tables C14a-b:

Factor loadings from CFA for the endogenous latent variables (total sample)

Tab. C14a Promotion Focus			Tab. C14b Avoidance		
	Factor loading	Sig. (2-tailed)		Factor loading	Sig. (2-tailed)
Pm2/Psyched	.557	<.001	Av1/Nervous	.817	<.001
Pm3/DoWell	.763	<.001	Av2/Worry	.785	<.001
Pm4R/NoPerform	.510	<.001	Av3/Anxiety	.868	<.001
Pm5/Progress	.240	<.001	Av4/BadExperiences	.716	<.001
			Av5/Escape	.677	<.001
			Av6/ImagineBad	.707	<.001

The factorial structure of the two revised exogenous latent variables is illustrated in Figures C18-C19.

Figure C18:

Dimensional structure of the revised latent variable of promotion focus and its four observed indicators with one error covariance

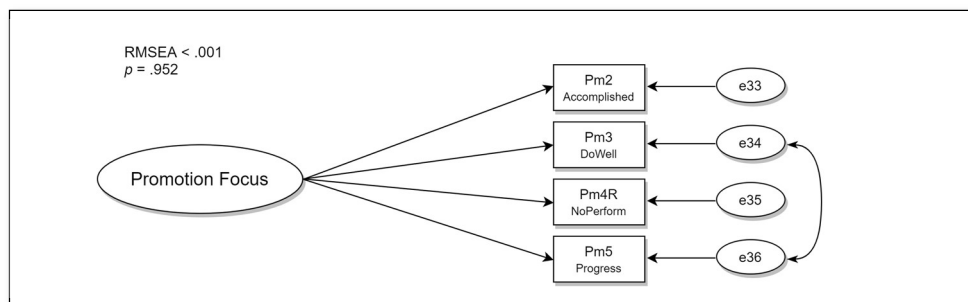
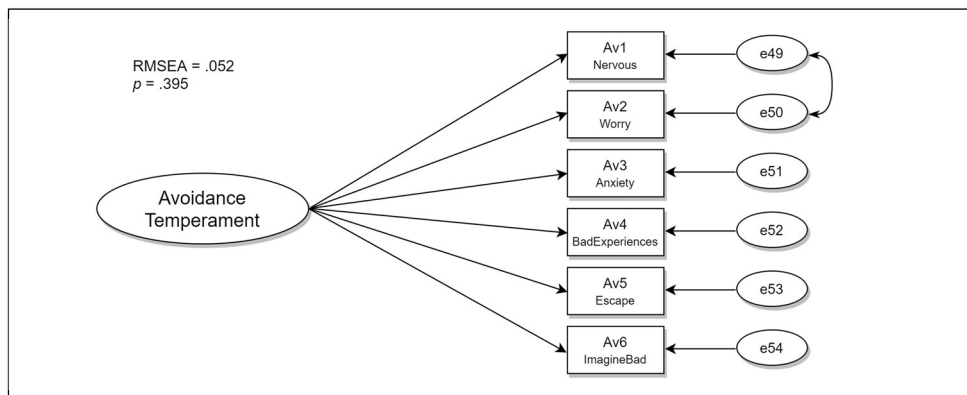


Figure C19:

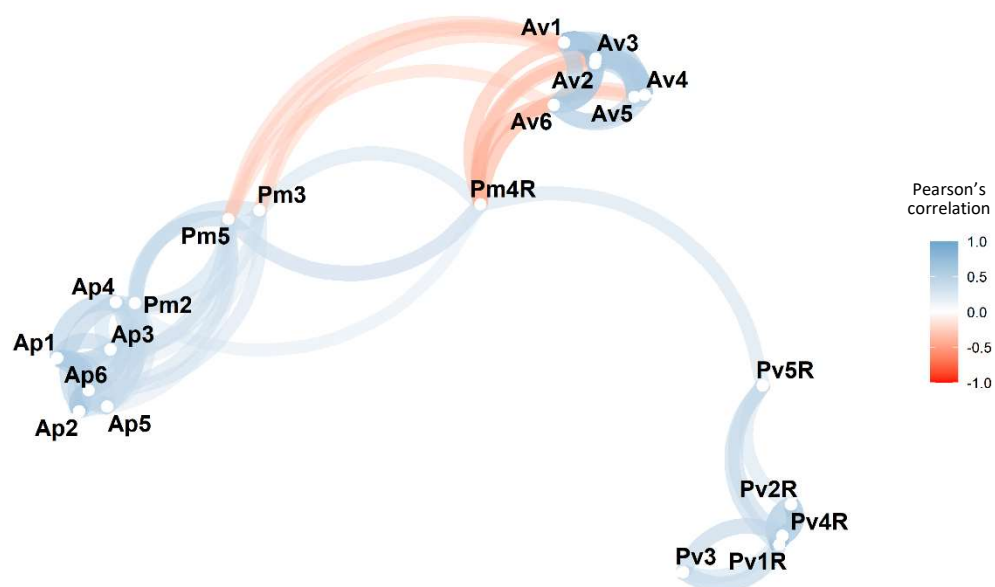
Dimensional structure of the revised latent variable of avoidance temperament and its six observed indicators with one error covariance



The structure of the correlations across all the retained twenty-one motivational indicators at total sample is graphically depicted in Figure C20.

Figure C20:

Graphical illustration of the zero-order correlations across the twenty-one retained motivational dispositions indicators (total sample)



Note: the graph shows Pearson's correlation coefficients greater than .30 in absolute value.

The two constellations of prevention focus and avoidance stand apart on the right-hand side of the graph, whereas promotion focus and approach tend to stick together to form one single constellation on the left-hand side. Even after the removal of the two items with poor loadings, promotion focus appears to be the least cohesive construct among the motivational dispositions, with its four remaining items clearly spreading wide in the graph with comparably weaker links: Pm2 very close to the approach items, Pm4 (reverse scored) on the opposite side forming a bridge towards prevention (positive correlation) and avoidance (negative correlation), and Pm3 and Pm5 (the two items with correlated errors) standing in between.

In sum, the *revised* CFA of the latent variables provided evidence of the validity and reliability of the *single* measurement model (at total sample level, common for the two vignette) for the six unidimensional latent constructs of the conceptual model. This measurement model was utilised to test the full structural models, separately for the two vignettes and together at total sample.

Testing the Structural Model: Full Structural Equation Modelling

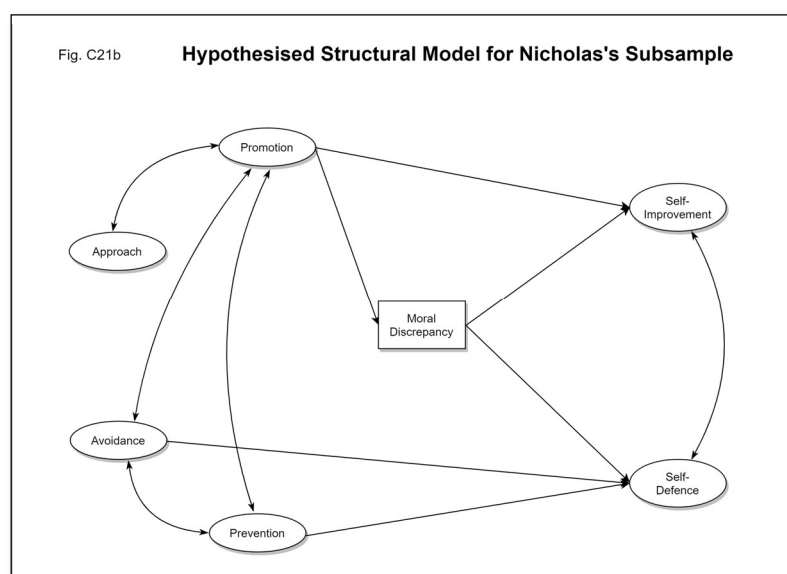
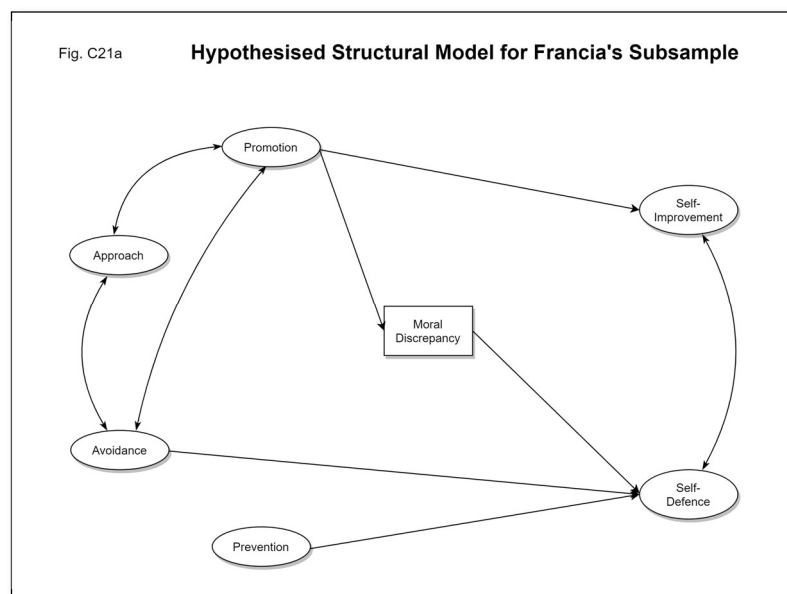
Initial Structural Models. Given lack of structural invariance in Study 2, *separate* structural models were hypothesised for each of the two vignettes. Two initial structural models were fitted using the *single* measurement model that emerged from CFA in the present study and specifying the significant pathways identified in Study 2. The hypothesised models are illustrated in Figures C21a-b. They differ from the models obtained as the final output of Study 2 in that here *latent* variables (instead of observed) were used for moral self-regulation and motivational dispositions⁵⁶; additionally, to make the confirmatory tests more stringent, in this initial step all the latent variables

⁵⁶ In this thesis, the usual convention adopted in SEM diagrams was followed, using rectangles to indicate observed variables and ellipses to indicate latent variables.

were fitted without allowing any covariance between error terms. The absence of a pathway linking moral discrepancy and self-improvement in Francia's vignette descends from the Study 2 findings, but the coefficient was expected to be significant in Study 3, as discussed in the previous chapter.

Figures C21a-b:

Hypothesised models for the two vignettes following Study 2. Indicators and errors/disturbances not displayed to facilitate legibility



The results showed that the goodness-of-fit of the fitted models was considerably improved compared to Study 2, yet it remained unsatisfactory. The RMSEA was close to the cutoff point for excellence for both vignettes, but the p -values were still significant; CFI and TLI were below the expected cutoff point of .900 (Tab. C15).

Table C15:

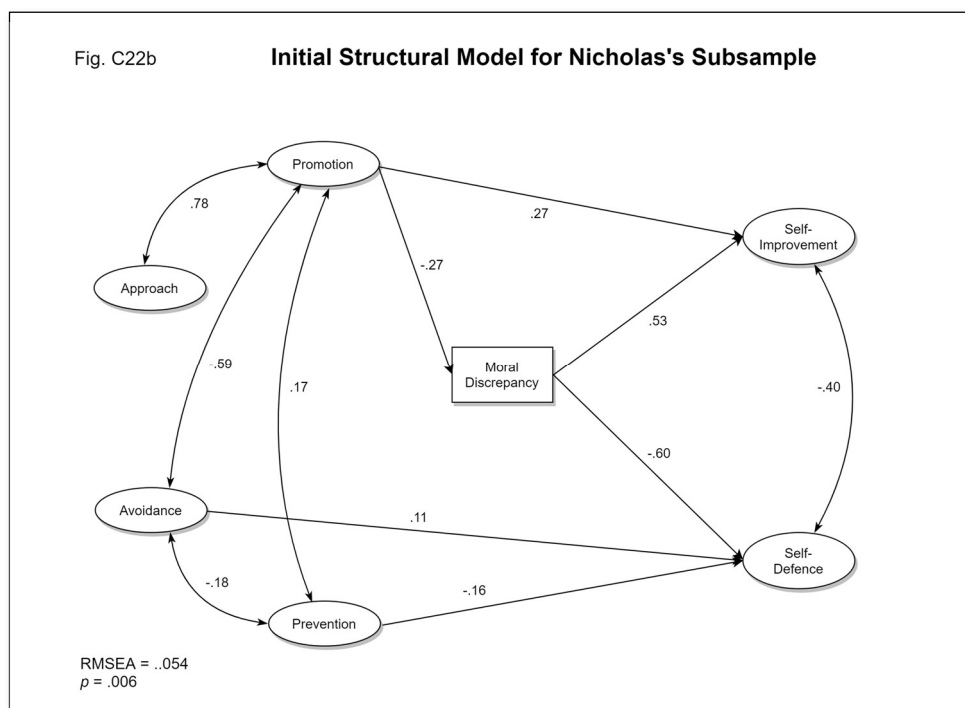
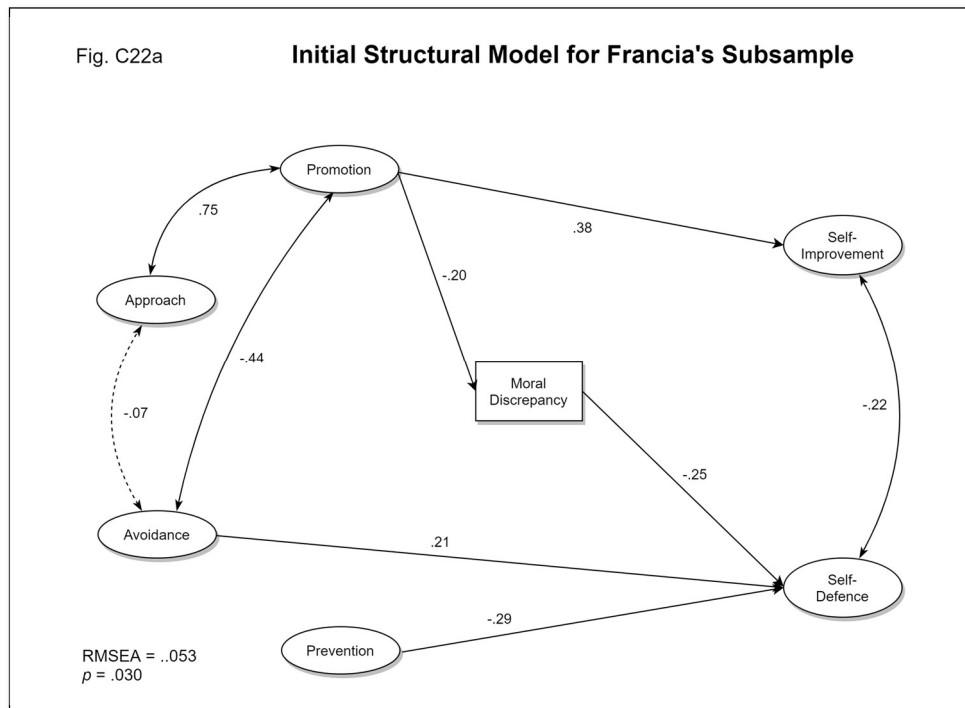
Main fit indices of the initial structural models for the two vignettes

	Francia	Nicholas
Chi-Square (scaled)	2119.883	2168.247
df (scaled)	891	890
p -value (scaled)	<.001	<.001
RMSEA (scaled)	.053	.054
p -value (scaled)	.030	.006
CFI (scaled)	.884	.893
TLI (scaled)	.877	.886
GFI	.956	.944
AGFI	.949	.935

A more detailed inspection of the fitted models confirmed the expected overall pattern of the regression weights, as illustrated in Figures C22a-b. The main inconsistency between hypothesised and fitted models was the non-significance of the covariance between approach and avoidance in Francia's vignette. The details of the full SEM output are available in SMC12.

Figures C22a-b:

Initial structural models for the two vignettes. Estimator: robust maximum likelihood (MLR). Standardised regression weights. Indicators and errors/disturbances not displayed to facilitate legibility



Note: the dotted line in panel "a" indicates a pathway hypothesised to be significant that turned out to be non-significant.

Enhancing the Structural Models. As a whole, the main indices still revealed that the data did not adequately fit the hypothesised models; however, a few simple theory- and data-driven modifications proved to be effective.

First and foremost, as expected, in Francia's vignette moral discrepancy significantly predicted self-improvement, so this pathway was added to the model. Second, in Francia's subsample a few covariances across the motivational dispositions were modified to adhere to the data: the non-significant covariance between approach and avoidance was removed; covariances between prevention focus and promotion focus, approach, and avoidance were added. No major changes were needed to the pathways in Nicholas's subsample, but one error covariance was allowed within two self-defence error terms (SD16-SD18). Third, the substitutability of promotion focus and approach in the prediction of self-improvement emerged again, but this time the larger sample showed a slightly stronger effect size for approach (with negligible impact on the fit indices), whereas in Study 2 they were equivalent; therefore, approach was fitted in the models as a direct predictor of self-improvement (replacing promotion), while promotion remained the better predictor of moral discrepancy.

With this rationale and the support of the modification indices, revised models were fitted for the two vignettes. The main fit indices are reported in Table C16, the path diagrams are illustrated in Figures C23a-b, and the full output of the SEM is available in SMC13.

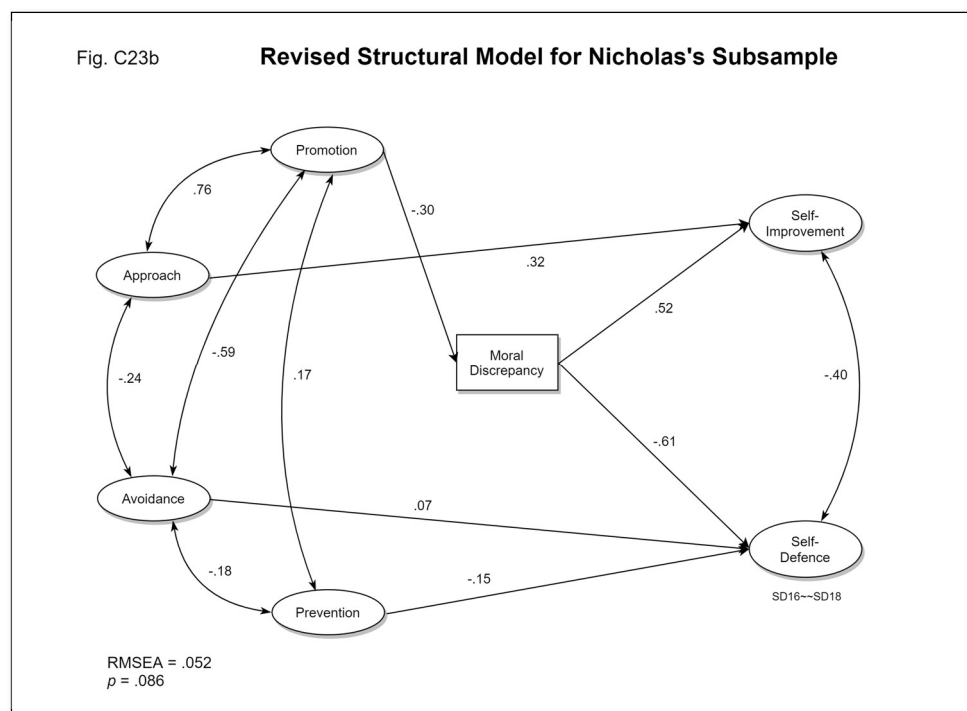
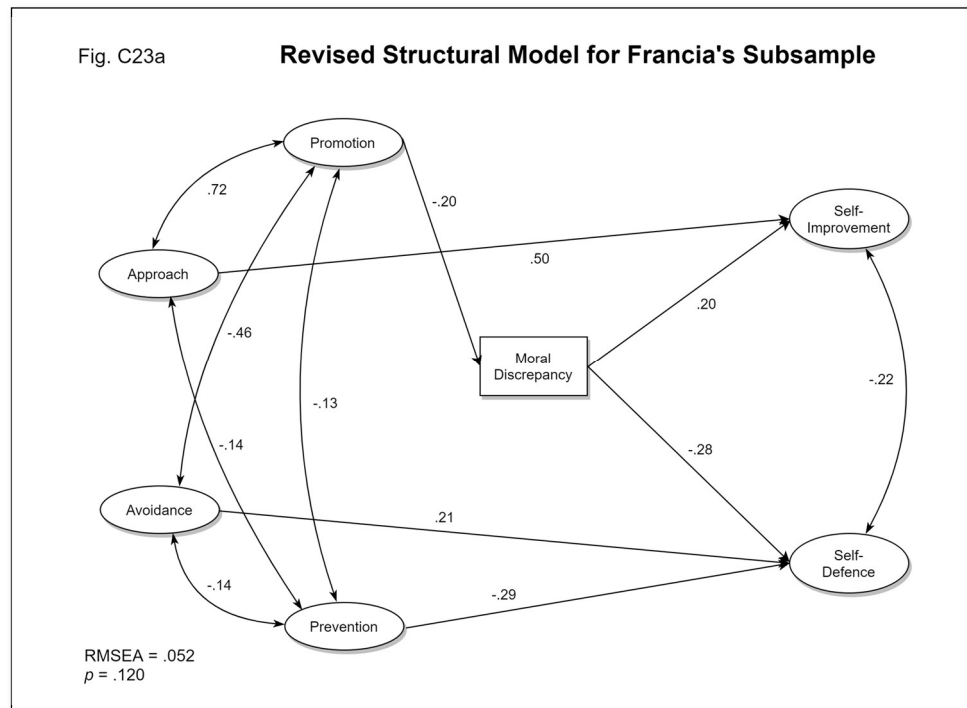
Table C16:*Main fit indices of the revised structural models for the two vignettes*

	Francia	Nicholas
Chi-Square (scaled)	2065.060	2081.104
df (scaled)	890	889
<i>p</i> -value (scaled)	<.001	<.001
RMSEA (scaled)	.052	.052
<i>p</i> -value (scaled)	.120	.086
CFI (scaled)	.889	.900
TLI (scaled)	.882	.894
GFI	.957	.946
AGFI	.950	.937

The fit indices of the revised models were satisfactory: the RMSEA was very close to the excellence cutoff point and its *p*-value non-significant; the other indices were for the most part adequate. However, the models still showed a critical weakness: the covariances between the two modes of moral self-regulation. These fairly large covariances of the outcome variables (-.26 for Francia and -.57 for Nicholas) must be intended as *residual* covariances, that is, covariance of the latent variable *disturbances*. The inclusion of further predictors currently not modelled could potentially enhance the models. Residual covariances were already present in the final output of path modelling in Study 2, but at that stage the use of observed variables and the smaller sample size suggested not drawing conclusive inferences. The fact that here the full structural models fitted among a larger sample using latent variables revealed these covariances again, with approximately the same effect sizes, unequivocally meant that the search for parsimony somewhat compromised the quality of the models, which is always the result of a balancing act between theory, parsimony, and fit.

Figures C23a-b:

Revised structural models for the two vignettes. Estimator: robust maximum likelihood (MLR). Standardised regression weights. Indicators, errors/disturbances and covariances across error terms are not graphically displayed to facilitate legibility



To integrate the models with relevant variables, it was necessary to go back to the theoretical framework: when narrowing down the conceptual model and nomological network, the key predictors of moral self-regulation were identified in characteristic adaptations and moral comparisons. Among the characteristic adaptations, self-esteem was abandoned due to evidence of redundancy. Moral comparisons based on *ability* were prioritised (operationalised with the moral discrepancy indicator), whereas those based on *opinions* were omitted from the models for the sake of parsimony; it is precisely this choice that now needed to be reconsidered. Seeking to match this theoretical consideration with empirical findings from the research, when re-examining the multiple linear regression in Study 2, it appeared that two influential variables predicting moral self-regulation *not* included in the current models were the judgments of the goodness and the propriety of the deeds (see Tab. B17a-b). As shown in Figure A14, these variables are those that subsume *opinion-based* comparisons. The convergence of these empirical and theoretical considerations clearly suggested that these variables could be the best candidates for inclusion in new *integrated* structural models.

From a conceptual perspective, in these integrated structural models opinion-based judgments/comparisons were hypothesised to precede ability-based comparisons. This is because it could be argued that agreement on the goodness and propriety of the deed could be viewed as a prerequisite for engaging in ability-comparison of one's moral virtue with that of the agent. In the absence of agreement, people would have fewer reasons to feel compelled to ask themselves if they would be so virtuous as the moral agent to perform the same deeds. Consequently, they would be less likely to self-regulate on that basis and, rather, would self-defend as a direct effect of opinion-based comparisons.

This conceptual framework was put to the test in new *integrated* structural models inclusive of goodness and propriety of the deeds as predictors of moral discrepancy and moral self-regulation. The plausibility of these assumptions would be confirmed by the emergence of significant pathways, improved fit indices, and reduced (or levelled) residual covariances between the outcome variables. The results showed that the hypothesised pathways between the moral comparisons were significant, the residual covariances of the self-regulatory constructs became very small and non-significant (see Fig. C24a-b), and the overall fit indices remained essentially the same (Tab. C17). Because satisfactory models not only account for goodness-of-fit but also limited error (Beatty et al., 2015), these new integrated models must be considered substantively and statistically better models than the previous ones even if the main fit indices remained largely unchanged.

Table C17:

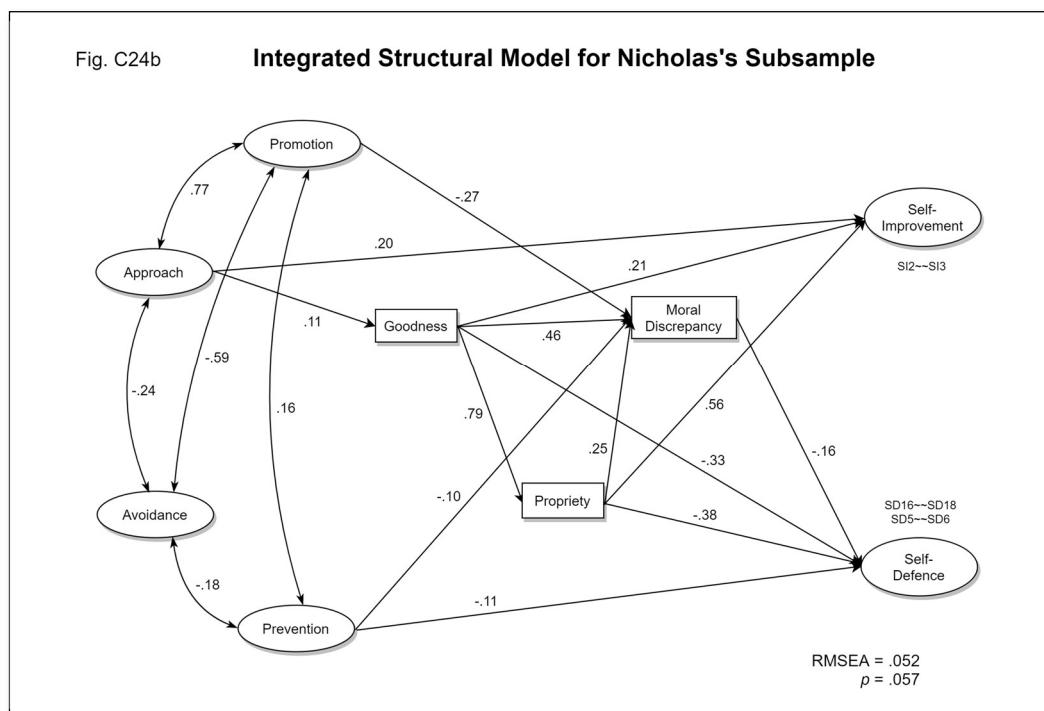
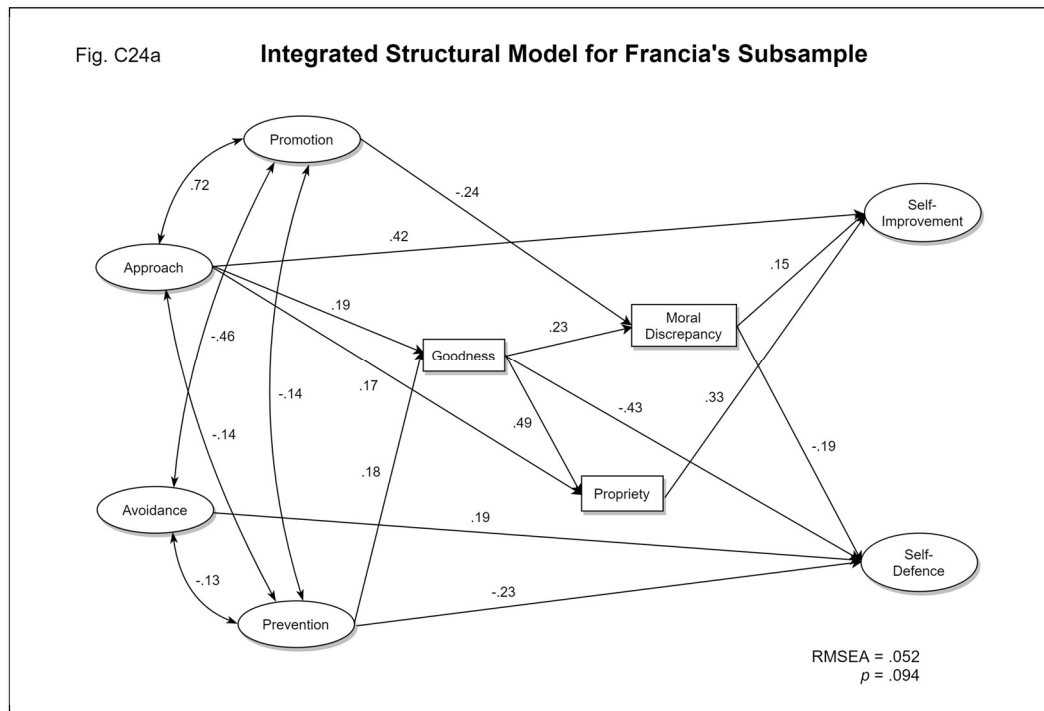
Main fit indices of the integrated structural models for the two vignettes

	Francia	Nicholas
Chi-Square (scaled)	2262.164	2281.365
df (scaled)	972	969
<i>p</i> -value (scaled)	<.001	<.001
RMSEA (scaled)	.052	.052
<i>p</i> -value (scaled)	.094	.057
CFI (scaled)	.884	.900
TLI (scaled)	.876	.893
GFI	.975	.958
AGFI	.971	.952

The complete output of the integrated structural equations models is available in SMC14.

Figures C24a-b:

Integrated structural models for the two vignettes (with goodness and propriety of the deed). Estimator: robust maximum likelihood (MLR). Standardised regression weights. Indicators, errors/disturbances and covariances across error terms not graphically displayed to facilitate legibility



A meaningful consideration emerged comparing the more parsimonious models in Figures C23a-b (without opinion-based comparisons) and the integrated models in Figures C24a-b (with opinion-based comparisons), particularly for Nicholas's vignette. When in Nicholas's integrated model the goodness and propriety of the deed are taken into account, moral discrepancy (i.e., ability-based comparison) loses a substantial share of its power to predict moral self-regulation in favour of judgments underlying opinion-based comparisons. In fact, in Nicholas's subsample, the regression weight of the pathway linking moral discrepancy with self-improvement was non-significant, and the regression weight of the pathway linking it with self-defence dropped from $-.61$ to $-.16$. At the same time, the total variance explained of the dependent variables roughly doubled from about 30% to about 60% thanks to the strong contribution of opinion-based comparisons (Tab. C18). These phenomena could be explained considering that a sizeable portion of participants who viewed Nicholas's vignette did *not* believe the deed was admirable and worth imitating; if these participants had no intention to be and behave like the moral agent (as they did not judge the action positively), they most likely did not engage in self-regulatory processes instigated by ability-based comparisons, but self-defended mainly as a direct result of the less positive judgment of the deed. By contrast, Francia's subsample judged the deed nearly universally praiseworthy and admirable, so the inclusion in the model of judgments underlying opinion-based comparisons only slightly reduced the size of the effect of moral discrepancy as a predictor of self-improvement and self-defence, but without dramatically altering the balance or greatly increasing the total variance explained of the dependent variables.

Table C18:

Total variance explained of the dependent variables for the two vignettes in the revised models (more parsimonious) and the integrated models (with the addition of opinion-based comparisons)

	Francia	Nicholas
Self-improvement:		
revised model	.253	.302
integrated model	.366	.610
Self-defence:		
revised model	.204	.376
integrated model	.394	.615

The structural model integrated with judgments underlying opinion-based comparisons was also fitted to the total sample (full output in SMC15). This model was fitted using the robust maximum likelihood (MLR) estimator, with the mean structure, and the variable vignette was dummy coded as in Study 2. The fit indices are shown in Table 19 and the structural model is illustrated in Figure 25. This final model fitted to the total sample documents the fundamental regulatory mechanisms across vignettes, that is, those processes that appear to exist regardless of the moral scenario.

Table C19:

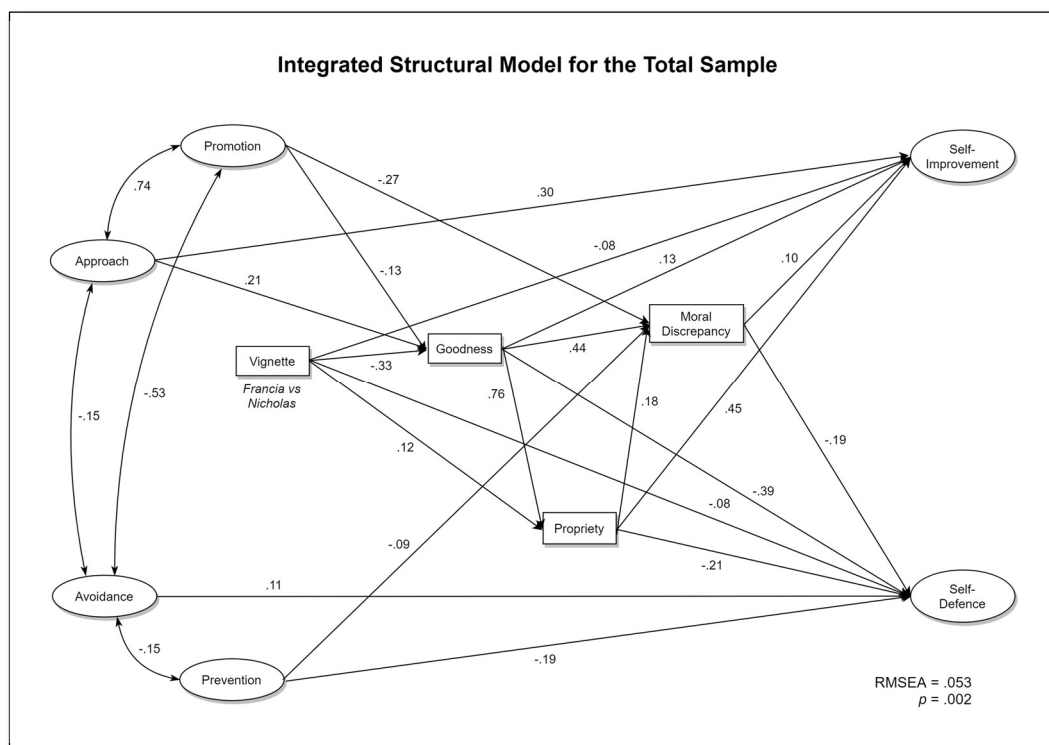
Main fit indices of the integrated structural model for the total sample

	Integrated model (total sample)
Chi-Square (scaled)	3782.408
df (scaled)	1011
<i>p</i> -value (scaled)	<.001
RMSEA (scaled)	.053
<i>p</i> -value (scaled)	.002
CFI (scaled)	.885
TLI (scaled)	.877
GFI	.966
AGFI	.961

As in Study 2, the negative sign of the regression weight between the variables vignette and goodness indicates that Francia's vignette elicited significantly more positive judgments about the goodness of the deed relative to Nicholas's vignette.

Figures C25:

Integrated structural model for the total sample. Estimator: robust maximum likelihood (MLR). Standardised regression weights. Indicators and errors/disturbances not graphically displayed to facilitate legibility



Secondary Analyses

The secondary analyses concerned both measurement and structural models, and sought to: a) provide evidence of the psychometric properties of the models (measurement and structural invariance); b) develop a person-centred analysis (latent profile modelling) to complement the variable-centred analyses based on CFA and SEM.

Measurement Invariance of the Endogenous Latent Variables

The construction of a new measurement instrument for investigating the self-regulation of virtue was conducted assuming that *one* single instrument could be identified and utilised to measure the response to moral exemplars, independent of the content of their acts. To assess the extent to which the new self-improvement and self-defence scales measured the same constructs across different groups, analysis of measurement invariance was conducted.

Three sets of analyses were carried out: the first evaluated whether participants exposed to Francia's vignette interpreted the moral self-regulatory constructs in a similar way to those exposed to Nicholas's vignette (invariance *between vignettes*); the second assessed whether participants who engaged in upward moral comparisons interpreted them in a similar way to those who engaged in downward and lateral comparisons (invariance *across moral comparison types*); the third tested whether male participants interpreted them in a similar way to female participants (invariance *between genders*).

Of these three sets, analysis of invariance between vignettes returned more unbiased results, thanks to the equivalent size of the two groups ($n = 490$ for Francia, 492 for Nicholas), whereas for the moral comparison types the analyses were partly affected by the large differential in group sizes ($n = 692$ for upward comparison, 212 for downward comparison, 78 for lateral comparison), which to a certain extent also affected the gender contrasts ($n = 517$ for male, 460 for female⁵⁷).

Four kinds of tests were carried out for the above-mentioned contrasts, each with increasing levels of parameter restrictions:

⁵⁷ The sample for the computation of the invariance tests across genders was reduced to 977, with the deletion of 5 participants who declared "other" to the gender question (a group of 5 participants would be too small to conduct any kind of invariance test).

