**The Psychological and Social Factors That Influence Moral Transference.**

Katherine O’Lone

Thesis submitted in fulfilment of the degree of

Doctor of Philosophy

Royal Holloway, University of London

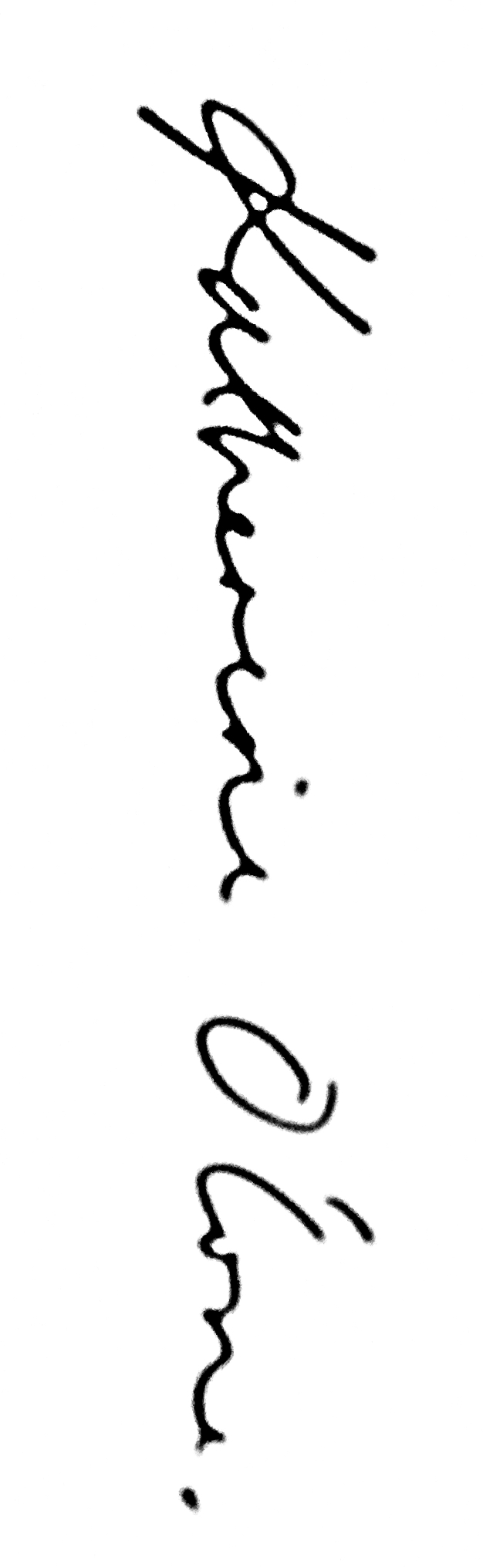
Department of Psychology

July 2018

# Declaration of Authorship

I, Katherine O’Lone, hereby declare that this work was carried out in accordance with the Regulations of the University of London. I declare that this submission is my own work, and to the best of my knowledge does not represent the work of others, published or unpublished, except where duly acknowledged in the text. No part of this thesis has been submitted for a higher degree at another university or institution.

Signed:



Date: 27/07/2018

# Abstract

Research has revealed that our moral dispositions and preferences are dynamic, not static. Our previous decisions, as well as the decisions of others like us, influence our subsequent moral activities. For example, the performance or recollection of past transgressions often compels us to compensate by engaging in good deeds, i.e., moral “cleansing”. On the other hand, after performing or recalling our own good deeds – or those of in-group members – we may feel “licensed” to act immorally. These two separate but complimentary mechanisms (i.e., moral *licensing* and *cleansing*) form a model of moral behaviour termed moral compensation*.* The Moral compensation model allows us to entertain a “debits and credits” conceptualisation of moral behaviour but it cannot adequately account for the flexibility of our day-to-day moral behaviour. The purpose of this thesis is to investigate instances of *moral transference*, i.e., episodes where past (im)moral behaviour is displaced into the present, where past in-group behaviour is displaced onto us, and where moral imperatives are displaced onto other (e.g., supernatural) agents.

The methods used in this research are varied and include self-report, memory recall, implicit association testing and perceptual judgement. Using a novel confession paradigm, Chapter 3 explores whether the performance of a false confession can alleviate guilt for an unrelated past transgression. Chapter 4 investigates whether the recollection of a past unethical behaviour can trigger increased sensitivity to cues of social surveillance. Chapter 5 probes the effect of surveillance cues on confessional behaviour and seeks to discover whether such cues trigger voluntary confession as a means of mitigating possible punishment. Chapter 6 considers the role of forgiving God concepts on people’s endorsement of state-sanctioned punishment, in particular whether such concepts restrict the tendency for individuals to “outsource” the responsibility of punishment to God. Chapter 7 explores whether information about ingroup (im)morality triggers vicarious moral compensation effects.

This thesis concludes that – contra the moral compensation model – our everyday moral behaviours are influenced by more than a process of moral self-regulation. I further conclude that religious affiliation and beliefs are important factors influencing the mechanism of moral transference.

# List of Publications

**Chapter 6 accepted for publication:**

**O’Lone, K.,** & McKay, R. T. (2018). Divine forgiveness and moral support for state-sanctioned punishment*. Journal for the Cognitive Science of Religion*, (in press).

I co-authored two other articles during my PhD, which reflect my thinking on certain issues and are cited throughout this thesis. However, these are not included as chapters as I was not first author:

McKay, R. T., Jong, J., & **O’Lone, K.** (2018). Idolizing the indexical: commentary on Van Leeuwen and van Elk. [commentary] *Religion, Brain and Behavior,* https://doi.org/[10.1080/2153599X.2018.1453533](https://doi.org/10.1080/2153599X.2018.1453533)

McKay, R. T., Ross, R. M., **O’Lone, K.,** & Efferson, C. (2017). The error of God revisited. [commentary] *Religion, Brain and Behavior, 8*(3), 306-310.

# Acknowledgements

First and foremost, I would like to thank my supervisor, Ryan McKay, for his support throughout. I am forever grateful for the tremendous encouragement and advice he has given me. I have no doubt that without his guidance and incredible insight this thesis would never have come to fruition. As well as for his academic support, I would also like to thank him for the emotional support he gave me in the aftermath of my brother’s death. I will remember his kindness and wisdom for the rest of my life.

I would also like to express my eternal gratitude to Cumberland Lodge. Without their generous scholarship I would never have been able to undertake the PhD. As well as giving me invaluable financial support they opened up their doors and within that house I spent some of the best three years of my life. In particular, I would like to thank Dr Owen Gower and Canon Dr Edmund Newell, who not only gave me such a wonderful opportunity but also unwavering support and advice.

There are several individuals to whom I owe particular gratitude. First Dr Olivier Morin and Dr Eyal Pe’er for their invaluable suggestions and advice, as well as their generosity in sharing materials. Second, Ali Bajwa QC, who was so generous with his time and allowed me to tap into his astonishing legal brain. My research is the richer because of them.

And my wonderful friends, who I value so much that it is impossible to even attempt to quantify; thank you. You never doubted me. You encouraged and supported me throughout. And of course, Andy, thank you forever, you have no idea of the part you played.