1. *configural* invariance tested the hypothesis of equal form of the latent variables across groups, or in other words, the hypotheses of the same number of factors and of their association with the same indicators across groups;
2. *weak* invariance (or metric invariance) tested the hypothesis of equal factor loadings of the indicators to the latent variable across groups;
3. *strong* invariance (or scalar invariance) tested the hypothesis of equal means for the indicators of the latent variable across groups;
4. *strict* invariance tested the hypothesis of equal residual variances of the observed scores not accounted for by the factors across groups.

The criterion used here to assess measurement invariance across groups followed the suggestions reported by Kline (2016) and Byrne (2016), who recommend using changes in CFI (ΔCFI) equal to or less than .010 as reasonable evidence of invariance. Other criteria were also considered and reported in SMC17, for example the differences in RMSEA (ΔRMSEA). The traditional significance of the chi-square difference was also tested and reported in the full output in SMC17 for the sake of completeness, but not adopted in the formal assessment, due to its well-known oversensitivity to sample size (Rutkowski & Svetina, 2014) and violations of distributional assumptions (Chen, 2007).

Analysis of ΔCFI provided an accurate outlook (Tab. C20). All invariance analyses were conducted with the R package semTools ver. 0.5-2 (Jorgensen et al., 2019).

Table C20:

Measurement invariance tests (ΔCFI) across groups for moral self-improvement and self-defence

	Invariance between vignettes		Invariance across comparison types		Invariance between genders	
	ΔCFI Self-Improvement	ΔCFI Self-Defence	ΔCFI Self-Improvement	ΔCFI Self-Defence	ΔCFI Self-Improvement	ΔCFI Self-Defence
1. Configural invariance	-	-	-	-	-	-
2. Weak invariance	.008*	.005*	.008*	.010*	<.001*	.001*
3. Strong invariance	.006*	.007*	.012	.009*	<.001*	.001*
4. Strict invariance	.004*	.001*	.013	.021	.001*	.002*

Note: * cutoff criterion for reasonable invariance: $\Delta CFI \leq .010$.

The results showed strict invariance between vignettes and genders for both self-improvement and self-defence. Invariance across comparison types was weak for self-improvement and strong for self-defence. Considering that the cutoff point of .010 for ΔCFI is regarded by some scholars as too stringent⁵⁸, and the high frequency in the literature of violations of strict invariance (Kline, 2016) and even strong invariance (Vandenberg & Lance, 2000), the measurement invariance of the moral self-regulation inventory appears to be remarkable.

An examination of the data *by vignette* allowed to identify the items that exhibited the largest discrepancies between vignettes in terms of factor loadings (slopes) and mean structures (intercepts) (see Tab. C21a-b for self-improvement and Tab. C22a-b for self-defence).

⁵⁸ Based on their Monte Carlo simulation studies, Rutkowski and Svetina (2014) found deficiencies in the traditional cutoff points, especially for weak invariance, and therefore recommend more liberal criteria, particularly for large group sizes.

Tables C21a-b:

Factor loadings (slopes) and means (intercepts) from CFA for self-improvement by vignette

Tab. C21a	Factor loadings		
	Self-improvement		
	Francia	Nicholas	Δ
SI1/Admirable	.581***	.776***	-.195
SI2/Awakened	.781***	.798***	-.017
SI3/Humanity	.791***	.835***	-.044
SI4/BeBetter	.707***	.778***	-.071
SI6/BeLike	.707***	.790***	-.083
SI8/Uplifted	.847***	.905***	-.058
SI9/Inspired	.883***	.907***	-.024
SI10/Moved	.872***	.882***	-.010
SI11/Proud	.787***	.886***	-.099
SI12/Happy	.854***	.889***	-.035

Tab. C21b	Mean structure		
	Self-improvement		
	Francia	Nicholas	Δ
SI1/Admirable	5.904***	3.230***	2.674
SI2/Awakened	2.797***	2.262***	.535
SI3/Humanity	3.053***	2.365***	.688
SI4/BeBetter	2.184***	1.999***	.185
SI6/BeLike	2.466***	2.150***	.316
SI8/Uplifted	2.639***	2.044***	.595
SI9/Inspired	2.664***	2.097***	.567
SI10/Moved	2.742***	2.044***	.698
SI11/Proud	2.900***	2.309***	.591
SI12/Happy	3.003***	2.077***	.926

Note: *** $p < .001$

Delta with the highest absolute value in **boldface**

Note: *** $p < .001$

Tables C22a-b:

Factor loadings (slopes) and means (intercepts) from CFA for self-defence by vignette

Tab. C22a	Factor loadings		
	Self-defence		
	Francia	Nicholas	Δ
SD1/MeGreater	.726***	.725***	.001
SD2/Devalue	.797***	.733***	.064
SD3/Ordinary	.663***	.723***	-.060
SD5/Praise	.764***	.755***	.009
SD6/Ulterior	.668***	.693***	-.025
SD8/Uneasy	.799***	.739***	.060
SD10/NoPraise	.657***	.751***	-.094
SD11/Superior	.775***	.810***	-.035
SD12/Seriously	.777***	.769***	.008
SD16/Threatened	.834***	.603***	.231
SD18/Resentful	.769***	.687***	.082
SD19/Irritated	.805***	.739***	.066

Tab. C22b	Mean structure		
	Self-defence		
	Francia	Nicholas	Δ
SD1/MeGreater	.671***	.749***	-.078
SD2/Devalue	.580***	.734***	-.154
SD3/Ordinary	.611***	.793***	-.182
SD5/Praise	.766***	.803***	-.037
SD6/Ulterior	.828***	.980***	-.152
SD8/Uneasy	.561***	.640***	-.079
SD10/NoPraise	.904***	.928***	-.024
SD11/Superior	.714***	.711***	.003
SD12/Seriously	.653***	.758***	-.105
SD16/Threatened	.479***	.511***	-.032
SD18/Resentful	.453***	.522***	-.069
SD19/Irritated	.455***	.544***	-.089

Note: *** $p < .001$

Delta with the highest absolute value in **boldface**

Note: *** $p < .001$

Analysis of slope deltas for self-improvement and self-defence allowed to identify the items with the largest differentials, and these items were tested for measurement invariance. They were constrained one by one to be equal across vignettes and then compared to an unconstrained model. The results are displayed in Tables C23a-b.

Tables C23a-b:

Measurement invariance tests by vignette: models with items with largest differentials between vignettes constrained to equality and tested against an unconstrained model

Table C23a: Self-improvement

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	66	84360	84673	273.34			
SI01 constrained	68	84496	84800	413.35	78.358	2	<.001***
SI11 constrained	67	84399	84708	314.48	29.372	1	<.001***
SI06 constrained	67	84401	84709	315.56	33.723	1	<.001***
SI04 constrained	67	84407	84715	322.04	37.652	1	<.001***
SI08 constrained	67	84415	84723	329.94	27.83	1	<.001***

Table C23b: Self-defence

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	100	100927	101318	530.94			
SD16 constrained	101	100937	101324	543.66	5.9348	1	0.015*
SD10 constrained	101	100926	101312	532.03	1.2549	1	0.263
SD18 constrained	101	100930	101316	536.34	3.1885	1	0.074
SD19 constrained	101	100926	101312	532.12	0.62095	1	0.431
SD02 constrained	101	100929	101315	535.23	4.9277	1	0.026*
SD03 constrained	101	100928	101314	534.20	3.419	1	0.064

Note: *** $p < .001$ (2-tailed)

* $p < .05$ (2-tailed)

For self-improvement, five indicators out of ten were not invariant, especially item SI01/Admirable, yet without jeopardising the overall strict invariance of the model. By contrast, only two were not invariant in the self-defence latent variable, especially item SD16/Threatened. This is consistent with the stronger metric invariance achieved by the self-defence scale (see Tab. C20).

Structural Invariance Between Vignettes

Tests of invariance were also conducted at the structural level to provide further evidence of the distinctiveness of the individual structural models for each vignette. The analysis followed the same procedure as in Study 2, with the exception that now latent variables were used. First, an unconstrained model was fitted, then it was compared to a fully constrained model (slopes and intercepts constrained to equality) and the significance of the chi-square difference for the corresponding difference in degrees of

freedom was assessed. The result was significant (Tab. 23a), indicating – as expected – lack of invariance. At that point, the most dissimilar pathways were tested for structural invariance one by one against the unconstrained model (Tab. 23b). The full output is available in SMC18.

Tables C23a-b:

Structural invariance tests by vignette: models with pathways with largest differentials between vignettes constrained to equality and tested against an unconstrained model

Table C23a: structural invariance test – unconstrained vs fully constrained models

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	1936	390415	391970	5550.0			
Fully constrained	1993	390797	392073	6045.5	373.01	57	<.001***

Table C23b: structural invariance tests – unconstrained model vs models with most dissimilar pathways constrained to equality

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	1936	390415	391970	5550.0			
Path A constrained	1937	390415	391965	5551.9	1.617	1	0.204
Path B constrained	1937	390417	391967	5554.1	2.7595	1	0.097
Path C constrained	1937	390421	391971	5557.5	5.3814	1	0.020*
Path D constrained	1937	390422	391972	5558.2	5.2318	1	0.022*
Path E constrained	1937	390415	391965	5552.1	2.3028	1	0.129
Path F constrained	1937	390428	391978	5564.9	14.178	1	<.001***
Path G constrained	1937	390418	391968	5555.1	4.491	1	0.034*
Path H constrained	1937	390418	391968	5554.9	5.0498	1	0.025*

Note: * $p < .05$ level (2-tailed)

*** $p < .001$ (2-tailed)

Path A = moral discrepancy / self-improvement
 Path B = goodness of deed / self-improvement
 Path C = propriety of deed / self-defence
 Path D = propriety of deed / moral discrepancy
 Path E = prevention focus / goodness of deed
 Path F = promotion focus / prevention focus
 Path G = approach / avoidance
 Path H = approach / prevention focus

Only a few pathways that appear to be different between vignettes (significant in one and non-significant in the other) reached statistical significance; those that were truly divergent were especially the covariances between the characteristic adaptations. This finding shows that, all in all, the covariances across exogenous latent variables

contribute more than the pathways across endogenous variables to determine lack of invariance of the structural models; the divergence in the regression weights linking moral discrepancy and self-improvement, which is significant in Francia's vignette but not in Nicholas's, is not so large as to cause the models to produce a significant difference in the overall model fit.

A Person-Centred Analysis: Latent Profile Modelling

The next analysis moved the focus from the variables to the *participants*, with the aim to identify clusters of individuals who are similar within themselves (homogeneity criterion) and different from others (separation criterion). There are various kinds of “person-oriented approaches” – as Bergman and Magnusson (1997) defined them – that generate clusters of individuals; the family of methods chosen for this specific analysis, latent variable mixture models, uses probabilistic models which have stronger properties than other cluster analyses (e.g., K-means) that use distance algorithms (He & Fan, 2019). Within the family of latent variable mixture models, latent profile analysis (LPA) enables to aggregate individuals with similar response patterns to certain sets of quantitative questionnaire items that underlie specific latent constructs. One of the critical assumptions of LPA is that the observed indicators are measured through *continuous* variables⁵⁹, as is the case for the variables that were analysed here: moral self-regulation and motivational dispositions.

Clustering participants with one of the latent variable mixture models was a preregistered secondary analysis and was performed as an additional exploration of the data from Study 3, in order to shed further light on nuances of the self-regulation of virtue and its antecedents. Although these methods have become popular only recently,

⁵⁹ A conceptually similar, but distinct, type of latent variable mixture modelling (latent class analysis) can be performed in the presence of categorical item indicators.

the literature in the social, behavioural, and health sciences provides several examples. Collins and Lanza (2010) report the case of modelling the latent variable of drinking motivation from a sample of high school seniors, which allowed to identify the four clusters of experimenters, thrill-seekers, relax-seekers, and all-round drinkers who are simultaneously motivated by all the previous factors (Coffman et al., 2007). It is easy to see how distinguishing these independent clusters might lead to more effective targeted interventions.

In LPA, participants' clusters are referred to as "profiles". The idea behind LPA is that latent profiles are "organising principles" that *cause* participants to respond differentially to certain observed indicators, much like in CFA latent variables *cause* the observed indicators to aggregate according to distinct covariance patterns. And like with CFA, thanks to LPA it is possible to reduce the dimensionality of the phenomenon at hand, isolating and highlighting the essential elements that inspection of the individual observations across indicators would make painstakingly slow and complex (if possible at all) to identify.

CFA and LPA are not mutually exclusive or competing methods (Robins et al., 1996). One of the reasons why they could be employed in the analysis of the same dataset is that the phenomenon under investigation can have both continuous and categorical features (Collins & Lanza, 2010): indeed, self-improvement and self-defence can be viewed as mechanisms that can be experienced as a "quantity", on a spectrum from weak hints to strong manifestations, but can also be observed as a set of various expressive "categories", each qualitatively distinct from the others. Thus, these operationalisations allow answering a wider set of research questions.

As is often the case with LPA (see Marsh et al., 2009), no specific predictions about the nature and number of profiles could be made in the preregistration, because too

limited information was available prior to data collection. The analysis was performed with the R package tidyLPA ver. 1.0.5 (Rosenberg et al., 2018). This package, which is based on the tidyverse code collection, does not directly handle the modelling, but it interfaces with mclust ver. 5.4.5 (Scrucca et al., 2016), which computationally carries out the mixture modelling. tidyLPA uses the expectation-maximisation (EM) algorithm to generate the maximum likelihood solution.

The analysis consisted of three steps⁶⁰, through which three kinds of models – each with specific types of parametrisations for means, variances, and covariances – were fitted and compared. While the means were estimated allowing them to freely vary across all models, variances and covariances were modelled fixing or constraining them in various ways. Models 1 and 2 meet the classical assumption of local independence, based on which – conditional on the latent variable – the observed indicators are independent (Collins & Lanza, 2010); that is, within the latent profile, covariances are set to zero. This assumption is relaxed in Model 3, where covariances (as well as variances) are constrained to be equal⁶¹. Thus, Models 1-3 were fitted in a sequence with decreasing levels of parameter restrictions:

- Step 1, Model 1: class-invariant parametrisation estimates profiles with equal variances and covariances constrained to zero (the most restrictive);
- Step 2, Model 2: class-varying diagonal parametrisation estimates profiles with varying variances and covariances constrained to zero;
- Step 3, Model 3: class-invariant unrestricted parametrisation estimates profiles

⁶⁰ Full LPA usually consists of *six* steps and *six* corresponding models. However, Models 4 and 5 were not computed because they can only be fitted by tidyLPA in association with the MPlus software, and Model 6 was dropped because with large datasets it requires hardware capabilities beyond those of an ordinary personal computer. Model numbers from 1 to 6 are a specific feature of the tidyLPA package in R.

⁶¹ According to Marsh et al. (2009), only the models that meet the assumption of local independence can be considered “classical” latent profile models, while the others should be viewed as other forms of latent variable mixture models.

with equal variances and equal covariances (the least restrictive).

At each step, a varying number of profiles was extracted. Because the optimal number of profiles was unknown a priori, the procedure was repeated as many times as necessary, until the size of the last profile became so small that too few observations could be assigned to it⁶². Within each step, models with varying numbers of profiles were fitted and compared through multiple indices/information criteria, for example the log-likelihood, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the sample size-adjusted Bayesian information criterion (SABIC), and others. Thanks to its correction for sample size and the penalty for parameter complexity, in this analysis the SABIC was regarded as the most diagnostic criterion, although all of them were considered. There is no cutoff point for information criteria such as the SABIC; when comparing models, the smaller the value, the better the fit.

Once the models with the ideal number of profiles were extracted from each step of the procedure, those models were compared using the information criteria. After analysis of all the indices, particularly the SABIC, the preferred model was selected and interpreted qualitatively. It must be noted that models cannot be chosen purely based on information criteria; the model with the best fit must also be *meaningful* from a substantive viewpoint and contribute to explaining the phenomenon at hand (Marsh et al., 2009; He & Fan, 2019).

The above procedure was carried out twice among the total sample: first to fit profiles of moral self-regulation, collapsing all the broadening and defensive items into one single inventory of twenty-two items, and then to fit profiles of motivational dispositions, collapsing the regulatory focus and hedonic orientation items into one

⁶² The maximum number of profiles in a single step for the current dataset was ten for self-regulation and nineteen for motivational dispositions.

single motivation inventory of twenty-one items. Moral comparison types were not extracted from LPA, as they were already available (see Fig. C6).

Moral Self-Regulation Profiles. The best model from the LPA on moral self-regulation yielded three self-regulatory profiles from class-varying diagonal parametrisation (Model 2, see Tab. C24). The proportion of participants who fell into each profile varied from 26% to 43% (each participant could only belong to *one* profile).

Table C24:

Overall fit indices and information criteria for the moral self-regulation profiles

Fit index / Information criterion	Model 1 (step 1) <i>10 profiles</i>	Model 2 (step 2) <i>3 profiles</i>	Model 3 (step 3) <i>2 profiles</i>
Log Likelihood	-91991.281	-89609.401	-91654.886
AIC	184484.561	179486.802	183905.772
BIC	185711.849	180142.007	185362.871
SABIC	184914.668	179716.420	184416.417
AWE	188192.272	181465.255	188308.024
CLC	183984.425	179220.758	183311.717
KIC	184738.561	179623.802	184206.772

The three profiles are illustrated in Figure C26, which also indicates their relative size in percentage.

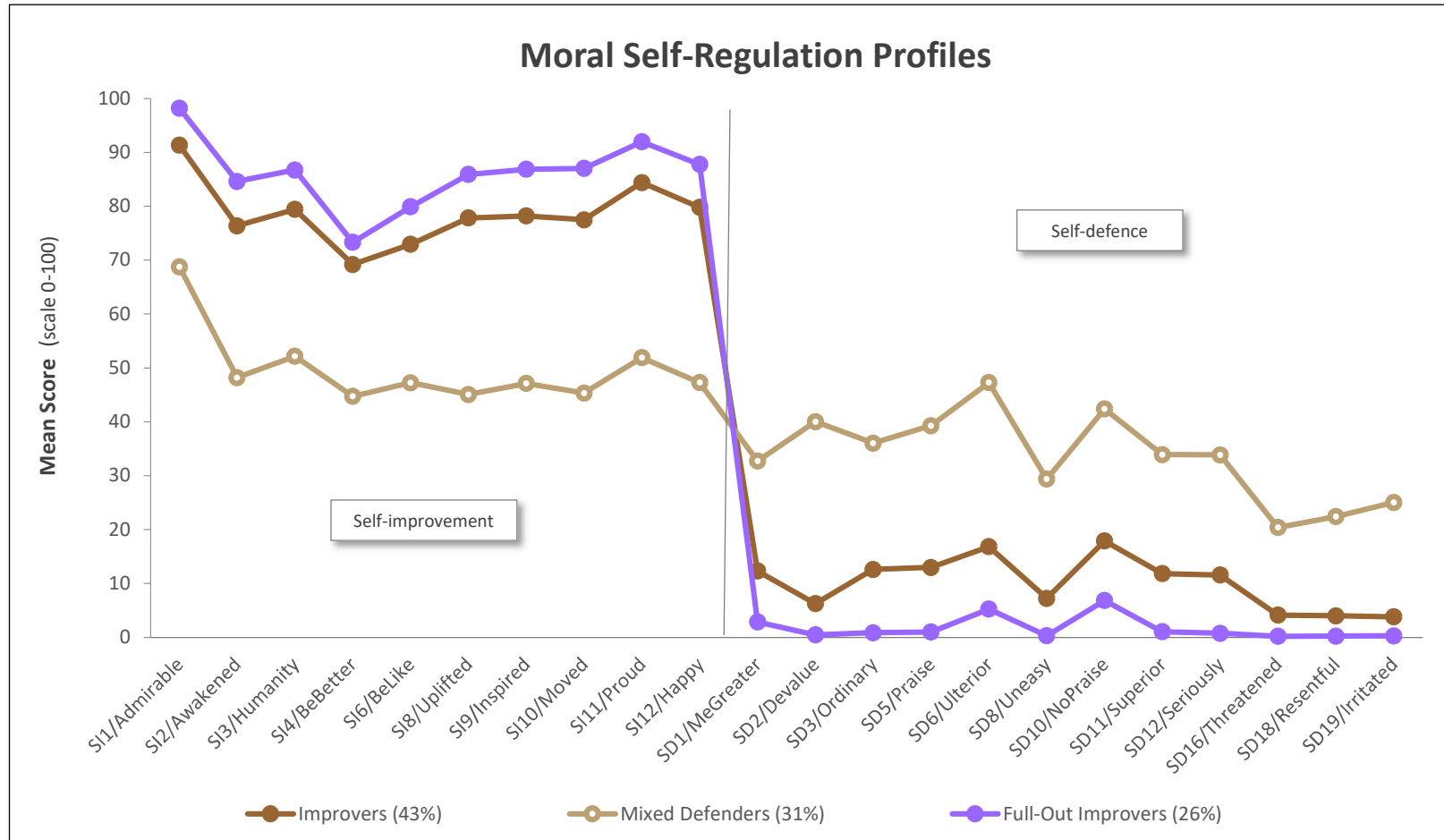
The three clusters of the model with the best fit were labelled based on a qualitative evaluation of their profile shape across the indicators: full-out improvers, improvers, mixed defenders. Their main characteristics are briefly described below:

- *full-out improvers*: this group displays the highest degree of self-improvement and at the same time the minimum level of self-defence (null or close to null for most items). It is the smallest of the three self-regulation clusters, with 26% of participants;
- *improvers*: this group is characterised by strong self-improvement, although to a lesser degree than the previous cluster. In addition, it endorses self-defence items

more strongly than full-out improvers. It is the largest cluster, composed by 43% of participants;

- *mixed defenders*: this group, despite an overall slight prevalence of self-improvement over self-defence, is the cluster that comparatively shows (by a margin) the highest level of defensive mechanisms. It is a sizeable profile, aggregating 31% of participants.

Figure C26:
Moral self-regulation profiles from LPA (total sample) with relative profile sizes



Overall, it can be noticed that the two high self-improvement clusters have similar profile shapes, except that the full-out improvers score on average about 10 points higher on the self-improvement items and 10 points lower on self-defence items (relative to the improvers). The mixed defenders are a heterogeneous cluster of participants who on average score around the midpoint of the scale for most self-improvement items, but at the same time endorse slightly below the midpoint of the scale also the self-defence items, especially SD6/Ulterior and SD10/NoPraise.

The full LPA output for moral self-regulation can be found in SMC19, while the descriptive statistics of the three self-regulatory profiles are available in SMC20.

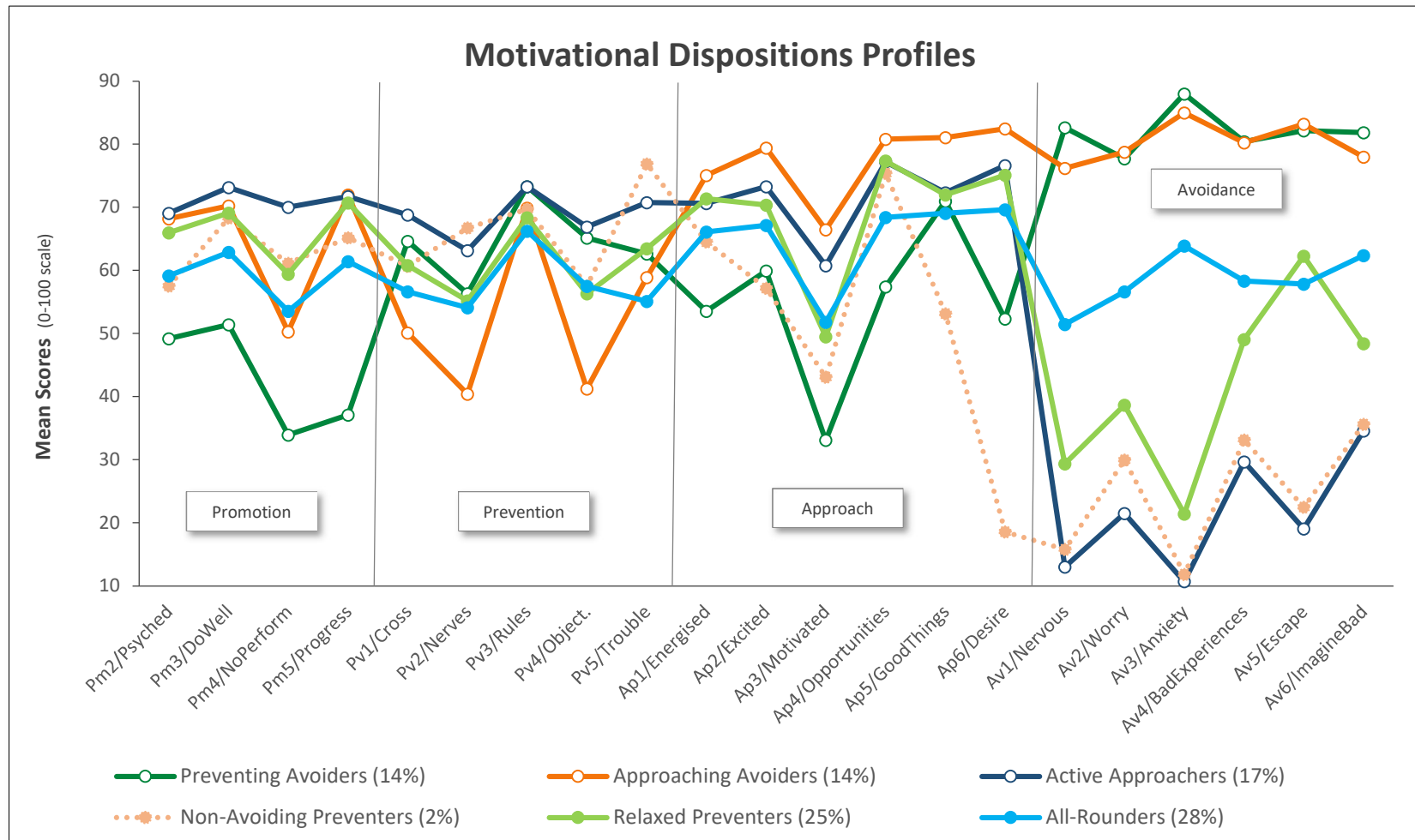
Motivational Dispositions Profiles. The best model from the LPA on regulatory focus and hedonic orientation yielded six motivational profiles from class-invariant unrestricted parameterisation (Model 3, see Tab. C25). The proportion of participants who fell into each profile varied from 2% to 28% (again, each participant could only belong to *one* profile). The six profiles are illustrated in Figure C27, which also indicates their relative size in percentage.

Table C25:

Overall fit indices and information criteria for the motivational dispositions profiles

Fit index / Information criterion	Model 1 (step 1) <i>19 profiles</i>	Model 2 (step 2) <i>10 profiles</i>	Model 3 (step 3) <i>6 profiles</i>
Log Likelihood	-90502.304	-90416.827	-89824.114
AIC	181880.608	181691.655	180372.228
BIC	184022.249	183789.289	182142.260
SABIC	182631.153	182426.777	180992.541
AWE	188352.092	188030.076	185720.659
CLC	181006.407	180835.503	179649.861
KIC	182321.608	182123.655	180737.228

Figure C27:
Motivational dispositions profiles from LPA (total sample) with relative profile sizes



The six clusters of the model with the best fit were labelled based on a qualitative evaluation of their profile shape across the indicators: active approachers, relaxed preventers, non-avoiding preventers, approaching avoiders, preventing avoiders, all-rounders. Their main characteristics are described below:

- *active approachers*: this group is composed of participants who score high on approach, promotion focus and prevention focus (around 70 on the 0-100 scale), while at the same time endorsing the lowest levels of avoidance (around 20 on the 0-100 scale). It is a medium-sized profile made up of 17% of participants;
- *relaxed preventers*: this group has a similar pattern to the previous one, with a prevalence of approach, promotion focus and prevention focus (between 60 and 69 on the 0-100 scale, and higher on prevention focus relative to the other groups) and at the same time low avoidance, particularly the items referring to anxiety, worry and nervousness (hence the label “relaxed”). They make up a total of 25% of the sample;
- *non-avoiding preventers*: this is the smallest group (only 2% of participants). Participants in this profile exhibit a prevalence of prevention focus and low avoidance, while maintaining relatively robust levels of promotion focus;
- *approaching avoiders*: this medium-sized group (14%) has the peculiarity of showing at the same time high levels of approach and avoidance (around 80 on the 0-100 scale), which among the total sample are weakly inversely correlated;
- *preventing avoiders*: this group (14% of participants) shows the highest level of avoidance (above 80 on the 0-100 scale) and at the same time scores relatively high on prevention focus (about 65). Participants in this profile are also considerably lower than average on promotion focus and approach;
- *all-rounders*: the remaining group is the largest (28%) and shows average levels of

all the four motivational traits, with a relatively flat profile.

The full LPA for motivational dispositions is available in SMC21, while the descriptive statistics for the six motivational profiles are in SMC22.

Associations Between Profiles. In the variable-centred analysis, the structural equation models had provided initial evidence of a causal chain of comparative and self-regulatory effects induced by the experimental manipulation of vignettes portraying moral exemplars; the models also included associations of exogenous motivational dispositions with moral comparisons and moral self-regulation. Now, latent profile modelling enabled the aggregation of specific groups of participants characterised by higher probabilities of endorsing certain motivational traits and experience certain moral self-regulatory modes when exposed to acts of virtue; in addition, previous analysis of moral evaluations of the self and the moral character of the agents had also allowed to distinguish three further groups of participants who engaged in upward, downward, and lateral comparisons. If through the variable-centred analysis specific correlation patterns emerged between motivational, comparison, and self-regulation *variables*, at this point it was of interest to analyse if any significant associations existed between motivational, comparison, and self-regulation *clusters* of participants.

To analyse these associations, chi-square and likelihood ratio tests were performed: first between motivational groups (the six motivational profiles from LPA) and comparison groups (the three clusters of upward, downward, and lateral “comparers”), then between comparison groups (the same three categories mentioned above) and self-regulation groups (the five self-regulatory profiles from LPA), and lastly between motivational groups (the six motivational profiles from LPA) and self-regulation groups (the five self-regulatory profiles from LPA).

These analyses were performed among the total sample of 982 participants. The results are reported in Table C26. All the associations were highly significant, and the effect size measured through Cramér's V (equivalent to the percentage of the maximum possible variation of the variables) was of moderate strength for one of them (association between comparison and regulatory groups) and weaker for the others.

Table C26:

Tests of associations between motivational dispositions profiles, comparison type groups, and moral self-regulation profiles, with corresponding effect sizes (total sample)

N of valid cases: 982	Pearson's Chi-Square	Likelihood Ratio	Cramér's V
Motivational dispositions profiles * moral comparison groups			
Value	30.675 ^a	30.313	.125
df	10	10	
Asymptotic Significance (2-sided)	.001	<.001	
Moral comparison groups * moral self-regulation profiles			
Value	172.407 ^b	180.794	.296
df	4	4	
Asymptotic Significance (2-sided)	<.001	<.001	
Motivational dispositions profiles * moral self-regulation profiles			
Value	33.887 ^c	33.936	.131
df	10	10	
Asymptotic Significance (2-sided)	<.001	<.001	

Note: a. 2 cells have expected count less than 5. The minimum expected count is 1.43.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 18.19.

c. 1 cell has expected count less than 5. The minimum expected count is 4.20.

Pearson's chi-square and the likelihood ratio are *omnibus* tests that indicate whether the null hypothesis of a uniform distribution of participants across groups (the independence hypothesis) should be rejected. Results showed that the independence hypothesis should be rejected across the board.

When the null hypothesis is to be rejected, as in this case, it is of interest to detect which group pairings have a stronger influence and drive the overall significant effect.

Analysis of the standardised residuals provided the answer. In a nutshell, what emerged is the following (in brackets the value of the standardised residuals⁶³):

- *association between comparison groups and self-regulation profiles*: downward comparers tended to be more than proportionally represented among the mixed defenders (8.8) and less than proportionally represented among the full-out improvers (-6.2) and improvers (-3.7). Conversely, upward comparers tended to be more than proportionally full-out improvers and improvers (2.4 both profiles), and less than proportionally mixed defenders (-4.4);
- *association between motivational profiles and self-regulation profiles*: non-avoiding preventers tended to fall more than proportionally in the mixed defenders profile (2.1), whereas relaxed preventers were more than proportionally unlikely to fall in the full-out improvers profile (-2.0);
- *association between motivational profiles and comparison groups*: approaching avoiders were more than proportionally likely to engage in lateral comparisons (3.1), while preventing avoiders were more than proportionally unlikely to engage in lateral comparisons (-2.2).

The full results of the chi-square and likelihood ratio tests, including the residuals, are available in SMC23.

Flow Analysis. An attractive way to represent graphically the *flow* of participants from motivational profiles to self-regulation profiles via comparison groups is represented by network diagrams, for instance Sankey or alluvial diagrams. They enable to visualise response patterns across the key variables in the form of “transition trajectories” that cluster participants’ experience along the causal pathways (see Fig.

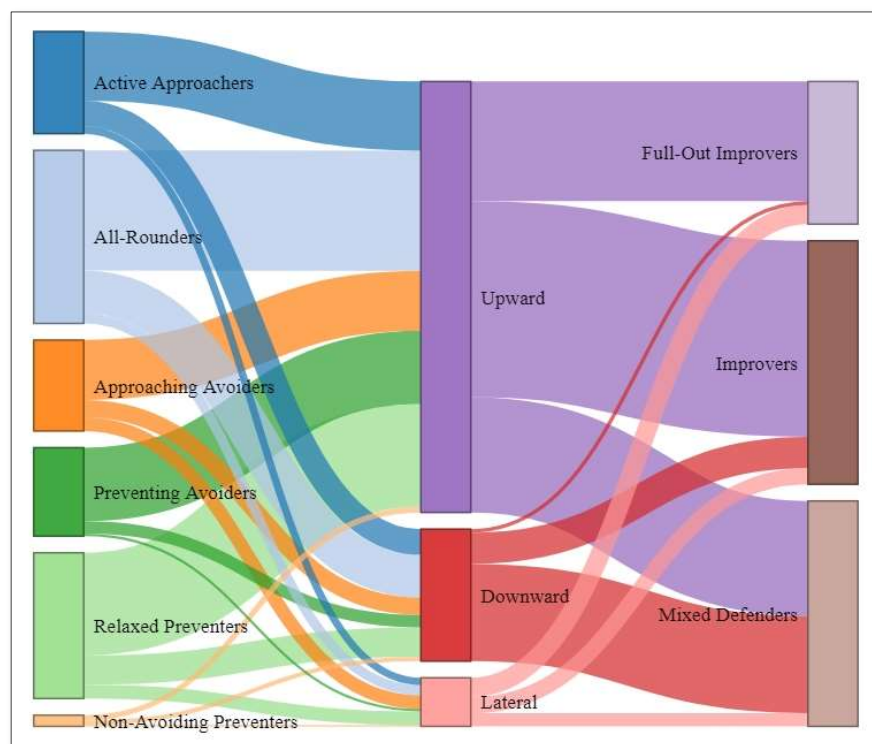
⁶³ Standardised residuals are z-scores and therefore they can be considered significant if greater in absolute value than 1.96 at the .05 level (see Field, 2013).

C28). In this study, Sankey diagrams can effectively provide an overview of the proportion of individuals that flow across the various stages of the psychological process of the response to moral exemplars. In these plots, the boxes are the nodes and the arrows are the links that connect the nodes; the thicker the boxes and the links, the higher their value (number of individuals at each stage transitioning along the respective trajectories).

The Sankey diagram in Figure C28 was created in R with the package *flipPlots* ver. 1.2.0 (Displayr, 2019), which uses in the background the package *networkD3* ver. 0.4 (Allaire et al., 2017) to generate the network graph. The contingency tables across the variables are available in SMC24.

Figure C28:

Sankey diagram illustrating participants' flow across clusters along the causal pathways of the self-regulation of virtue (total sample)



Discussion

Study 3 was designed to provide a conceptual replication of the results from Study 2. The provisional measurement and structural models were further improved and the identification of specific participants' profiles with common response patterns added new perspectives to the findings. In this respect, Study 3 accomplished what it was designed to deliver.

Measurement Model

Endogenous Latent Variables

With a few modifications, the measurement model for the moral self-regulation inventory yielded satisfactory fit, with the main indices within or close to the excellence cutoff points. The measurement model hypothesised based on Study 2 required two kinds of interventions: the removal of a few self-regulation items (two from the self-improvement and six from the self-defence scales) and the specification of a few error covariances (two for self-improvement and four for self-defence⁶⁴).

The removal of a few indicators from the scales was due to factor loadings equal to or smaller than .7, which were causing undesirable measurement error. These items were not necessarily poorly designed. The main issue consisted in the lack of measurement invariance between vignettes; in other words, they did not function the same way across the two vignettes, as slightly different reactions were elicited by the two moral deeds. This was reflected in the factor loadings, higher for one vignette and lower for the other. Two key examples are items SI5/Values and SD17/Guilty. Item SI5 had a relatively strong loading for Nicholas's subsample (.764), but weaker for Francia's (.604), determining a lower-than-expected performance among the total sample. Even

⁶⁴ Importantly, not all these error covariances were necessary in the structural models.

more striking is the case of item SD17, whose factor loading for Francia's subsample was acceptable, although not fully satisfactory (.609), but was far too weak for Nicholas's subsample (.333), causing a poor performance among the total sample. From a substantive viewpoint, beyond the statistics, shared values with the moral agent (SI5) and guilty feelings (SD17) were not common experiences across the two vignettes and therefore could not be considered *generalisable* experiences related to the self-regulation of virtue. Therefore, these items were dropped from their respective scales.