Above all, I want to thank my family. There’s no way I could ever put into words how much you’ve done for me, so I won’t even try. Instead I’ll say this: I love you more than words can say. And lastly, I want to thank my brother Dan, who never got to see his sister complete a PhD. He would no doubt delight in telling me “just because you have a PhD doesn’t mean you’re not an idiot”. He was, of course, completely correct.

# Contents

Declaration of Authorship 2

Abstract 3

List of Publications 5

Acknowledgements 6

Contents 8

List of Tables 20

List of Figures 22

I. Overview of Thesis 23

1. Introduction. 23

**1.1. Labile Morality: More Than Self-Regulation.** 24

**1.2. Religious Conceptions of Moral Transference.** 25

**1.3. Psychological Equivalents.** 27

2. Methodological Overview. 29

**2.1. The Use of Online Crowdsourcing Services.** 29

**2.2. Participant Inattention on MTurk.** 30

**2.3. Participant Non-naïveté and Dishonesty.** 31

**2.4. A Word about Sample Sizes.** 32

3. A Field in Crisis? 33

**3.1. Meta-analysis.** 34

**3.2. The Robustness of Religious Priming.** 35

**3.3. The Use of Pre-registration.** 38

**3.4. Replication.** 39

4. Structure of this Thesis: A Roadmap. 42

II. A Theoretical Framework of Moral Transference. 44

1. Introduction. 44

**1.1. A Dynamic Theory of Moral Behaviour.** 45

**1.2. Moral Licensing.** 46

**1.3. Meta-Analytic Overview of Moral Licensing.** 48

1.3.1. Pre-registered direct replications. 51

**1.4. Moral Cleansing.** 52

**1.5. Meta-Analytic Overview of Moral Cleansing.** 54

**1.6. Limitations of the Moral Compensation Model.** 55

1.6.1. Insufficient evidence to support a balancing mechanism. 56

1.6.2. Moral Consistency. 56

2. Reputation and Surveillance. 58

**2.1. The Effects of Being Watched.** 58

**2.2. Effects of Artificial Surveillance Cues On Behaviour.** 60

**2.3. Meta-Analytical Overview of Artificial Surveillance Cues Effect.** 60

3. Agency Detection Capacities. 63

**3.1. Illusory Agency Perception.** 65

**3.2. Boundary Conditions of the HADD.** 66

**3.3. HADD and the Formation of Religious Beliefs.** 68

**3.4. An Alternative to the HADD: The Interactive Religious Experience Model.** 70

3.4.1. Fleshing out the formation of personal (religious) beliefs. 71

3.4.2. The Distinction between General and Personal Religious Beliefs. 73

3.4.3. The role of transcendent experiences in religious belief. 74

4. Religion and Prosociality. 76

**4.1. Religion and Antisocial Effects.** 78

**4.2. Literal Prosocial Behaviour: Punishment as a Case in Point.** 79

5. Supernatural Surveillance, Punishment and Morality. 80

**5.1. Watchful Gods and the Fear of Punishment.** 82

**5.2. The Supernatural Punishment Hypothesis.** 82

5.2.1. Evidence for the Supernatural Punishment Hypothesis. 83

6. Powerful Gods and Individual Moral Responsibility. 85

**6.1. Divine Moral Licensing.** 87

**6.2. Religiosity and Punishment.** 88

7. Group Affiliation and Moral Transference. 90

**7.1. The Importance of Morality to Ingroup Identification.** 91

**7.2. Vicarious Self-Perception.** 91

**7.3. Vicarious Moral Licensing.** 92

**7.4. Vicarious Moral Cleansing.** 95

**7.5. Collective Guilt.** 96

7.5.1. Group Identification and Collective Guilt: A Paradox. 97

7.5.2. Asymmetrical effects of (im)moral information on vicarious moral cleansing. 99

8. Literature Synthesis. 102

III. STUDY 1: DOES A FALSE CONFESSION ALLEVIATE GUILT FOR A PAST TRANSGRESSION? 107

1. Introduction. 107

**1.1. Voluntary False Confession.** 108

**1.2. Moral Transference.** 110

**1.3. Incidental Guilt.** 111

**1.4. The Current Study.** 112

2. Study 1. 113

**2.1. Methods.** 113

2.1.1. Participants. 113

2.1.2. Measures. 114

2.1.3. Design and Procedure. 117

3. Results. 118

4. Discussion. 124

**4.1. The Ambiguity of The Source of Guilt.** 124

**4.2. Disingenuous Confession.** 125

**4.3. Moral Indignation as a Distancing Response.** 126

**4.4. Evolved Reputational Concerns Trigger Aversive Emotions Response.** 126

**4.5. An Illusion of Innocence.** 127

**4.6. Religious Affiliation and Guilt.** 128

5. Limitations. 128

6. Conclusion. 129

IV. STUDIES 2 - 4: DO RECOLLECTIONS OF PAST TRANSGRESSION TRIGGER INCREASED SURVEILLANCE PERCEPTION? 131

1. Introduction. 131

**1.1. Heightened Social Awareness as a Response to Reputation Threat.** 132

**1.2. Intuitive Direct Eye Gaze Detection.** 133

**1.3. Direct Eye gaze Judgements in the Face of Uncertainty.** 134

**1.4. Eye gaze Aversion and Guilt.** 135

**1.5. The Current Studies.** 136

2. Study 2. 137

**2.1. Methods** 137

2.1.1. Participants. 137

2.1.2. Measures. 138

2.1.3. Design and Procedure. 138

**2.2. Results and Discussion.** 139

3. Study 3. 144

**3.1. Methods.** 144

3.1.1. Participants. 144

3.1.2. Measures. 145

3.1.3. Design and procedure. 146

**3.2. Results and Discussion**. 147

4. Study 4. 149

**4.1. Methods.** 150

4.1.1 Participants. 150

4.1.2. Measures. 150

4.1.3. Design and procedure. 152

**4.2. Results and Discussion**. 153

5. General Discussion. 156

**5.1. “God is watching”: Religion and Surveillance.** 156

**5.2. Watchful Gods and Social Expansion.** 158

6. Limitations. 160

7. Conclusion. 162

V. STUDY 5: DO SURVEILLANCE CUES INCREASE THE LIKELIHOOD OF VOLUNTARY FALSE CONFESSION? 164

1.Overview. 164

**1.1. The Puzzle of Confession.** 164

**1.2. Confession as Strategic Reputation Management.** 165

**1.3. Confession and Punishment.** 166

**1.4. Voluntary Confession in Response to Imminent Social Exposure.** 167

**1.5. Moral Outrage as a Response to Reputational Threat.** 169

**1.6. The Current Study.** 169

2. Study 5. 171

**2.1. Methods.** 171

2.1.1. Participants. 171

2.1.2. Measures. 172

2.1.3. Design and Procedure. 174

3. Results. 175

4. Discussion. 178

**4.1. Time Constraints and Intuitive Processing.** 179

**4.2. Justifiable Dishonesty.** 180

**4.3. Moral Outrage as a Distancing Strategy.** 181

5. Conclusion. 182

VI. STUDIES 6-9: DIVINE FORGIVENESS AND MORAL SUPPORT FOR STATE-SANCTIONED PUNISHMENT 184

1. Introduction. 184

**1.1. The Role of Punishment in Large-scale Cooperation.** 185

**1.2. Supernatural Punishers.** 185

**1.3. Supernatural Punishment and Cooperation.** 186

**1.4. Punitive and Forgiving God Concepts.** 187

**1.5. Divergent Effects of Forgiving and Punitive God Concepts.** 188

**1.6. Priming Methodologies: Delineating the Dimensions of Religious Cognition.** 189

**1.7. God as a Behavioural Model.** 190

**1.8. Current Research.** 191

2. Study 6. 192

**2.1. Method.** 193

2.1.1. Participants. 193

2.1.2. Measures. 193

2.1.3. Design and procedure. 195

**2.2. Results and Discussion.** 195

3. Study 7. 200

**3.1. Methods.** 200

3.1.1. Participants. 200

3.1.2. Measures. 201

3.1.3. Design and procedure. 201

**3.2. Results and Discussion.** 202

4. Studies 8 and 9. 205

**4.1. Method (Study 8).** 206

4.1.1. Participants. 206

4.1.2. Measures. 206

4.1.3. Design and procedure. 207

**4.2. Results and Discussion.** 207

**4.3. Methods (Study 9).** 211

4.3.1. Participants. 211

4.3.2. Measures. 211

4.3.3. Design and procedure. 211

**4.4. Results and Discussion.** 211

5. General Discussion. 217

**5.1. God as a Behavioural Model.** 217

**5.2. Second Order Free-rider Problem and the Supernatural Punishment Hypothesis.** 218

**5.3. A Mutualistic Moral Framework.** 218

**5.4. The Divine Origins of State-sanctioned Punishment.** 219

**5.5. Complex Systems of Law as a Product of Large-scale Societies.** 220

**5.6. Impersonal versus Personal Punitive Decision-making.** 220

**5.7. Forgiving Gods and Cooperation.** 221

6. Limitations. 222

7. Conclusion. 224

VII. STUDIES 10-11: DOES INGROUP IMMORALITY TRIGGER VICARIOUS MORAL CLEANSING? 225

1. Introduction. 225

**1.1. Theoretical Background.** 226

**1.2. Vicarious Guilt.** 226

**1.3. Vicarious Moral Cleansing.** 227

**1.4. The Moderating Effect of Identification.** 228

**1.5. Identity Fusion.** 228

**1.6. Overview of The Current Studies.** 229

2. Study 10. 230

**2.1. Methods.** 230

2.1.1. Participants. 230

2.1.2. Measures. 231

2.1.3. Design and procedure. 232

**2.2. Results.** 233

**2.3. Discussion.** 238

**3. Study 11.** 239

**3.1. Methods.** 239

3.1.1. Participants. 239

3.1.2. Measures. 240

3.1.3. Design and procedure. 240

**3.2. Results.** 241

**3.3. Discussion.** 242

4. General Discussion. 243

**4.1. Vicarious Moral Licensing.** 244

**4.2. The Absence of Vicarious Moral Cleansing.** 244

4.2.1. Overview of subsequent vicarious cleansing studies. 245

**4.3. Boundary Conditions of Vicarious Moral Cleansing.** 246

**4.4. Protecting the Group Moral Image.** 247

5. Limitations. 248

**5.1. Deontological Considerations.** 249

**5.2. Problems with the Job-Hiring Scenario*.*** 249

**5.3. The Conflation of Egalitarianism and Morality.** 250

6. Conclusion. 250

VIII. Conclusion. 252

1. Revisiting Moral Transference. 252

**1.1. The “Direct Eye Gaze Effect”.** 253

**1.2. Religion and Surveillance.** 254

**1.3. Forgiving Gods and Cooperation.** 257

**1.4. Vicarious Moral Licensing.** 258

2. In Support of the Null Hypothesis. 260

3. Limitations. 261

4. Concluding Remarks. 267

References. 270

Appendices. 365

Appendix A: Recommended safeguards against participant inattention on MTurk. 365

Appendix B: Morality-purity metaphors. 367

Appendix C: Religious affiliation measure. 369

Appendix D: Positive and negative affect scale (Watson et al., 1988). 370

Appendix E: False confession narrative presented to participants in Study 1. 371

Appendix F: Recall instructions for participants in the false confession condition (Study 1). 372

Appendix G: Control narrative presented to participants in Study 1. 373

Appendix H: Recall instructions for participants in the control condition (Study 1). 374

Appendix I: Examples of portraits presented to participants in Studies 2-4. 375

Appendix J: Transgression recall instructions presented to participants in Study 4. 378

Appendix K: Neutral recall instructions presented to participants in Study 4. 379

Appendix L: The set of studies included in Zhou and Fishbach (2016). 380

Appendix M: (high/low) power prime (taken from Study 5, May & Monga, 2014). 381

Appendix N: Information and Consent form shown to participants on MTurk at the start of Study 4. 382

Appendix O: Confession probe presented to participants in Study 5. 383

Appendix P: Forgiving God prime (Studies 6 & 7). 384

Appendix Q: Punishing God prime (Studies 6 & 7). 385

Appendix R: Original punishment scenario shown to participants in Study 6 (taken from Carlsmith et al, 2002). 386

Appendix S: Victim-directed punishment scenario (Study 7). 387

Appendix T: Forgiving God scale presented to participants (Studies 8 & 9). 388

Appendix U: Job hiring task shown to participants in Studies 10 & 11 (taken from Kouchaki, 2011). 389

Appendix V: Verbal fusion measure (taken from Gómez et al., 2011). 390

Appendix W: Moral/immoral/neutral in-group information presented to participants (Study 11). 391

Appendix X: Moral/immoral/neutral salient out-group information presented to participants (Study 11). 392

# 

# List of Tables

**Table 1.** A non-exhaustive list of examples of Catholic moral transference **26**

**Table 2.** Synthesis matrix of the literature **105**

**Table 3.** Summary of multiple regression (predicting implicit guilt from self-reported religious affiliation, false confession dummy and their interaction) **121**

**Table 4.** Multiple hierarchical regression (predicting implicit guilt from strength of beliefs in God, false confession dummy and their interaction). **123**

**Table 5.** Means for interaction in Study 2 **140**

**Table 6.** Multiple regression (predicting average response score from religious affiliation, surveillance dummy and their interaction). **142**

**Table 7.** Correlations and descriptive statistics for Study 2 (N = 321). **148**

**Table 8.** Multiple regression analysis (predicting direct eye gaze from religious affiliation, transgression dummy and their interaction). **155**

**Table 9.** Moderated multiple regression (predicting moral outrage from honesty, surveillance dummy and their interaction). **177**

**Table 10.** Multiple hierarchical regression (predicting endorsement of state-sanctioned punishment from self-reported religious affiliation, dummy forgiveness and their interaction). **197**

**Table 11.** Hierarchical multiple regression (predicting endorsement of state-sanctioned punishment from strength of belief in God, dummy forgiveness and their interaction). **199**

**Table 12.** Multiple regression (predicting endorsement of state-sanctioned punishment from self-reported religiosity, dummy forgiveness and their interaction). **204**

**Table 13.** Hierarchical multiple regression (predicting state-sanctioned punishment from forgiving God score, dummy salience and their interaction). **208**