The other intervention to improve the fit of the endogenous latent variables consisted in allowing a few covariances between the error terms of some observed indicators. In theory, this represents a violation of the principle of local independence that underlies the construction of latent variables, that is, the assumption that the indicators should not be correlated after conditioning on the latent variable (Borsboom et al., 2003). Researchers have provided conflicting evidence on existence and entity of a potential bias introduced by allowing such correlations, for example in studies that used the multitrait-multimethod approach, first introduced by Campbell and Fiske (1959). Some scholars reported negligible bias (see e.g., Marsh & Bailey, 1991; Tomás et al., 2000) and others cautioned against underestimating the distortions in the factor loadings, especially under specific circumstances (when the product of the method loadings and method correlations is relatively large; see Conway et al., 2004).

Even when the method is a single one (experimental survey-based self-reports, as in the present study), the introduction of error covariances in a latent factor model implies that the latent variable itself cannot account for the entirety of the variance of its indicators. While part of this unexplained variance may be due to random measurement error, covariation among indicators could underlie unspecified systematic causes of the observed variables which are independent of and left unanalysed by the

latent model. This generally unwanted component is sometimes referred to as “construct-irrelevant variance” (Baird, 2010).

According to Kline (2012), the assumption of local independence is restrictive and unrealistic in the behavioural sciences. Its violation probably occurs in empirical research more often than usually acknowledged. He recommends including correlated errors in model specification if they are substantively justifiable. Introducing them without clear substantive reasoning could result in overfitting and the solution could be too sample-specific, failing to generalise to the whole population (see e.g., MacCallum et al., 1992). At the same time, Cole and colleagues (2007) warn that failing to specify justifiable correlated errors could have even more harmful consequences in terms of biased estimates of the factor loadings.

Research has suggested that sources of construct-irrelevant variance can be both methodological and substantive. Brown (2015) lists among the sources of bias the kind of assessment (self-reports or observer ratings), the data collection method (in-person interviews or self-administered online or pen-and-paper questionnaires), the kind of questionnaire items (reversed or similarly worded), the content of the response set, demand characteristics, susceptibility to socially desirable or acquiescent responding, reading disabilities or other cognitive biases. Bocell (2015) mentions other method sources, for example item order effects (especially when questionnaire items are not fully randomised), use of mixed response scales (multiple-choice, true/false, free-recall responses), use of different item stem wording (negatively- and positively-worded stems), context or priming effects (prior questions affecting the following ones), language inconsistencies (use of common/simple versus unusual/complex words). It could be argued that in the present research, certain items of the moral self-regulation

inventory carried the same introductory words (“I felt...”, “I was...”, “It made me feel...”, “Actions...”, “Francia/Nicholas may have...”).

In sum, method effects depend on the type of assessment instrument or item formatting/placement in a survey, whereas substantive effects could be related to personal characteristics or propensities of the participants. All of them could easily result in correlated errors and it is often quite difficult to identify and specify them before data collection or determine their individual contribution, singling out the most impactful.

In Study 3, six error covariances were allowed in CFA between items of the moral self-regulation inventory (Fig. C15-C16), but the final integrated structural models (Fig. C26a-b) only accommodated three in Nicholas’s subsample and none in Francia’s⁶⁵. The items involved in these error covariances are summarised in Table C27.

Table C27:

Moral self-regulation: item pairs with correlated errors in the final integrated structural equation models (Nicholas’s subsample)

Item code	Item wording
SI2	When I read these stories, I feel awakened to the good in the world
SI3	This story strengthens my faith in humanity

Item code	Item wording
SD5	Francia may have done a good deed, but I bet she is seeking the praise of others
SD6	Francia may have had ulterior motives for doing this

Item code	Item wording
SD16	I felt as if I was threatened by something
SD18	I felt resentful

⁶⁵ The revised structural models without opinion-based comparison variables (Fig. C23a-b) only had one correlated error between items SD16 and SD18 in the self-defence latent variable in Nicholas’s subsample.

These item pairs carry repetitive or very similar expressions at the beginning of each statement. Therefore, without ruling out other explanations, one of the possible sources for the emergence of construct-irrelevant variance could be attributed to method effects related to item wording.

Other than that, because analysis of measurement invariance showed negligible differential item functioning (DIF⁶⁶) for the twenty-two indicators of the moral self-regulation inventory, Study 3 could be said to have provided initial evidence of validity and reliability of the dimensional structure and item composition of these endogenous latent variables.

Exogenous Latent Variables

Analysis of the exogenous latent variables presented specific challenges. The measurement instruments chosen for regulatory focus and approach/avoidance were validated scales that have been widely used in psychological research for decades. Despite evidence of validity and reliability in the motivation science literature, in Study 3 these constructs exhibited relatively low internal consistency, in some cases clearly suboptimal. Promotion focus was the latent variable with the weakest reliability, due to the low loadings of two items: Pm1/Unable (.452) and Pm6/Hobbies (.373). This phenomenon caused measurement errors within the latent variable, reducing the overall fit. For that reason, those items were dropped from the latent variable. Notably, the error covariances included in CFA between two promotion focus items and two avoidance items were *not* specified in the full structural models.

Structural Models

⁶⁶ DIF consists in unequal slopes or intercepts of a latent variable indicator across different sample groups (see e.g., Kline, 2016).

Path modelling from Study 2 provided the fundamental specification criteria to carry out full SEM for the total sample and each of the two vignettes in Study 3. Since the path models developed in Study 2 used observed instead of latent variables, the initial structural models in Study 3 did not yield satisfactory fit and required addressing a few issues. First, for the reasons discussed in the previous chapter, a positive correlation between moral discrepancy and self-improvement was expected in Francia's vignette even if it did not appear in the final path model from Study 2. A few parameters in the motivational variables were also modified: a) their covariance structure was adjusted to reflect the findings from the larger sample; b) approach was preferred to promotion focus for the prediction of self-improvement due to a stronger regression weight and a slightly better overall fit.

Although the resulting revised models achieved nearly excellent fit, a few modifications were made to reduce residual error and further improve the substantive theory underlying them. The search for parsimony had led to the prioritisation of moral comparisons based on ability, overlooking the contribution of opinion-based comparisons for the explanation of the dependent variables variance. This narrower specification of the perimeter of the models partly compromised their accuracy, especially in Nicholas's condition, where the impact of less positive judgments of the deed was stronger. From a statistical standpoint, the integrated models inclusive of judgments underlying opinion-based comparisons achieved two important improvements while retaining the same goodness-of-fit: they allowed to increase the total variance explained of the outcome variables (more than doubling it in Nicholas's subsample) and levelled to non-significance the covariance of the disturbances of the outcome variables. This way, from a substantive perspective, the integrated models provided a more generalisable account of the self-regulation of virtue, putting forward a

theoretical explanation that is better equipped to elucidate the mechanisms of the response to moral exemplars not only when their deeds are unequivocally judged as morally excellent (e.g., Francia), but also when they are liable to a wider range of moral judgments, from more to less positive (e.g., Nicholas).

Whereas the measurement model was invariant between vignettes, the integrated structural models were not, replicating the findings from path modelling in Study 2. Analysis of the integrated models by vignette (Fig. 24a-b) and for the total sample (Fig. C25) yielded critical insights. At the level of the total sample (Fig. C25), the two moral self-regulatory modes were predicted by both opinion- and ability-based comparisons. By contrast, at the level of the individual vignette opinion- and ability-based comparisons exerted a comparatively different level of impact on moral self-regulation depending on the nature of the moral action: opinion-based comparisons (judgments about the goodness and the propriety of the deed) were more influential in Nicholas's than in Francia's vignette. They are probably essential to explain the variance in all moral scenarios where the deed results in conflicting opinions, but when comparison processes remain more in the background, as in Francia's vignette, motivational processes acquire comparatively greater relevance, directly affecting moral self-regulation with greater intensity. Indeed, the size of the approach and prevention focus regression weights for the prediction of respectively self-improvement and self-defence was twice as large in Francia's compared to Nicholas's subsample.

Complementary Insights

The secondary analyses provided further discernment about some of the mechanisms that govern the psychological processes experienced by individuals when they witness virtuous deeds performed by others. Latent profile analysis allowed to identify three moral self-regulation profiles. Two profiles of improvers were identified:

one characterised by very strong self-improvement and virtually no self-defence (full-out improvers), and another with strong self-improvement along with mild defensive regulation (improvers) limited to dismissal of the exceptionalism of the moral agent and the attribution of ulterior motives. The third profile (mixed defenders) was a cluster that on average displayed a blend of broadening and defensive states, with a considerable variation around the means with which each self-regulation item was endorsed. This profile did not feature participants who completely rejected acknowledgement of the morality of the agents or denied any form of positive inspiration, but at the same time they questioned the agents' motives and the generalisability of the behaviour as an ethical norm to adhere to; these defensive reactions were probably the result of feelings of either vulnerability and guilt when the agent was seen as almost inaccessibly outstanding (the organ donor), or feelings of threat and resentment when the behaviour implied ethical standards that strongly conflicted with those of the participants (homosexuality).

The association pattern of the moral comparison clusters with the self-regulation profiles closely mirrored the results from the structural models: notably, participants in the upward comparison cluster associated with participants in the two self-improvement profiles, and participants in the downward comparison cluster associated with participants in the mixed defender profile. Moreover, the Sankey diagram (Fig. C29) graphically illustrated that, for the most part, participants engaging in upward comparison were the same who self-improved, and participants engaging in downward comparison were those who self-defended more intensely. The convergence of the results from the variable-centred and the person-centred analyses offered strong evidence of the robustness of the main findings. As Loken and Molenaar (2008) emphasised, much can be learned by applying one approach alongside the other.

One further consideration concerns the size of the positive versus negative regulatory response: overall, the full-out improvers and improvers clusters were much larger than the mixed defenders cluster (69% versus 31%), making moral self-improvement a much more common reaction (relative to self-defence) to the exposure to acts of moral goodness. Interestingly, a recent study by Sun and Goodwin (2020) highlighted that people do not seem to value moral self-improvement goals (e.g., in terms of honesty, fairness, compassion) as much as non-moral improvement targets (e.g., sociability, productiveness, anxiety). Following Hudson and Fraley (2016), the authors attribute this finding to well-being concerns: they argued that becoming more moral results in fewer personal benefits and is often accompanied by personal costs, thus becoming, for instance, less anxious or more sociable turns out to be a safer and more desirable target (Sun & Goodwin, 2020).

This explanation raises a question. If personality changes such as becoming less anxious or more sociable were primarily motivated by the desire to improve one's own rather than others' life, why exposure to others' displays of virtue instigated in the present research such strong self-improvement responses? Although this work does not speak directly to this point, the reasons might be found in the fact that moral self-improvement was measured here as a transient *state* induced by an experimental manipulation, whereas Sun and Goodwin's study refers to changes of more permanent moral *traits*. Study 4 addressed the issue of the extent to which a temporary state of self-improvement translates into actual prosocial behaviour.

Limitations

Although Study 3 accomplished its goal of identifying substantively meaningful solutions with satisfactory psychometric properties for a single measurement model and two similar but distinct structural models (one for each vignette) of the self-regulation of

virtue, the results should be considered *preliminary*, and a few clear limitations must be acknowledged.

First, modifications were introduced *after* fitting the initial hypothesised models in CFA and SEM. The input models (those obtained from Study 2) were clearly provisional, having been developed with exploratory/generative techniques, such as EFA and path modelling with observed variables. Those provisional models did not yield satisfactory fit in Study 3 and therefore *post-hoc* modifications were introduced. Although all of them were theoretically grounded and the specific rationale was discussed in detail, *post-hoc* modifications require replication through further testing among new samples from the same population. In defence of the final models, though, it could be argued that the fundamental relationships across the key variables was repeatedly observed across the first three studies of the present research.

An aspect that would benefit from further research is the measurement of the motivational dispositions, so that their impact on moral self-regulation can rest on firmer ground. Not only did these latent variables exhibit lower reliability than the new moral self-regulation constructs, but also they showed some inconsistent relational patterns. While a smaller sample could have partly explained this phenomenon in Study 2, the large sample size of almost a thousand participants in Study 3 should have ensured stable results. Independent samples *t*-tests between vignettes for each of the four motivational constructs returned non-significant results (see SMC4); yet, in the integrated structural models, they behaved differently between vignettes. The correlation between promotion and prevention focus, which in the literature is usually non-significant or weakly positive, was found to be weakly positive in Nicholas's subsample and weakly negative in Francia's subsample. Inconsistencies also appeared in the correlation between approach and avoidance, which in Study 3 was negative in

Nicholas's subsample and non-significant in Francia's subsample. The effect size of these correlations was always small, regardless of their sign, but these inconsistencies are still worth noting. Additionally, LPA returned six motivational profiles that were complex to interpret; unlike the three very clear self-regulation profiles, they showed a certain level of conceptual overlap and were clearly differentiated only by the avoidance construct. The six-profile solution emerged as the best among others with slightly different parametrisations that yielded ten or even nineteen profiles. These results testify to the intricacy of participants' response to the motivational dispositions, but could also hide possible methodological issues due to the known deficiencies in the respective measurement instruments, particularly chronic regulatory focus, as discussed in the introduction to Study 2.

Beyond Self-Regulation: Measuring Social Behaviour

The structural model fitted to the total sample provided a measure of the effectiveness of the experimental manipulation, highlighting the chain of causal effects induced by the presentation of moral vignettes and experienced in the form of two modes of moral self-regulation, through the mediation of moral comparisons and with the additional contribution of correlated motivational traits. The size of these effects varied from moderate to strong or very strong, and the RMSEA was excellent, although with a significant p -value (persisting lack of invariance). One thing to notice here is that the fairly large regression weight between the variables vignette and goodness of the deed ($r = -.33$; $p < .001$) was obtained from two stimuli whose individual structural models shared more commonalities than differences.

The following study (Study 4) was designed to move beyond pitting vignettes with different (but relatively equipollent) moral deeds against each other, and instead directly compare an experimental condition (Francia, a highly praiseworthy moral deed)

against a control condition functioning as a near-neutral benchmark (a mildly positive moral act). This approach allowed to test whether the experimental condition, via the elicitation of stronger self-improvement through a greater deed, could inspire a higher degree of helping behaviour than the control condition. After all, because self-regulation is in the service of actual behaviour and represents one of the critical predictors of success in life (Baumeister, 2007), the integration of overt social behaviour in the research was an important aspect highlighted in the initial conceptual model.

Study 4: Model Extension

Introduction

The moral self-regulation inventory and the integrated structural models of the self-regulation of virtue obtained with Study 3 represented a fundamental step in the research plan delineated in the introductory chapter (Fig. IV). However, the downstream processes remained to be investigated. Indeed, the original conceptual model (Fig. III) also included the behavioural outcomes of moral self-regulation. The integration in the model of actual social behaviour was therefore the key aim of Study 4, which also provided an opportunity for the exploration of other related constructs.

In more detail, beyond retesting essential aspects of the models, Study 4 was designed to meet three key objectives:

1. measure to what extent prosocial behaviour (specifically, helping behaviour) is affected by different levels of self-improvement experimentally induced by two distinct deeds: an outstanding and a mild moral action;
2. measure mechanisms and strength of the relationship between moral self-improvement and helping behaviour;
3. explore the relationship between moral self-regulation and other related constructs (e.g., regulatory mode, social desirability), providing directions for further research.

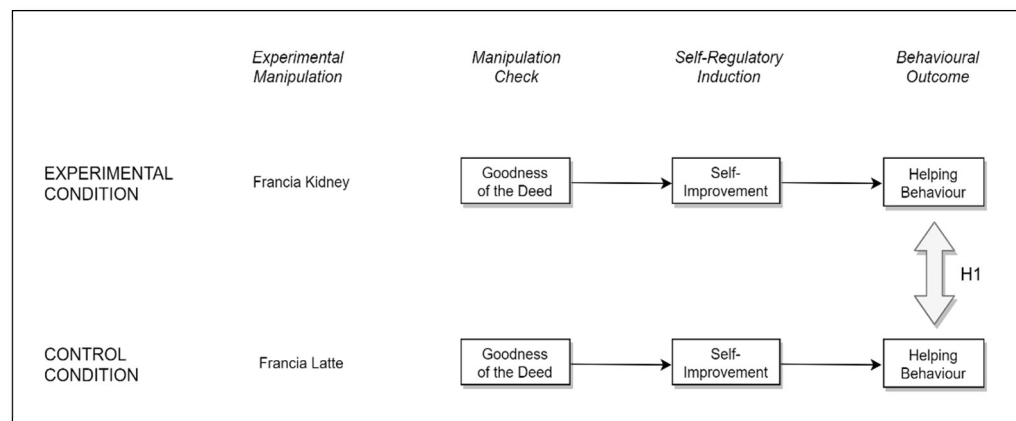
Objective 1: Behavioural Effects

With regard to the first objective, Study 4 was designed to compare measures related to the deed in Francia's vignette to baseline measures obtained from a new moral scenario, with the same protagonists and essentially the same contextual features, except for one fundamental element: the level of goodness of the moral action. Whereas in the experimental condition Francia donated a kidney to a friend who

was at risk of organ failure (a remarkable moral deed), in the control condition Francia bought a latte to a friend who forgot her wallet and could not pay for her drink (a mild act of everyday kindness). Based on the structural model, the higher level of goodness of the deed in the experimental condition was hypothesised to lead to a higher degree of helping behaviour (relative to control) through the elicitation of a higher level of self-improvement (hypothesis 1 – H1; see Fig. D1).

Figure D1:

Objective 1 – Hypothesis 1 (H1): measurement of differential behavioural effects between conditions



The manipulation of the goodness of the deed was clearly intended to provide further evidence to corroborate confidence about its causal impact on self-improvement and indirectly on helping behaviour.

Objective 2: Mechanisms of the Behavioural Effects

Although social cognitive models of self-regulation include overt behaviour in the domain of self-regulation (e.g., Zimmerman, 2005), in the present research the definition of moral self-regulation left behaviour outside. Yet, an understanding of the behavioural implications of moral self-regulation was considered integral to the full

model, representing a measure of concurrent validity. This relationship was precisely the focus of Study 4.

The link between self-regulation and actual behaviour is not necessarily straightforward, since several factors can interfere with people's standards, goals, and strivings. For example, effective self-regulation of behaviour can be impaired if the standards implicated in the regulatory process are unclear or in conflict with each other, or in the presence of poor monitoring mechanisms, or when emotion regulation hinders instead of supporting goal pursuit (Baumeister, 2007). Additionally, plans and implementation intentions can play an important role as facilitators or inhibitors in the pursuit of desired end-states (Gollwitzer & Brandstätter, 1997), enabling the link between thought and action (Haggard & Lau, 2013).

The moral self-regulation inventory from Study 3 accounted for some of the above-mentioned elements: certain items directly or indirectly captured goals and standards in the moral domain (e.g., SI1, SI4, SD2, SD10), others referred to desires and intentions (e.g., SI6, SD8). However, a few indicators initially designed to tap into these components were dropped from the final inventory, because they behaved differently across the two vignettes of Francia and Nicholas (in other words, they were not invariant, but dependent on the moral content of the stimuli). These items, although no longer part of the final moral self-regulation inventory, were still tested in the present study for confirmatory or exploratory purposes. One of them was particularly important, item SI7 ("I feel like I want to do something good for others", originally in the moral self-improvement scale), because it measured the desire to enact altruistic behaviours, a critical (although not necessary, see e.g. Kossowska et al., 2020) pathway to actual behaviour.

In previous experimental research (Schnall et al., 2010), a direct causal link between the emotion of moral elevation (which partly overlaps with moral self-improvement) and helping behaviour was found; there, the measure of elevation included the action tendency item “want to help others”, which is very similar to item SI7. In Study 4, in the absence of a similar action tendency measure within the self-regulation construct, the link between moral self-improvement and helping behaviour could still exist, since other items include motivational tendencies, for instance SI4 (“Francia has shown me how to be a better person”) and SI6 (“I want to be more like Francia”), but it could potentially be weaker or marginally non-significant, as at least some of the variance could be absorbed by the desire to enact helping behaviours (item SI7).

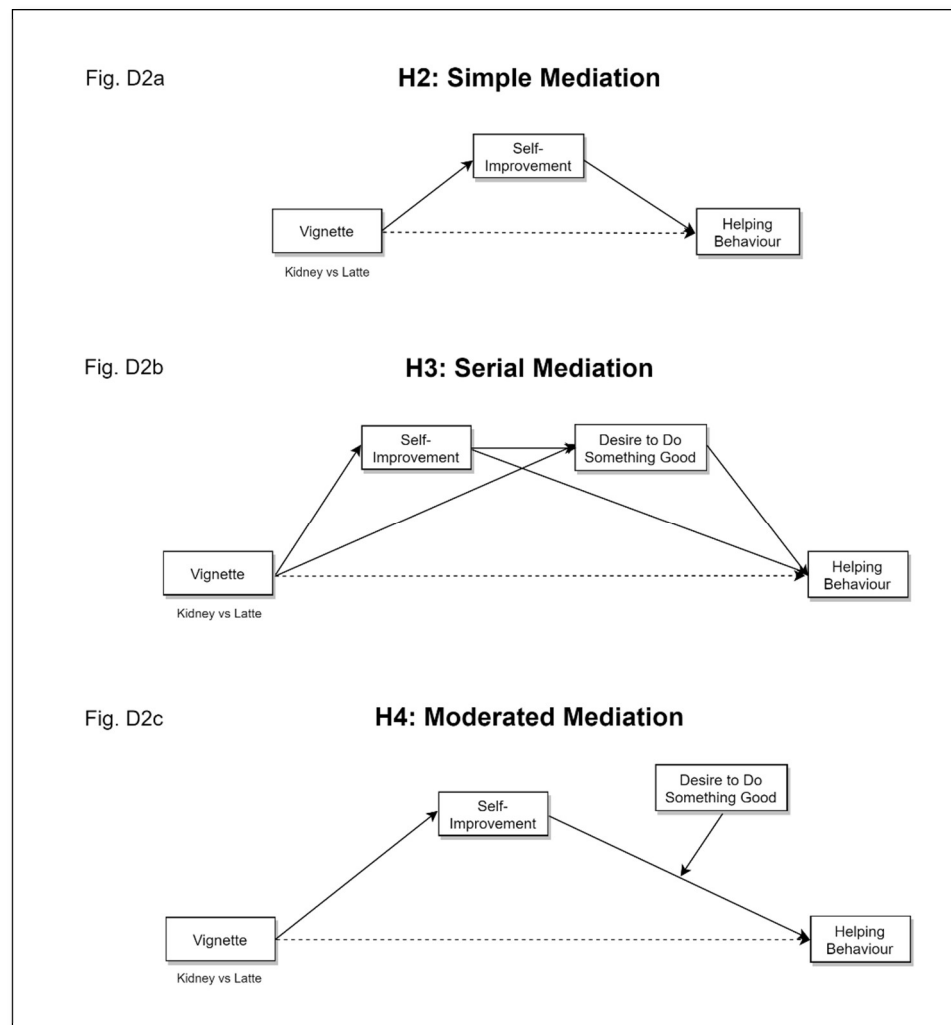
For this reason, in Study 4 the relationship between the final self-improvement scale and helping behaviour could potentially take different forms, each corresponding to alternative hypotheses. The relationship could be direct, with item SI7 having no influence on the outcome. However, item SI7 could also play a decisive role in bridging self-improvement and helping behaviour; therefore, in addition to the primary hypothesis of a *direct* relationship between self-improvement and helping behaviour, a secondary *indirect* hypothesis was formulated, with the desire to do something good for others functioning as a mediator (partial or full). A further alternative hypothesis was also tested, according to which self-improvement could exert a stronger or weaker influence on helping behaviour depending on the strength of the desire to do something good for others; in this case, the action tendency item SI7 would function as a moderator.

With the addition of the variable vignette as predictor, the three hypothesised mechanisms of simple mediation (hypothesis 2 – H2), serial mediation (hypothesis 3 –

H3), and moderated mediation (hypothesis 4 – H4) are graphically illustrated in Figures D2a-c.

Figures D2a-c:

Objective 2 – Hypotheses 2-4: three alternative mechanisms of simple mediation (H2), serial mediation (H3), and moderated mediation (H4)



Note: the dotted lines represent direct pathways that might or might not be significant, depending on whether the mediation process is partial or full (not a critical pathway for these tests).

If Study 4 results supported any of these three hypotheses, evidence would be obtained that not only certain emotional states (e.g., moral elevation, kama muta) but also self-regulatory states (moral self-improvement) would effectively lead to prosocial

behaviour, either directly or through conditional processes. This finding would complement respectively Schnall et al.'s (2010) and Blomster Lyshol et al.'s (2020) effects of moral elevation and kama muta on prosociality.

Objective 3: Further Exploration

Beyond the two objectives described above, Study 4 was also designed to analyse the relationship between moral self-improvement with other constructs of interest, such as motivational drivers (regulatory mode), social desirability, social comparison orientation, locus of control, personality traits (humility, narcissism), and self-growth constructs (e.g., hedonia and eudaimonia motives, growth motivation, etc.). These constructs were measured in the second part of the questionnaire in a *post-task* whose primary function was to provide a measure of helping behaviour (the outcome variable of this study), but also offered the opportunity to conduct further exploratory work with a view to informing future research. The methodological aspects will be clarified in the next section of this chapter. Some of the aforementioned constructs, briefly discussed in the introductory chapter, had been identified as pertinent to the project since its inception: regulatory mode, humility, narcissism. Others emerged in response to the results of Studies 1-2: socially desirable responding and social comparison orientation. Others still were envisaged while designing Study 4: internal locus of control and self-growth constructs. As a whole, the investigation of these additional variables was intended to provide: a) preliminary information about their nature of predictors of moral comparison and self-regulation; b) specific psychometric properties of moral self-improvement (and partly self-defence), namely convergent and discriminant validity. The rationale for their investigation is further discussed next.

Regulatory Mode

In the context of the self-regulation of virtue, regulatory mode was expected to function as an exogenous predictor, like regulatory focus and hedonic orientation, with which it constitutes a prominent triad in motivation science (Higgins, 2014). Like the other two measures mentioned earlier, regulatory mode operates in a cybernetic framework. The two regulatory modes of assessment and locomotion consist in distinct but complementary concerns respectively for truth and control (Higgins, 2014). Assessment mode revolves around needs for accurate evaluation and scrupulous decision making through meticulous pondering of strengths and weaknesses of alternative choices and courses of action; by contrast, locomotion mode is rooted in psychological needs for constant motion, ongoing change, and state shift (Kruglanski et al., 2000). Both are critical to attain an effective link between thought and action, assessment stressing goal setting (deliberative phase) and locomotion emphasising goal striving (implementation phase), as delineated in the Rubicon model of action phases in goal-orientated behaviour (Gollwitzer et al., 1990; Heckhausen & Gollwitzer, 1987). The Rubicon model posits a clear separation between these two phases: the deliberative phase is focused on choices regarding the likelihood that a particular goal could bring about a desired end-state, or regarding the worth of a particular goal as opposed to an alternative goal (consistent with assessment mode); the implementation phase entails execution of the intended course of action and maintenance of the effort to sustain action (consistent with locomotion mode). According to the Rubicon model, once a decision is made, further deliberative reflection is usually inhibited and psychological resources converge toward the effective implementation of the planned actions, echoing Julius Caesar's principle expressed in his famous claim "Alea iacta Est" (literally, *the die has been cast*) as he crossed the river Rubicon, declaring the reflective phase

over and inciting his legions to initiate the fight in the imminent civil war (Heckhausen & Gollwitzer, 1987).

Empirical research on regulatory mode applied to the moral domain is scant. Recently, Cornwell and Higgins (2014) have explored the links between locomotion and assessment with the moral foundations, identifying a positive association, particularly among liberals, between locomotion and the binding moral foundations, primarily based on concerns for social *usefulness*; this link can be severed when participants are experimentally induced into an assessment mode, which shifts the focus to moral *truth* (Cornwell & Higgins, 2014).

Regarding the relationship between regulatory mode and prosocial behaviour, newly published research (Baldner et al., 2020) has found that a locomotion orientation, thanks to the high importance placed on goal attainability, contributes to the formation of helping goals (along with sympathy toward a specific target in need), which is a critical antecedent of helping behaviour.

Despite the scarcity of empirical research in this area, locomotion mode seems conceptually related to self-improvement strivings, given its focus on action and change. High locomotors crave for movement and show an impatience with any barriers or delays; they worry about standing still, and so they readily embrace opportunities for change, rarely look back or engage in counterfactual thinking, and usually experience positive affect (Kruglanski et al., 2013). If locomotors are presented with an act that they deem morally admirable, their favourable judgment is likely to trigger positively valenced conation and action, and for this reason locomotion mode was expected to positively correlate with moral self-improvement, and indirectly with prosocial behaviour.

By contrast, assessment mode seems conceptually related to defensive self-regulation. High assessors are concerned with epistemic needs and go to great lengths to find out the “right” option, making comparisons between several alternatives and evaluating “counterfactual ‘might have beens’ or imagined futures” (Kruglanski et al., 2013, p. 81), which make them more susceptible to negative affect, particularly guilt or regret (Kruglanski et al., 2013). Various studies have also associated assessment with neuroticism (Kruglanski et al., 2000) and depressive moods (Hong et al., 2004). This penchant to rumination and volatility predisposes high assessors to actively seek and detect flaws in their own and others’ behaviour; when presented with another person’s good deed, they might be inclined to thoroughly examine it (along with its context) until they identify imperfections, weak spots, faults, ulterior motives, as they would do with their own behaviour (see Komissarouk et al., 2018); the more they weigh up strengths and weaknesses, the more likely they are to devalue the action and defensively dismiss the positive character of the moral agent. For these reasons, assessment mode was anticipated to be positively correlated with moral self-defence.

In addition, given the prominence of accurate evaluation concerns, assessment mode was also expected to bear a strong positive correlation with social comparison orientations (Higgins et al., 2003), which will be considered next.

Social Comparison Orientation

In the discussion of Study 2, it was observed that individual differences exist among people in their tendency to engage in social comparisons with others (Gibbons & Buunk, 1999). It was argued that participants’ disposition to engage in comparisons with the moral agents in the vignettes might have correlated with distinct self-regulatory patterns. More specifically, participants with stronger social comparison orientations could be more prone to self-defensive reactions, as they usually tend to see others as

“competitors”, particularly when they engage in ability-based comparisons. Therefore, in Study 4 a positive correlation was anticipated to exist between self-defence and social comparison orientation (driven by ability-based comparisons).

Additionally, as noted earlier, social comparison orientation was also expected to be positively correlated with assessment mode, due to the shared disposition to engage in ongoing evaluations of the social environment.

Social Desirability

As previously discussed, research participants sometimes show a tendency to under-report socially undesirable thoughts and feelings, and over-report socially desirable ones, especially when the topic is somehow sensitive (Krumpal, 2013). Even in the absence of any concerns for impression management, participants may still be subject to a non-conscious self-deception bias and respond in ways that are more socially desirable (Paulhus, 1984).

In the discussion of Study 1, it was noted that some form of social desirability bias could have potentially been at play in the present research: participants might have self-reported their response to the good deeds presented to them in ways that, to a certain extent, inflated self-improvement, and especially deflated self-defence. Therefore, in Study 4 it was anticipated that a positive correlation would be found between self-improvement and social desirability, and a negative correlation would be found between self-defence and social desirability.

Internal Locus of Control

The notion of locus of control was introduced in psychology by Rotter in the context of his social learning theory of personality back in 1954 (published over a decade later). It was defined as a generalised expectancy for internal as opposed to external control of reinforcements (Rotter, 1966). He believed that people differ in the degree to which

they tend to think of events as under their own influence (internal locus) or under the control of outside influences (external locus); he also specified that internality and externality should not be considered typologies, but two ends of a continuum (Rotter, 1975).

In the 1980s the literature on ego actions started to investigate in more depth the relationships between the loci of control and coping/defence mechanisms. Vickers and colleagues (1983) found a significant positive correlation between externality and self-defence, and a positive correlation (albeit weaker) between internality and coping mechanisms⁶⁷. Additionally, Furnham and Steele (1993), in their review on locus of control measures, discuss defensive externality in terms of external attributions motivated by expected failures.

Because moral self-defence can be considered a subset of wider ego defensive mechanisms and moral self-improvement a specific kind of coping mechanism, it was expected that in Study 4 internal locus of control would be negatively associated with moral self-defence, and possibly positively associated with moral self-improvement⁶⁸. Further, because of the respectively high concerns for inner control and low concerns for external verification, internal locus of control was expected to correlate positively with locomotion mode and negatively with social comparison orientation.

Narcissism and Modesty/Humility

As discussed in the introductory chapter, narcissism could be potentially related to downward comparison and defensive self-regulation, and modesty/humility to upward

⁶⁷ In the literature on ego actions, Haan (1985, 1986) defines *coping* as a set of adaptive, purposeful, conscious behaviour choices that adhere to reality and are morally superior, and *defences* as compelled, maladaptive, rigid behaviours that distort reality and hinder mature moral behaviour (see also Paulhus et al., 1997).

⁶⁸ *Possibly* positively associated with moral self-improvement because of the weaker correlation between internality and coping mechanisms (see above and in Vickers et al., 1983).

comparison and self-improvement. Although narcissism and humility were left out of the main model due to parsimony concerns, Study 4 provided the opportunity to explore their correlation patterns. It was expected that both facets of narcissism would positively correlate with self-defence, especially vulnerable narcissism, since the wide majority of moral self-defence items pertain to self-protection mechanisms, which the literature found to be associated more strongly with low self-esteem and vulnerable narcissism (Hart et al., 2018; Rohmann et al., 2012). Further, because of the tendency by modest/humble people to hold moderate and accurate views of one's strengths and weaknesses (Davis et al., 2016), a positive correlation was expected between these personality traits and moral discrepancy (through upward comparisons) and self-improvement.

The two facets of narcissism have also been investigated in the context of regulatory mode theory. In a recent study, Hanke and colleagues (2019) found that the need to “get ahead” of others, typical of grandiose narcissists (see also Morf & Rhodewalt, 2001), can be fuelled by strong locomotion concerns, which emphasise movement and advancement. By contrast, the sense of insecurity and fear of failure that characterises vulnerable narcissists can be sustained by greater sensitivity to the evaluation of information from the social environment, which is typical of assessment mode. Therefore, and in accordance with Boldero and colleagues (2015), these correlation patterns between narcissism and regulatory mode were expected to be replicated in Study 4.

Self-Growth Constructs

Lastly, the voluntary task in the second part of the Study 4 questionnaire allowed to measure a few constructs related to the growth of the self that were recently developed in the moral and humanistic psychology literature.

The first of them was the hedonia and eudaimonia motives for activities (HEMA; Huta & Ryan, 2010), which the authors define as two distinct (and partly overlapping) motivational sources of subjective well-being: hedonia motives concern seeking pleasure or comfort, and eudaimonia motives seeking development or use of the best in oneself. Both are conceptualised as *motives* for acting with specific orientations, and for this reason both can be seen as predictors of subjective well-being broadly defined as “one or more subjectively experienced states or evaluations of one’s life” (Huta & Ryan, 2010, p. 736). Hedonic and especially eudaimonic motives encompass the search for elevating experiences, such as awe, transcendence, connection with a greater whole and other broader levels of functioning (Huta & Ryan, 2010) and therefore should be associated with moral self-improvement processes, which are characterised by the elevating sense of inspiration experienced when witnessing outstanding moral deeds. Consequently, in Study 4 it was expected that eudaimonic motives for activities correlated positively with moral self-improvement.

The second self-growth measure explored in Study 4 was the growth motivation index (GMI; Bauer et al., 2019), assessing the desire to foster personal growth. The authors describe it as a eudaimonic motive of the good life centred around personally meaningful self-growth pathways, which could manifest in two forms: one focused on the cultivation of happiness and well-being (experiential eudaimonia), and the other on the cultivation of wisdom and psychosocial maturity (reflective eudaimonia) (Bauer et al., 2008). Because of the clear reference to self-growth and improvement in areas pertinent to virtue and the good life, in Study 4 it was expected that the growth motivation index correlated positively with moral self-improvement.

The third self-growth construct explored in Study 4 was the desire and commitment to self-improvement (Breines & Chen, 2012), which assesses the motivation to improve

one's weakness and grow as a person. Again, this measure was expected to correlate positively with moral self-improvement.

If moral self-improvement showed the expected correlation patterns with eudaimonic motives for activities, experiential eudaimonia of the growth motivation index, and desire and commitment to self-improvement, these correlations would provide initial evidence of *convergent validity* for the new construct of moral self-improvement. Also, if no significant correlations were found between these three constructs and moral self-defence, they would provide initial evidence of *discriminant validity* for the new construct of moral self-defence.

Methods

Data Collection

Study 4 was initially designed to be conducted during the months of February and March 2020 in the Social Processes Laboratory at Lancaster University (UK) using Qualtrics self-administered questionnaires run on personal computers. When data collection was nearly halfway through, it had to be halted due to the COVID-19 outbreak and the subsequent lockdown. The dataset was too small to be analysed and since there was no sign of a full reopening of campus for the foreseeable future, in April it was decided to reset, switching to online data collection⁶⁹. A few minor changes were made to the script of the Qualtrics questionnaire to make it suitable to run online in the US through CloudResearch/MTurk (for consistency with Studies 2-3). The sample was also increased to account for the new data collection method. At its completion, it was noticed that data quality was poor, as nearly half of participants had to be removed from the sample; analysis of their answers to the check questions suggested that they might have not read the questions or might have answered randomly, in violation of the guidelines set up by MTurk. This behaviour could have been caused by the difficult material and psychological conditions that participants could have been experiencing during the lockdown. Given the unusual loss of participants, CloudResearch supplemented a top-up of new participants to make up for the discarded questionnaires. At the end of data collection, the intended sample was approximately achieved (further details in the Participants section).

Experimental Design and Materials

⁶⁹ The dataset collected in the laboratory was *not* analysed (except for data exclusions).