**Table 14.** Hierarchical multiple regression (predicting state-sanctioned punishment from punishing and forgiving God average, salience and their interactions). **210**

**Table 15.** Hierarchical multiple regression (predicting state-sanctioned punihsment from forgiving God score, dummy condition and their interaction). **213**

**Table 16.** Hierarchical multiple regression (predicting state-sanctioned punishment from punishing and forgiving God average, dummy salience and their interaction). **215**

**Table 17.** Correlations and descriptive statistics for Study 9. **216**

**Table 18.** Cell means for the effect in Study 10 **235**

**Table 19.** Results of moderated multiple regression analysis regressing job suitability ratings on fusion, a dummy variable for the moral condition, a dummy for the immoral condition, and the interaction of these dummy variables with fusion. **237**

# List of Figures

**Figure 1.** An example of pareidolia: Donald Trump's face in a cyst found in a dog's ear. **66**

**Figure 2.** Screenshot of coin toss task presented to participants in Study 5. **173**

**Figure 3.** Surveillance cue presented to participants in the surveillance condition in Study 5. **174**

**Figure 4.** Plot of estimated marginal means for job suitability ratings, Study 11. **242**

**Figure 5.** The Eye of Providence in a stained-glass window at St Raphael's Catholic church, Springfield, Ohio. **255**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# I. Overview of Thesis

*"And Aaron shall lay both his hands on the head of the live goat and confess over it all the iniquities of the people of Israel, and all their transgressions, all their sins. And he shall put them [their sins] on the head of the goat and send it away into the wilderness by the hand of a man who is in readiness."*

*~* (Leviticus 16:21, King James Version).

## 1. Introduction.

In 2010 the villagers of a small village in Shropshire, England, raised over £1000 to restore the grave of Richard Munslow, a local man who died in 1906. Munslow holds the curious accolade of being the last known “sin-eater” in England (“Last ‘Sin-eater’ to be celebrated”, 2010). Sin-eaters were paid to consume bread over corpses in the belief that they would take on the un-confessed sins of the deceased and thus absolve them (Puckle, 1926). With each repetition of the ritual a sin-eater was believed to consume further quantities of sin. Although this practice may seem rooted in superstition, at its heart is a pervasive intuition about how certain actions can compensate for unrelated bad deeds[[1]](#footnote-1). This folk notion, in turn, reflects the core theme of this thesis: that moral status is malleable and that certain (im)moral actions, imperatives and emotions can be subject to a kind of moral *transference.*

### **1.1. Labile Morality: More Than Self-Regulation.**

In January 2018 a furore erupted in the media over a black-tie, men-only charity dinner at an exclusive hotel in London (Marriage, 2018). The event was organised by a charitable organisation called The Presidents’ Club, which counts numerous businessmen, millionaires and politicians as members. Several of the guests at the fundraiser made donations entering into the millions for a children’s hospital. Yet what emerged in the press was the widespread sexual harassment of female hostesses who were working at the event. Did an extravagant altruistic gesture make these men feel entitled to engage in dubious moral behaviour? A leading theory of moral behaviour, *moral compensation,* would suggest as much (Nguyen, 2018).

According to the moral compensation model, moral behaviour is highly labile (Brañas-Garza, Bucheli, Espinosa, & García-Muñoz, 2013; Merritt, Effron & Miller, 2010; Sachdeva, Iliev & Medin, 2009) and our moral behaviours are informed and shaped by previous choices (Brañas-Garza et al., 2013; Loewenstein & Prelac, 1993; Read, Loewenstein & Kalyanaraman, 1999; Simonson, 1990; Zhong, Liljenquist & Cain, 2009). Having performed a moral act, we may feel licensed to perform a subsequent unethical behaviour. Conversely, having committed an immoral act we may seek to cleanse ourselves by engaging in reparative moral behaviours. Under this view moral behaviour involves a process of self-regulation that is comprised of two separate but complimentary mechanisms: moral licensing and moral cleansing (Brañas-Garza et al., 2013; Sachdeva et al., 2009). A key contention of this thesis is that this compensation model is inadequate to fully account for the malleability of our moral behaviour. I suggest that our moral behaviours and judgements are subject to more than a homeostatic process of self-regulation. This suggestion is informed by research that indicates we can acquire moral credentials from others; compensate for the misdeeds of others; outsource certain moral responsibilities to other agents; and alter our moral behaviour according to reputational concerns. The overarching aim of this thesis is to elucidate the cognitive mechanisms that underpin such instances of *moral transference* by articulating the social and psychological factors that trigger and constrain it.

### **1.2. Religious Conceptions of Moral Transference.**

The transference of morality is a common feature of many religions and is well illustrated in Catholicism. The Catholic Church developed a framework for the consequences of sin as subject to complex systems of loaning, borrowing and offsetting[[2]](#footnote-2) and many doctrinal concepts and associated religious practises reflect metaphysical notions of moral transference (see Table 1).

Table 1. A non-exhaustive list of examples of Catholic moral transference

|  |  |
| --- | --- |
| **Religious concept or practise** | **Description** |
| Treasury of Merit | The accumulated “capital” of the good deeds of Jesus and his faithful which can be used as a common pool to pay for the sins of others. Similarly, in early Christianity the accumulated suffering of the Martyrs could pay for the sins of the apostates, or Lapsi (Metzger, 1997). |
| The administration of Indulgences | An indulgence is a way to reduce the amount of punishment one must undergo for one’s sins by paying a fee or performing a good work (Peters, 2008; Swanson, 2007). |
| Inherited sin | An inherited “debt” of sin that all people are born into, which is transmitted down from Adam after his expulsion from the Garden of Eden. |
| Mortification of the flesh | Mortification of the flesh: This refers to a physical ordeal or punishment that a group or an individual undergoes to purify sin (e.g., fasting, self-flagellation, extended genuflection, the wearing of a hair shirt). |

The Catholic Church’s conceptual framework of sin and penance is arguably so comprehensive because it reflects intuitions about moral transference. The religious examples above have equivalents in the psychological literature. Furthermore, I suggest that these religious notions share the same underlying cognitive processes as their psychological counterparts.

### **1.3. Psychological Equivalents.**

Just as the accumulated good deeds of Jesus can be used as a common pool that people draw from to pay off individual sins, research has found that people can incorporate the moral credentials of ethical ingroup members into their own moral self-concept and use these to license subsequent unethical behaviour (Kouchaki, 2011; Meijers, Noordewier, Verlegh, Zebregs & Smit, 2018). The Catholic Martyrs and the everyday ethical ingroup members can be thought of as moral exemplars of the group. Their good deeds provide a pool of moral credits, which less ethical group members may use to pay for their own misdeeds.

The Catholic Church banned the offering of indulgences in 1567. Yet the practise continued, and indeed does to this day[[3]](#footnote-3). Arguably, its persistence is because it is such an intuitive notion. For example, a similar principle of “buy now and sin later” that governed the medieval sale of indulgences has been shown to operate in individuals, who deliberately accumulate moral credentials through the expression of socially preferable attitudes in advance of unethical activity (Effron, 2014; Merritt et al, 2012; Thai, Hornsey & Barlow, 2016). Indeed, just as a medieval knight could engage in a crusade knowing that he could buy full absolution for his sins upon his return, research indicates that merely thinking about the good deeds we plan to perform has a licensing effect (Cascio & Plant, 2015).