Study 4 had a between-subjects design (Fig. D1): participants were randomly assigned to one of two independent conditions. In each condition, participants were presented with one of two moral vignettes: Francia Kidney or Francia Latte (Appendix 3). The initial study in the laboratory was preregistered on the OSF website; the revised online version was registered again just after the first part of data collection, before the extra-sampling and, crucially, before data analysis. A structured questionnaire was made available to potential participants in return for payment. Repeat participants from the previous two studies were automatically excluded via the MTurk participant ID.

Questionnaire Structure and Procedure

The study had received prior ethical approval by the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University (UK), and was divided in two parts. In part one, the initial instructions informed participants that the study was about recall and reactions to social situations (see SMD1). To minimise demand characteristics, the cover story asked participants to carefully read a vignette describing a social situation for an episodic memory experiment, followed by some questions about the situation itself. After the socio-demographic questions, the moral self-evaluation question, the measurement of critical predictors (regulatory mode), and a few faking items related to participants' perceived mnemonic abilities (to reinforce the cover story), participants were presented with a vignette and carried out the ostensive memory task (free recall of the story, used as attention check). Next, they answered the usual response questions involving moral judgments, self-regulation items, and an open-ended question to allow them to freely share comments.

Next, the study was declared finished, and participants were thanked and debriefed. However, at that point – unbeknownst to participants – the second part of the study started. Before receiving the MTurk code for payment, participants were

offered to voluntarily take part in an additional unpaid task. They were told that they were under no obligation to participate, but any number of questions they would answer from the ensuing personality scales would have greatly helped the experimenter. They were also told that they could interrupt the task at any time (without being penalised) by simply clicking the “End” button at the bottom of each screen. The personality scales in the post-task comprised the constructs for the exploratory analyses described earlier.

Lastly, an open-ended question allowed participants to describe their perception of the purpose of the study, after which they were thanked for their effort, provided with their MTurk code for payment, and fully debriefed, including an explanation of the post-task and its rationale.

Participants

Sample Size Determination

The sample size for the online study was predetermined based on power analysis. Two kinds of power analysis were conducted: one for an independent samples *t*-test or Mann-Whitman *U*-test, and one for mediation analysis. G*Power ver. 3.1 was used for the first analysis, while the second was run using Schoemann et al.'s (2017) application for Monte Carlo power analysis for mediation models in R.

Power analysis for a *t*-test or *U*-test was based on the effect size of the helping behaviour mean difference across experimental and control conditions estimated in the second experiment by Schnall et al. (2010). As a conservative strategy, here the lower limit of the confidence interval of Schnall et al.'s effect size was taken ($d = .5402$). The total sample size to detect this effect was determined to be 88 and 92 participants, respectively for a one-tailed *t*-test and a one-tailed *U*-test, assuming $\alpha = .05$, power = .80, and equal sample size across conditions (see SMD4 for further details).

Power analysis for mediation analysis was based on a model with self-improvement as predictor, desire to do something good as mediator, and helping behaviour as outcome (corresponding to hypothesis 3). Assuming $\alpha = .05$, power = .80, and a standardised correlation coefficient of .30 between self-improvement and helping behaviour (using a more conservative estimate of the original correlation coefficient of .49 measured in the second experiment by Schnall et al., 2010), the required sample size for the experimental condition was determined to be 120 participants, hence a total sample of 240 participants (see SMD4 for further details).

Because the sample size requirement for mediation analysis was more stringent than for a *t*-test or *U*-test, that sample size (240 participants) was regarded as more appropriate and augmented to 300 participants in consideration of the objective to analyse the post-task data and run more complex path models.

Data Collection and Cleaning (Step 1)

At the end of the first part of online data collection, the dataset comprised a total of 430 subjects. Of these, 30 did not provide consent or dropped out before providing consent, 26 declared to have poor English language skills, 70 dropped out before completing the questionnaire, 60 failed the two instructed response questions, 82 failed the attention check based on the free recall of the story, and 2 correctly guessed the purpose of the study. Having removed these subjects from the sample (following the preregistration plan), the remaining sample size was only 160, well below the expected 300 participants. The loss of power was caused for the most part by the removal of participants who failed the attention checks based on the instructed responses and the free recall (142 participants). The analysis of the open-ended responses to the free recall questions allowed to determine the clear random nature of the responses, bearing no relation whatsoever with the stories in the vignettes.

Data Collection and Cleaning (Step 2)

The top-up sample provided 191 additional participants, of whom 7 did not give consent or dropped out before providing consent, 38 dropped out before completing the questionnaire, 9 failed the two instructed response questions, 15 failed the attention check based on the free recall of the story, and 9 correctly guessed the purpose of the study. Having removed these participants from the top-up sample, 120 participants remained and were added up to the previous 160 to form the new total sample of 280, a much closer number to the expected 300 participants.

Multivariate Outliers

At that point, outlier analysis was conducted, reproducing the same methodology applied in the previous studies using the combined analysis of centred leverage values, Mahalanobis distance, and Cook's distance. The observations that exceeded two out of three of the cutoff points determined by those values were considered extreme multivariate outliers. This analysis enabled the detection of 6 outliers, which were subsequently deleted from the sample, following the preregistered plan.

Final Sample Composition

After the above-mentioned exclusions, the final valid sample was composed of 274 participants: 141 respondents allocated to the Kidney condition (experimental) and 133 to the Latte condition (control). It comprised 125 females (45.6%), 148 males (54.0%), and 1 participant (0.4%) who self-reported "other" (non-binary) to the gender question. Age ranged from 20 to 68 years⁷⁰, with median of 34 and mean of 37 years ($SD = 11$). All participants were US residents, of which 269 (98.2%) were US nationals and the remaining 5 (1.8%) from other nationalities. The median completion time was 15

⁷⁰ One participant's answer to the age question was 1993. Given the absence of other issues on the participant's record, it was considered to indicate the year of birth and converted to 27 years of age.

minutes. There were no significant differences in the socio-demographical composition of the sample across the two conditions (see SMD5).

Measures

Many of the measures in Study 4 were the same as in Study 3 (e.g., socio-demographic questions, moral judgment questions, moral self-regulation inventory), whereas others were new (helping behaviour, regulatory mode, scales measured in the post-task). The new measures will be described next in more detail, but a couple of important considerations about the moral self-regulation inventory must be discussed first.

Moral Self-Regulation

Two critical (and related) points concerning the measurement of moral self-regulation must be emphasised for Study 4. First, moral self-regulation was measured through the scales consisting of ten self-improvement and twelve self-defence items. However, the two self-improvement and the seven self-defence items that were dropped from the respective final scales were also included in the questionnaire to test further hypotheses. For instance, the inclusion of item SI7/ForOthers was justified to enable the test of hypotheses 3 and 4 (Fig. D2b-c), and the inclusion of item SD4/People was related to the substantiation of expectations about the functioning of the self-defence scale in the control condition.

The latter point leads to the second important consideration about the moral self-regulation inventory. Effectively, Study 4 was designed to investigate primarily self-improvement and helping behaviour, thus focusing on the positive response to greater or smaller good deeds; the development of the control vignette (Francia Latte) and the formulation of the key hypotheses reflected this specific aim. Therefore, the negative response to good deeds (moral self-defence) was only considered for a restricted set of

secondary exploratory analyses limited to the experimental condition. In fact, moral self-defence was expected to be a valid measure only for the experimental condition (Kidney), but *not* for the control condition (Latte). This is because its items were designed to measure self-serving defensive reactions which apply to virtuous acts of a certain entity, such as the donation of a kidney in the experimental vignette; however, many of them were *not* appropriate for small-scale acts of ordinary goodness (e.g., paying a latte for a friend in the control vignette), as they would *not* constitute instantiations of defensiveness. A couple of key examples can help clarify this important point. Scoring high on item SD1 (“In many ways, I have done greater deeds than Francia”) can be legitimately considered an overly flattering self-enhancing response when referred to the outstanding moral act of donating a kidney, but cannot be regarded as defensiveness when referred to an everyday act of kindness like paying a latte for a friend. For most participants it will be probably true that they have performed a more significant moral action than that in their lives; as such, scoring high on this item simply constitutes the recognition of a true fact, not a manipulation of the reality to favour the self beyond what objective facts would warrant, which is precisely how self-enhancement/self-protection are defined according to Sedikides and Alicke (2012). Similarly, scoring high on SD3 (“It is not such an extraordinary action”) should be true for most people when referred to paying a latte for a friend (a truly ordinary action), while it would clearly indicate self-defence in the case of donating a kidney, which *is* an extraordinary action. A similar reasoning can apply to other items, such as SD10 (“Everyone occasionally does something really good, so Francia isn’t more praiseworthy than anybody else”) and even items dropped from the final scale, such as SD4 (“I know people who have done greater deeds than Francia”), which for this reason were included in the Study 4 questionnaire. For all these items (SD1, SD3, SD10, and SD4), in

the control condition the scoring expectancy for the measurement of self-defence is opposite to what it should be; while in general self-defence should be lower (or equivalent) for accessible everyday good deeds compared to outstanding deeds, for these items a significantly higher score was anticipated to be found in the control condition because they clearly reflect an objective reality for most people. If evidence for this expected pattern were obtained from the data, then the self-defence scale would not be analysed among the Latte subsample and the total sample, and would be regarded as a valid measure only for the experimental condition in the secondary exploratory analyses⁷¹.

New Measures

Helping Behaviour. Amongst the new measures, helping behaviour was the most important, since it represented the dependent variable through which behavioural outcomes of moral self-regulation were measured and added to the model. It is important to emphasise here that the behaviour measured in the present study was *actual* behaviour, not merely a self-reported behavioural intention or propensity. The measure of actual helping behaviour was obtained thanks to a voluntary task offered to participants in the second part of the questionnaire, as described earlier in this chapter.

When the study was designed, a deliberate attempt was made to devise a measure of overt helping behaviour that was *not* categorical (e.g., *helped* versus *did not help*). This was achieved by offering participants an extra-task consisting of 122 questions, divided into multiple personality scales. The number of questions answered would count as a measure of helping behaviour: the greater the number, the higher the degree of

⁷¹ The identical list of moral self-regulation items (those of final inventory and the dropped items) was tested for the two conditions also for consistency reasons (to retain the same questionnaire regardless of the stimuli). For the complete list of the moral self-regulation inventory tested in Study 4, see SMD2.

help voluntarily offered by the participant to the experimenter. Thus, helping behaviour was designed as a continuous variable ranging from 0 (declined to participate in the post-task) to 122 (answered to all the post-task questions). The measure was not anticipated to be normally distributed, but it was designed to capture the nuances of a phenomenon that was expected to be graded, not dichotomous.

While the order of the items within the scales in the post-task was randomised for all scales, the scales themselves were kept in the same order for all participants, so that a minimum sample size was achieved at least for the most critical measures (that were asked first).

Regulatory Mode. Other than helping behaviour, the other important new measure of Study 4 was regulatory mode. The measure was collected in the first part of the questionnaire, before the presentation of the vignettes, following the flow adopted for regulatory focus and hedonic orientation in the previous studies.

The regulatory mode questionnaire was developed by Kruglanski and colleagues (2000) and consists of twelve items for the measurement of assessment mode (three reverse-coded) and twelve items for the measurement of locomotion mode (two reverse-coded). Examples of assessment items are “I often compare myself to other people” and “I spend a great deal of time taking inventory of my positive and negative characteristics”; examples of locomotion items are “I enjoy actively doing things, more than just watching and observing” and “When I decide to do something, I can’t wait to get started”. The original version also comprises six faking items, which have not been included in the present study.

The twenty-four items were presented in randomised order and participants answered using scales ranging from 0 to 100 (with slider bars), whose anchor points

were worded as in the original scales: 0 = *strongly disagree*, 20 = *moderately disagree*, 40 = *slightly disagree*, 60 = *slightly agree*, 80 = *moderately agree*, 100 = *strongly agree*.

Social Desirability. Socially desirable responding was measured through the short social desirability scale (SDS-S; Reynolds, 1982). The original social desirability scale was developed by Crowne and Marlowe (1960) and consisted of a thirty-three-item scale that over the years became widely used in social psychology. However, due to its length, it proved to be difficult to administer in questionnaires that already include multiple measures, a problem shared with other popular social desirability scales, such as the balanced inventory of desirable responding (BIDR; Paulhus, 1994). Strahan and Gerbasi (1972) proposed shorter versions of Crowne and Marlowe's questionnaire, and a decade later Reynolds validated three different scales, of which the third (a thirteen-item scale) showed the best psychometric properties. That scale (with eight reverse-coded items) was adopted in Study 4. Item examples are "No matter who I'm talking to, I'm always a good listener" and "I have never deliberately said something that hurt someone's feelings". The items were presented in randomised order and participants answered using unipolar scales ranging from 0 to 100 (with slider bars), whose anchor points were worded as in the original scale: 0 = *not at all true*, 100 = *very true*. Higher values on the scale correspond to stronger socially desirable responding.

Social Comparison Orientation. In the social comparison literature, there was not much to choose from to measure social comparison orientation, as very few attempts were made to assess individual differences in tendencies to engage in social comparisons; the Iowa-Netherlands comparison orientation measure (INCOM; Gibbons & Buunk, 1999) was the obvious choice. It consists of a bidimensional scale of two correlated factors, one with six items measuring the tendency toward ability-based comparisons (one reverse-coded) and another with five items measuring the tendency

toward opinion-based comparisons (one reverse-coded). Given the correlation expected between the two factors, only the two sub-scales were anticipated to be used in the analysis.

Item examples are “If I want to find out how well I have done something, I compare what I have done with how others have done” and “I always like to know what others in a similar situation would do”. The items were presented in randomised order and participants answered using scales ranging from 0 to 100 (with slider bars), whose anchor points were worded as in the original scales: 0 = *strongly disagree*, 25 = *disagree*, 50 = *neutral (neither agree nor disagree)*, 75 = *agree*, 100 = *strongly agree*.

Locus of Control. Locus of control was measured as a trait through the internal control index (ICI; Duttweiler, 1984). This scale has replaced the original twenty-nine-item scale developed by Rotter (1966), which was criticised for its multidimensionality, forced choice format, low item total-score correlation, and inclusion of items not representative of the construct (Duttweiler, 1984). The ICI possesses stronger reliability and validity (confirmed by subsequent studies, e.g. Goodman & Waters, 1987; Meyers & Wong, 1988) and consists of twenty-eight items (half of them reverse-coded) that map onto one single factor (Jacobs, 1993).

The items were introduced by the statement “Thinking of your normal or usual attitudes, feelings, or behaviours, with reference to the following statements, to what extent...” and items examples are “...do you like jobs where you can make decisions and be responsible for your own work?” and “...does what other people think have strong influence on your behaviour?” (reverse-coded).

The items were presented in randomised order and participants answered using unipolar scales ranging from 0 to 100 (with slider bars), whose anchor points were

worded as in the original scale: 0 = *rarely*, 25 = *occasionally*, 50 = *sometimes*, 75 = *frequently*, 100 = *usually*.

Narcissism. The scales to measure the two facets of narcissism are very common. Grandiose narcissism was measured through the short version of the narcissistic personality inventory (NPI-16; Ames et al., 2006), while the hypersensitive narcissism scale (HSNS; Hendin & Cheek, 1997) was used to measure vulnerable narcissism. The former consists of a unidimensional scale comprising sixteen items that cover the aspects of exhibitionism, entitlement/exploitativeness, and leadership/authority, whereas the latter is a unidimensional scale that includes ten items covering the aspects of vulnerability, hypersensitivity, and entitlement (Crowe et al., 2019).

The original NPI-16 has a forced choice format, but in Study 4 the items were measured through a semantic differential of 101 points; an example of the two opposing statements is “I am an extraordinary person” (highest narcissistic response corresponding to 100) and “I am much like everybody else” (highest non-narcissistic response corresponding to 0). They were presented in randomised order.

Answers to the HSNS items were provided by participants with slider bars using scales ranging from 0 to 100, whose anchor points were worded as in the original scale: 0 = *strongly disagree*, 25 = *disagree*, 50 = *neither agree nor disagree*, 75 = *agree*, 100 = *strongly agree*. Item examples are “My feelings are easily hurt by ridicule or by the slighting remarks of others” and “I dislike sharing the credit of an achievement with others”. They were presented in randomised order.

Modesty/Humility. Modesty was measured through the four questions (two reverse-coded) that make up the modesty sub-trait in the Honesty/Humility dimension of the HEXACO questionnaire (Lee & Ashton, 2004). An item example is “I wouldn’t want people to treat me as though I were superior to them”.

For the assessment of humility, Study 4 used the intellectual humility scale developed by McElroy and colleagues (2014), modified from informant-report to self-report. The scale includes sixteen items mapping onto two correlated factors: intellectual arrogance (e.g., “I value winning an argument over maintaining a relationship”) and intellectual openness (e.g., “I am good at considering the limitations of my perspective”).

The four modesty items and the sixteen intellectual humility items were mixed and presented in randomised order. Participants answered using scales ranging from 0 to 100 (with slider bars), whose anchor points were: 0 = *rarely*, 25 = *occasionally*, 50 = *sometimes*, 75 = *frequently*, 100 = *usually*.

Self-Growth Constructs. The first of the three self-growth constructs explored in Study 4 was hedonia and eudaimonia motives for activities (HEMA; Huta & Ryan, 2010). This measure consists of five items assessing hedonic motives (e.g., “Seeking enjoyment” and “Seeking pleasure”) and four items assessing eudaimonic motives (e.g., “Seeking to pursue excellence or a personal ideal” and “Seeking to use the best in yourself”). The nine items were presented in randomised order and participants answered using unipolar scales ranging from 0 to 100 (with slider bars), whose anchor points were worded as in the original scales: 0 = *not at all*, 100 = *very much*.

The second self-growth construct, the growth motivation index, was assessed through the GMI measure developed by Bauer and colleagues (2019), who tested it in the U.S., Japan, Guatemala, and India. It consists of four items measuring the motive to cultivate critical self-reflection and intellectual development (GMI-reflective) and four items measuring the motive to cultivate personally meaningful activities and relationships (GMI-experiential). According to the authors, the former reflects wisdom (thinking well), the latter maturity (feeling good). A GMI-reflective item example is “I

actively seek new conceptual or philosophical perspectives from which to think about life, even if they mean I've been wrong all along". A GMI-experiential item example is "The important activities in my life are activities that involve the people I love". The eight items were presented in randomised order and participants answered using unipolar scales ranging from 0 to 100 (with slider bars), whose anchor points were worded as in the original scale: 0 = *never*, 50 = *periodically*, 100 = *always*.

The third self-growth construct, desire and commitment for self-improvement (Breines & Chen, 2012), consists of seven items, for instance "I want to find opportunities that will challenge me and help me grow as a person" and "It's up to me whether or not I continue to have certain weaknesses". The items were presented in randomised order and participants answered using scales ranging from 0 to 100 (with slider bars), whose anchor points were worded as in the original scale: 0 = *strongly disagree*, 100 = *strongly agree*.

Analytic Approach

The final dataset was analysed using R ver. 3.6, RStudio ver. 1.3, IBM SPSS Statistics ver. 26, and Jamovi ver. 1.2.27 (Jamovi Project, 2020). Data analysis comprised the usual preliminary inspection of descriptive statistics and assumptions for statistical testing and estimation, as well as primary and secondary analyses, following the preregistered plan:

- *primary analyses* encompassed first and foremost null hypothesis significance testing (and Bayes factors⁷²) for the comparison of helping behaviour across the two conditions (objective 1), and also path modelling for the analysis of the possible conditional processes regulating the relationship between vignette and helping behaviour (objective 2), with the inclusion of other intervening variables in wider

⁷² Not preregistered.

models;

- *secondary analyses* comprised exploration of the correlation patterns of moral self-regulation (primarily self-improvement) with the new measures collected in the first part of the questionnaire (regulatory mode) and in the post-task (social desirability, social comparison orientation, locus of control, narcissism, modesty/humility, and the three self-growth constructs) (objective 3).

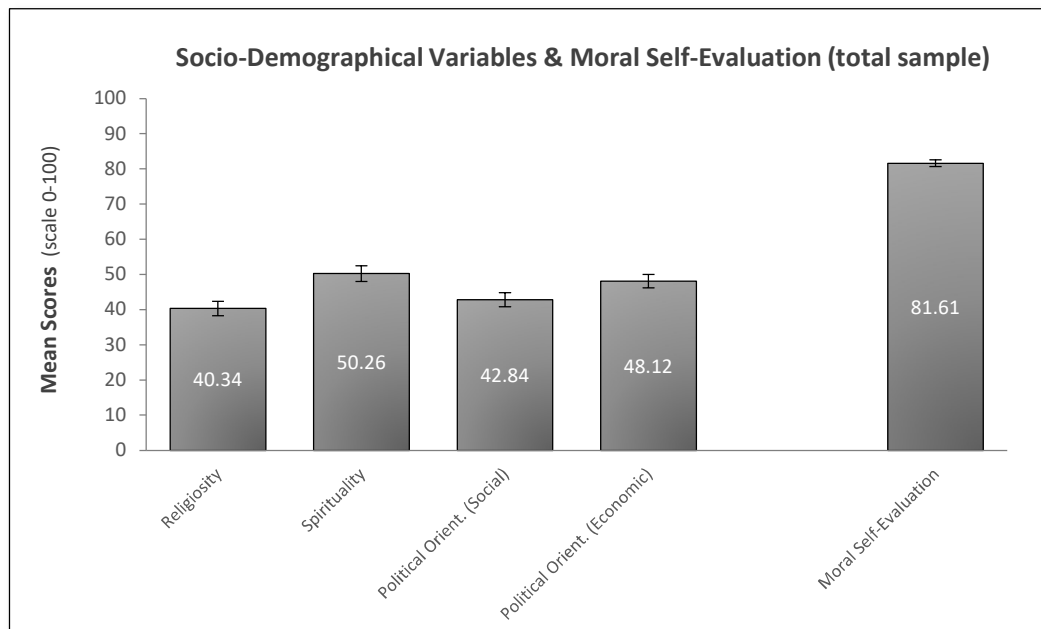
Results and Preliminary Reflections

Descriptive Statistics

The means of the socio-demographical variables, illustrated in Figure D3 along with the moral self-evaluation, were slightly more leaning toward the religious and conservative side of the spectrum than in the previous studies. There were no significant differences between the two conditions (see SMD5).

Figure D3:

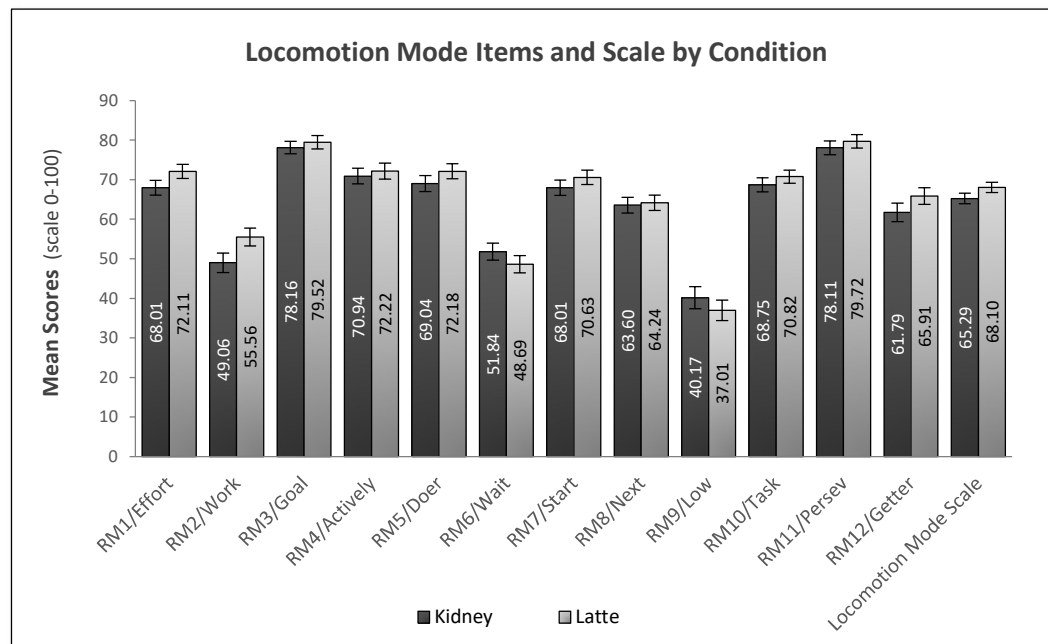
Mean scores for the socio-demographical variables and the moral self-evaluation (total sample), with bootstrap standard errors (BCa 95% CI)



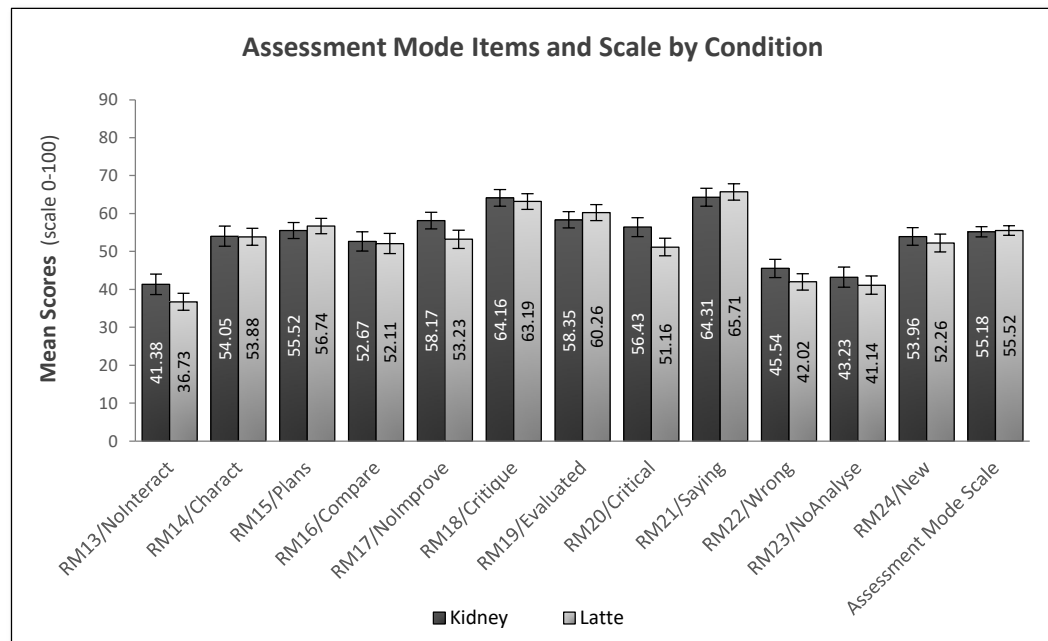
For what concerns regulatory mode, the means of the items and the scale composite means are illustrated in Figures D4-D5. The distributions of all the items as well as the scales showed non-normality, with no significant difference between conditions (see SMD6).

Figure D4:

Mean scores for locomotion mode by condition, with bootstrap standard errors (BCa 95% CI)

**Figure D5:**

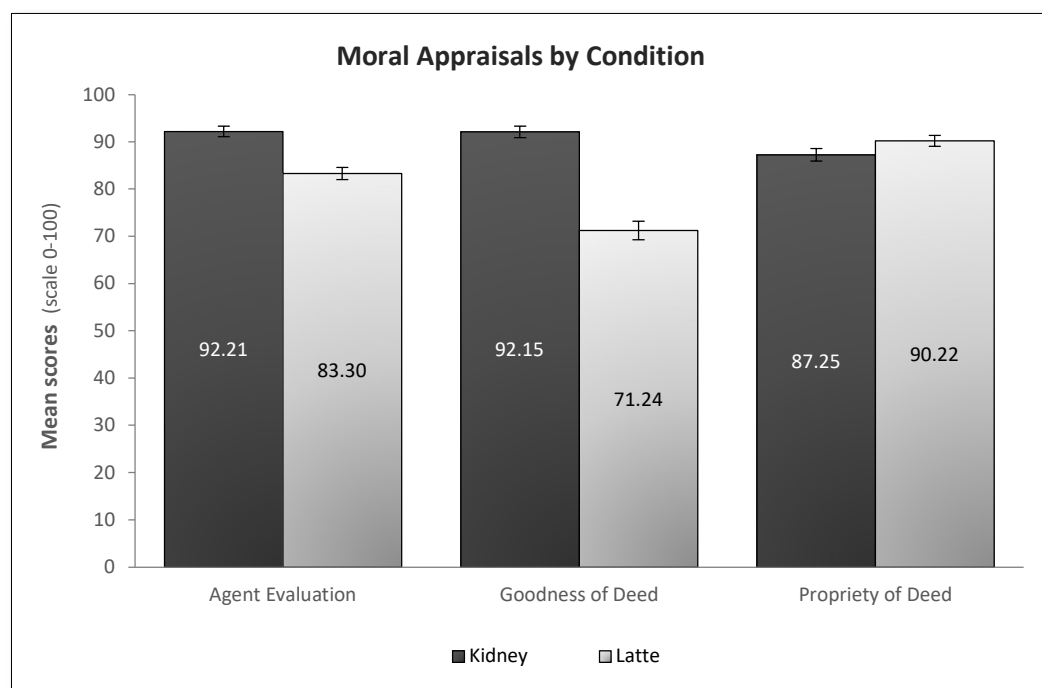
Mean scores for assessment mode by condition, with bootstrap standard errors (BCa 95% CI)



With regard to the moral appraisals (Fig. D6), one thing to notice is the discrepancy in the judgments between the two conditions, particularly in terms of goodness versus propriety of the deeds: comparatively, paying a latte for a friend was deemed to be more of an *obligatory* than a good action (very high score on propriety, i.e. the *right* thing to do), whereas donating a kidney was considered an outstanding deed in a more *supererogatory* sense (markedly lower score on propriety). The full output can be found in SMD7 (including significance testing).

Figure D6:

Mean scores for the moral appraisal variables by condition, with bootstrap standard errors (BCa 95% CI)

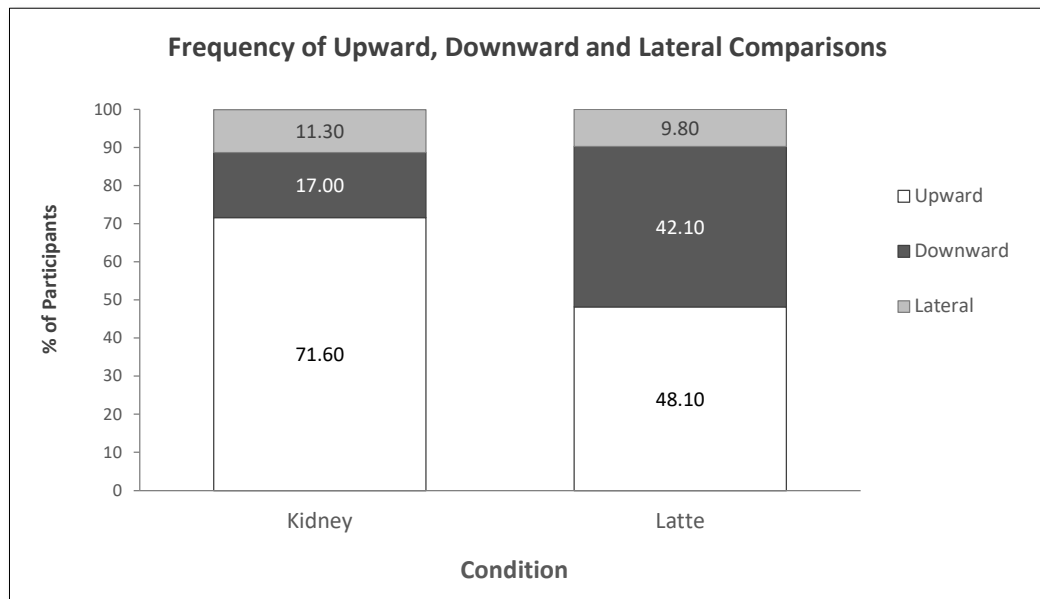


The mean moral discrepancy was positive for both conditions and significantly higher for Kidney: 11.27 ($SD = 18.09$) for Kidney and 0.98 ($SD = 18.20$) for Latte. Only 17.0% of participants engaged in downward comparison in the Kidney condition as

opposed to 42.1% in the Latte condition (Fig. D7; full analysis in SMD7, including significance testing).

Figure D7:

Frequency of moral comparison type by condition: clustered bar chart of frequency distributions (%)

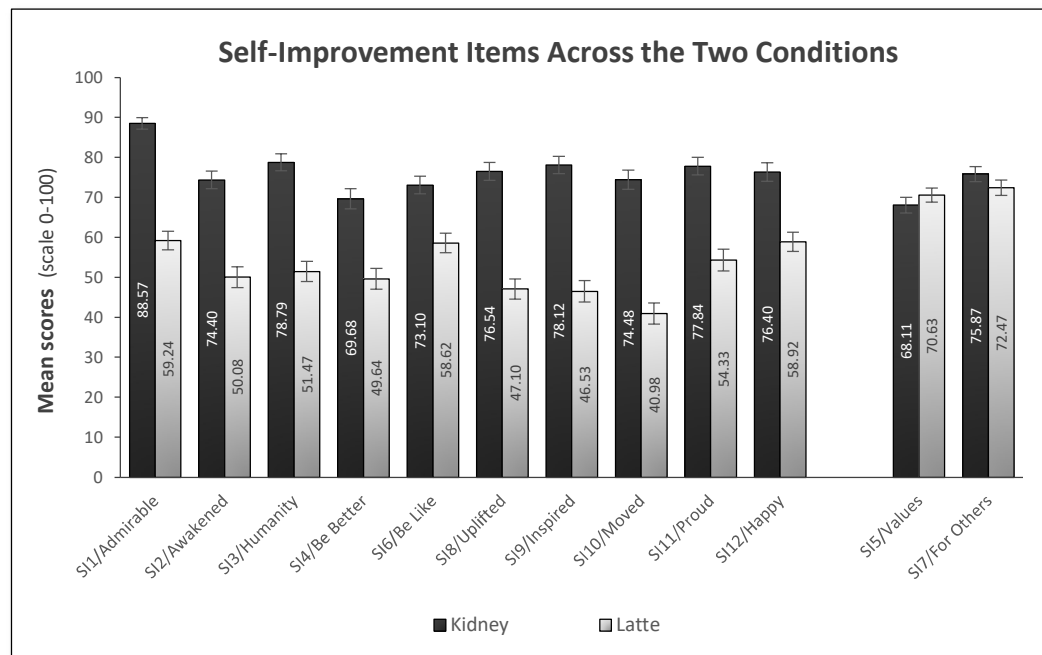


Moving on to the moral self-regulation measures, the composite means of the self-improvement scales were respectively 64.61 ($SD = 25.96$) for the total sample, 76.79 ($SD = 21.04$) for the Kidney condition, and 51.69 ($SD = 24.44$) for the Latte condition. Regarding the self-defence scale, the composite mean for the Kidney condition was 20.51 ($SD = 24.78$)⁷³. The means of the moral self-regulation items of the two scales are shown in Fig. D8-D9. As in the previous studies, all the distributions were not normal (see SMD8).

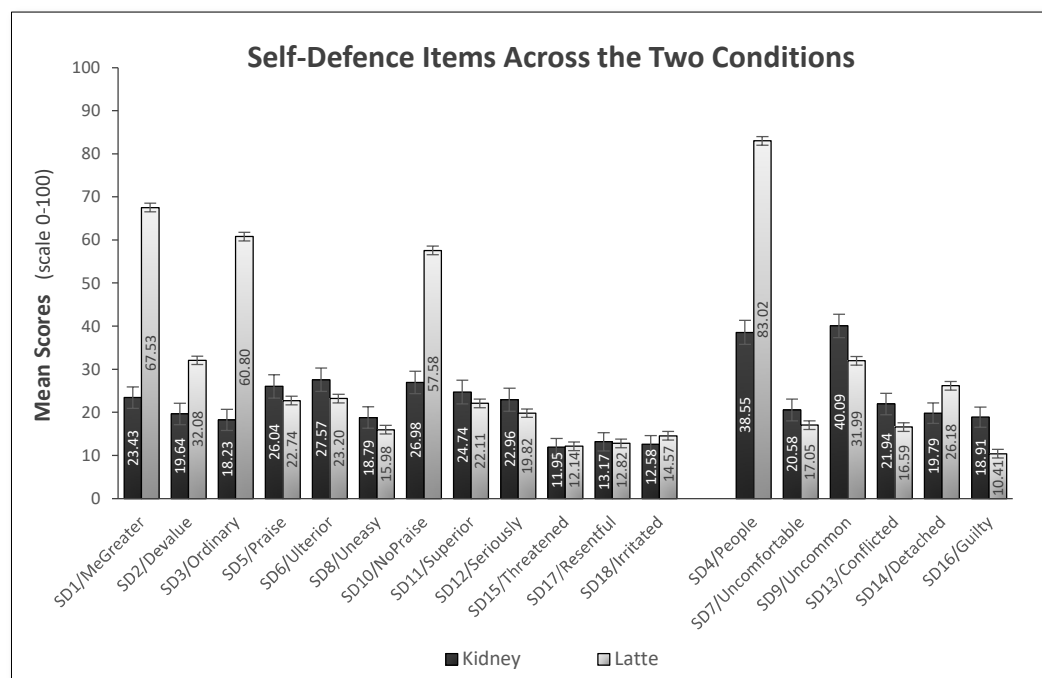
⁷³ The composite means of the self-defence scale for the Latte condition and the total sample are not reported for the reasons explained earlier in this chapter.

Figure D8:

Mean scores of the ten self-improvement items (plus two dropped items) for the two conditions, with bootstrap standard errors (BCa 95% CI)

**Figure D9:**

Mean scores of the twelve self-defence items (plus six dropped items) for the two conditions, with bootstrap standard errors (BCa 95% CI)



As hypothesised, items SD1, SD3, SD4, SD10 (and also SD2) were rated significantly higher in the control condition, offering support to the notion that the self-defence scale cannot be considered a valid measure for the Latte vignette (and consequently the total sample). Therefore, the planned exploratory analyses on self-defence were carried out only for the experimental condition.

Regarding the second part of the questionnaire, 181 participants (65.7%) chose not to help the experimenter and dropped out of the post-task without answering any questions; the remaining 93 participants (33.9%) who took part in the post-task answered an average of 18 questions ($SD = 36$): 22 questions in the Kidney condition ($SD = 40$) and 14 questions in the Latte condition ($SD = 31$). Eighteen participants (7%) answered all the 122 questions.

Normality tests for the dependent variable (Tab. D1) showed that the distributions were not normal (see also frequency distribution and density plots in Fig. D10-D11; full output in SMD9)⁷⁴.

Table D1:

Normality tests for the outcome variable helping behaviour (total sample and by vignette)

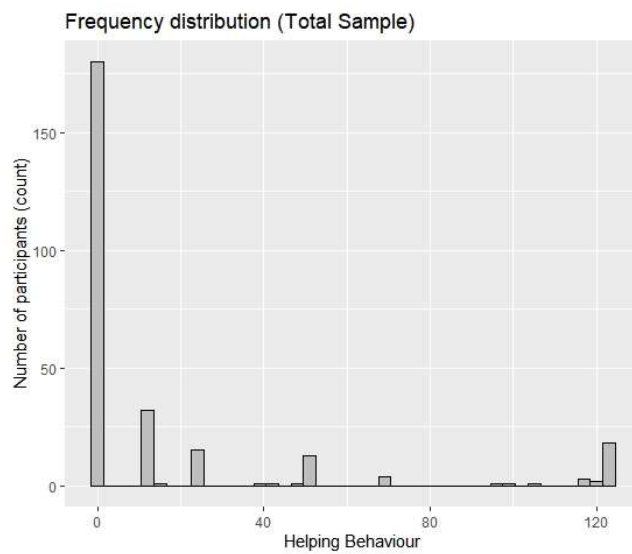
Tab. D1a: normality tests		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Total sample		.350	274	<.001	.554	274	<.001
Vignette	Kidney	.343	141	<.001	.602	141	<.001
	Latte	.356	133	<.001	.494	133	<.001

Note: a. Lilliefors significance correction

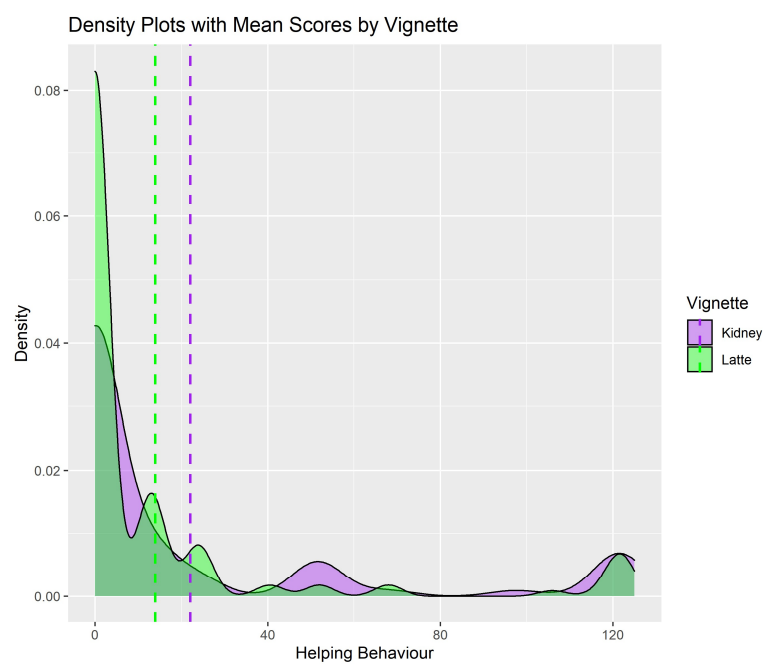
⁷⁴ Attempts to transform the data to remove the bias (logarithmic, square root, and reciprocal transformations) were unsuccessful (see SMD9).

Figure D10:

Frequency distribution of the outcome variable helping behaviour (total sample)

**Figure D11:**

Density plots with mean scores (by vignette) of the outcome variable helping behaviour



Because only 33.9% of participants helped the experimenter through participation in the post-task, the measures collected for exploratory purposes achieved a relatively

low sample size, especially those toward the end of the task. Consequently, some of the analyses reported later in this chapter are underpowered and should be taken with caution, while other planned analyses have not been reported at all. Sample sizes and key descriptive statistics for the post-task constructs are documented in Table D2.

Table D2:

*Sample size and key descriptive statistics for the post-task constructs (total sample).
Caveat: low sample size*

Construct	Sample size	Mean	Standard deviation	Skewness	Kurtosis
• Short Social Desirability Scale (SDS-S)	93	52.36	15.48	0.370	0.805
• Iowa-Netherlands Comparison Orientation Measure (INCOM):					
- INCOM Opinion-Based Comparisons	62	63.67	17.71	-0.560	0.435
- INCOM Ability-Based Comparisons	62	54.58	20.89	-0.517	-0.251
• Internal Locus of Control (ICI)	47	60.90	12.72	0.859	-0.176
• Grandiose Narcissism (NPI-16)	30	39.61	16.36	-0.364	-0.358
• Vulnerable Narcissism (HSNS)	26	56.47	20.64	0.199	-1.297
• Modesty (from HEXACO)	26	66.68	21.21	0.348	-1.515
• Intellectual Humility	26	61.24	16.08	0.511	-1.292
• Hedonia & Eudaimonia Motives for Activities (HEMA):					
- HEMA Hedonic Motives	24	71.52	18.79	-0.851	-0.188
- HEMA Eudaimonic Motives	24	73.99	18.37	-0.618	-0.129
• Growth Motivation Index (GMI)	23	71.59	14.78	-0.423	-0.925
• Desire and Commitment for Self-Improvement	23	74.06	15.44	-0.354	-0.062

Manipulation Check

A manipulation check was carried out to ensure the deeds in the two vignettes were associated with significantly different perceptions of goodness, which in turn were hypothesised to elicit significantly different levels of moral self-improvement.

To account for violations of normality of the dependent variables, one-tailed independent samples *t*-tests with 95% bootstrap confidence intervals were carried out:

- on average, participants in the Kidney condition judged the deed to be morally better ($M = 92.15$, $SD = 14.87$) than those in the Latte condition ($M = 71.24$, $SD = 23.39$); without assuming equal variances, this difference, 20.91, BCa 95% CI [16.29,

25.75], was statistically significant, $t(221.500) = 8.773$, $p = .001$, representing a large effect size of $d = 1.220^{75}$;

- on average, participants in the Kidney condition experienced a higher self-improvement state ($M = 76.79$, $SD = 21.04$) than those in the Latte condition ($M = 51.69$, $SD = 24.44$); without assuming equal variances, this difference, 25.10, BCa 95% CI [19.85, 30.59], was statistically significant, $t(260.851) = 9.085$, $p = .001$, representing a large effect size of $d = 1.10$.

Therefore, the manipulation was successful.

Test of H1 – Differential Prosocial Effects of the Manipulation

Following the preregistration plan, due to the bias in the distribution of helping behaviour, the differential effect in helping behaviour between conditions was assessed not only through a parametric test *with bootstrapping* (one-tailed independent samples t -test), but also a non-parametric test (one-tailed independent samples Mann-Whitney U -test):

- on average, participants in the Kidney condition offered the experimenter more help by answering more questions ($M = 22.06$, $SD = 39.52$) than those in the Latte condition ($M = 13.90$, $SD = 31.33$); without assuming equal variances, a t -test revealed that this difference, 8.15, BCa 95% CI [0.75, 16.22], was statistically significant, $t(260.851) = 9.085$, $p = .027$, representing a small effect size of $d = 0.229$;
- however, a Mann-Whitney U -test showed that there was *no* significant difference in the degree of participants' helping behaviour in the Kidney condition ($Mdn = 0$,

⁷⁵ According to Cohen's (1988) classic guidelines, the size of an effect d can be assessed thanks to the three cutoff points of 0.2, 0.5, and 0.8, respectively demarcating small, medium, and large effects.

mean rank = 71) and in the Latte condition ($Mdn = 0$, mean rank = 67), $U = 8642$, $p = .093$.

As it appears, while the t -test would reject the null hypothesis, the U -test failed to reject it. In the presence of conflicting results, accepting the outcome of a non-parametric tests usually represents a more conservative and reliable strategy, because non-parametric tests do not rely on any specific distributional assumptions and are less subject to bias with lower sample sizes and outliers (Leech & Onwuegbuzie 2002). Based on this strategy, the null hypothesis of a non-significant difference in helping behaviour between the two conditions should *not* be rejected. However, to assess the relative strength of evidence for the null versus the alternative hypothesis, a Bayesian one-tailed Mann-Whitney U -test was conducted using Jamovi, with computation of Bayes factors. A Bayes factor is a ratio that contrasts the likelihood of the data under the alternative hypothesis and the null hypothesis (BF_{+0}) (or vice versa: BF_{0+}), enabling to assess how much more likely the data are to occur if the alternative hypothesis is true, compared to if the null hypothesis is true (or vice versa) (Jarosz & Wiley, 2014).

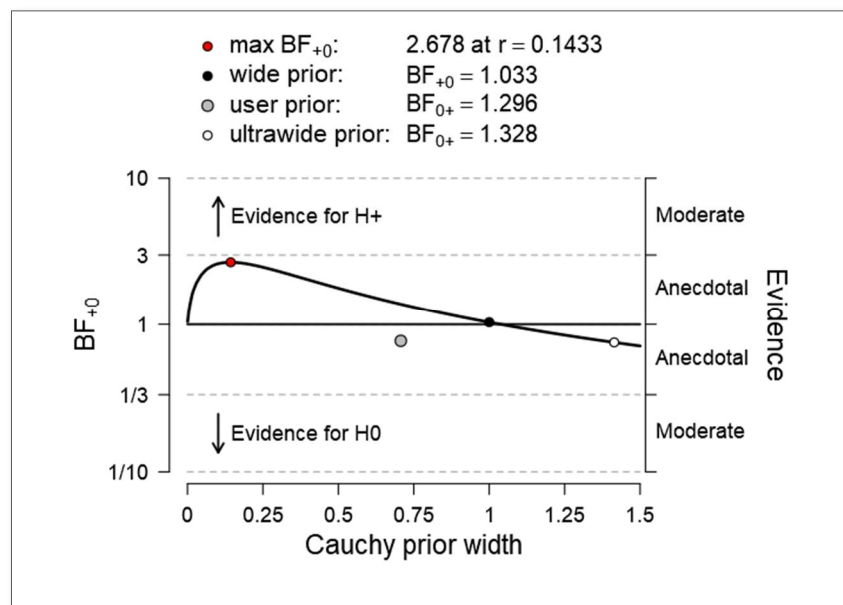
Following Jeffreys's (1961) recommendation, given the paucity of reliable information available for the analysis, a safe prior to adopt was the Cauchy distribution (using 0.707 times the standard deviation). The Bayes factors of $BF_{+0} = 0.772 / BF_{0+} = 1.296$ *weakly* supported the null hypothesis (Tab. D3); however, based on the guidelines and thresholds for the interpretation of Bayes factors recommended by Jarosz and Wiley (2014) and van Doorn et al. (2020), the evidence remains anecdotal and inconclusive (confirmed by a robustness check using a range of priors; see Fig. D12).

Table D3:*One-tailed Bayesian Mann-Whitney U-Test with Bayes factors*

	BF_{+0}	BF_{0+}	W
Helping Behaviour	0.772	1.296	10111

Note: the alternative hypothesis specifies that condition Kidney is greater than Latte.

Result based on data augmentation algorithm with 5 chains of 1000 iterations.

Figure D12:*One-tailed Bayesian Mann-Whitney U-test: robustness check*

Relationship and Mechanism: Mediation Tests

Hypotheses H2, H3, and H4 identified three potential mechanisms for the transmission of a significant effect of moral self-improvement to helping behaviour: a direct effect (H2), a mediated effect through the desire to do something good for others (H3), and a moderated effect with an interaction between self-improvement and the desire to do something good for others (H4). These hypotheses were tested among the total sample, fitting path models with observed variables with the MLR estimator across all of them (for consistency reasons) using the R package lavaan, ver. 0.6-7. In this case,

the above-mentioned approach was preferred to bootstrapping mediation analysis based on regression for two main reasons: a) path models (and SEM in general) incorporate and test causal assumptions, whereas regression methods do not (see Bollen & Pearl, 2013); b) path models utilise the maximum likelihood estimator instead of ordinary least squares, thus prioritising accuracy of the coefficient estimates (rather than of the predicted values), which is recommended for theory testing (see Kline, 2016). Additionally, model fitness measures (unavailable with regression) can be obtained.

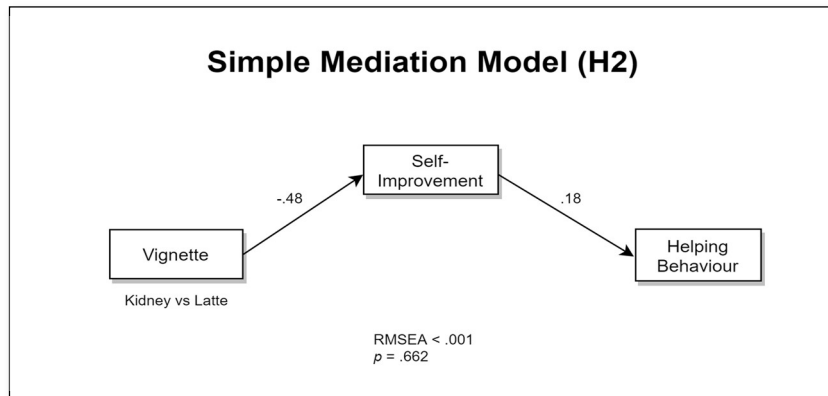
Test of H2 – Simple Mediation Model

To test hypothesis 2, a simple mediation model was fitted, with vignette as predictor, moral self-improvement as mediator, and helping behaviour as outcome. The revised path diagram⁷⁶ with standardised regression weights is illustrated in Figure D13 and the main fit indices are reported in Table D4 (full output in SMD10).

⁷⁶ The models portrayed in Figures D13-D15 slightly differ from the hypothesised models in Figures D2a-c as they only include significant pathway coefficients (essentially, the non-significant direct relationship between vignette and helping behaviour was omitted, and revised models were fitted and reported with adjusted regression weights and fit indices).

Figure D13:

Results from path analysis with observed variables, test of H2 – simple mediation. Revised model with only significant pathways (standardised regression weights). Estimator: robust maximum likelihood (MLR)

**Table D4:**

Main fit indices of the path model to test simple mediation among the total sample

	Tot. Sample
Chi-Square (scaled)	.312
df (scaled)	1
p -value (scaled)	.577
RMSEA (scaled)	<.001
p -value (scaled)	.662
CFI (scaled)	1.000
TLI (scaled)	1.023
GFI	.999
AGFI	.994
AIC	5226.767
BIC	5241.220

The relationship between moral self-improvement and helping behaviour was positive and significant, with a regression weight of .18 ($p < .001$), supporting H2. When analysed by group (i.e., by vignette), the regression weights for both vignettes are positive and significant: .16 ($p < .001$) for the Kidney condition, and .13 ($p = .048$) for the Latte condition. The mediation effect was -.09 ($p < .001$) (full output in SMD10).

Test of H3 – Serial Mediation Model

To test hypothesis 3, a serial mediation model was fitted, with vignette as predictor, moral self-improvement and desire to do something good (item S17) as sequential mediators, and helping behaviour as outcome. The revised path diagram with significant standardised regression weights is illustrated in Figure D14 and the main fit indices reported in Table D5 (full output in SMD10).

Figure D14:

Results from path analysis with observed variables, test of H3 – serial mediation. Revised model with only significant pathways (standardised regression weights). Estimator: robust maximum likelihood (MLR)

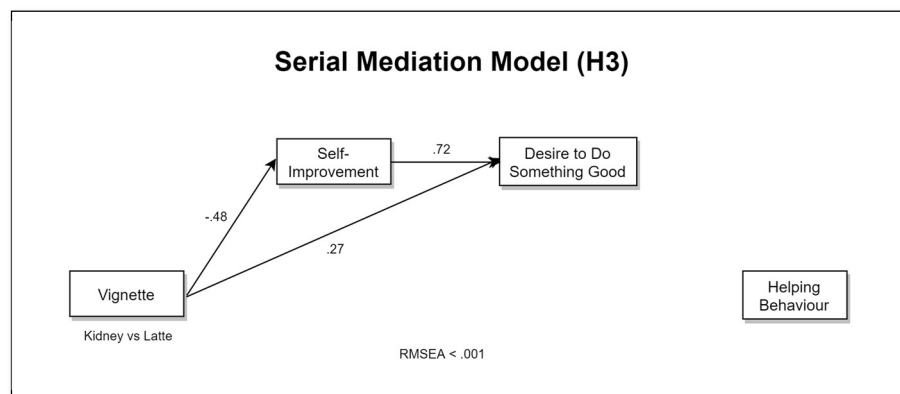


Table D5:

Main fit indices of the path model to test serial mediation among the total sample

	Tot. Sample
Chi-Square (scaled)	.312
df (scaled)	1
p-value (scaled)	.577
RMSEA (scaled)	<.001
p-value (scaled)	NA
CFI (scaled)	1.000
TLI (scaled)	1.000
GFI	1.000
AGFI	1.000
AIC	7515.693
BIC	7548.145

There was no significant relationship between the desire to do something good and helping behaviour among the total sample (and by vignette), and therefore the action tendency item SI7 did *not* function as a mediator between moral self-improvement and helping behaviour (H3 was not supported by the data, full output in SMD10). This finding could imply that the composition of the moral SI scale, even without item SI7, retained other action tendency items capable per se to promote prosocial behaviour, although future research will have to further investigate this effect.

Test of H4 – Moderated Mediation Model

To test hypothesis 4, a moderated mediation model was fitted, with vignette as predictor, moral self-improvement as mediator, desire to do something good (item SI7) as moderator, and helping behaviour as outcome. The revised path diagram with significant standardised regression weights is illustrated in Figure D15 and the main fit indices, all poor, are reported in Table D6 (full output in SMD10).

Figure D15:

Results from path analysis with observed variables, test of H4 – moderated mediation. Revised model with only significant pathways (standardised regression weights). Estimator: robust maximum likelihood (MLR)

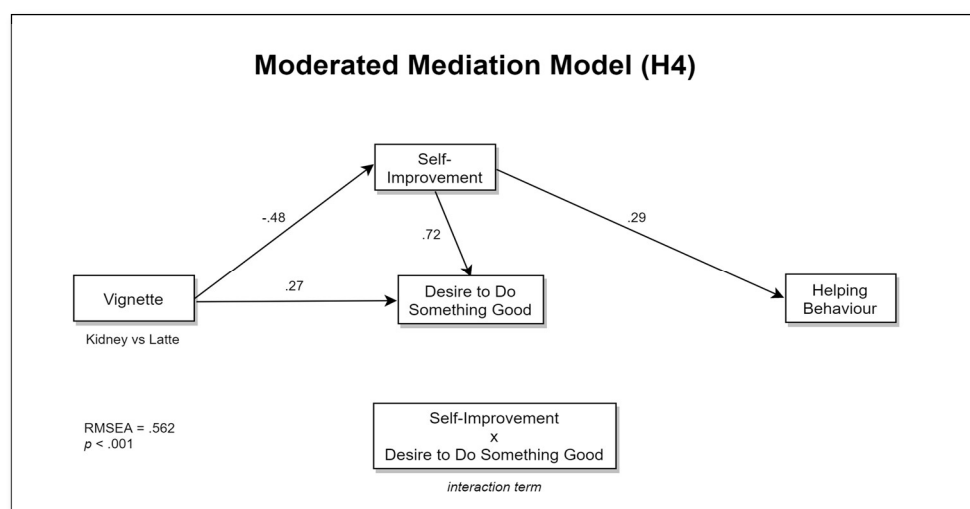


Table D6:*Main fit indices of the path model to test moderated mediation among the total sample*

	Tot. Sample
Chi-Square (scaled)	262.812
df (scaled)	3
<i>p</i> -value (scaled)	<.001
RMSEA (scaled)	.562
<i>p</i> -value (scaled)	<.001
CFI (scaled)	.582
TLI (scaled)	-.255
GFI	.628
AGFI	-.859
AIC	7566.837
BIC	7599.355

There was no significant interaction between moral self-improvement and the desire to do something good among the total sample ($p = .214$), and the action tendency item SI7 did not function as a moderator between moral self-improvement and helping behaviour. Thus, hypothesis 4 was not supported by the data: no evidence was found that the level of helping behaviour induced by a moral deed had different intensity depending on the desire to do something good for others (full output in SMD10).

Regulatory Mode: Correlation Patterns

Following the preregistered analysis plan, Pearson's zero-order correlation coefficients were computed and tested one-tailed⁷⁷ for locomotion and assessment mode among the total sample. Table D7 documents correlations between the two regulatory modes and between each of them and moral appraisal and self-regulation variables.

⁷⁷ One-tailed tests were chosen due to the directional nature of the related predictions.

Table D7:

Zero-order correlations between the regulatory mode and moral appraisals (total sample)

			Locomotion	Assessment	Goodness of Deed	Propriety of Deed	Agent Evaluation	Moral Discrep.
Locomotion Mode	Pearson Correlation		1	.001	.129*	.246***	.228***	-.116*
	Sig. (1-tailed)			.494	.016	<.001	<.001	.028
	N		274	274	274	274	274	274
	Bootstrap ^c	Bias	0	-.001	-.001	<.001	.001	.001
		Std. Error	0	.067	.052	.055	.054	.061
		BCa Lower	.	-.123	.032	.124	.125	-.224
		95% C.I. Upper	.	.128	.230	.357	.335	.013
Assessment Mode	Pearson Correlation		.001	1	.044	.031	.012	.036
	Sig. (1-tailed)		.494		.232	.305	.421	.275
	N		274	274	274	274	274	274
	Bootstrap ^c	Bias	-.001	0	-.002	.001	-.002	<.001
		Std. Error	.067	0	.058	.052	.054	.053
		BCa Lower	-.123	.	-.062	-.066	-.088	-.076
		95% C.I. Upper	.128	.	.150	.134	.116	.138

Note: * $p < .05$ (1-tailed) *** $p < .001$ (1-tailed).
c. Bootstrap results are based on 1000 bootstrap samples.

The first finding concerns the non-significance of the correlation between locomotion and assessment among the total sample. This finding is consistent with the literature (Kruglanski et al., 2000). The second finding concerns the non-significance of the correlations between assessment mode and all the main moral appraisal variables; by contrast, locomotion mode was positively correlated with the agent evaluation, as well as with the goodness and propriety of the deed, and negatively correlated with moral discrepancy.

The correlations between regulatory mode and moral self-regulation were assessed among participants in the experimental condition (Kidney) to obtain a valid measure of association for self-defence. As expected, results exhibited positive correlations between locomotion mode and self-improvement, and between assessment mode and self-defence (Tab. D8). No significant correlations were found between locomotion and self-defence, and between assessment and self-improvement.

Table D8:

Zero-order correlations between regulatory mode and moral self-regulation (Kidney condition)

				Locomotion	Assessment	Self-improvement	Self-defence
Locomotion Mode	Pearson Correlation			1	.058	.434***	.015
	Sig. (1-tailed)				.248	<.001	.430
	N			141	141	141	141
	Bootstrap ^c	Bias		0	-.002	-.001	-.003
		Std. Error		0	.093	.080	.068
		BCa 95% C.I.	Lower	.	-.127	.259	-.123
			Upper	.	.229	.586	.144
Assessment Mode	Pearson Correlation			.058	1	.034	.234**
	Sig. (1-tailed)			.248		.344	.003
	N			141	141	141	141
	Bootstrap ^c	Bias		-.002	0	.001	-.002
		Std. Error		.093	0	.077	.060
		BCa 95% C.I.	Lower	-.127	.	-.115	.113
			Upper	.229	.	.180	.343

Note: ** $p < .01$ (1-tailed) *** $p < .001$ (1-tailed).
c. Bootstrap results are based on 1000 bootstrap samples.

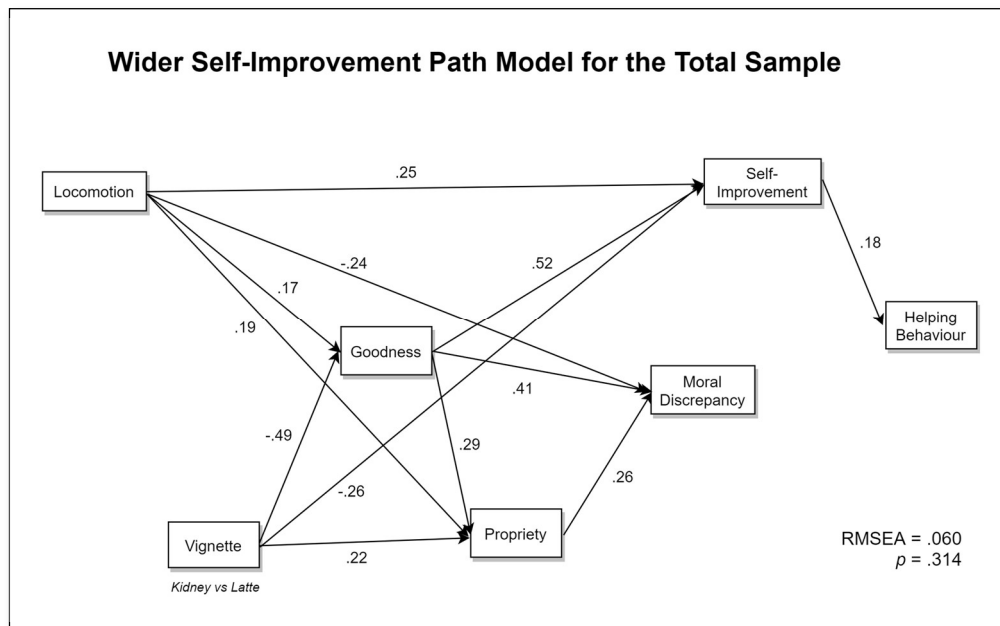
A Wider Behavioural Model of Moral Self-Improvement

Building on all these findings, a wider path model of moral self-improvement was fitted among the total sample, inclusive of helping behaviour as outcome variable and motivational and moral appraisal variables as predictors and mediators.

The fitted model is depicted in Figure D16 and its main fit indices are documented in Table D9 (full output in SMD11).

Figure D16:

*Wider self-improvement path model with observed variables for the total sample.
Estimator: robust maximum likelihood (MLR). Standardised regression weights*

**Table D9:**

Main fit indices of the wider self-improvement path model for the total sample

	Tot. Sample
Chi-Square (scaled)	13.802
df (scaled)	7
p-value (scaled)	.055
RMSEA (scaled)	.060
p-value (scaled)	.314
CFI (scaled)	.983
TLI (scaled)	.951
GFI	.999
AGFI	.996
AIC	12022.228
BIC	12105.330

The model showed good fit across all the main indices. Moral self-improvement was the only conduit to helping behaviour, which was not predicted by any other variable in the model. In turn, self-improvement was predicted by vignette (negatively),

goodness of the deed and locomotion mode (positively), but not by moral discrepancy, presumably due to the numerically low incidence of downward comparisons, particularly in the Francia Kidney subsample (as in Study 2). Moral discrepancy was predicted positively by the goodness and propriety of the deed, but also negatively by locomotion, consistent with the correlation coefficient examined earlier. Locomotion was also positively associated with goodness and propriety of the deed, thus exhibiting a similar pattern to that observed for approach in Study 3. The variable vignette negatively predicted the goodness of the deed, but contrary to Study 3, positively predicted the propriety of the deed. This phenomenon can be better understood by examining the results from structural invariance analysis by group (by vignette). An unconstrained model, whereby all parameters were freely estimated for each condition, was compared to a fully constrained model, whereby all parameters (regression weights and intercepts) were constrained to be equal between conditions. The result of the structural invariance test showed that the two conditions, at a model-wide level, were significantly different (see Tab. D10; full output in SMD11). What this finding implies is that the condition (vignette) behaved as a model-wide moderator of the relationship between the predictors and the behavioural outcome (helping behaviour) (see e.g., Sarstedt et al., 2011).

Table D10:
Structural invariance test by condition (total sample)

Model	df	AIC	BIC	χ^2	χ^2 diff.	df diff.	Sig.
Unconstrained	10	11948	12092	23.737			
Fully constrained	24	12094	12188	198.603	167.84	14	<.001***

Note: *** $p < .001$ (2-tailed)

Having ascertained the lack of structural invariance, two distinct models were fitted, one for each condition. Both achieved satisfactory fit (Tab. D11).

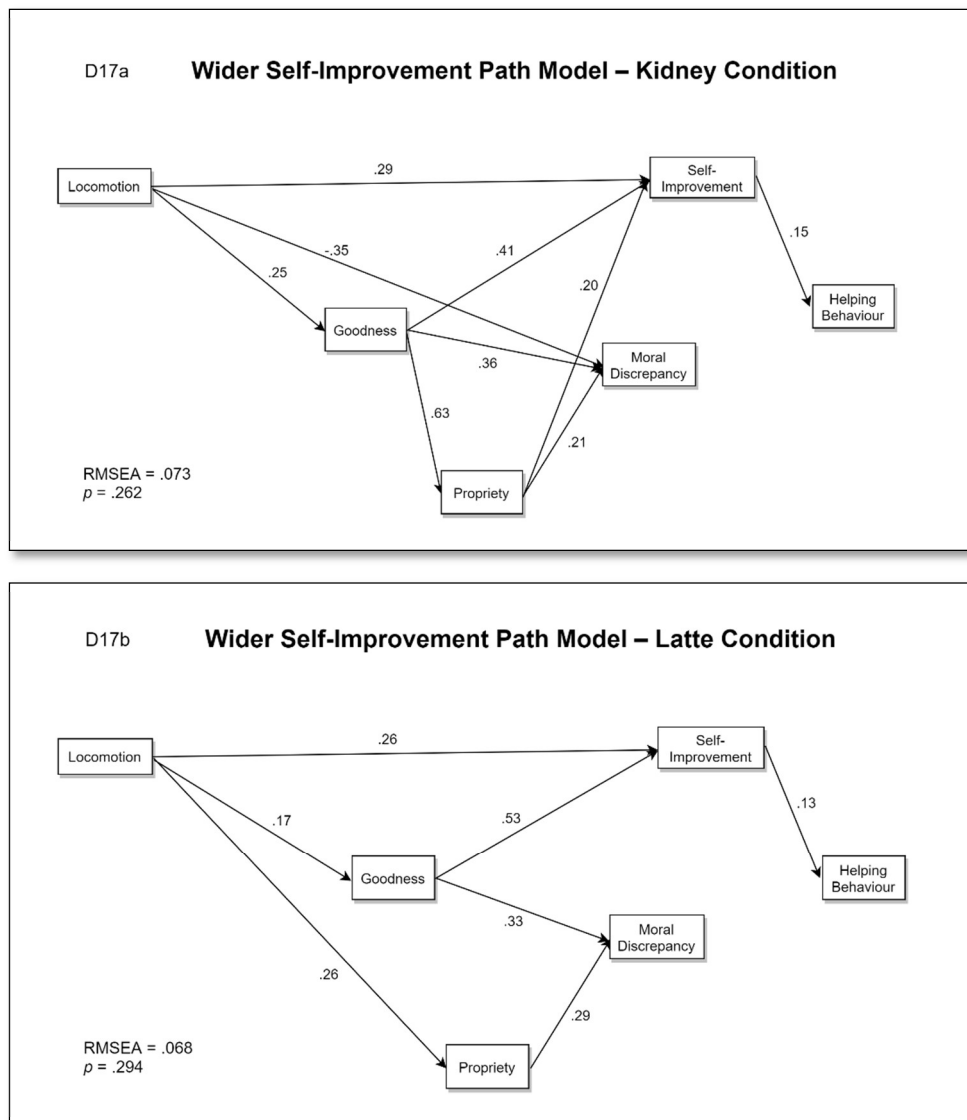
Table D11:
Main fit indices of the wider path models for the two vignettes

	Kidney	Latte
Chi-Square (scaled)	8.761	11.248
df (scaled)	5	7
<i>p</i> -value (scaled)	.119	.128
RMSEA (scaled)	.073	.068
<i>p</i> -value (scaled)	.262	.294
CFI (scaled)	.981	.963
TLI (scaled)	.942	.920
GFI	.999	.999
AGFI	.994	.995
AIC	6067.135	5873.418
BIC	6126.110	5925.444

The fitted models with their significant pathways are illustrated in Figure D17a-b (full output in SMD11).

Figures D17a-b:

Wider self-improvement path models with observed variables for the two conditions. Estimator: robust maximum likelihood (MLR). Standardised regression weights



The fundamental driver of the lack of structural invariance between the two models was represented by the relationship between the goodness and the propriety of the deed. The non-significance of the pathway in the control condition reflects the construal of the deed (paying a latte for a friend who forgot her wallet) as a morally *due* action (obligatory) and not necessarily a remarkable one. By contrast, the perception of the

deed in the experimental condition (donating a kidney) underlies its construal as an outstanding action well *beyond* the call of duty (supererogatory), as also confirmed by the analysis of the means (Fig. D6).

What was invariant between the two vignettes, though, was the association between self-improvement and helping behaviour (regression weights of .15 vs .13, respectively in the Kidney and Latte conditions). Thus, this finding confirms that, although donating a kidney was viewed as a significantly better moral deed that evoked a significantly higher level of self-improvement, there was no conclusive evidence that the two actions inspired significantly different degrees of helping behaviour.

Exploratory Analyses (Post-Task Constructs)

Due to the low sample sizes, the measures collected in the post-task allowed merely partial and preliminary understanding. Only correlation analysis for the constructs of social desirability and social comparison orientation are reported here (further information in SMD12), as the samples are comparatively higher than for the other constructs (see Tab. D2). All the variable associations were tested one-tailed⁷⁸ and expressed through zero-order Pearson's correlation coefficients with bootstrap confidence intervals. More complex modelling was not possible.

Social Desirability

Socially desirable responding was anticipated to be associated positively with self-improvement and negatively with self-defence. The results of the analysis confirmed this expectation: the SDS-S correlated positively with self-improvement ($r = .243$, $p = .009$;

⁷⁸ One-tailed tests were chosen due to the directional nature of the related predictions.

among total sample) and negatively with self-defence ($r = -.417, p = .001$; among Kidney subsample⁷⁹).

Social Comparison Orientation

Social comparison orientation was expected to be positively associated with defensive regulation, particularly driven by ability-based comparisons. The results of the analysis showed that ability-based comparison tendencies did correlate positively with self-defence ($r = .421, p = .005$; among Kidney subsample), but also with self-improvement, albeit to a lesser degree ($r = .343, p = .003$; among total sample). Opinion-based comparison tendencies correlated positively with self-improvement ($r = .323, p = .006$; among total sample), but not with self-defence ($r = .082, p = .314$; among Kidney subsample).

The expected positive correlation between social comparison orientation and assessment mode was also found among the total sample, especially strong for ability-based comparison tendencies ($r = .509, p < .001$). Locomotion did not correlate with either ability- or opinion-based comparison orientations (respectively, $r = -.068, p = .301$, and $r = .125, p = .169$).

⁷⁹ The correlation with self-defence was computed among the Kidney subsample to ensure validity of the self-defence measure.

Discussion

Study 4 was designed as an integration to the structural models obtained from Study 3. Its main purpose was to add a behavioural component to the models, investigating the prosocial effects of moral self-improvement, thus supplementing a measure of concurrent validity. To achieve this goal, the study assessed the extent an outstanding moral deed elicited a significantly different degree of helping behaviour from an ordinary act of kindness (objective 1). In addition, it measured alternative potential mechanisms (and the magnitude) of the relationship between moral self-improvement and helping behaviour under the described experimental conditions (objective 2).

The design was inspired by two experiments published by Schnall et al. (2010), modified to achieve two further goals: first, by using in the control condition a *moral* stimulus (a mild act of everyday kindness) instead of a non-moral one (a nature documentary in Schnall's experiment 1 or a funny story in experiment 2), the *full* moral self-improvement model, inclusive of the moral judgments and self-regulatory variables, could be re-tested and integrated with a behavioural outcome⁸⁰. Second, by adapting the post-task and using relevant personality scales instead of a boring mathematical exercise (Schnall's experiment 2), extra-measures could be obtained for exploratory purposes (objective 3).

Impact of the COVID-19 Pandemic

Data collection took place under unusual conditions related to the COVID-19 pandemic. The initial study in the laboratory in the UK had to be interrupted due to the lockdown and adapted to become suitable to an online environment. This solution did

⁸⁰ Measuring moral outcomes following the presentation of a non-moral stimulus would have been less relevant.

not prevent further issues from emerging soon after: about half of the participants recruited from the CloudResearch/MTurk platform did not follow the questionnaire guidelines and had to be removed from the sample. Given the exceptional circumstances, a top-up of participants was implemented so that the intended sample size was approximately achieved.

As a general consideration, it is fair to say that Study 4 was impacted by the COVID-19 pandemic just as much as many other online studies conducted between the end of winter and the beginning of summer 2020. The impact could be due to structural and behavioural changes occurred during that time span among popular crowdsourcing platforms. A recent analysis (Arechar & Rand, 2020) has provided evidence of a shift in the composition of MTurk samples during the lockdown, with new workers characterised by different socio-demographic profiles entering the platform (younger, more male, non-white, and conservative). At the same time, a behaviour change was noticed, as the new workers were found to be up to twenty percentage points more likely to fail attention checks, respond carelessly and inconsistently. This general trend was reflected in the issues experienced with Study 4, where nearly half of participants in the first part of online data collection failed the attention checks; additionally, the final sample was indeed more male, conservative, and religious/spiritual than, for example, in Study 3.