The concept of inherited sin is reflected in our intuitive reasoning about the transmission of immorality. Research has shown that people perceive immorality as contagious (Carnes & Winer, 2017; Eskine, Novreske & Richards, 2013; Gino, Ayal and Ariely, 2009). For example, after coming into contact with a transgressor, participants experienced higher levels of state guilt, an effect moderated by feelings of disgust (Eskine et al., 2013). Likewise, biblical concepts of the “stain of sin” being passed down from generation to generation may be underpinned by intuitive reasoning about the transference of *im*moral essence (see Eskine et al., 2013).

The practise of mortification of the flesh relates to research that demonstrates a link between guilt and self-punishment (Bastian, Jetten & Fasoli, 2011; Inbar, Pizarro, Gilovich, & Ariely, 2013; Nelissen & Zeelenberg, 2009; Roseman, Wiest & Swartz, 1994; Wallington, 1973; Wertheim & Schwartz, 1983). Here feelings of guilt are offset by (physical) behaviours outside of the moral domain, which are unrelated to the original transgression and not intended to make reparations or amends towards a specified victim (West & Zhong, 2015).

The literature that provides the theoretical and conceptual background for this thesis will be discussed at length in the next chapter. My purpose at this stage has been to introduce the notion of moral transference by highlighting culturally available examples that are, as I will discuss at length in the next chapter, supported by psychological equivalents. This thesis calls upon a large and diverse body of literature; the methodologies used are also varied. I will turn now to a discussion about the methodology of this research and then the replication “revolution” that took place during this PhD and subsequently informed many of the research practises I used. Finally, I will set out a roadmap for this thesis.

## 2. Methodological Overview.

Of the eleven studies in this thesis, two were conducted in the lab (Studies 1 and 2). This was necessary for logistical reasons as the study designs required face-to-face participation. The majority of studies, however, were conducted online (Studies 3-11). Due to my use of internet-mediated research, I feel that a brief discussion of some relevant issues pertaining to the reliability of online psychological research is warranted.

### **2.1. The Use of Online Crowdsourcing Services.**

Mechanical Turk (MTurk) is a web-based research platform increasingly used in online psychological research (see Buhrmester, Kwang & Gosling, 2011; Buhrmester, Talaifar & Gosling, 2018; Paolacci & Chandler, 2014). Although originally designed for computational tasks rather than behavioural experiments, (e.g., retrieving data from images, transcribing audio data and filtering adult content; Mason & Suri, 2012), MTurk has been validated as a tool for cognitive behavioural research (e.g., Crump, McDonnel & Gureckis, 2013; Mason & Suri, 2012; Mason & Watts, 2009; Suri & Watts, 2011). Several studies have found no difference between the behaviour of laboratory participants and MTurk workers (Horton, Rand, & Zeckhauser, 2011; Paolacci, Chandler, & Ipeirotis, 2010). Moreover, studies have demonstrated the quality of data gathered on MTurk is comparable to, or in some cases exceed, the standard required by published research (Buhrmester, Kwang, & Gosling, 2011; Paolacci et al, 2010).

The most obvious benefit for researchers is that they can gather data in a very short period of time; for a relatively small cost; from large populations (Buhrmester et al., 2011; Mason & Suri, 2012) which are often more diverse than the WEIRD[[4]](#footnote-4) undergraduate populations typically used in laboratory psychology studies (Burhmester et al., 2011; Casler, Bickel & Hackett, 2013). There is, however, some debate as to the extent of the diversity of the MTurk population (see Stewart et al., 2015).

### **2.2. Participant Inattention on MTurk.**

Though questions have been raised about participant inattention and lack of experimenter control on MTurk (Chandler, Mueller & Paolacci, 2014; Clifford & Jerit, 2014; Goodman, Cryder & Cheema, 2013), subsequent research has found MTurk participants’ attention to be equal to, or better than, that of traditional subject pools (e.g., Berinsky, Hubert & Lenz, 2012; Chandler & Shapiro, 2016; Goodman et al., 2013; Hauser & Schwarz, 2016; Klein et al., 2014). Furthermore, various techniques have been developed to safeguard against these issues (e.g., Oppenheimer, Meyris & Davidenko, 2009; Pe’er, Vosgerau & Aquisti, 2014; Ramsey, Thompson, McKenzie, & Rosenbaum, 2016). For example, researchers can use checks to remind inattentive participants to pay more attention (though see Pe’er et al., 2014) or employ “catch trials” that can identify participants paying insufficient attention. For all the MTurk studies conducted in this thesis (Studies 3, 4, 6, 7, 8, 9, 10, 11) attention checks were used to identify inattentive participants (see Appendices A).

### **2.3. Participant Non-naïveté and Dishonesty.**

One risk associated with MTurk studies is that due to high usage or “microlabour” (Marder & Fritz, 2015) workers are becoming savvy to behavioural research methods, which can potentially reduce effect sizes of known research findings (Chandler, Paolacci, Pe’er, Mueller & Ratcliff, 2015; Pe’er, Samat, Brandimarte & Acquisti, 2017). Although recent research provides evidence that MTurker overfamiliarity with well-known tasks[[5]](#footnote-5) is no barrier to its effectiveness (Myer, Zhou & Frederick, 2018; Stagnaro, Pennycook & Rand, 2018). However, researchers have suggested steps to take to guard against non-naïve participants, such as pre-screening questions for example (Chandler, Mueller, & Paolacci, 2014). Alternatively, there are other online platforms available whose users have reported less familiarity with common experimental tasks and higher levels of naivety, for example, Prolific Academic (see Pe’er, Samat, Brandimarte & Acquisiti, 2017). I used this platform instead of MTurk to collect data for Study 5 for precisely this reason, as the experimental paradigm required both high levels of naiveté and honesty in participants.

With regards to levels of honesty in MTurk participants, the picture is somewhat unclear. Dishonest behaviour has been noted, especially when participants lie to meet inclusion criteria for a study (Chandler & Paolacci, 2017; Rand, 2012) though in response to some general knowledge questions the dishonesty levels are negligible (Clifford & Jerit, 2016). Dishonest behaviour can be reduced by the explicit request to behave honestly (Clifford & Jerit, 2016; Lowry, D’Arcy, Hammer, & Moody, 2016) and by certain pre-screening measures such as eligible participants being re-routed to a second survey or contacted at a later date to complete the studies they qualify for (see Chandler & Paolacci, 2017). In all studies that were conducted on MTurk in this thesis I included explicit requests for honesty and the reminder of the need for high quality data at the outset. In studies that required a specific population (e.g., US participants in Studies 10 & 11) I included several a masked pre-screening question before the study that effectively eliminated all ineligible participants.

### **2.4. A Word about Sample Sizes.**

G\*Power (version 3.1; Faul, Erdfelder, Buchner & Lang, 2009) was used throughout this thesis to determine sample sizes. This was done to ensure that all studies were sufficiently powered. The effect sizes included as input parameters to a priori power analyses were based, where possible, on effect sizes previously reported in the literature. Alpha levelswere set throughout, as per convention, at .05, and power (1 – *β*), was set at 0.8 according to Cohen (1988), unless financial considerations precluded this.