Despite these issues, it should be recognised that, given the existing constraints, the actions undertaken to minimise disruption, convert the study to an online platform, and achieve satisfactory power represented effective countermeasures.

Significance of the Main Findings

Thanks to those corrective actions, the results of Study 4 seem for the most part to be valid and reliable. *Preliminary* evidence was found that both an excellent and a lesser

moral action were associated with helping behaviour in the sample. At the same time, the hypothesis that the two conditions were significantly different in terms of helping behaviour was not upheld by the data: the mild act of ordinary kindness in the control condition, although judged as significantly less remarkable than the truly outstanding moral deed in the experimental condition, ended up generating a similar degree of help even in the presence of significantly different levels of moral self-improvement. It is possible that the level of self-improvement in the experimental and control conditions was high enough to elicit a comparable amount of help, or that an exogenous variable, left outside of the remit of Study 4, was the common cause across the two conditions. Either way, a clear causal link between moral self-improvement and helping behaviour could not be established and should be further investigated, because the evidence in favour of the null hypothesis was anecdotal and more research will be necessary to ascertain the true nature of the effect in the population (if any).

This result may have been affected by the design and some residual demand characteristics. Despite the cover story, participants knew that they were under experimental conditions. The presentation of the vignettes *in isolation* within each condition (between-subjects design) might have produced a slightly inflated degree of prosociality in the control group. In future research, it could be considered to show participants across conditions an initial vignette to warm them up and frame their response; reactions to the following experimental stimuli would then be anchored to a common set of baseline measures, which could be used to calibrate the ensuing variables. This kind of pre-task, common to the two conditions, could enhance the accuracy of the comparison between them, but would affect the measurement of the “absolute” effect of the experimental stimulus on the outcome variable.

Future research could also consider designs with alternative control conditions. To retain the same framework adopted in the present research, a comparison of the experimental condition depicting a remarkable moral action with a control condition showing another action seems necessary⁸¹. In addition, to justify a questionnaire centred on morality (moral appraisal questions, moral self-regulation inventory) the control deed should be, as in Study 4, as small as possible, but still morally relevant⁸²; it should preferably be normative⁸³, and perhaps in a domain leaving others in the background and implicating the self more directly (e.g., purity).

The kind of prosocial behaviour measured in Study 4 was helping behaviour. It was an *exogenous* measure of *actual* behaviour: exogenous because helping the experimenter by filling out personality scales did not bear any resemblance to the moral deed that elicited help, and actual behaviour because it was real overt behaviour, not just dispositions or intentions to help. Indeed, the desire to do something good for others (a measure of action tendencies dropped from the moral self-improvement inventory in Study 3) neither was conducive to helping behaviour nor interacted with self-improvement in the prediction of helping behaviour; thus, it did not function as either a mediator or a moderator, despite its strong correlation with self-improvement. The self-improvement construct, thanks to the indicators that shape it, *directly* translated into helping behaviour, without any conditional processes related to helping

⁸¹ Designing a control condition with the behavioural task immediately following the introductory sociodemographic questions (i.e., with no moral stimulus and no appraisal and regulatory questions) would likely lead to high demand characteristics, endangering the validity of the measurement: participants' knowledge of being in an experimental setting could make them behave (if only non-consciously) in a way that is consistent with what they might assume is the experimenter's intent, resulting in a paradoxical higher degree of help in the control than in the experimental condition.

⁸² A control condition depicting a non-moral action would likely cause a high rate of participant confusion and drop-out, due to the irrelevance of the appraisal and self-regulatory questions specifically centred on morality.

⁸³ A normative action is less likely to elicit reactance based on unfavourable judgments.

desire (limited to the measures collected in Study 4). This finding addressed the question of what sort of mechanism affects the relationship between self-improvement and helping behaviour⁸⁴.

When considering the wider nomological network of moral self-regulation and its prosocial effects, Study 4 offered at least two novel insights. The first concerned the results from the analysis of regulatory mode, another important predictor from the motivation triad that also includes hedonic orientation and regulatory focus. As expected, locomotion mode was found to correlate positively with self-improvement, and assessment mode to correlate positively with self-defence. In the self-improvement portion of the path model, locomotion worked similarly to promotion focus and approach temperament, positively correlating with self-improvement and downward comparison.

The second insight highlighted how judgments of propriety and goodness of moral deeds were underpinned by their fundamental construal in terms of obligation and supererogation. This learning was already partly brought to life by the results of Studies 2-3 through the comparison of Nicholas's and Francia's deeds, but became more evident in Study 4 thanks to the sharper contrast between a mild act of everyday kindness (control condition) and an outstanding moral action (experimental condition): paying a latte for a friend who forgot her wallet was "the right thing to do" (obligation) to a greater extent than donating a kidney, which in turn was a greater moral deed, far beyond the call of duty (supererogation). Beside this analysis of the means of the moral judgments, consideration of the regression weights (and their sign) lent further credit to this interpretation: among the total sample, the variable vignette negatively predicted

⁸⁴ Leaving aside considerations about the inconclusive results of the comparison between conditions (H1).

the goodness of the act (i.e., the deed in the Kidney condition was morally better), but positively predicted the propriety of the act (i.e., the deed in the Latte condition was more morally due). The analysis of the path model by vignette made this finding even more striking: whereas in the Latte condition there was no significant correlation between goodness and propriety (a considerable number of participants judged the deed obligatory and not necessarily particularly good), a very strong one was found in the Kidney condition (the deed was remarkable just as much as it was the right thing to do). Critically, whereas propriety alone was unable to induce self-improvement processes in the Latte condition, the coincidence of judgments of high goodness and high propriety did elicit high self-improvement states in the Kidney condition. Further, at total sample level, the vignette predicted the goodness/propriety of the deed and (directly and indirectly) self-improvement, functioning as a model-wide moderator of the relationship between all the model predictors and helping behaviour⁸⁵.

Although future research will have to solidify our understanding of the causal links between moral self-improvement and prosocial behaviour, the findings overall provide at least correlational evidence of the social relevance of the SR of virtue. The integration of a behavioural component in the study of the self-regulation of virtue remains one of the crucial features of future endeavours in the field.

Limitations

The conditions under which Study 4 was carried out were more challenging compared to the previous studies due to the impact of the COVID-19 pandemic. In the face of all the difficulties, a lot of care was devoted to putting in place measures that could effectively address the problems that arose. The data collected in the laboratory

⁸⁵ The absence of structural invariance between vignettes was essentially driven by the different relationships between goodness and propriety of the deed across the two conditions.

had to be discarded, so the study was converted to an online methodology. The initial online sample was affected by unusual non-compliant participant behaviour, so a top-up of participants was supplemented to complete the sampling plan and achieve approximately the intended power. The effect size of the relationship between moral self-improvement and helping behaviour was considerably smaller than the expectation based on a conservative estimate from previous research; consequently, the exploratory analysis of the measures collected in the post-task was underpowered and must be taken with caution. Where the sample was large enough, a few correlational patterns were reported because they could be reasonably utilised to inform predictions for future confirmatory research, but more complex modelling could not be undertaken. Where modelling was feasible, for the most important measures, it was conducted with path analysis using observed variables instead of full structural modelling with latent variables. This solution was adopted for consistency reasons, so that all the hypotheses were modelled with the same approach, including the moderated mediation model, for which notoriously interactions with latent variables with many indicators are extremely challenging with full SEM. This simplification was deemed reasonable in light of the fact that in the wider models examined in Study 4 only locomotion and self-improvement could have been treated as latent variables (recall that, by contrast, in Study 3 most of the variables were latent). Therefore, not modelling the error in path analysis should have had limited impact on the regression weight estimates and the fit indices. Lastly, another solution adopted for consistency reasons was the fitting of all the path models with the MLR estimator⁸⁶.

⁸⁶ The one and only categorical variable was vignette in the analyses among the total sample, and the variable was dummy-coded as in the previous studies.

General Discussion and Conclusive Remarks

In the ancient world, in both Western and Eastern societies, moral teaching revolved around the characterisation of virtue through noteworthy human beings of impeccable moral character (see e.g., Solomon et al., 2008). The works and the life of philosophers such as Socrates, Plato, and Aristotle in the West, Lao-Tzu, Confucius, and the Buddha in the East, testify this commitment to moral education through virtue and virtuous exemplars.

The centrality of virtue was eclipsed in Western thought starting from the scientific revolution and especially during the age of enlightenment, as the focus gradually shifted from moral character to right behaviour, with the rise of ethical codes based on the moral imperative (Kant) and utilitarian reasoning (Bentham and Mill). The return to prominence of virtue began in the late 1950s in philosophy (Anscombe, 1958) and was embraced soon after by the behavioural and social sciences with the movements of, among others, humanistic and positive psychology. The present research can be considered part of this line of inquiry: through the investigation of the reactions to the noble actions of virtuous people, moral excellence has been put back centre stage, offering new food for thought about how psychology can contribute to better understanding what role the bright side of human nature can still have in contemporary societies to inspire people to better themselves and behave in ways that are socially desirable.

Methods

Capitalising on the recommendations by Dubin (1969), the methodological approach was designed to make large of use generative/exploratory techniques before moving on to confirmatory methods, thus consolidating enough knowledge to develop

testable hypotheses based on a sound blend of deductive (theory-based) and inductive (data-driven) reasoning.

In an open science project that included four studies in the US and UK (N=1,814), participants were presented with uplifting stories of human goodness in the form of vignettes, where virtuous people performed good actions to the benefit of others and through various degrees of personal cost. These vignettes were created specifically for this research, building on the recommendations by Christensen and Gomila (2012) to facilitate comparability and replicability.

The questionnaire, and more generally the research design and procedure, were structurally similar across the four studies (Fig. V). Participants, following a few general questions about themselves, including a moral self-evaluation, made judgments about the level of goodness and propriety of the deeds, and drew inferences about the moral character of the protagonists of the stories, making attributions from situational and personal cues, and reporting the feelings they experienced. As a result of their cognitive, affective, and conative states, they also accepted or declined to participate in an unpaid voluntary task to help the experimenter, filling out a few additional personality scales, thus providing a measure of helping behaviour.

Hypotheses and Conceptual Model

In moral psychology, research and theorising on moral goodness have already shed light on some of the psychological processes implicated in the reactions to others' virtuous actions. These were mostly independent endeavours that investigated either the uplifting positive emotions of moral elevation (e.g., Haidt, 2000, 2003) and *kama muta* (Blomster Lyshol et al., 2020), or the dysfunctional phenomenon of moral resentment and do-gooder derogation (e.g., Minson & Monin, 2012). Partly inspired by this existing work, the present research developed for empirical testing a unifying

theoretical model capable of integrating these perspectives, identifying the workings of the *self* (see Leary & Tangney, 2012) as the central psychological function that could provide a more thorough explanation of why some people react positively and others negatively when they are exposed to displays of human virtue. The research hypothesised the existence of two distinguishable kinds of moral responses, each characterised by distinctive sets of self-regulatory mechanisms: one socially adaptive defined “moral self-improvement” and the other ultimately dysfunctional defined “moral self-defence”. These mechanisms were anticipated to be the result of moral comparative processes, to be associated with specific motivational dispositions, and to impact individuals’ social behaviour.

Main Findings

Analysis of the response to the moral exemplars in the vignettes unearthed how complex and multifaceted the human nature can be; as expected, all participant samples across the four studies experienced both positive and negative reactions to others’ moral excellence, as a result of the processes through which the self negotiates its standing and standards *vis-à-vis* the moral agents and their deeds.

Participants exhibited the two hypothesised moral self-regulatory modes. Some praised the moral agents, felt uplifted by their actions and inspired to better themselves (moral self-improvement); others felt threatened, denigrated the moral agents and dismissed the goodness of their actions, inferring ulterior motives and experiencing resentment (moral self-defence).

Among the critical antecedents of these self-regulatory modes were specific judgments about the moral deeds and the agents, which subsumed moral comparisons based on opinion and ability, in line with Festinger’s (1954) original social comparison theory. Upward and downward ability-based comparisons (Wood, 1996) explained a

significant portion of the variance of respectively self-improvement and self-defence. Opinion-based comparisons were strong predictors for moral deeds susceptible to a wider spectrum of judgments (e.g., Nicholas's fight against discrimination), including lower ratings on the goodness and propriety of the deed, which explained defensive regulation.

Although comparative processes are ubiquitous (Dunning, 2000; Hoorens, 2011), the present research showed that people exhibited different degrees of proneness to engage in moral comparisons, in line with Gibbons & Buunk (1999): individuals with strong externalities reverted more frequently to comparisons than those who relied predominantly on internal standards (Duttweiler, 1984; Rotter, 1975). Simple everyday acts of kindness tended to feed moral comparisons more than truly outstanding acts of goodness, due to the greater perceived self-other similarity (Wills, 1991) and reassuring accessibility of the deed.

Moral judgments/comparisons and self-regulation were also associated with specific motivational factors. The present research examined the impact of approach/avoidance (Elliot & Thrash, 2002), regulatory focus (Higgins, 1997), and regulatory mode (Kruglanski et al., 2000), identifying distinctive patterns through which they affected the moral response, directly or indirectly. All other things being equal, motivational drivers such as approach (or promotion focus) and prevention focus, showed greater predictive power than moral comparisons when the deed was perceived more universally as highly positive (e.g., Francia's organ donation).

In line with extant social cognitive theories (Bandura, 1991; Zimmerman, 2005), there was evidence that moral self-improvement, far from being an "inert" internal process, was rather *in the service* of behaviour: self-improvement was linked with actual helping behaviour, although with a more modest effect size ($r = .18$ among the total

sample) than hypothesised based on previous studies on moral elevation (Schnall et al., 2010). This relationship was not mediated or moderated by the desire to do something good for others (a critical action tendency item dropped from the final inventory in Study 3), providing evidence that the structure of the final self-improvement scale includes action tendencies that are in direct relationship with prosocial behaviour.

Surprisingly and against predictions, a smaller deed (paying a latte for a friend who forgot her wallet) and a greater deed (donating a kidney to a friend at risk of organ failure) did not reveal any *conclusive* evidence of a significant difference in the degree of prosociality that they triggered. Therefore, this particular result is deserving of further investigation before a clear causal relationship can be claimed.

Theoretical Contributions

The study of virtue – and morality more broadly – has its roots in philosophy, from the legacy of the ancient Greek forefathers (Socrates, Plato, and Aristotle) up to the more recent contributions by Anscombe, MacIntyre, and Rawls in the twentieth century. Psychology owes a great deal to philosophy, but over time has grown as an independent discipline, branching into different strands of research and theorising, among which *moral* psychology retains a lot of common ground. The present research project has borrowed multiple concepts originally defined and refined in moral philosophy (virtue, moral character, eudaimonia, supererogation, just to mention a few), investigating them as part of an inquiry into the psychological processes elicited by moral goodness. To do so, a self-regulation approach was adopted, enabling the integration of the above-mentioned independent lines of research, and injecting new life into them.

Self-regulation has been conceived of here in its broader meaning, encompassing both explicit and implicit processes (Carver & Scheier, 2016; Forgas et al., 2009), thus beyond the restrictive definition to which some social psychologists have confined it,

that is, self-control, willpower, and other executive functions (see e.g., Baumeister et al., 2006). The moral regulatory processes examined in the present research (i.e., moral self-improvement and self-defence) include respectively broadening self-growth processes (Sedikides & Hepper, 2009) and self-serving enhancement and protection mechanisms (Sedikides & Alicke, 2012), which map onto the two most fundamental regulatory functions that evolved for the expansion and the preservation of the self (Pyszczynski et al., 2012).

The literature on self-improvement comprises two theoretical traditions: the humanistic and the achievement strands (Sedikides & Hepper, 2009). The construct of moral self-improvement examined here is rooted in ethical strivings, and therefore is more closely linked with the schools of humanistic (Maslow, 1954) and positive psychology (Peterson & Seligman, 2004), which emphasise the growth of the self as a eudaimonic endeavour aiming at human flourishing and social harmonisation (Bauer et al., 2015). On the other hand, the literature on self-defence was initially developed in a psychodynamic framework (Freud, 1894), but the present research has drawn primarily on social psychology theorising on self-enhancement and self-protection (Alicke & Sedikides, 2009), social desirability and self-deception (Crowne & Marlowe, 1960; Paulhus et al., 1997).

Analysing the response to virtuous acts through the lens of self-regulation allowed the current research to embrace but also broaden the perspective taken up by existing studies on emotion, whether positive emotions such as moral elevation (Schnall et al., 2010) and *kama muta* (Blomster Lyshol et al., 2020), or negative emotions such as resentment and do-gooder derogation (Minson & Monin, 2012). The present research has shown that it is not just an emotional state elicited by a good deed that relates to certain social behaviours, but rather broader and more complex processes whereby the

self, energised to approach desired end-states and avoid undesired end-states through promotion and prevention strategies, negotiates its moral position against other people and referential standards (beliefs, values, ideals, norms). The self was also found to be at the centre of the upstream processes that elicit self-improvement and self-defence: comparative and motivational mechanisms. These cover a lot of ground in the social psychology literature, gravitating toward the more dated work on social comparison by Festinger (and his students and followers), and the vast opus in contemporary motivation science by psychologists such as Higgins, Kruglanski, Elliot, and colleagues on regulatory focus, regulatory mode, and approach/avoidance.

In a nutshell, by observing from a self-regulation viewpoint the processes elicited by exposure to acts of virtue performed by people of admirable moral character, it was possible to combine in a single model the cybernetic mechanisms of key motivational drivers and comparative processes that – taken together – trigger regulatory functions of the self associated with social behaviour.

While acknowledging certain methodological limitations and substantive ambiguities in the findings, the harmonisation of all these literatures and conceptual perspectives under a coherent unified self-regulatory framework constitutes the most crucial substantive theoretical contribution of the present research.

Methodological Advances

Stimuli

The use of vignettes is not new in moral psychology, but those developed for the present research programme share specific features that are relatively novel and worth noting. Prior studies on positive emotions (elevation, awe, etc.) have used mainly videos (e.g., Algoe & Haidt, 2009; Erickson et al., 2017; Lai et al., 2014; Schnall et al., 2010; Silvers & Haidt, 2008), but this research followed the example of Freeman et al. (2009),

Aquino et al. (2011), and Thomson and Siegel (2013), who used written stories. The text told *true* stories, and visuals were juxtaposed to make the narrative more realistic and engaging. Compared to videos, this kind of stimulus has the advantage of being easier to administer in online research among participants who own simpler technology or are less computer savvy, and are more flexible to develop, pre-test, modify, and re-test. Furthermore, to make research with moral vignettes more comparable and reproducible, the present investigation followed recommendations (Christensen & Gomila, 2012; Clifford et al., 2015) to standardise as much as possible format and content elements, so that more rigorous control could be achieved.

The result was satisfactory, as the twelve vignettes created and tested in Study 1 delivered on the objectives of the research, particularly three of them (Francia, Nicholas, and Ruxandra). Francia and Nicholas were used again in Studies 2-3 to refine the measurement and structural models, and a new one was created and tested as control stimulus in Study 4 against Francia; Ruxandra was put aside due to resource constraints, but still represents a valuable option for future research.

Moral Self-Regulation Inventory

The present research has also developed and assessed a new measurement instrument for two modes of moral self-regulation. The inventory encapsulates cognitive, conative, and affective indicators that measure type and level of the moral response to acts of virtue performed by moral exemplars.

The moral self-improvement scale partly overlaps with some measures of moral elevation used, for example, by Schnall et al. (2010), Aquino et al. (2011), and Thomson & Siegel (2013). However, whereas these are just a set of items collated to assess various components of the related moral emotion, the moral self-improvement scale is a true psychological measurement instrument assessing a latent variable through item

indicators that have been developed and pre-screened (Study 1), and subsequently assessed in exploratory factor analysis (Study 2) and confirmatory factor analysis (Study 3), exhibiting good psychometric properties. A measure of concurrent validity (helping behaviour) was also provided, while measures designed to provide convergent validity (HEMA, growth motivation, etc.) will require larger sample sizes.

The moral self-defence scale is new altogether. The item indicators that compose it have been inspired by disparate sources, indirectly by the psychoanalytic tradition (Freud, 1894) and more directly by the theoretical frameworks of self-enhancement and self-protection (Alicke & Sedikides, 2009; Hepper et al., 2010). Again, across Studies 1-3, the instrument was pre-screened, and then assessed through EFA and CFA, showing good psychometric properties. Together, the moral self-improvement and self-defence scales form the moral self-regulation inventory, which represents a new methodological tool available to scholars who wish to further research on the response to virtue or test whether their stimuli or intervention materials have the desirable effect, that is, inspire self-improvement and not trigger strong self-defence.

Independence of Moral Comparison and Self-Regulation Measures

Another contribution of the present research is the assessment of moral comparison and moral self-regulation as conceptually independent measures. A large portion of the literature on social comparison is based on the rank-order paradigm (Gerber, 2018), which tends to conflate them by measuring the choice of the comparison target, which is interpreted as a regulatory mechanism to improve or defend the self.

By contrast, in the present research ability-based moral comparisons were measured through the difference between moral evaluations of the agent and the self (moral discrepancy), and opinion-based comparisons were assumed to be underlying

judgments about the goodness and propriety of the deeds (against personal norms or beliefs). These measures were completely distinct from (albeit correlated with) the measures of moral self-regulation, represented by the new inventory described above. The conceptual clarity deriving from distinguishing these measures from each other could be claimed to be another strength of the research, further advancing similar approaches introduced by Crocker and colleagues (1987), while adapting them to the domain of virtue investigated here.

Causal Inferences

Although the interpretation of the structural equation models made use of causal inference, by no means the four studies that make up this research project were intended to provide “definitive proof” of causality. Rather than an output, causality was an input to structural equation modelling. Pearl (2012) clearly states that SEM is an inference engine that converts causal assumptions, queries, and data into logical implications, conditional claims, and data-fitness indices. If the qualitative causal assumptions used as input are grounded in theory and/or results from empirical studies (Kline, 2016), then “passing these tests does not *prove* the validity of the causal assumptions, but it lends *credibility* to them” (Bollen & Pearl, 2013, p. 9, emphases added).

It still holds true that the qualitative causal assumptions of a specified model remain just a single test away from rejection (Bollen & Pearl, 2013). The three conditions traditionally required for causal inference are: a) association (variables must be correlated); b) temporal ordering (a cause must unambiguously precede an effect); c) isolation (ruling out extraneous variables) (see e.g., Bollen, 1989; Bullock, Harlow, & Mulaik, 1994). While the first two conditions are fulfilled in the presence of respectively significant SEM regression weights and experimental designs (conditions substantiated

in the present research), providing evidence of the third is more arduous; indeed, it is possible that inclusion in the model of an additional variable might reveal that a certain association interpreted causally is spurious, and would disappear if only that variable were accounted for in the model specification. The condition of perfect isolation is in practice an “unobtainable ideal” (Bollen, 1989), but can be replaced by *pseudo-isolation*, which occurs when the error terms of the outcome variables, which summarise all omitted determinants, are independent of the predictors. Under this assumption and in conjunction with the other fundamental principles of causality, the structural equation models that achieved satisfactory fit to the data in the present research cannot be denied the role of “contributors” to the determination of “causal and counterfactual expressions” (Pearl, 2012) and represent a first step toward further credibility that could be achieved thanks to future testing among new samples.

Directions for Future Research

The current research project has set foot in the territory of virtue with a self-regulation perspective which promises plenty of possible further developments.

Personality Traits

To begin with, replicating some of the exploratory work conducted in Study 4 has the potential to expand the theoretical model, for example adding personality predictors of moral comparison and self-regulation, such as humility and narcissism. Their analysis in Study 4 was hindered by the low sample size, but appropriately powered studies could validate the correlation of humility with upward comparison and self-improvement, and of narcissism with downward comparison and self-defence.

Moral Scenarios

Beyond the top vignettes of Francia and Nicholas, Study 1 pointed to Ruxandra’s vignette (the vegetarian) as a third interesting scenario. Testing Ruxandra’s vignette

would represent a sort of “stress test” for the model. This scenario garnered comparatively higher levels of self-defence following poorer judgments of the goodness of the deed. This is the same phenomenon that was observed, to a lesser degree, in Nicholas’s vignette, suggesting in Study 3 the addition of opinion-based comparisons to the model. For Ruxandra’s vignette, it is likely that a larger sample will replicate the finding that a considerable portion of the variance of self-defence is explained by the low ratings of the goodness of the deed (which subsume opinion-based comparisons), with moral discrepancy (which is underpinned by ability-based comparisons) playing a more marginal role. Should the model fit be satisfactory and the error negligible, the theoretical framework would be further solidified, having been successfully applied to *three* vignettes – depicting decreasing levels of perceived moral goodness – capable of explaining differential gradients of moral self-improvement and self-defence.

Ideally, a wide variety of further moral scenarios could be crafted and tested, allowing the systematic manipulation of relevant contextual factors that are hypothesised to be implicated in the self-regulation of virtue. Indeed, these situational variables are known to interact with individual differences in generating complex response patterns. Modelling these patterns in a person-by-situation framework could expand even further the remit of the research.

Trait Measures of Moral Self-Regulation

The moral self-regulation inventory measures *state* latent variables. Because during human development the ongoing prevalence of certain states ends up consolidating in more stable traits, it would be interesting to develop and validate a *trait* moral self-regulation inventory that assesses ingrained dispositions to engage in moral self-improvement and self-defence. This new inventory could be used to investigate the

effects of systematic moral self-regulatory tendencies on happiness, well-being, self-actualisation, eudaimonia and hedonia.

Cross-Cultural Validation

The moral self-regulation inventory, as well as the full structural models of the self-regulation of virtue, could also be investigated cross-culturally among non-WEIRD⁸⁷ samples, for example testing them in Japan, China, Brazil, South Africa, and so on. Particularly interesting would be their validation in more communal and less individualistic societies. Communal orientation (Clark et al., 1987) and interdependent self-construal (Markus & Kitayama, 1991; Singelis, 1994) were found to be associated⁸⁸ with social comparison orientation (which should increase defensive regulation), but also with prosocial tendencies (which should stem from moral self-improvement). The coexistence of these opposite regulatory forces could lead to suppression effects (which could potentially cancel each other out), making this investigation worth undertaking. However, it would be important to tightly control demand characteristics in the experimental design and carefully model a measure of socially desirable responding, since individuals in communal interdependent cultures tend to conform to the assumed expectations of the experimenter and the perceived social norm (Buunk & Gibbons, 2007).

Manipulation of Motivational Variables

In the structural models in Studies 2-4, the motivational predictors were treated as exogenous traits. To strengthen causal inference, future research could design experiments that manipulate those variables, treating them as transient states. The

⁸⁷ WEIRD is the popular acronym of Western, Educated, Industrialized, Rich, Democratic (see Henrich et al., 2010a, 2010b).

⁸⁸ For a review, see Buunk & Gibbons, 2007.

literature on regulatory focus, regulatory mode, and approach/avoidance provides a wealth of related experimental inductions⁸⁹. As examples of regulatory focus and approach/avoidance inductions adapted to the context of moral goodness, participants could be assigned the task of writing/typing a memory of when they performed a good deed that reflected an ideal enabling them to achieve an important aspiration in their life (promotion induction in approach motivation), or a memory of when they performed a good deed that reflected a moral duty enabling them to avert a critical threat in their life (prevention induction in avoidance motivation).

Antisocial Behaviour

As mentioned in the introductory chapter, the primary form of social behaviour investigated in the present research was prosocial behaviour, due to the prevalence of positive reactions (moral self-improvement) in response to the moral exemplars portrayed in the vignettes. The integration of social behaviour in the theoretical model was accomplished in Study 4 through a helping task that tested the prosocial effects of moral self-improvement. Because of the non-conclusive findings, further research will have to clarify this link.

An additional area of investigation could measure possible relationships between moral self-defence and antisocial behaviour, not covered in the present research. Despite the scarcity of specific evidence, the literature has examined associations between defence mechanisms and antisocial personality (see e.g., Bowins, 2004; Bryhn Nørbech et al., 2013; Presniak et al., 2010; Vaillant, 1994), the link between self-threat and antisocial personality and behaviour (see e.g., Aquino & Douglas, 2003; Kumari et

⁸⁹ For regulatory focus see e.g., Freitas et al., 2002; Higgins et al., 1994; Higgins et al., 2001; for regulatory mode see e.g., Avnet & Higgins, 2003; Komissarouk et al., 2018; Mannetti et al., 2009; for approach/avoidance see e.g., Muise et al., 2017; Strachman & Gable, 2006.

al., 2009; McGregor et al., 2009), and the links between moral disengagement (and its predictors and moderators) and antisocial behaviour (see e.g., Caprara et al., 2014; Hardy et al., 2015; Hyde et al., 2010). This body of research could inform hypotheses about the possible relationship between moral self-defence and antisocial behaviour, provided a valid measure of moral self-defence is warranted to assess self-shielding reactions to moral goodness in a “morally minimal” control condition, and a specific task is devised to measure antisocial behaviour (for example a cheating task, see e.g., Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008). Should no consistent relationship emerge, future research could examine at least to what extent moral self-defence hampers prosocial behaviour.

Neurophysiological Measures

Future research on the self-regulation of virtue could also include some neurophysiological measures, extending the methodological horizon beyond self-reports. In the literature on moral elevation, the study by Piper, Saslow, and Saturn (2015) stands out as one of the few in which the researchers attempted measurements of heart rate (HR), respiratory sinus arrhythmia (RSA), and medial prefrontal cortex (mPFC) activity. What they found is that, during elevation, HR and RSA increased, indicative respectively of sympathetic and autonomic activation. These functions are recruited simultaneously in a narrow range of situations, typically when both arousal and social engagement are required; they also found increased mPFC activity, which is often measured at the emergence of the emotions of compassion (see e.g., Immordino-Yang et al., 2009) and awe (see e.g., Keltner & Haidt, 2003). Given the similarity of the stimulus content used for the study by Piper et al. (2015) and Study 4, as well as the partial conceptual overlap between the measures of moral elevation and moral self-

improvement, a similar neurophysiological response pattern should be expected for moral self-improvement elicited by stories of moral goodness.

At a first glance, it might appear that another theoretical framework could provide physiological measures useful to expand the methodological remit of the current research: the biopsychosocial model of challenge and threat (BPS; Blascovich, 2008; Blascovich & Tomaka, 1996). The model links specific psychological states with patterns of physiological responses in situations of *social facilitation*, or in other words, when the presence of others (co-actors, observers, etc.) induces effects of performance enhancement and impairment, which are triggered by the interplay of affective and cognitive processes (Blascovich et al., 1999). In the BPS model, challenge and threat are conceptualised as person-situation motivational states: challenge occurs when the perceived availability of resources meets or exceeds the perceived situational demands, and vice versa for threat, when demands exceed resources (Seery, 2013). A perceived challenge in a self-relevant domain causes the heart to pump more blood while arteries are more dilated, which translates into measures of high cardiac output (CO) and low total peripheral resistance (TPR); by contrast, a perceived threat in a self-relevant domain causes the heart to pump less blood as the arteries constrict (low CO and high TPR). It is important to highlight that, according to the model, excess of or lack of resources in relation to situational demands represent evaluations that occur mainly at an implicit level, outside of conscious awareness, and are constantly updated (Quigley et al., 2002).

It might appear that the BPS notions of challenge and threat could roughly correspond conceptually to what in the present research was referred to as “opportunity” and “threat” to the self. However, despite similarities in collateral aspects of the theories, there is not a precise equivalence between the corresponding constructs

of the two models. An opportunity to improve as a person does not imply evaluations that resources exceed demands, but rather a tension toward an ideal moral aspiration, whatever the personal resources available at the moment (even if currently insufficient, one might believe that, as a result of improving, additional resources will become available in the future); a threat to one's moral standing does not imply evaluations that demands exceed resources, but rather the perception of an attack to cherished self-views, independent of the defensive armoury of resources currently available to the self (an attack to moral self-beliefs is unpleasant regardless of the comparative moral stature attributed to self and other). Furthermore, the BPS applies to a *performance* context (Seery, 2013), which is not directly applicable to the moral domain.

Another related theory that links conceptual neurophysiological systems to specific psychological functions is Gray's reinforcement sensitivity theory (RST). In its revised formulation (Gray & McNaughton, 2000), RST defines three brain-behaviour systems: 1) the behavioural approach system (BAS), a system that motivates approach toward all appetitive stimuli (Pickering & Smillie, 2008); 2) the fight-flight-freeze system (FFFS), an avoidance mediator of all aversive stimuli⁹⁰ (Smillie et al., 2011); 3) the behavioural inhibition system (BIS), a goal-conflict resolver, for instance in the simultaneous presence of BAS-approach and FFFS-avoidance (Pickering & Corr, 2008). The BAS is related to *desire* (approach), the FFFS to *fear* (defensive avoidance), and the BIS to *anxiety* (defensive approach). Initially, RST was proposed as a state theory, but soon after it was suggested that the functioning of the three neurobiological systems could manifest themselves as stable dispositions (Corr & McNaughton, 2008), thus making RST

⁹⁰ The FFFS elicits defensive attack (fight) in the presence of *proximal* threats, and flight and freeze concern responses to *distal* threats, depending respectively upon the availability and unavailability of escape (Smillie et al., 2011).

a biological theory of personality in its own right (Corr, 2015). The possibility to use neuroimaging techniques (e.g., fMRI) for these neurophysiological measures of approach and avoidance motives opens new opportunities to complement self-reports with other indicators. However, the complexity of the interpretation of these measurements should not be underestimated, since the three neurobiological systems are widely distributed, each of them mapping onto multiple brain organs⁹¹. Another issue related to this kind of neuroimaging studies is the availability and cost of large samples required to embed them into wider structural models; a practical solution could be to design studies that isolate smaller portions of the wider moral self-regulation model, so that adequately powered research can be conducted in the laboratory among reasonably achievable samples.

Recommended Priorities

Of all the possible directions examined above for future research, two seem to be critical: first, expanding the analysis by investigating a wider repertoire of moral actions, for example testing Ruxandra's vignette or other vignettes expected to generate a sizeable amount of less positive reactions; second, validating the measurement and structural models among non-WEIRD samples. It must be added that cross-cultural validation would also assume the integration of a measure of social desirability, modelling it based on the results of the exploratory analysis conducted in Study 4. These integrations would ensure coverage of a wider set of situational variables as well as invariance across cultures, improving the theory generalisability.

Practical Implications

⁹¹ For the interested reader, a synthesis can be found in Pickering & Corr (2008).

Beyond its theoretical contributions to the literature in moral psychology and adjacent disciplines, the current research provides insights that could be useful for the design and implementation of various kinds of interventions aimed at maximising the benefits of moral self-improvement and minimising the drawbacks of self-defence. Although beyond the scope of the present endeavour, a few examples will be briefly discussed in the next paragraphs to elucidate the practical implications of the research; the focus will be primarily on the collective rather than the individual level, thanks to the potential wide-ranging benefits of related interventions and policies.

Amplifying the Upsides of Moral Self-Improvement

In the Media. Stories and exemplars of human goodness do appear in the present media landscape, but the news is often said to be burdened by overwhelmingly negative information, presumably due to the endemic “negativity bias” (Rozin & Royzman, 2001), which makes people naturally attend more closely to negative than positive information. In this context, expanding the presence of positive moral stories in the mass and social media would inspire uplifting prosocial responses among the audiences, as evidenced by this research project. This is precisely the mission of organisations such as Greater Good and research centres such as The Greater Good Science Center at the University of California, Berkeley⁹².

Moral Education. Exemplars of human virtue could also be viewed as part of a moral education journey that could potentially inform revised curricula across a variety of developmental stages and disciplines. This could be done by using dialogical and critical engagement techniques (see Hart et al., 2019) while framing inspiring moral stories of praiseworthy heroes and heroines in a narrative biographical form in primary

⁹² Further information can be found respectively visiting the following websites: <https://greatergood.com>, and <https://ggsc.berkeley.edu>.

and secondary schools for children and adolescents (e.g., in the study of history, philosophy, but also the sciences), or in more sophisticated forms in academic institutions and other organisations for adults (e.g., in undergraduate and graduate modules, courses and training sessions for entrepreneurs, social workers, medical staff, civil servants, etc.). If execution and delivery of these programmes are accurately designed for the respective audiences, their potential to positively affect all aspects of social, political, and economic life should not be underestimated.

These new strategies differ from traditional approaches to moral education limited to the presentation and discussion of moral dilemmas (Kohlberg, 1981). These are useful to improve moral reasoning and judgment about right behaviour, but remain silent about other aspects of morality, such as empathy, moral motives, moral affect, which are more conducive to building positive character traits (Han, 2019). Indeed, wise moral judgment could remain idle if not accompanied by character strengths that help translate sound moral reasoning into actual moral behaviour (Bebeau, 2002). It is precisely for this reason that character traits are becoming increasingly central to a new moral education paradigm that goes beyond rational rule-based ethics and emphasises motivation and virtue as essential means toward the good life and human flourishing (Han, 2015).

Containing the Downsides of Moral Self-Defence

The results of the present research also provide hints to devise actionable interventions capable of reducing the negative effects of moral self-defence. The related theoretical framework is offered by self-affirmation theory (Steele, 1988), which proposes that people can put aside their need to deflect self-threats by affirming the self in another domain, unrelated to the one under threat and at the same time central to their identity, thus reaffirming an overall perception of integrity and worth. Through this

mechanism, self-affirmations function as part of a “psychological immune system” (Gilbert et al., 1998) and promote more effective social adaptation: by shifting attention towards a different domain, individuals see the larger context of who they are, refocusing on the values by which they holistically define themselves as worthy (Sherman & Cohen, 2006). Self-affirmation interventions have the potential to counter the need to activate defensive strategies that in the long run end up compromising the person’s ability to cultivate and nurture constructive social relations (Sedikides, 2009; Sedikides & Luke, 2007).

Self-affirmations can be induced for instance by asking individuals to write/type, reflect, and engage in a task evocative of a value that is personally relevant to them. There is substantial empirical evidence (for a research review, see e.g., Sherman & Hartson, 2011) that these procedures generate the effect of boosting global self-resources, reducing defensive mechanisms. Although they have not been extensively employed in the context of moral self-threats, it is plausible to assume that they can be effective in mitigating the pitfalls of moral self-enhancement and self-protection.

At a collective level, these interventions could become part of moral education programmes. They could be particularly important for individuals who, due to the kind of socialisation that they were exposed to during their childhood, have developed ingrained dispositions to defensive regulation. Self-affirmations could help sustain their sense of self-worth by leveraging cherished values and experiences, limiting the maladaptive potential inherent in self-serving defensive mechanisms.

A Closing Note

Putting the self at the centre of an investigation of the response to others’ virtuous actions and explaining moral self-regulation in terms of how the self negotiates its standing and standards *vis-à-vis* moral exemplars allowed the current research to

provide evidence that virtue *can* be a double-edged sword. The findings shed new light on why most people feel uplifted and inspired to better themselves (*moral self-improvement*), but others feel resented, derogate the moral agents, and minimise or deny the goodness of their actions (*moral self-defence*). Empirical testing across a series of studies led to the consolidation of a new instrument (the moral self-regulation inventory) to measure the two above-mentioned self-regulatory modes, and new structural models that identify the nomological network of critical predictors and behavioural consequences. Analysis of the moral typologies through latent profile modelling provided an additional person-centred view at various levels of the models; the associations across the motivational, comparison, and self-regulatory types mirrored the relational pathways obtained through structural equation modelling, reinforcing the robustness of the findings.

The self-regulatory profiles indirectly showed the relative size of the positive versus negative responses to acts of human virtue. Although they speak to the bittersweet nature of the human response to moral goodness, they also revealed the strong prevalence of the sweet side, thanks to the widely predominant positive reactions experienced by participants. Theoretical/methodological advances and promising new directions for research were discussed alongside limitations in the interpretation of the findings; further, practical guidelines for possible strategies in education and media policies have been suggested for the maximisation of the benefits of moral self-improvement and the minimisation of the drawbacks of moral self-defence.

References

- Adams, R. M. (2006). *A theory of virtue: excellence in being for the good*. Oxford University Press
- Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: the “other-praising” emotions of elevation, gratitude, and admiration. *Journal of Positive Psychology, 4* (2), 105-127
- Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 85-106). Psychology Press
- Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: what they are and what they do. *European Review of Social Psychology, 20* (1), 1-48
- Alicke, M. D., Guenther, C. L., & Zell, E. (2012). Social self-analysis: constructing and maintaining personal identity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd Ed., pp. 291-308). Guilford Press
- Allaire, J. J., Gandrud, C., Russell, K., & Yetman, C. J. (2017). *networkD3: D3 JavaScript network graphs from R. R package version 0.4*. <https://cran.r-project.org/web/packages/networkD3/index.html>
- Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: the Muhammad Ali effect. *Social Cognition, 7* (3), 275-296
- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality, 40*, 440-450
- Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy, 33* (124), 1-19
- Aquino, K. & Douglas, S. (2003). Identity threat and antisocial behavior in organizations: the moderating effects of individual differences, aggressive modeling, and

- hierarchical status. *Organizational Behavior and Human Decision Processes*, 90, 195-208
- Aquino, K., McFerran, B., & Laven, M. (2011). Moral identity and the experience of moral elevation in response to acts of uncommon goodness. *Journal of Personality and Social Psychology*, 100 (4), 703-718
- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83 (6), 1423-1440
- Archer, A. (2018). Supererogation. *Philosophy Compass*, 13 (3), 1-9
- Arechar, A. A., & Rand, D. G. (2020). Turking in the time of COVID. *PsyArXiv*.
<https://doi.org/10.31234/osf.io/vktqu>
- Aruguete, M. S., Huynh, H., Browne, B. L., Jurs, B., Flint, E., & McCutcheon, L. E. (2019). How serious is the “carelessness” problem on Mechanical Turk? *International Journal of Social Research Methodology*, 22 (5), 441-449
- Asendorpf, J. B. (2012). Bias due to controlling a collider: a potentially important issue for personality research. *European Journal of Personality*, 26, 391-392
- Avnet, T., & Higgins, E. T. (2003). Locomotion, assessment, and regulatory fit: value transfer from “how” to “what”. *Journal of Experimental Social Psychology*, 39, 525-530
- Baird, J.-A. (2010). Construct-irrelevant variance sometimes has consequential validity. *Assessment in Education: Principles, Policy & Practice*, 17 (4), 339-343
- Baldner, C., Pierro, A., Kruglanski, A. W. (2020). Moving toward helping behavior: the roles of sympathy, helping goal attainability, and locomotion orientation. *Basic and Applied Social Psychology*, 42 (3), 133-149
- Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall

- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50 (2), 248-287
- Banker, C. C., & Leary, M. R. (2019). Hypo-egoic nonentitlement as a feature of humility. *Personality and Social Psychology Bulletin*, 46 (5), 1-16
- Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences*, 42, 815-824
- Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making". In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 478-515). Wiley
- Bashir, N. Y., Lockwood, P., Chasteen, A. L., Nadolny, D., & Noyes, I. (2013). The ironic impact of activists: negative stereotypes reduce social change influence. *European Journal of Social Psychology*, 43, 614-626
- Bastos, J. L., Celeste, R. K., Faerstein, E., & Barros, A. J. D. (2010). Racial discrimination and health: a systematic review of scales with a focus on their psychometric properties. *Social Science and Medicine*, 70, 1091-1099
- Bauer, J. J., & McAdams, D. P. (2010). Eudaimonic growth: narrative growth goals predict increases in ego development and subjective well-being three years later. *Developmental Psychology*, 46, 761-772
- Bauer, J. J., McAdams, D. P., & Pals, J. L. (2008). Narrative identity and eudaimonic well-being. *Journal of Happiness Studies*, 9, 81-104
- Bauer, J. J., Park, S. W., Kamide, H., Pesola, N. V., Kamble, S. V., Graham, L. E., DeBrosse, J., & Waddar, M. S. (2019). Growth motivation and well-being in the U.S., Japan, Guatemala, and India. *Journal of Happiness Studies*, 21 (3), 899-919

- Bauer, J. J., Park, S. W., Montoya, R. M., & Wayment, H. A. (2015). Growth motivation toward two paths of eudaimonic self-development. *Journal of Happiness Studies*, 6, 185-210
- Bauman, C. W., McGraw, A. P., & Bartels, D. M. (2014). Revisiting external validity: concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8 (9), 536-554
- Baumeister, R. F. (2007). Self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 841-845). Sage
- Baumeister, R. F., Bratslawsky, E., Finkenauer C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5 (4), 323-370
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, 4, 1-44
- Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2006). Self-regulation and the executive function: the self as controlling agent. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: handbook of basic principles* (2nd edition, pp. 516-539). Guilford Press
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspectives on Psychological Science*, 2 (4), 396-403
- Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., Macleod, C., Miller, L. C., Read, S. J., Roberts, B., Robinson, M. D., Wood, D., Wrzus, C. (2017). Integrating Personality Structure, Personality Process, and Personality Development. *European Journal of Personality*, 31 (5), 503-528

- Beatty, M. J., Pascual-Ferra, P., & Levine, T. R. (2015). Two studies examining the error theory underlying the measurement model of the Verbal Aggressiveness Scale. *Human Communication Research, 41*, 55-81
- Bebeau, M. J. (2002). The defining issues test and the four component model: contributions to professional education. *Journal of Moral Education, 31* (3), 271-295
- Behrend, T., Sharek, D., Meade, A., & Wiebe, E. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43* (3), 800-813
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. Wiley
- Bennett, W. J. (2010). *The book of virtues: a treasury of great moral stories*. Simon and Schuster
- Berger, S. M. (1977). Social comparison, modeling, and perseverance. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: theoretical and empirical perspectives* (pp. 209-234). Hemisphere
- Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology, 9*, 291-319
- Blascovich, J. (2008). Challenge and threat. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 431-445). Psychology Press
- Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology, 77*, 68-77
- Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 1-51). Academic Press

- Blomster Lyshol, J. K., Thomsen, L., & Seibt, B. (2020). Moved by observing the love of others: kama muta evoked through media fosters humanization of out-groups. *Frontiers in Psychology, 11* (1240). <https://doi.org/10.3389/fpsyg.2020.01240>
- Blunch, N. J. (2013). *Introduction to structural equation modelling using IBM SPSS Statistics and AMOS* (2nd Ed.). Sage
- Bocell, F. D. (2015). The impact of unmodeled error covariance on measurement models in structural equation modeling [unpublished doctoral dissertation]. University of Washington, Seattle, WA, USA
- Boldero, J. M., Higgins, E. T., & Hulbert, C. A. (2015). Self-regulatory and narcissistic grandiosity and vulnerability: common and discriminant relations. *Personality and Individual Differences, 76*, 171-176
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley
- Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S.L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301-328). Springer
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110* (2), 203-219
- Bowins, B. (2004). Psychological defense mechanisms: a new perspective. *The American Journal of Psychoanalysis, 64* (1), 1-26
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy, 25* (1), 49-59
- Breines, J. G., & Chen, S. (2012). Self-compassion increases self-improvement motivation. *Personality and Social Psychology Bulletin, 38* (9) 1133-1143

- Brickman, P., & Bulman, R. J. (1977). Pleasure and pain in social comparison. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: theoretical and empirical perspectives* (pp. 149-186). Hemisphere
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd Ed.). Guilford Press
- Bryhn Nørbech, P. C., Crittenden, P. M., & Hartmann, E. (2013). Self-protective strategies, violence, and psychopathy: theory and a case study. *Journal of Personality Assessment*, 95 (6), 571-584
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6 (1), 3-5
- Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal*, 1 (3), 253-267
- Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: a tutorial. *Advances in Methods and Practices in Psychological Science*, 2 (1) 77-101
- Busse, C., Kach, A. P., & Wagner, S. M. (2017). Boundary conditions: what they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*, 20 (4), 574-609
- Buunk, B. P., & Gibbons, F. X. (2007). Social comparison: the end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, 102, 3-21
- Buunk, B. P., & Mussweiler, T. (2001). New directions in social comparison research. *European Journal of Social Psychology*, 31, 467-475

- Buunk, B. P., & Ybema, J. F. (2003). Feeling bad, but satisfied: the effects of upward and downward comparison upon mood and marital satisfaction. *British Journal of Social Psychology, 42*, 613-628
- Byrne, B. M. (2016). *Structural equation modeling with Amos: basic concepts, applications, and programming* (3rd Ed.). Routledge
- Cain, A. (2005). Books and becoming good: demonstrating Aristotle's theory of moral development in the act of reading. *Journal of Moral Education, 34* (2), 171-183
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-106
- Campbell, J. D. (1986). Similarity and uniqueness: the effects of attribute type, relevance, and individual differences in self-esteem and depression. *Journal of Personality and Social Psychology, 50* (2), 281-294
- Campbell, J.D. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology, 59*, 538-549
- Campbell, W. K., Reeder, G. D., Sedikides, C., & Elliot, A. J. (2000). Narcissism and comparative self-enhancement strategies. *Journal of Research in Personality 34*, 329-347
- Caprara, G. V., Alessandri, G., Fida, R., Tisak, M. S., Fontaine, R. G., & Paciello, M. (2014). The contribution of moral disengagement in mediating individual tendencies toward aggression and violence. *Developmental Psychology, 50* (1), 71-85
- Carr, D. (2005). On the contribution of literature and the arts to the educational cultivation of moral virtue. *Journal of Moral Education, 34* (2), 137-151
- Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion, 30*, 105-110

- Carver, C. S., & Scheier, M. F. (2002). Control processes and self-organization as complementary principles underlying behavior. *Personality and Social Psychology Review, 6* (4), 304-315
- Carver, C. S., & Scheier, M. F. (2016). Self-regulation of action and affect. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: research, theory, and applications* (3rd Ed., pp. 3-23). Guilford Press
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319-333
- Cates, D. F. (1998). Ethics, literature, and the emotional dimension of moral understanding. *Journal of Religious Ethics, 26* (2), 409-431
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: consequences and solutions for behavioral researchers. *Behavioral Research, 46*, 112-130
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14* (3), 464-504
- Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11* (4), 1-10
- Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review. *Neuroscience and Biobehavioral Reviews, 36*, 1249-1264
- Chung, E. K., Kim, S. J., & Sohn, W. Y. (2014). Regulatory focus as a predictor of omission bias in moral judgment: mediating role of anticipated regrets. *Asian Journal of Social Psychology, 17*, 302-311

- Clark, M. S., Oullette, R., Powell, M. C., & Milberg, S. (1987). Recipient's mood, relationship type, and helping. *Journal of Personality and Social Psychology*, 53, 94-103
- Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: a standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research*, 47, 1178-1198
- Cline, E. M. (2004). Two interpretations of de in the Daodejing. *Journal of Chinese Philosophy*, 31 (2), 219-233
- Coffman, D. L., Patrick, M. E., Palen, L. A., Rhodes, B. L., & Ventura, A. K. (2007). Why do high school seniors drink? Implications for a targeted approach to interventions. *Prevention Science*, 8, 241-248
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Erlbaum
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155-159.
- Cohen, T. R., & Morse, L. (2014). Moral character: what it is and what it does. *Research in Organizational Behavior*, 34, 43-61
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12, 381-398
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with application in the social, behavioral, and health sciences*. Wiley
- Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling*, 11 (4), 535-559
- Cornwell, J. F. M., & Higgins, E. T. (2014). Locomotion concerns with moral usefulness: when liberals endorse conservative binding moral foundations. *Journal of Experimental Social Psychology*, 50, 109-117

- Cornwell, J. F. M., & Higgins, E. T. (2015a). Approach and avoidance in moral psychology: evidence for three distinct motivational levels. *Personality and Individual Differences, 86*, 139-149
- Cornwell, J. F. M., & Higgins, E. T. (2015b). The “ought” premise of moral psychology and the importance of the ethical “ideal”. *Review of General Psychology, 19* (3), 311-328
- Corr, P. J. (2015). Personality, biological models of. *International Encyclopedia of the Social & Behavioral Sciences* (2nd Ed., Vol. 17, pp. 857-862). Elsevier
- Corr, P. J., & Cooper, A. J. (2016). The reinforcement sensitivity theory of personality questionnaire (RST-PQ): development and validation. *Psychological Assessment, 28* (11), 1427-1440
- Corr, P. J., & McNaughton, N. (2008). Reinforcement sensitivity theory and personality. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 155-187). Cambridge University Press
- Costa, P. T., & McCrae, J. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13* (6), 653-665
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10* (7), 1-8
- Cramer, P. (1998). Defensiveness and defense mechanisms. *Journal of Personality, 66* (6), 879-894
- Crisp, R., & Slote, M. (1997). Introduction. In R. Crisp & M. Slote (Eds.), *Virtue ethics* (pp. 1-25). Oxford University Press
- Critcher, C. R., Helzer, E. G., & Tannenbaum, D. (2020). Moral character evaluation: testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology, 87* (103906). <https://doi.org/10.1016/j.jesp.2019.103906>

- Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, 130, 392-414
- Crocker, J., Thompson, L. L., McGraw, K. M., & Ingerman, C. (1987). Downward comparison, prejudice, and evaluations of others: effects of self-esteem and threat. *Journal of Personality and Social Psychology*, 52 (5), 907-916
- Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*, 69 (2), 117-132
- Crowe, M., Carter, N. T., Campbell, W. K., & Miller, J. D. (2016). Validation of the narcissistic grandiosity scale and creation of reduced item variants. *Psychological Assessment*, 28 (12), 1550-1560
- Crowe, M. L., Lynam, D. R., Campbell, W. K., & Miller, J. D. (2019). Exploring the structure of narcissism: toward an integrated solution. *Journal of Personality*, 87 (6), 1-19. <https://doi.org/10.1111/jopy.12464>
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24 (4), 349-354
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4-19
- Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, 78, 106-124
- Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, 6, 97-103
- Cushman, F., Young, L., & Greene, J. D. (2010). Multi-system moral psychology. In J. M. Doris (Ed.), *The moral psychology handbook* (pp. 47-71). Oxford University Press

- Darwall, S. (2003). Introduction. In S. Darwall (Ed.), *Virtue ethics* (pp. 1-4). Blackwell
- Davis, D. E., & Hook, J. N. (2014). Humility, religion, and spirituality: an endpiece. *Journal of Psychology and Theology*, 42, 111-117
- Davis, D. E., McElroy, S. E., Rice, K. G., Chloe, E., Westbrook, C., Hook, J. N., Van Tongeren, D. R., DeBlaere, C., Hill, P., Placares, V., & Worthington, E. L. (2016). Is modesty a subdomain of humility? *Journal of Positive Psychology*, 11 (4), 439-446.
<http://dx.doi.org/10.1080/17439760.2015.1117130>
- de Saussure, F. (1959). *Course in general linguistics* (W. Baskin, Trans.). McGraw-Hill
(original work published in 1916)
- Dedeke, A. (2015). A cognitive-intuitionist model of moral judgment. *Journal of Business Ethics*, 126, 437-457
- Dent, N. J. H. (1984). *The moral psychology of the virtues*. Cambridge University Press
- DeVellis, R. F. (2003). *Scale development: theory and applications* (2nd Ed.). Sage
- DeYoung, C. G. (2015). Cybernetic Big Five theory. *Journal of Research in Personality*, 56, 33-58
- Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, 17 (3), 188-207
- Displayr (2019). *flipPlots: creates plots. R package version 1.2.0*.
<https://github.com/Displayr/flipPlots>
- Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: a meta-analysis. *Journal of Experimental Psychology: General*, 147 (11), 1713-1727
- Dubin, R. (1969). *Theory building*. The Free Press

- Dufner, M., Gebauer, J. E., Sedikides, C., & Denissen, J. J. A. (2018). Self-enhancement and psychological adjustment: a meta-analytic review. *Personality and Social Psychology Review, 23* (1), 48-72
- Dunning, D. (2000). Social judgment as implicit social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: theory and research* (pp. 353-378). Springer
- Dunning, D. (2007). Self. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 785-787). Sage.
- <http://dx.doi.org.ezproxy.lancs.ac.uk/10.4135/9781412956253>
- Duttweiler, P. C. (1984). The internal control index: a newly developed measure of locus of control. *Educational and Psychological Measurement, 44*, 209-221
- Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: a review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review, 23* (4), 1-35
- Elliot, A. J. (2008). Approach and avoidance motivation. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 3-14). Psychology Press
- Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology, 82* (5), 804-818
- Elliot, A. J., & Thrash, T. M. (2010). Approach and avoidance temperament as basic dimensions of personality. *Journal of Personality, 78* (3), 865-906
- Erickson, T. M., McGuire, A. P., Scarsella, G. M., Crouch, T. A., Lewis, J. A., Eisenlohr, A. P., & Muresan, T. J. (2017). Viral videos and virtue: moral elevation inductions shift affect and interpersonal goals in daily life. *Journal of Positive Psychology, 13* (6), 643-654

- Exline, J. J., & Geyer, A. L. (2004). Perceptions of humility: a preliminary study. *Self and Identity, 3*, 95-114
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191
- Feldman Barrett, L., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2005). Interoceptive sensitivity and self-reports of emotional experience. *Journal of Personality and Social Psychology, 87* (5), 684-697
- Fellner, B., Holler, M., Kirchler, E., Schabmann, A. (2007). Regulatory focus scale (RFS): development of a scale to record dispositional regulatory focus. *Swiss Journal of Psychology, 66* (2), 109-116
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th Ed.). Sage
- Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. Sage
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise: an experimental study on cheating. *Journal of the European Economic Association, 11* (3), 525-547
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences, 11* (2), 77-83
- Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: from brains to culture* (2nd Ed.). SAGE
- Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality, 56*, 82-92
- Fleeson, W., & Nofhle, E. (2008). The end of the person-situation debate: an emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass, 2* (4), 1667-1684

- Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble: overfitting in the assessment of internal structure and some editorial thoughts on it. *European Journal of Psychological Assessment, 33* (6), 399-402
- Foot, P. R. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review, 5*, 1-5
- Forgas, J. P., Baumeister, R. F., & Tice, D. M. (2009). The psychology of self-regulation: an introductory review. In J. P. Forgas, R. F. Baumeister, & D. M. Tice (Eds.), *Psychology of self-regulation: cognitive, affective and motivational processes* (pp. 1-17). Psychology Press
- Fournier, M. A., Di Domenico, S. I., Weststrate, N. M., Quitasol, M. N., & Dong, M. (2015). Toward a unified science of personality coherence. *Canadian Psychology / Psychologie Canadienne, 56* (2), 253-262
- Fox, J. (2015). *Applied regression analysis and generalized linear models* (3rd Ed.). Sage
- Fredrickson, B. (2001). The role of positive emotions in Positive Psychology: the broaden-and-build theory of positive emotions. *American Psychologist, 56* (3), 218-226
- Fredrickson, B., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 19* (3), 313-332
- Freeman, D., Aquino, K., & McFerran, B. (2009). Overcoming beneficiary race as an impediment to charitable donations: social dominance orientation, the experience of moral elevation, and donation behavior. *Personality and Social Psychology Bulletin, 35* (1), 72-84
- Freitas, A. L., Liberman, L., Salovey, P., & Higgins, E. T. (2002). When to begin? Regulatory focus and initiating goal pursuit. *Personality and Social Psychology Bulletin, 28* (1), 121-130

- Freud, A. (1937). *The ego and the mechanisms of defense* (Cecil Baines, Trans.). Hogarth
(original work published 1936)
- Freud, S. (1894). *The neuro-psychoses of defence. The Standard Edition of the Complete Psychological Works of Sigmund Freud, 3*, 41-61. Hogarth
- Freud, S. (1896). *Further remarks on the neuro-psychoses of defence. The Standard Edition of the Complete Psychological Works of Sigmund Freud, 3*, 157-185. Hogarth
- Friend, R. M., & Gilbert, J. (1973). Threat and fear of negative evaluation as determinants of locus of social comparison. *Journal of Personality, 41*, 328-340
- Frijda, N. H. (1986). *The emotions*. Cambridge University Press
- Furnham, A., & Steele, H. (1993). Measuring locus of control: a critique of general, children's, health- and work-related locus of control questionnaires. *British Journal of Psychology, 84*, 443-479
- Gawronski, B., & Creighton, L. A. (2013). Dual process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282-312). Oxford University Press
- Gerber, J. P. (2018). Social comparison theory. In V. Zeigler-Hill, & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Springer.
<https://doi.org/10.1007/978-3-319-28099-8>
- Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76* (1), 129-142
- Giesler, R. B., Josephs, R. A., & Swann, W. B. (1996). Self-verification in clinical depression. *Journal of Abnormal Psychology, 105*, 358-368
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78

- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75, 617-638
- Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology*, 73 (1), 186-199
- Gollwitzer, P. M., Heckhausen, H., & Stellar, B. (1990). Deliberative vs. implemental mind-sets: cognitive tuning toward congruous thoughts and information. *Journal of Personality and Social Psychology*, 59, 1119-1127
- Gomez-Lavin, J., Prinz, J., Strohminger, N., & Nichols, S. (2020). Expansive interdisciplinarity and the moral self. In N. E. Snow & D. Narvaez (Eds.), *Self, motivation and virtue: innovative interdisciplinary research* (pp. 25-41). Routledge
- Goodman, S. H., & Waters, L. K. (1987). Convergent validity of five locus of control scales. *Educational and Psychological Measurement*, 47, 743-747
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48, 250-256
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106 (1), 148-168
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P.H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101 (2), 366-385
- Gray, J. A. (1982). *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system*. Oxford University Press
- Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system* (2nd Ed.). Oxford University Press

- Gray, K., Young, L., & Waitz, A. (2012). Mind perception Is the essence of morality. *Psychological Inquiry*, 23, 101-124
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg self-esteem scale: do they matter? *Personality and Individual Differences*, 35, 1241-1254
- Greene, J. D. (2015). The rise of moral cognition. *Cognition*, 135, 39-42
- Gregg, A. P., & Sedikides, C. (2018). Essential self-evaluation motives: caring about who we are. In M. van Zomeren & J. F. Dovidio (Eds.), *The Oxford handbook of the human essence* (pp. 59-70). Oxford University Press
- Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, 15 (5), 1-13
- Haan, N. (1985). Processes of moral development: cognitive or social disequilibrium? *Developmental Psychology*, 21 (6), 996-1006
- Haan, N. (1986). Systematic variability in the quality of moral action, as defined in two formulations. *Journal of Personality and Social Psychology*, 50, 1271-1284
- Haggard, P., & Lau, H. (2013). What is volition? *Experimental Brain Research*, 229, 285-287
- Haidt, J. (2000). The positive emotion of elevation. *Prevention & Treatment*, 3 (3), 1-5
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108 (4), 814-834
- Haidt, J. (2003). Elevation and the positive psychology of morality. In C. L. M. Keyes & J. Haidt (Eds.), *Flourishing: positive psychology and the life well-lived* (pp. 275-289). American Psychological Association

- Haidt, J., & Joseph, C. (2004). Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133 (4), 55-66
- Haidt, J., & Kesebir, S. (2010). Morality. In S. Fiske, G. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th Ed., pp. 797-832). Wiley
- Hakemulder, F. (2000). *The moral laboratory: experiments examining the effects of reading literature on social perception and moral self-concept*. John Benjamin's Publishing Company
- Han, H. (2015). Virtue ethics, positive psychology, and a new model of science and engineering ethics education. *Science and Engineering Ethics*, 21, 441-460
- Han, H. (2019). The VIA inventory of strengths, positive youth development, and moral education. *Journal of Positive Psychology*, 14 (1), 32-40
- Hankmiller, K. L. (1966). Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology*, 2 (Suppl. 1), 32-39
- Hardy, S. A., Bean, D. S., & Olsen, J. A. (2015). Moral identity and adolescent prosocial and antisocial behaviors: interactions with moral disengagement and self-regulation. *Journal of Youth and Adolescence*, 44, 1542-1554
- Harman, G. (2000). The nonexistence of character traits. *Proceedings of the Aristotelian Society*, 100, 223-226
- Hart, P., Oliveira, G., & Pike, M. (2019). Teaching virtues through literature: learning from the "Narnian Virtues" character education research. *Journal of Beliefs and Values*. <https://doi.org/10.1080/13617672.2019.1689544>
- Hart, W., Richardson, K, & Tortoriello, J. K. (2018). Revisiting the interactive effect of narcissism and self-esteem on responses to ego threat: distinguishing between assertiveness and intent to harm. *Journal of Interpersonal Violence*, 1-26. <https://doi.org/10.1177/0886260518777551>

- Havlicek, L. L., & Person, N. L. (1976). Robustness of the Pearson correlation against violations of assumptions. *Perceptual and Motor Skills*, 43 (3), 1319-1334
- Haws, K. L., Dholakia, U. M., & Bearden, W. O. (2010). An assessment of chronic regulatory focus measures. *Journal of Marketing Research*, 47 (5), 967-982
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional processes: a regression-based approach* (2nd Ed.). Guilford Press
- Hayes, A. F., & Preacher, K. J. (2013). Conditional process modeling: using structural equation modeling to examine contingent causal processes. In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences: issues, research, and teaching. Structural equation modeling: a second course* (pp. 219-266). IAP Information Age Publishing
- He, J., & Fan, X. (2019). Latent Profile Analysis. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*. Springer
- Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion*, 11, 101-120
- Heimpel, S. A., Elliot, A. J., & Wood, J. V. (2006). Basic personality dispositions, self-esteem, and personal goals: an approach-avoidance analysis. *Journal of Personality*, 74, 1293-1320
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58 (1), 78-80
- Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: a reexamination of Murray's narcissism scale. *Journal of Research in Personality*, 31, 588-599

- Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, 466 (7302), 29
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61-135
- Hepper, E. G., Gramzow, R. H., & Sedikides, C. (2010). Individual differences in self-enhancement and self-protection strategies: an integrative analysis. *Journal of Personality*, 78 (2), 781-814
- Hernán, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, 108, 616-619
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319, 1362-1367
- Hickman, J. (1981). A new perspective on response to literature: research in an elementary school setting. *Research in the Teaching of English*, 15 (4), 343-354
- Higgins, E. T. (1996). Knowledge activation: accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: handbook of basic principles* (pp. 133-168). Guilford Press
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52 (12), 1280-1300
- Higgins, E. T. (2014). *Beyond pleasure and pain: how motivation works*. Oxford University Press
- Higgins, E. T., Bond, R. N., Klein, R., & Strauman, T. (1986). Self-discrepancies and emotional vulnerability: how magnitude, accessibility, and type of discrepancy influence affect. *Journal of Personality and Social Psychology*, 51 (1), 5-15

- Higgins, E. T., Cornwell, J. F. M., & Franks, B. (2014). "Happiness" and "The Good Life" as motives working together effectively. In A. J. Elliott (Ed.), *Advances in motivation science* (Vol. 1, pp. 135-179). Elsevier
- Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: promotion pride versus prevention pride. *European Journal of Social Psychology*, 31, 3-23
- Higgins, E. T., Kruglanski, A. W., & Pierro, A. (2003). Regulatory mode: locomotion and assessment as distinct orientations. *Advances in Experimental Social Psychology*, 35, 293-344
- Higgins, E. T., Roney, C. J., Crowe, E., & Hymes, C. (1994). Ideal versus ought predilections for approach and avoidance: distinct self-regulatory systems. *Journal of Personality and Social Psychology*, 66, 276-286
- Hoffarth, M. R., Azevedo, F., & Jost, J. T. (2019). Political conservatism and the exploitation of nonhuman animals: an application of system justification theory. *Group Processes and Intergroup Relations*, 22 (6), 858-878
- Hong, R. Y., Tan, M. S., & Chang, W. C. (2004). Locomotion and assessment: self-regulation and subjective well-being. *Personality and Individual Differences*, 37, 325-332
- Hook, J. N., Davis, D. E., Owen, J., Worthington, E. L., Jr., & Utsey, S. O. (2013). Cultural humility: Measuring openness to culturally diverse clients. *Journal of Counseling Psychology*, 60, 353-366
- Hoorens, V. (2011). The social consequences of self-enhancement and self-protection. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 235-257). Guilford Press

- Hornsey, M. J. (2005). Why being right is not enough: predicting defensiveness in the face of group criticism. *European Review of Social Psychology*, 16, 301-334
- Hoyle, R. H. (1995). The structural equation modeling approach: basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 1-15). Sage
- Hudson, N. W., & Fraley, R. C. (2016). Changing for the better? Longitudinal associations between volitional personality change and psychological well-being. *Personality and Social Psychology Bulletin*, 42, 603-615
- Hui, C. M., Bond, M. H., & Molden, D. C. (2012). Why do(n't) your partner's efforts at self-improvement make you happy? An implicit theories perspective. *Personality and Social Psychology Bulletin*, 38 (1), 101-113
- Huta, V., & Ryan, R. M. (2010). Pursuing pleasure or virtue: the differential and overlapping well-being benefits of hedonic and eudaimonic motives. *Journal of Happiness Studies*, 11, 735-762
- Hyde, L. W., Shaw, D. S., & Moilanen, K. L. (2010). Developmental precursors of moral disengagement and the role of moral disengagement in the development of antisocial behavior. *Journal of Abnormal Child Psychology*, 38, 197-209
- Hyland, P., Boduszek, D., Dhingra, K., Shevlin, M., & Egan, A. (2014). A bifactor approach to modelling the Rosenberg self esteem scale. *Personality and Individual Differences*, 66, 188-192
- IBM SPSS Statistics for Windows (2019). *IBM Corp. (Version 26.0)*. Armonk, NY, USA.
<https://www.ibm.com/analytics/spss-statistics-software> [computer software]
- Immordino-Yang, M. H., McColl, A., Damasio, H., & Damasio, A. (2009). Neural correlates of admiration and compassion. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (19), 8021-8026

- Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, 2 (3), 260-280
- Jackson, C. J. (2009). Jackson-5 scales of revised reinforcement sensitivity theory (r-RST) and their application to dysfunctional real world outcomes. *Journal of Research in Personality*, 43, 556-569
- Jacobs, K. W. (1993). Psychometric properties of the internal control index. *Psychological Reports*, 73, 251-255
- Jamovi Project (2020). *Jamovi (Version 1.1)*. <https://www.jamovi.org> [computer software]
- Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: two faces of moral regulation. *Journal of Personality and Social Psychology*, 96 (3), 521-537
- Janssen, D.-J., & Scheepers, P. (2019). How religiosity shapes rejection of homosexuality across the globe. *Journal of Homosexuality*, 66 (14), 1974-2001
- Jarosz, A. F., Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, 7 (1), 2-9
- Jayawickreme, E., Meindl, P., Helzer, E. G., Furr, R. M., & Fleson, W. (2014). Virtuous states and virtuous traits: how the empirical evidence regarding the existence of broad traits saves virtue ethics from the situationist critique. *Theory and Research in Education*, 12 (3), 283-308
- Jefferson, T. (1975). Letter to Robert Skipwith. In M. D. Peterson (Ed.), *The portable Thomas Jefferson* (pp. 349-351). Penguin
- Jeffreys, H. (1961). *Theory of Probability* (3rd Ed.). Oxford University Press

- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78-89.
- Jordan, A. H., & Monin, B. (2008). From sucker to saint: moralization in response to self-threat. *Psychological Science, 19* (8), 809-815
- Jöreskog, K., & Sörbom, D. (1996). *PRELIS2: user's reference guide*. Scientific Software International
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2019). *semTools: useful tools for structural equation modeling. R package version 0.5-2*.
<https://cran.r-project.org/web/packages/semTools/index.html>
- Kahane, G. (2015). Sidetracked by trolleys: why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience, 10* (5), 551-560
- Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). Utilitarian judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193-209
- Kam, C. C. S., & Chan, G. H.-H. (2018). Examination of the validity of instructed response items in identifying careless respondents. *Personality and Individual Differences, 129*, 83-87
- Kaplan, D. (1995). Statistical power in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 100-117). Sage
- Kaplan, D. (2009). *Structural equation modeling: foundations and extensions* (2nd Ed.). Sage
- Karney, B. R., & Frye, N. E. (2002). "But we've been getting better lately": comparing prospective and retrospective views of relationship development. *Journal of Personality and Social Psychology, 82*, 222-238

- Kawamura, Y., & Kusumi, T. (2020). Altruism does not always lead to a good reputation: a normative explanation. *Journal of Experimental Social Psychology, 90* (104021).
<https://doi.org/10.1016/j.jesp.2020.104021>
- Kelley, H. H. (1971). Moral evaluation. *American Psychologist, 26* (3), 293-300
- Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and Emotion, 17* (2), 297-314
- Kline, R. B. (2012). Assumptions in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 111-125). Guilford Press
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th Ed.). Guilford Press
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research*, (Vol. 1, pp. 383-432). Sage
- Kohlberg, L. (1976). Moral stages and moralization: the cognitive-developmental approach. In T. Lickona (Ed.), *Moral development and behavior: theory, research, and social issues* (pp. 31-53). Holt, Rinehart and Winston
- Kohlberg, L. (1981). *The philosophy of moral development: moral stages and the idea of justice*. Harper & Row
- Komissarouk, S., Chernikova, M., Kruglanski, A. W., & Higgins, E. T. (2018). Who is most likely to wear rose-colored glasses? How regulatory mode moderates self-flattery. *Personality and Social Psychology Bulletin, 45* (3), 327-341
- Koopman, E. M., & Hakemulder, F. (2015). Effects of literature on empathy and self-reflection: a theoretical-empirical framework. *Journal of Literary Theory, 9* (1), 79-111

- Kossowska, M., Szumowska, E., Szwed, P., Czernatowicz-Kukuczka, A., & Kruglanski, A. W. (2020). Helping when the desire is low: expectancy as a booster. *Motivation and Emotion*. <https://doi.org/10.1007/s11031-020-09853-3>
- Krizan, Z., & Bushman, B. J. (2011). Better than my loved ones: social comparison tendencies among narcissists. *Personality and Individual Differences*, 50, 212-216
- Krizan, Z., & Herlacher, A. D. (2018). The Narcissism Spectrum Model: a synthetic view of narcissistic personality. *Personality and Social Psychology Review*, 22 (1) 3-31
- Kruglanski, A. W., Pierro, A., Mannetti, L., Higgins, T. E. (2013). The distinct psychologies of "looking" and "leaping": assessment and locomotion as the springs of action. *Social and Personality Psychology Compass*, 7 (2), 79-92
- Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., & Chun, W. Y. (2002). A theory of goal systems. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 331-378). Academic Press
- Kruglanski, A. W., Thompson, E. P., Higgins, E. T., Atash, M N., Pierro, A., Shah, J. Y., & Spiegel, S. (2000). To "Do The Right Thing" or to "Just Do It": locomotion and assessment as distinct self-regulatory imperatives. *Journal of Personality and Social Psychology*, 79 (5), 793-815
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47, 2025-2047
- Kuhn, M., Jackson, S., & Cimentada, J. (2020). *corrr: correlations in R*. R package version 0.4.2. <https://CRAN.R-project.org/package=corrr>
- Kumari, V., Das, M., Taylor, P. J., Barkataki, I., Andrew, C., Sumich, A., Williams, S. C. R., & ffitch, D H. (2009). Neural and behavioural responses to threat in men with a history of serious violence and schizophrenia or antisocial personality disorder. *Schizophrenia Research*, 110, 47-58