## 3. A Field in Crisis?

This PhD was conducted in the midst of what has been termed the “replication crisis” in psychology (Spellman, 2015; see also Świątkowski & Dompnier, 2017). This is an umbrella term for a cluster of problematic research practices and issues that have brought the credibility of the field of psychology into question (Lilienfeld, 2012; 2017). Key amongst these is the failure to successfully replicate the results of published studies. A recent survey on the reproducibility of science, conducted by *Nature* in 2016*,* revealed that out of the 1,567 researchers who took part, 70% reported they had failed to successfully replicate previous findings. More than half reported failing to replicate their own earlier findings (Baker, 2016). A cursory glance at the literature on many widely cited effects in psychology reveals conflicting findings and failures to replicate (see Chapter 2, sections 1.3, 1.5 & 2.3). The crisis has prompted much soul searching and discussion within academia (Chambers, 2017; Ioannidis, 2014; Laws, 2013; Lilienfeld, 2017; Nosek, Spies & Motyl, 2012) and has also reached public awareness (see Chivers, 2017; Fielden, 2017; Ferguson & Beresin, 2017; Sample, 2015; Van Bavel, 2016).

The Open Science Collaboration was an initiative born out of the replication crisis to estimate the rate of reproducibility of psychological science. A body of researchers, led by Brian Nosek (the co-founder and director of the Center for Open Science), conducted 100 high-powered replications of experimental and correlational studies that had been published in three leading psychology journals[[6]](#footnote-6) (Nosek et al., 2015). There was a marked decline in the mean effect size of the replication attempts (*Mr* = 0.197, SD = 0.257) compared to the magnitude of the mean effect size of the original effects (*Mr* = 0.403, SD = 0.188). While 97% of the original studies reported significant results, only 36% of the replication studies yielded significant results (Nosek et al., 2015). In an attempt to address the methodological issues that trouble the field there has been a shift towards greater transparency and rigour in conducting psychological research (Munafò et al., 2017; Nosek et al, 2015), for example through the use of meta-analysis, pre-registration and direct replication.

### **3.1. Meta-analysis.**

In research areas affected by unsuccessful replications, and in some cases, questionable research practises (see John, Loewenstein & Prelec, 2012; Schimmack, 2012; Simmons, Nelson & Simonsohn, 2011) the use of high quality meta-analyses can be extremely important in not only reaching a more robust understanding of a phenomenon but also altering the way research is conducted (see Gurevitch, Koricheva, Nakagawa & Stewart, 2018). For example, they can provide a more comprehensive picture of the evidence for a particular effect than can any individual study and have increased attention to reporting results in a transparent way (Gurevitch et al., 2018). Despite these benefits, meta-analyses are not the panacea for all ills in psychological research. They are statistical techniques that can highlight areas in which evidence is insufficient, which is helpful for guiding future research efforts. What they cannot do is fix these gaps, and it is not their function to do so (Gurevitch et al., 2018).

Greater challenges to meta-analyses are posed by experimenter and publication bias[[7]](#footnote-7) (Anderson & Kichkha, 2017; Morey & Yadav, 2018; Pfeiffer, Bertram & Ioannidis, 2011; Rothstein, Sutton & Borenstein, 2005; Sánchez-Tójar et al., 2018; Thornton & Lee, 2000; van Elk et al. 2015), which can influence the meta-analytic mean and lead to erroneous interpretations. Moreover, different statistical techniques can yield different meta-analytic results (see van Elk et al., 2015). So while meta-analytic techniques can be an important strategy for improving the integrity of psychological research (Asendorpf et al., 2013; Cumming, 2014; Cumming, Fidler, Kalinowski & Lai, 2012), a growing body of researchers argue against the exclusive reliance on such techniques to determine the presence or absence of psychological effects (see Van Elk et al., 2015; Wagenmakers, Wetzels, Borsboom, Kievit & van der Maas, 2015). The recent, and on going, debate about the robustness of religious priming methodologies in the psychology of religion is a good example of the “not by meta-analyses alone” argument.

### **3.2. The Robustness of Religious Priming.**

The use of priming techniques has become prevalent in research on the psychology of religion; I used such techniques in this thesis (see Chapter 6). Researchers investigating the causal role of religion on human behaviour and cognition have turned to priming techniques as they avoid the limitations associated with correlational design and quasi-experimental studies (Shariff, Willard, Anderson & Norenzayan, 2016).

In a highly cited paper, Shariff and Norenzayan (2007) found that participants implicitly primed with God concepts were more generous in a subsequent economic game than participants who had neutral (or no) concepts activated. This paper paved the way for a wave of publications claiming to have found effects of religious primes (both implicit and explicit) on prosociality (e.g., Aveyard, 2014; Carpenter & Marshall, 2009; Ginges, Hansen, & Norenzayan, 2009; LaBouff, Rowatt, Johnson, & Finkle, 2012; Sagioglou & Forstmann, 2013; Schumann, McGregor, Nash, & Ross, 2014; Xygalatas, 2013; see Galen, 2012, for a review). However, the robustness of priming techniques in general was left in some doubt after the publication of two failures to replicate well-known unconscious priming effects (e.g., Doyen, Klein, Pichon & Cleeremans, 2012; Shanks et al., 2013).

In response, Shariff, Willard, Anderson & Norenzayan (2016) conducted a series of meta-analyses (that first appeared online in 2015) to examine, among other things, the reliability of religious priming. One analysis which included 92 religious priming studies (*N* = 11, 608) revealed an average effect size for the difference between religiously primed and control groups of *g* = .40, *p* < .0001, 95% CI = [0.34, 0.46], indicating a small to medium effect of religious priming. To assess the specific effect of religious priming on prosociality a separate effect size meta-analysis was conducted. Across 25 studies (*N* = 4,825) they found a small to moderate effect of religious priming on prosociality, *g* = 0.27, p < .001, 95% CI = [0.15, 0.40]. Shariff et al. (2016) concluded that religious priming does have robust effects across a variety of measures, including prosociality.

However, a reappraisal of Shariff et al. (2016) by van Elk et al. (2015) identified various methodological shortcomings with the original analyses that restrict the conclusions that can be drawn about the efficacy of religious priming techniques. A re-analysis of Shariff et al. (2016) by van Elk et al. (2015) using two different meta-analytic techniques, both of which attempt to correct for publication bias, revealed contradictory results. The first suggested that the effect of religious priming is driven by publication bias. The second, using Bayesian bias correction, did suggest a religious priming effect even after controlling for publication bias. A meta-analysis conducted on studies that have not been preregistered yields a mixture of possible true effects and effects due to experimenter bias (van Elk et al., 2015). More recent, pre-registered and direct replication attempts have failed to replicate previous findings of religious priming on prosociality (see Billingsley, Gomes & McCullough, 2018; Gomes & McCullough, 2015; Miyatake & Higuchi, 2017). Moreover, a small scale meta-analysis by Gomes and McCullough (2015), which included all known studies investigating the effect of religious priming on dictator game offers (including Shariff & Norenzayan, 2007) revealed a) the mean effect size was not different form zero, *g* = 0.37, *SE* = 0.18, *p* = .09, 95% CI [-0.09, 0.83] and b) evidence of publication bias (Gomes & McCullough, 2015).

The cluster of problems associated with the debate on the efficacy of religious priming strongly suggests that meta-analysis alone is not sufficient to reach a definitive conclusion about the reliability and replicability of psychological effects. This is due to the problem of both experimenter[[8]](#footnote-8) and publication bias. So, the debate rumbles on and some researchers claim a satisfactory conclusion will only be reached through the use of large-scale pre-registered replication attempts (see Chambers, 2013; Nosek and Lakens, 2014; Simons, Holcombe & Spellman, 2014; van Elk et al., 2015).

### **3.3. The Use of Pre-registration.**

Pre-registration involves researchers stating their research plans (i.e., main hypotheses, intended statistical analyses and data exclusions) before collecting any data[[9]](#footnote-9). This can help guard against both publication bias (in the case of pre-registered reports[[10]](#footnote-10)) and questionable research practices such as HARKing (i.e., hypothesising after the results are known, see Kerr, 1998) and *p*-hacking (see Simmons, Nelson & Simonsohn, 2011). Although pre-registration is more widely used in other fields (e.g., clinical medicine) it is now becoming a feature of psychological research (Munafò et al., 2017)[[11]](#footnote-11). In 2017 for example, an editorial in the journal *Psychological Science* announced the inclusion of a new category of articles; pre-registered direct replications of studies published previously in *Psychological Science*. *Psychological Science* and other journals such as the *International Journal for the Psychology of Religion* now publish articles with up to three badges, one of which signals that the study has been pre-registered (see van Elk, Rowatt & Streib, 2017).

### **3.4. Replication.**

Replication is a fundamental feature of empirical science (Asendorpf et al., 2013; Jasny, Chin, Chong, & Vignieri, 2011; Nosek, Spies, & Motyl, 2012; Rosenthal, 1990; Schmidt, 2009) yet psychology has, until recently, reported relatively few replications (see Makel, Plucker, & Hegarty, 2012). However, as part of the current drive toward greater transparency and integrity in psychology many journals now publish replications[[12]](#footnote-12).

In very broad terms, replication can be conceptual or direct. A conceptual replication involves trying to confirm previous results using altered or different sets of methods than those in the original study. While providing potentially theoretically interesting findings, some claim they are not best suited to validate an original effect (Earp, Everett, Madva & Hamlin, 2013; Harris, Coburn, Rohrer & Pashlwe, 2013; Simons, 2014; *cf.* Stroebe & Strack, 2014). Direct replications, on the other hand, use exactly the same (or as close as possible) materials and methods as those of the original study (Brandt et al., 2014; Zwaan, Etz, Lucas & Donnellan, 2018). Direct replications by multiple labs are the best way, some argue, to isolate an underlying psychological effect of interest from sampling error, and average across systematic error and measurement error. As Simons writes (2014): “The rejection or deferral of direct replication by other labs assumes that unspecified moderators or quirks of sampling are reliably biased against finding the original effect in a new setting” (p. 76). While the relative merits of direct replications (Chambers, 2013; Drummond, 2009; Simons, 2014) and conceptual replications (Crandall & Sherman, 2016; Stroebe & Strack, 2014) are weighed up by researchers, there are now many guidelines and online platforms available to facilitate researchers in conducting wide scale replication (see Asendorf et al., 2016; Brandt et al., 2014; Munafò et al., 2017; Nosek et al., 2015[[13]](#footnote-13)).

In a recent blog post, Tennie and Moore (2018), outlined a proposal of systematic replication, which suggests journals mark the difference between replicated and non-replicated in writing (e.g., “Smith et al., 2010- awaiting replication”) and moreover, that editors require this (Tennie & Moore, 2018). This system could be extended to count only replications from *outside* the original group (e.g., “Smith et al., 2000; 2001- currently awaiting independent replication”) and that authors explicitly acknowledge and discuss any unsuccessful replication attempts of their key claims (see Tennie & Moore, 2018). They also suggest the establishment of “PhD Replication Project”. Under this scheme PhD students would be required to conduct a replication of one finding in the psychological literature. Students could choose which unreplicated study to attempt to replicate from a centrally constructed database, as it pertained to their research interests (Tennie & Moore, 2018). This PhD replication project has obvious benefits and in principle would be an effective way of ensuring systematic replication. However, there is no suggestion of how to ensure that all replications attempts are conducted equally across the field; some topics are more popular than others. Therefore, unless students are randomly assigned an unreplicated study at the outset of their PhD, there is no way to ensure that every non-replicated study in the literature is given equal attention.

Some note that the replication process itself could induce bias: the “gotcha bias”.Here replications are more likely to be published if they overturn existing findings: “The gotcha bias is a concept that only applies to replication studies that attempt to reproduce results of previous, original studies using similar study designs. The hypothesized mechanism behind this bias is again a proclivity for novelty and sensationalism” (Berinsky, Druckman & Yamamoto, in press, p.2). This potentially leads to the scenario where the published literature over-emphasises replications that contradict previous findings and *mis-*incentivises researchers from conducting replications (see Berinsky et al., in press).

Researchers are aware that there is no single solution to the replication crisis and the most promising solutions may well involve multiple factors, such as combining insights from meta-analyses with both pre-registration and high-quality replications (see Carter & McCullough, in press; Lindsay, 2017; van Elk et al., 2015).

## 4. Structure of this Thesis: A Roadmap.

* In Chapter 2 I will critically examine the relevant literature that provides the theoretical framework for this thesis. I will discuss prominent models of moral compensation, the effect of surveillance on moral behaviour, the influence of religious belief and supernatural punishment on pro-sociality, and the influence of group alignment on morality.
* In Chapter 3, using a novel confession paradigm, I explore whether the performance of a false confession can alleviate guilt for a past, unrelated transgression.
* Chapter 4 investigates whether the recollection of past unethical behaviour can trigger increased sensitivity to cues of social surveillance in ambiguous facial stimuli.
* Chapter 5 probes the effect of surveillance cues on confessional behaviour and seeks to identify whether such cues trigger voluntary confession, as a means of mitigating possible punishment.
* In Chapter 6, I consider the role of forgiving God concepts on people’s endorsement of state-sanctioned punishment, in particular whether such concepts restrict the tendency of individuals to outsource the responsibility for punishment to God.
* Chapter 7 explores whether information about the (im)morality of one’s group triggers vicarious moral compensation. In particular, where information that one’s group is less moral than others triggers vicarious moral *cleansing*.
* In the final Chapter (8), I conclude by discussing several limitations of the current studies and outlining several avenues for future research.

With this overview in place, I now turn to a discussion of the literature that underpins this research.