- Kurman, J. (2006). Self-enhancement, self-regulation and self-improvement following failures. *British Journal of Social Psychology*, 45, 339-356
- Lai, C. K., Haidt, J., & Nosek, B. A. (2014). Moral elevation reduces prejudice against gay men. *Cognition and Emotion*, 28 (5), 781-794
- Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biological Psychology*, 84, 437-450
- Larsen, R. J., Diener, E., & Emmons, R. A. (1986). Affect intensity and reactions to daily life events. *Journal of Personality and Social Psychology*, 51 (4), 803-814
- Lazarus, R. S. (1999). The cognition-emotion debate: a bit of history. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 3-19). John Wiley & Sons
- Leary, M. R. (2004). Editorial: what is the self? A plea for clarity. *Self and Identity*, 3 (1), 1-3
- Leary, M. R., & Guadagno, J. (2013). The sociometer, self-esteem, and the regulation of interpersonal behaviour. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: research, theory, and applications* (3rd Ed., pp. 339-354). Guilford Press
- Leary, M. R., & Tangney, J. P. (2012). The self as an organizing construct in the behavioral and social sciences. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 1-20). Guilford Press
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39 (2), 329-358
- Leech, N. L. & Onwuegbuzie, A. J. (2002). *A call for greater use of nonparametric statistics*. Presented at Mid-South Educational Research Association Annual Meeting 2002. Retrieved October 15, 2020 from <https://www.learntechlib.org/p/95744/>

- Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1-10
- Liu, F., & Kong, Y. (2015). zoib: an R package for bayesian inference for beta regression and zero/one inflated beta regression. *The R Journal*, 7 (2), 34-51
- Lockwood, P., & Matthews, J. (2007). The self as a social comparer. In C. Sedikides & P. Spencer (Eds.), *The self* (pp. 95-113). Psychology Press
- Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive and negative role models: Regulatory Focus determines who will best inspire us. *Journal of Personality and Social Psychology*, 83, 854-864
- Loken, E., & Molenaar, P. (2008). Categories or continua? The correspondence between mixture models and factor models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 277-297). Information Age
- London, A. J. (2001). Moral knowledge and the acquisition of virtue in Aristotle's Nicomachean and Eudemian Ethics. *The Review of Metaphysics*, 54, 553-583
- Luco, A. (2014). The definition of morality: threading the needle. *Social Theory and Practice*, 40, 361-387
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin*, 111 (3), 490-504
- MacInnis, C. C., & Hodson, G. (2017). It ain't easy eating greens: evidence of bias toward vegetarians and vegans from both source and target. *Group Processes and Intergroup Relations*, 20 (6), 721-744
- MacIntyre, A. (1981). *After virtue: a study in moral theory*. University of Notre Dame Press

- Mannetti, L., Leder, S., Insalata, L., Pierro, A., Higgins, E. T., & Kruglanski, A. W. (2009). Priming the ant or the grasshopper in people's mind: how regulatory mode affects inter-temporal choices. *European Journal of Social Psychology, 39*, 1120-1125
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57* (3), 519-530
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review, 98* (2), 224-253
- Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: a comparison of alternative models. *Applied Psychological Measurement, 15*, 47-70
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*, 191-225
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110
- Maslow, A. H. (1954). *Motivation and personality*. Harper & Row
- Matthews, G., Schwan, V. L., Campbell, S. E., Saklofske, D. H., & Mohamed, A. A. R. (2005). Personality, self-regulation, and adaptation: a cognitive-social framework. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 171-207). Academic Press
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *Journal of Marketing Research, 45*, 633-644

- McAdams, D. P. (2008). Personal narratives and the life story. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: theory and research* (3rd Ed., pp. 242-262). Guilford Press
- McAdams, D. P., & Pals, J. L. (2006). A new Big Five: fundamental principles for an integrative science of personality. *American Psychologist*, 61 (3), 204-217
- McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52 (5), 509-516
- McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington, E. L., & Van Tongeren, D. R. (2014). Intellectual humility: scale development and theoretical elaborations in the context of religious leadership. *Journal of Psychology and Theology*, 42, 19-30
- McGregor, I., Gailliot, M. T., Vasquez, N. A., & Nash, K. A. (2007). Ideological and personal zeal reactions to threat among people with high self-esteem: motivated promotion focus. *Personality and Social Psychology Bulletin*, 33 (11), 1587-1599
- McGregor, I., Nash, K. A., & Inzlicht, M. (2009). Threat, high self-esteem, and reactive approach-motivation: electroencephalographic evidence. *Journal of Experimental Social Psychology*, 45, 1003-1007
- McIntyre, J. (2012). The idea of the self in the evolution of Hume's account of the passions. *Canadian Journal of Philosophy*, 42 (1), 171-182
- Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology*, 57, 487-503
- Mendelson-Maoz, A. (2007). Ethics and literature introduction. *Philosophia*, 35, 111-116
- Meyers, L. S., & Wong, D. T. (1988). Validation of a new test of locus of control: the internal control index. *Educational and Psychological Measurement*, 48, 753-761

- Minson, J. A., & Monin, B. (2012). Do-gooder derogation: disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, 3 (2) 200-207
- Monin, B. (2007). Holier than me? Threatening social comparison in the moral domain. *International Review of Social Psychology*, 20, 53-68
- Monin, B., & Jordan, A. H. (2009). The dynamic moral self: a social psychological perspective. In D. Narvaez & D. Lapsley (Eds.), *Personality, identity, and character: explorations in moral psychology* (pp. 341-354). Cambridge University Press
- Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: resenting those who do the right thing. *Journal of Personality and Social Psychology*, 95, 76-93
- Monni, A., Olivier, E., Morin, A. J. S., Olivetti Belardialli, M., Mulvihill, K., & Scalas, L. F. (2020). Approach and avoidance in Gray's, Higgins', and Elliot's perspectives: a theoretical comparison and integration of approach-avoidance in motivated behavior. *Personality and Individual Differences*, 166, 110163, 1-10
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: state of the art and future development. *Emotion Review*, 5 (2), 119-124
- Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: a dynamic self-regulatory processing model. *Psychological Inquiry*, 12 (4), 177-196
- Muise, A., Boudreau, G. K., & Rosen, N. O. (2017). Seeking connection versus avoiding disappointment: an experimental manipulation of approach and avoidance sexual goals and the implications for desire and satisfaction. *Journal of Sex Research*, 54 (3), 296-307

- Muthén, L. K., & Muthén, B. O. (2002) How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9 (4), 599-620
- Myers, R. (1990). *Classical and modern regression with applications* (2nd Ed.). Duxbury
- Nelson, M. T. (2019). The is/ought fallacy. In R. Arp, S. Barbone, & m. Bruce (Eds.), *Bad arguments: 100 of the most important fallacies in Western philosophy* (pp. 360-363). Wiley Blackwell
- O'Meara, T. F. (1997). Virtues in the theology of Thomas Aquinas. *Theological Studies*, 58, 254-285
- Oatley, K. (1994). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23, 53-74
- Olson, E. T. (1998). There is no problem of the self. *Journal of Consciousness Studies*, 5 (5-6), 645-657
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press
- Ouschan, L., Boldero, J. M., Kashima, Y., Wakimoto, R., && Kashima, E. S. (2007). The regulatory focus strategies scale (RFSS): A measure of individuals differences in the endorsement of regulatory strategies. *Asian Journal of Social Psychology*, 10, 243-257
- Pagliaro, S., Ellemers, N., Barreto, M., & Di Cesare, C. (2016). Once dishonest, always dishonest? The impact of perceived pervasiveness of moral evaluations of the self on motivation to restore a moral reputation. *Frontiers in Psychology*, 7, 586, 1-12
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411-419

- Papies, E. K., & Aarts, H. (2016). Automatic self-regulation. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: research, theory, and applications* (3rd Ed., pp. 203-222). Guilford Press
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46 (3), 598-609
- Paulhus, D. L. (1994). *Balanced inventory of desirable responding: reference manual for BIDR version 6*. Unpublished manuscript, University of British Columbia, Vancouver, Canada
- Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: a mixed blessing? *Journal of Personality and Social Psychology*, 74 (5), 1197-1208
- Paulhus, D. L., Fridhandler, B., & Hayes, S. (1997). Psychological defense: contemporary theory and research. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of Personality Psychology* (pp. 543-579). Academic Press
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). Guilford Press
- Pearl, J. (2009). *Causality: models, reasoning, and inference* (2nd Ed). Cambridge University Press
- Pearl, J. (2012). The causal foundations of structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 68-91). Guilford Press
- Pearl, J., & MacKenzie, D. (2018). *The book of why: the new science of cause and effect*. Allen Lane

- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: a handbook and classification*. Oxford University Press
- Pfattheicher, S., & Keller, J. (2013). Vigilant self-regulation and costly punishment in public goods situations. *European Journal of Personality, 27*, 346-354
- Phillips, N. (2017). *yarr: a companion to the e-book 'YaRrr!: the pirate's guide to R'*. R package version 0.1.5. <https://CRAN.R-project.org/package=yarr>
- Piazza, J., Sousa, P., Rottman, J., & Syropoulos, S. (2019). Which appraisals are foundational to moral judgment? Harm, injustice, and beyond. *Social Psychological and Personality Science, 10* (7) 903-913
- Pickering, A. D., & Corr, P. J. (2008). J. A. Gray's reinforcement sensitivity theory (RST) of personality. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment* (pp. 239-256). Sage
- Pickering, A. D., & Smillie, L. D. (2008). The behavioural activation system: challenges and opportunities. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 120-154). Cambridge University Press
- Piper, W. T., Saslow, R. L., & Saturn, S. R. (2015). Autonomic and prefrontal events during moral elevation. *Biological Psychology, 108*, 51-55
- Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: how the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: exploring the causes of good and evil* (pp. 91-108). American Psychological Association

- Pleasant, A., & Barclay, P. (2018). Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. *Psychological Science*, 29, 868-876
- Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89 (4), 344-350
- Polman, E. (2012). Effects of self-other decision making on regulatory focus and choice overload. *Journal of Personality and Social Psychology*, 102 (5), 980-993
- Presniak, M. D., Olson, T. R., MacGregor, M. W. (2010). The role of defense mechanisms in borderline and antisocial personalities. *Journal of Personality Assessment*, 92 (2), 137-145
- Pyszczynski, T., Greenberg, J., & Arndt, J. (2012). Freedom versus fear revisited: an integrative analysis of the dynamics of the defense and growth of the self. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd Ed., pp. 378-404). Guilford Press
- Quigley, K. S., Barrett, L. F., & Weinstein, S. (2002). Cardiovascular patterns associated with threat and challenge appraisals: A within-subjects analysis. *Psychophysiology*, 39, 292-302
- R Core Team (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org> [computer software]
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, 118 (1), 57-75
- Rawls, J. (1975). The independence of moral theory. *Proceedings and Addresses of the American Philosophical Association*, 48, 5-22

- Reeder, G. D. (2009). Mindreading: judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, 20, 1-18
- Reeder, G. D., Pryor, J. B., Wohl, M. J. A., & Griswell, M. L. (2005). On attributing negative motives to others who disagree with our opinions. *Personality and Social Psychology Bulletin*, 31 (11), 1498-1510
- Revelle, W. (2019). *psych: procedures for personality and psychological research*. Northwestern University, Evanston, IL, USA. <https://cran.r-project.org/package=psych>
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*, 38 (1), 119-125
- Rihmer, Z., Akiskal, K. K., Rihmer, A., & Akiskal, H. S. (2010). Current research on affective temperaments. *Current Opinion in Psychiatry*, 23, 12-18
- Robbins, B. D. (2008). What is the good life? Positive psychology and the renaissance of humanistic psychology. *The Humanistic Psychologist*, 36, 96-112
- Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, 80, 340-352
- Robins, R. W., John, O. P., Caspi, A., Moffitt, T. E., & Stouthamer-Loeber, M. (1996). Resilient, overcontrolled, and undercontrolled boys: three replicable personality types. *Journal of Personality and Social Psychology*, 70, 157-171
- Rogers, C. R. (1956). What it means to become a person. In C. E. Moustakas (Ed.), *The self: explorations in personal growth* (pp. 195-211). Harper
- Rohmann, E., Neumann, E., Herner, M. J., & Bierhoff, H. W. (2012). Grandiose and vulnerable narcissism: self-construal, attachment, and love in romantic relationships. *European Psychologist*, 17 (4), 279-290

- Rohrer, J. (2018). Thinking clearly about correlations and causation: graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1 (1), 27-42
- Rosenberg, J. M., Beymer, P. N., Anderson, D. J., & Schmidt, J. A. (2018). tidyLPA: an R package to easily carry out Latent Profile Analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, 3 (30), 978.
<https://doi.org/10.21105/joss.00978>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. University Press
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1-36. <https://www.jstatsoft.org/article/view/v048i02>
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80 (1), 1-28
- Rotter, J. B. (1975). Some problems and misconceptions related to the construct of internal versus external control of reinforcement. *Journal of Consulting and Clinical Psychology*, 43 (1), 56-67
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5 (4), 296-320
- RStudio Team (2020). *RStudio: integrated development for R*. RStudio, PBC, Boston, MA, USA. <http://www.rstudio.com> [computer software]
- Ruch, W., Gander, F., Wagner, L., & Giuliani, F. (2019). The structure of character: on the relationships between character strengths and virtues. *Journal of Positive Psychology*. <https://doi.org/10.1080/17439760.2019.1689418>
- Runes, D. D. (1983). *The dictionary of philosophy*. Philosophical Library
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39 (6), 1161-1178

- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74* (1), 31-57
- Ryff, C. D. (1991). Possible selves in adulthood and aging: a tale of shifting horizons. *Psychology and Aging, 6*, 286-295
- Saariluoma, P. (2020). Hume's guillotine resolved. In M. Rauterberg (Ed.), *Culture and computing. HCII 2020. Lecture Notes in Computer Science* (Vol. 12215). Springer, Cham. https://doi-org.ezproxy.lancs.ac.uk/10.1007/978-3-030-50267-6_10
- Salerno, L., Ingoglia, S., Lo Coco, G. (2017). Competing factor structures of the Rosenberg self-esteem scale (RSES). *Personality and Individual Differences, 113*, 13-19
- Sarstedt, M., Henseler, J., & Ringle, C. M. (2011). Multigroup analysis in partial least squares (PLS) path modeling: alternative methods and empirical results. *Measurement and Research Methods in International Marketing: Advances in International Marketing, 22*, 195-218
- Savalei, V. (2018). On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. *Multivariate Behavioral Research, 53*, 419-429
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44* (4), 695-729
- Scherer, K. R. (2009). The dynamic architecture of emotion: evidence for the component process model. *Cognition and Emotion, 23* (7), 1307-1351
- Scherer, K. R. (2019). Studying appraisal-driven emotion processes: taking stock and moving to the future. *Cognition and Emotion, 33* (1), 31-40
- Schnall, S., & Roper, J. (2012). Elevation puts moral values into action. *Social Psychological and Personality Science, 3* (3), 373-378

- Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruistic behavior. *Psychological Science, 21* (3), 315-320
- Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8* (4), 379-386
- Scholer, A. A., & Higgins, E. T. (2012). Too much of a good thing? Trade-offs in promotion and prevention focus. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 65-84). Oxford University Press
- Scholer, A. A., Ozaki, Y., & Higgins, E. T. (2014). Inflating and deflating the self: sustaining motivational concerns through self-evaluation. *Journal of Experimental Social Psychology, 51*, 60-73
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*, 609-612
- Schönbrodt, F. D., & Perugini, M. (2018). Corrigendum to "At what sample size do correlations stabilize?" [J. Res. Pers. 47 (2013) 609–612]. *Journal of Research in Personality, 74*, 194
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99*, 323-337
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd Ed.). Psychology Press
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8* (1), 205-233

- Sedikides, C. (2009). On self-protection and self-enhancement regulation: the role of self-improvement and social norms. In J. P. Forgas, R. F. Baumeister, & D. M. Tice (Eds.), *Psychology of Self-Regulation: cognitive, affective and motivational processes* (pp. 73-92). Psychology Press
- Sedikides, C. (2012). Self-protection. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd Ed., pp. 327-353). Guilford Press
- Sedikides, C., & Alicke, M. D. (2012). Self-enhancement and self-protection motives. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 303-322). Oxford University Press
- Sedikides, C., & Hepper, E. G. D. (2009). Self-improvement. *Social and Personality Psychology Compass*, 3 (6), 899-917
- Sedikides, C., & Luke, M. (2007). On when self-enhancement and self-criticism function adaptively and maladaptively. In E. C. Chang (Ed.), *Self-criticism and self-enhancement: theory, research and clinical implications* (pp. 181-198). American Psychological Association
- Seery, M. D. (2013). The biopsychosocial model of challenge and threat: using the heart to measure the mind. *Social and Personality Psychology Compass*, 7 (9), 637-653
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *American Psychologist*, 55 (1), 5-14
- Shah, J. Y., & Kruglanski, A. W. (2000). Aspects of goal networks: implications for self-regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 85-110). Academic Press
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: self-affirmation theory. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, Vol. 38 (pp. 183-242). Guilford Press

- Sherman, D. K., & Hartson, K. A. (2011). Reconciling self-protection with self-improvement: self-affirmation theory. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 128-151). Guilford Press
- Silvers, J. A., & Haidt, J. (2008). Moral elevation can induce nursing. *Emotion*, 8 (2), 291-295
- Simms, L. J., & Watson, D. (2007). The construct validation approach to personality scale construction. In R. W. Robins, C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 240-258). Guilford Press
- Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, 20 (5), 580-591
- Smillie, L. D., Loxton, N. J., & Avery, R. E. (2011). Reinforcement sensitivity theory: research, applications, and future. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 101-131). Wiley-Blackwell
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48 (4), 813-838
- Smith, R. H., & Insko, C. A. (1987). Social comparison choice during ability evaluation. *Personality and Social Psychology Bulletin*, 13, 111-122
- Smith, R. H., Parrott, W. G., Ozer, D., & Moniz, A. (1994). Subjective injustice and inferiority as predictors of hostile and depressive feelings in envy. *Personality and Social Psychology Bulletin*, 20 (6), 705-711
- Solomon, R. C., Martin, C. W., & Vaught, W. (2008). *Morality and the good life: an introduction to ethics through classical sources* (5th Ed.). McGraw-Hill
- Steele, C. M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Advances in Experimental Social Psychology*, 21, 261-302

- Sterelny, K., & Fraser, B. (2016). Evolution and moral realism. *British Journal for the Philosophy of Science*, 68 (4), 981-1006
- Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, 95 (2), 334-344
- Stevens, S. S. (1946). On the theory of scales of measurement, *Science*, 103, 677-680
- Strachman, A., & Gable, S. L. (2006). What you want (and do not want) affects what you see (and do not see): avoidance social goals and social events. *Personality and Social Psychology Bulletin*, 32 (11), 1446-1458
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*, 28, 191-193
- Summerville, A., & Roese, N. J. (2008). Self-report measures of individual differences in regulatory focus: a cautionary note. *Journal of Research in Personality*, 42 (1), 247-254
- Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral? *Psychological Science*, 31 (3), 243-257
- Swann, W. B., & Read, S. J. (1981). Self-verification processes: how we sustain our self-conceptions. *Journal of Experimental Social Psychology*, 17, 351-372
- Tasimi, A., Dominguez, A., & Wynn, K. (2015). Do-gooder derogation in children: the social costs of generosity. *Frontiers in Psychology*, 6 (1036).
<https://doi.org/10.3389/fpsyg.2015.01036>
- Taylor, S. E., Neter, E., & Wayman, H. A. (1995). Self-evaluation processes. *Personality and Social Psychology Bulletin*, 21 (12), 1278-1287
- Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: an assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, 77, 184-197

- Thomson, A. L., & Siegel, J. T. (2013). A moral act, elevation, and prosocial behavior: moderators of morality. *Journal of Positive Psychology, 8* (1), 50-64
- Tice, D. M. (1993). The social motivations of people with low self-esteem. In R. F. Baumeister (Ed.), *Self-esteem: the puzzle of low self-regard* (pp. 37-53). Plenum Press
- Tice, D. M., & Masicampo, E. J. (2008). Approach and avoidance motivations in the self-concept and self-esteem. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 505-519). Psychology Press
- Tomás, J. M., Hontangas, P. M., & Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: the effects of number of indicators and correlation among methods. *Multivariate Behavioral Research, 35*, 469-499
- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64* (1), 231-255
- Tompkins, J. P. (1980). The reader in history: the changing shape of literary response. In J. P. Tompkins (Ed.), *Reader-response criticism: from formalism to post-structuralism* (pp. 201-232). Johns Hopkins University Press
- Torrubia, R., Ávila, C., Moltó, J. & Caseras, X. (2001). The sensitivity to punishment and sensitivity to reward questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences, 31*, 837-862
- Trapnell, P. D., & Paulhus, D. L. (2012). Agentic and communal values: their scope and measurement. *Journal of Personality Assessment, 94* (1), 39-52
- Tsai, W., Chiang, J. J., & Lau, A. S. (2016). The effects of self-enhancement and self-improvement on recovery from stress differ across cultural groups. *Social Psychological and Personality Science, 7* (1), 21-28

- Tsakiris, M., & Critchley, H. D. (2016). Interoception beyond homeostasis: affect, cognition and mental health. *Philosophical Transactions of the Royal Society B*, 371, 1-6
- Twain, M. (1999). *Pudd'nhead Wilson*. Dover (original work published 1894)
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10 (1), 72-81
- Uhlmann, E. L., Zhu, L. L., & Diermeier, D. (2014). When actions speak volumes: the role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology*, 44, 23-29
- Vaillant, G. E. (1994). Ego mechanisms of defense and personality psychopathology. *Journal of Abnormal Psychology*, 103 (1), 44-50
- van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., ..., Wagenmakers, E. (2020). The JASP Guidelines for Conducting and Reporting a Bayesian Analysis. *Psychonomic Bulletin and Review*.
<https://doi.org/10.3758/s13423-020-01798-5>
- Van Horn, E. K. (2017). *The little book of Buddhist virtue*. CreateSpace Independent Publishing Platform
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3 (1), 4-70
- Van Laer, T., De Ruyter, K., Visconti, L. M., & Wetzels, M. (2013). The extended transportation-imagery model: a meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research*, 40 (5), 797-817

- Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 47 (1), 65-72
- Vickers, R. R., Conway, T. L., & Haight, M. A. (1983). Associations between Levenson's dimensions of locus of control and measures of coping and defense mechanisms. *Psychological Reports*, 52, 323-333
- von Wright, G. H. (1963). *Varieties of goodness*. Routledge and Kegan Paul
- Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, 93 (5), 845-860
- Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: just, brave, and caring. *Journal of Personality and Social Psychology*, 86 (4), 629-647
- Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods*, 5 (2), 344-346
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54 (6), 1063-1070
- Wheeler, L. (1966). Motivation as a determinant of upward social comparison. *Journal of Experimental Social Psychology*, 2 (Suppl. 1), 27-31
- Wheeler, L. (1991). A brief history of social comparison theory. In J. Suls & T. A. Wills (Eds.), *Social comparison: contemporary theory and research* (pp. 3-21). Lawrence Erlbaum Associates
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer
- Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: the interpersonal domain. *Journal of Personality and Social Psychology*, 37 (3), 395-412
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti & W. M.

- Grove (Eds.), *Thinking clearly about psychology: essays in honor of Paul E. Meehl* (Vol. 2, pp. 89-113). University of Minnesota Press
- Williams, M. N. (2020). Levels of measurement and statistical analyses. *PsyArXiv*.
<https://doi.org/10.31234/osf.io/c5278>
- Wills, A. T. (1991). Similarity and self-esteem in downward comparison. In J. Suls, & A. T. Wills (Eds.), *Social comparison: contemporary theory and research* (pp. 51-78). Lawrence Erlbaum Associates
- Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology*, 61 (4), 590-597
- Wispé, L. G. (1972). Positive forms of social behavior: an overview. *Journal of Social Issues*, 28 (3), 1-19
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73 (6), 913-934
- Wood, J. V. (1996). What is social comparison and how should we study it? *Personality and Social Psychology Bulletin*, 22 (5), 520-537
- Wood, J. V., & Taylor, K. L. (1991). Serving self-relevant goals through social comparison. In J. Suls & T. A. Wills (Eds.), *Social comparison: contemporary theory and research* (pp. 23-49). Lawrence Erlbaum Associates
- Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ*, 3, e1058.
<https://doi.org/10.7717/peerj.1058>
- Zbikowski, J., & Collins, J. (1994). *Literature as the laboratory of the moral life: building moral communities through literature study*. Paper presented at the 20th Annual

Conference of the Association for Moral Education (Banff, Alberta, Canada,
November 11, 1994), ERIC

Zimmerman, B. J. (2005). Attaining self-regulation: a social cognitive perspective. In M.
Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-
39). Academic Press.

Appendices

1. The twelve Study 1 vignettes:

Cory

In April 2012 Cory Booker, the Mayor of Newark, New Jersey, was returning home after taping a television interview with a local news programme, when he saw flames and smoke pouring from the second floor of the small apartment building next to his home.

Seeing no residents outside, without thinking twice he rushed into the burning building after fighting off the members of his security staff who tried to restrain him.

He heard a woman's cries from the second floor, managed to reach her in a back bedroom through the flames, and whipped her out of the bed. While his security officers helped first-floor residents to safety, Cory and the woman made their way out of the building as quickly as they could and collapsed when they reached the street.

They were both taken to a nearby hospital: Cory was treated for smoke inhalation and second-degree burns on his hand and discharged soon after, while the woman recovered in a few days.



Wesley

In January 2007 Wesley Autrey was waiting on a train in the New York City subway with his two daughters, when suddenly a man standing beside him started having seizures and fell on to the train tracks. After considering what he could do, Wesley jumped down to lift the man off the tracks, but realised there was no time as the train was coming; the only chance was to lie on top of the man between the rails, covering him with his body and pressing him down until the train was gone.

The clearance was minimal, but the train rolled right over them as they lay still. At the end both emerged unscathed; the man was taken to hospital and made a quick recovery.

Wesley was honoured with the Carnegie Medal for Heroism and a visit to the White House. He also appeared on the David Letterman's Late Night Show. Two years after the event, the pair met again with kids and parents in a restaurant in New York, breaking bread as one happy family.



Arnaud

In March 2018 Arnaud Beltrame, a French policeman, was one of the first officers to respond to a terrorist attack in South-West France. An armed Islamic terrorist took hostages among the many who were shopping at a local supermarket.

Arnaud voluntarily swapped places with a female hostage. During the two-hour standoff, he managed to leave his mobile phone on a table with an open line, so that police outside could hear what was going on inside the supermarket.

When the extremist gunman opened fire, police stormed the building and eventually killed the terrorist. Three hostages lost their lives and more than a dozen others were wounded in the attack. Arnaud was shot four times; he was rushed to hospital, but died of his injuries. He was 44 and was supposed to get married in June.

Tributes came from all over France. The Interior Minister said that his country will never forget Arnaud's heroism, bravery and sacrifice.



Francia

Francia Raisa is an American actress and has been pop star Selena Gomez's close friend for many years. In 2017 Francia learnt that Selena had been suffering from lupus and now needed a kidney transplant, as she could experience organ failure at any moment. Francia was upset to discover that the waiting list for the transplant was 7 to 10 years long.

Francia promised Selena that she would get tested to see if she could donate a kidney. The results showed that she was a match. Despite her parents' and her own worries about the possible consequences to her health, Francia agreed to undergo surgery, and wrote a will in case she didn't pull through the procedure.

The surgery was a success and they are now both healthy. Selena said Francia saved her life by giving her the ultimate gift and sacrifice, but Francia claims that actually the experience transformed her life and made her and Selena family. They both cooperate to raise awareness on lupus and support a charity that funds research on this little known autoimmune disease.



Matthew

In April 2017 Matthew Rees was running the London Marathon. After the last turn, he saw another runner who was visibly exhausted, his legs crumbling before him. Matthew was worried that the runner might bang his head against the ground, and decided to compromise his own finish time and a potential award to help him.

Matthew saved the runner from falling by wrapping his arm around him. He then pointed the runner to the finish line to encourage him along, and they eventually crossed the line together.

Matthew and the man became friends and after a year they met again in London to train together for the 2018 Marathon. They crossed the finish line just a few seconds apart, and Matthew was the first to give his friend a hug and celebrate their achievement.



Sarah

In December 2017 American comedian Sarah Silverman tweeted a comment about an article. One of her followers responded by calling her a c***.

Instead of retaliating with equal aggression or blocking the user, Sarah took the time to read the Twitter feed of the man who had harassed her. She learnt that he had been dealing with back pain, and encouraged him to seek help and to see love in himself.

In response to Sarah's disarming gesture, the man apologised, explaining that he had been hardened by long-standing back injuries, which he couldn't treat because he didn't have the money.

Sarah not only accepted the apology, but mobilised her 12 million Twitter followers to find specialists who could treat his back. Within minutes, she had responses from clinics willing to help and one week later the man got an MRI which showed five herniated discs. Sarah offered to pay his back treatment in full, and a crowdfunding campaign also raised \$1,774. The legions of Sarah's fans praised her compassion and kindness, acknowledging she took the high road that very few would have taken.



Joey

In January 2016 Joey Resto, a paralegal from Brooklyn, saw on the train a shirtless man who was shivering with cold in the freezing temperatures of the New York winter.

All the other passengers were crowded at the other end of the car, as the man didn't smell nice. But Joey had a different reaction: he approached the man, took off his shirt and offered it to him. The man looked at him incredulously without moving, so Joey gently helped him put it on. Joey also noticed that the man had abrasions on his head, so he went back to his seat, took his hat and placed it on the man's head.

Joey later said he wanted to offer the man comfort with more warm clothes and a hot meal, but the man wanted to rest, so Joey returned to his seat, while the man curled up in a corner.

A passenger filmed the scene and put the video up on Facebook. The video quickly went viral, and soon after the news was in the media.



Markus

After obtaining a degree in Engineering in 2009, Markus Hachmöller got a good job in his home city in Germany. As he gained more experience, he moved on to higher positions within different companies. He worked hard and achieved a respectable career; however, he questioned himself about what he really wanted to do in his life. More than just a burnout or midlife crisis, Markus claimed that he wanted to make a difference to others.

In 2016 Markus felt like it was the right time to do something different and leave his comfort zone: he quit his job and left Germany to travel to South America with the intention of teaching young children. He joined a charity in Peru that works with local communities to alleviate poverty.

Markus has volunteered as a teacher of English since, and by helping children in need he has found what he was sorely missing. His quest for meaning led him to do something that truly impacts the life of other people in a way that he could not have achieved in his previous life.



Ruxandra

Ruxandra Micu is a freelance graphic and web designer who in 2013 became vegetarian.

When she was younger, she had a friend who became vegetarian, but at that time she thought she would never give up meat. However, years later she started reconsidering her position. She went through a difficult time in her life and was depressed. She found refuge in reading about nutrition, one of her interests, and realised how bad eating habits she had developed.

When she watched documentaries about the food industry and the treatment of animals who end up on our plate, that was the last straw: she realised that she had to do something to stop harming them.

Ruxandra is now vegetarian and loves to prepare her food. She is very active blogging to raise animal cruelty awareness and show other people how easy it is to adopt a vegetarian lifestyle that is healthy and cruelty-free.



Alvaro

In 1984 Alvaro Múnera was 18 years old and was training to become a professional bullfighter in Colombia. He was still young and unsure about his future career, but his manager recognised his talent and sent him abroad to make further experience. During a fight in Spain, the course of his life changed forever: the bull gored him in the leg and tossed him across the ring, fracturing his 5th cervical vertebra and leaving him permanently paraplegic. In addition, his best friend was gored to death a few months later.

The incidents elicited a thorough revision of Alvaro's moral code. He remembered the times he saw the bulls cry in the arena before being killed, and felt deeply guilty for having caused the death of 150 innocent animals during his brief career.

Alvaro is now a council member in his home town in Colombia and promotes anti-bullfighting campaigns. He has attracted the sympathy of animal-rights defenders all over the world, who praise his brave commitment in a country that is still largely pro-bullfighting.



Sunita

Sunita Narain is an Indian environmental activist. When she was a high school student, she witnessed the devastation of the Delhi's Ridge Forest caused by logging companies. She joined a pro-environment organisation, and campaigned to protect the forests and restore barren lands, supporting local communities.

Sunita has fought many battles since. In recent years she has raised the bar even further, claiming that fighting global warming is not just an environmental endeavour, but also a matter of justice towards the developing countries.

The economic growth of India, for example, has allowed to improve the living standards of millions of people. However, the country has now become the 3rd largest emitter of greenhouse gases in the world, heavily relying on fossil fuels that are responsible for air pollution and raising temperatures. She contends that the rich Western countries, historically the largest contributors to carbon emissions, must now provide substantial financial aid to the developing countries, so they can support environment-friendly growth policies otherwise unsustainable.

With her passionate and controversial activism, Sunita has been included on Time's list of the world's 100 most influential people.



Nicholas

Nicholas Opiyo is a lawyer in Uganda. Since the passing of the 2014 Anti-Homosexuality Act, which made homosexuality illegal in the country, his clients have been reporting a wave of acts of hostility and violence towards the gay and lesbian communities.

Nicholas decided to publicly challenge the Act on the basis of discrimination. For this reason, he became the target of threats and verbal aggression, in a country where religious leaders of all major faiths consider homosexuality immoral, often stirring up revolt among the population.

Nicholas led the legal challenge and after 2 years of proceedings in 2016 the Constitutional Court ruled the Anti-Homosexuality Act illegal. Despite this civil rights victory, Nicholas's fight for justice is not over: a risk exists that a new anti-homosexuality law is passed again and his legal services are still in demand among many Ugandans who are discriminated on the basis of their sexual orientation.



2. Sources of information for the twelve Study 1 vignettes:

Cory:

<https://www.politico.com/story/2012/04/mayor-booker-saves-woman-in-fire-075097>
https://www.nj.com/news/2012/04/newark_mayor_cory_booker_taken.html
<https://www.thedailybeast.com/cory-booker-rescues-a-freezing-dog-and-9-other-things-he-has-saved>

Wesley:

<http://beta.nydailynews.com/news/subway-hero-wesley-autrey-holiday-reunion-man-saved-article-1.420219>
<https://newyork.cbslocal.com/2012/02/21/5-years-later-new-york-city-subway-hero-wesley-autrey-is-still-the-man/>

Arnaud:

<https://www.nbcnews.com/news/world/french-police-officer-who-traded-places-hostage-during-terror-attack-n859696>
<https://www.nytimes.com/2018/03/23/world/europe/france-trebes-attack.html>

Francia:

<https://www.today.com/health/selena-gomez-francia-ra-sa-share-story-behind-dramatic-kidney-t118147>
<http://www.foxnews.com/entertainment/2018/03/15/selena-gomez-kidney-donor-francia-raisa-reveals-went-through-depression-after-transplant.html>

Matthew:

<https://www.walesonline.co.uk/news/wales-news/london-marathon-hero-runner-man-14561030>
<https://www.theguardian.com/global/shortcuts/2017/apr/24/london-marathon-man-runner-finish-matthew-rees-david-wyeth>

Sarah:

<https://qz.com/quartz/1173220/sarah-silvermans-powerful-response-to-a-sexist-troll/>
<https://mashable.com/2018/01/06/sarah-silverman-troll/?europa=true#RmOulyREiaqZ>

Joey:

<https://abcnews.go.com/US/good-samaritan-shirt-off-back-homeless-man-nyc/story?id=36223452>
<http://www.dailymail.co.uk/news/article-3392851/Heartwarming-moment-passenger-gives-shirt-homeless-man-New-York-subway-helps-caught-camera.html>

Markus:

<http://www.hoopperu.org/2016/quit-job-go-volunteering/>

Ruxandra:

<https://gourmandelle.com/vegetarian-story-avid-meat-eater-happy-vegetarian/>
<https://gourmandelle.com/about/>

Alvaro:

<https://longhaultrekkers.com/street-tails-alvaro-munera-and-esperanza/>
https://www.vice.com/en_us/article/wdz8vz/bullfighter-152-v15n10

Sunita:

<https://www.smithsonianmag.com/science-nature/indian-activists-path-climate-justice-180964866/>

Nicholas:

<https://www.vox.com/2014/8/21/6005841/meet-the-attorney-who-got-ugandas-antigay-law-overturned>

3. The two Study 4 vignettes:

Experimental condition: Francia Kidney

Francia

Francia Raísa is an American actress and has been pop star Selena Gomez's close friend for many years. In 2017 Francia learnt that Selena had been suffering from lupus and now needed a kidney transplant, as she could experience organ failure at any moment. Francia was upset to discover that the waiting list for the transplant was 7 to 10 years long.

Francia promised Selena that she would get tested to see if she could donate a kidney. The results showed that she was a match. Despite her parents' and her own worries about the possible consequences to her health, Francia agreed to undergo surgery, and wrote a will in case she didn't pull through the procedure.

The surgery was a success and they are now both healthy. Selena said Francia saved her life by giving her the ultimate gift and sacrifice, but Francia claims that actually the experience transformed her life and made her and Selena family. They both cooperate to raise awareness on lupus and support a charity that funds research on this little known autoimmune disease.



Control Condition: Francia Latte

Francia

Francia Raísa is an American actress and has been pop star Selena Gomez's close friend for many years. They both live in the same city and enjoy meeting each other for a coffee and shopping.

Selena promised Francia that they would meet up for coffee the next time they were both free. The following week their schedules aligned, and they met up at a local coffee shop for a latte.

While ordering, Selena realized that she had forgotten to put her wallet in her handbag and so she did not have money to pay. Francia had enough money for both of them and so did not hesitate to help out her friend, paying the bill with a smile. Selena was grateful and they spent an enjoyable hour chatting about their current jobs and relationships.

Before parting, the two agreed to meet up again soon, but next time they would go shopping.