# II. A Theoretical Framework of Moral Transference.

*“Often we do right, that we may later do wrong with impunity”*

*~* La Rochefoucauld, *Maxims,* (1665)

## 1. Introduction.

Morality matters to us. We consistently rate moral values as among the most important (Goodwin, 2015; Goodwin, Piazza & Rozin, 2014; Leach, Ellemers & Barreto, 2007; Schwartz, 1992; Schwartz & Bardi, 2001) and we perceive them to be indicative of the essence of a person’s identity both when they’re alive (Strohminger & Nichols, 2014) and of their souls after they die (Bering, 2006). Morality forms a central component of our self-concept (Aquino & Reed, 2002; Chaiken, Giner-Sorolla & Chen, 1996; Dunning, 2007) and we strive to maintain a positive moral self-image (Aronson, Cohen & Nail, 1999; Greenwald 1980; Griffin & Ross 1991; Sanitioso, Kunda, & Fong 1990; Rodriguez Mosquera, Manstead & Fischer, 2002; Schlenker, 1980; Steele, 1988), often going to extraordinary lengths to do so (Vonasch, Reynolds, Winegard & Baumeister, 2017). Yet despite this, we can be tempted by unethical behaviour (Mazar, Amir & Ariely, 2008). This results in a tension between desiring to maintain a positive moral-self concept and benefitting from unethical behaviour (Mazar et al., 2008)*.*

Behaving morally can bolster our self-image, which authorises us to engage in ethically dubious behaviour (Effron & Conway, 2015; Kouchaki & Jami, 2018; Monin & Miller, 2001; Merritt, Effron & Monin, 2010; Merritt et al, 2012; Sachdeva, Iliev & Medin, 2009) a phenomenon termed *moral licensing*. Conversely, having jeopardised or damaged our moral self-image, we often engage in subsequent, reparative moral behaviours (Jordan, Mullen & Murnighan, 2011; Sachdeva et al., 2009; Tetlock, Kristel, Elson, Lerner & Green, 2000; Zhong & Liljenquist, 2006), i.e., *moral cleansing.*

### **1.1. A Dynamic Theory of Moral Behaviour.**

Sachdeva et al. (2009) found that asking participants to affirm a moral identity by writing a self-relevant story containing positive trait words led to a decrease in charitable donations and fewer expressed intentions to engage in pro-environmental actions. In contrast, threatening an individual’s identity by getting them to write a self-relevant story containing *negative* trait words led to increased charitable donations and more expressed intentions to engage in pro-environmental activities (Sachdeva et al., 2009). This finding has been taken to suggest that these two separate but complementary mechanisms, (i.e., *licensing* and *cleansing*) act together as part of a broader system of moral self-regulation (e.g., Brañas-Garza, Bucheli, Espinosa, & García-Muñoz, 2013; Newman & Brucks, 2018; Nisan & Horenczyk, 1990; Sachdeva et al, 2009; Zhong, Ku, Lount, & Murnighan, 2010).

Brañas-Garza et al. (2013) propose a dynamic model of this moral self-regulation process. Moral decisions are made over a period of timeand an individual self-regulates in order to achieve an optimal level of moral self-worth. This self-regulation follows a predictable pattern: an individual’s good and bad decisions will eventually balance each other out. In a series of studies Brañas-Garza et al. (2013) had participants play a sequence of dictator games[[14]](#footnote-14) each with a randomly chosen recipient. The results revealed that participants’ behaviour in every round was negatively correlated with their behaviour in the previous round. Participants tended to balance an above average donation in one round of a dictator game with a below average donation in another, and vice versa. This suggests that participants were not donating on the basis of some set moral standard (if that were so then they would show a consistent preference for a donation amount) but rather a systematic process of “dynamic equalisation” leading to a pattern of moral licensing and cleansing (Brañas-Garza et al., 2013).

### **1.2. Moral Licensing.**

Moral licensing occurs when past moral behaviours render us more likely to engage in immoral activities without worrying about feeling or appearing immoral (Khan & Dhar, 2006; Kouchaki & Jami, 2018; Merritt, Effron & Monin, 2010). This phenomenon has been documented in a wide range of contexts including recruitment decision-making (Cascio & Plant, 2015; Monin & Miller, 2001; Effron, Cameron & Monin, 2009); racial attitudes (Effron, Cameron, & Monin, 2009; Effron, Monin, & Miller, 2012; Mann & Kawakami, 2012); consumer choice (Khan & Dhar, 2006; Mazar & Zhong, 2010); environmental judgements and behaviours (Meijers, Verlegh, Noordewier & Smit, 2015;

Nilsson, Bergquist & Schultz, 2016; Noblet & McCoy, 2017; Tiefenbeck, Staack, Roth & Sachs, 2013); food consumption (Eskine, 2013; Garvey & Bolton, 2017) and pro-social behaviour (Conway & Peetz, 2012; Jordan, Mullen & Murnighan, 2011).

For example, male participants who had the opportunity to establish non-sexist credentials by disagreeing with overtly sexist statements were subsequently more likely to indicate that a particular job was better suited for a man than a woman (Monin & Miller, 2001). In the racial domain, non-black participants who had been given feedback stating they were progressing on egalitarian goals were subsequently more likely to choose to sit further away from black individuals and closer to whites as well as indicate greater implicit racial prejudice (Mann & Kawakami, 2012). Similarly, participants who had the opportunity to display non-racist credentials by endorsing Barack Obama’s presidential candidacy (Effron, Cameron & Monin, 2009) or by selecting a black candidate for a category-neutral job (Monin & Miller, 2001) were more likely to exhibit preferential judgements of white candidates in a later job hiring scenario. In fact, simply indicating they would do a future good deed reduced charitable donations (Khan & Dhar, 2006). When asked to explain their choices participants made no reference to the effect of their prior choice on their subsequent behaviours, suggesting that people are unaware of the effect of previous decisions on subsequent moral preferences (Khan & Dhar, 2006; Simbrunner & Schlegelmilch, 2017).

### **1.3. Meta-Analytic Overview of Moral Licensing.**

Despite the prevalence of moral licensing studies in the literature, replication attempts have not always been successful (e.g., Blanken, van de Ven, Zeelenberg & Meijers, 2014; Moery & Calin-Jageman, 2016). Recently the focus has been on conducting meta-analyses to provide an objective measure of the strength of this effect. Blanken, van de Ven and Zeelenberg (2015) for example conducted a meta-analysis of 91 studies (*N* = 7,397) comparing a licensing condition with a control condition. The results yielded an average effect size of *d* = .31, 95% CI [0.23, 0.38]; a small to medium effect on Cohen’s (1988) classification. They also found that article status (published *vs.* unpublished) significantly influenced the estimated moral licensing effect size; published studies had larger effects (*d* = 0.43, *SE* = 0.04) than unpublished studies (*d* = 0.11, *SE* = 0.06), *QM(*1) = 20.66.[[15]](#footnote-15).

Simbrunner and Schlegelmilch (2017) conducted another meta-analysis (*N =* 106)[[16]](#footnote-16) and found a comparable generalised moral licensing effect size; *d* = .32, 95% [0.23, 0.41] (Simbrunner & Schlegelmilch, 2017). Similarly, their results revealed that published studies (*d* = 0.49, *SE* = 0.064) showed bigger licensing effects than unpublished ones (*d* = 0.09, *SE* = 0.056). Cultural background also emerged as a factor that explained a substantial amount of variation in moral licensing effect size (Simbrunner & Schlegelmilch, 2017). Notably, licensing effects were stronger in North America than in Western Europe (Simbrunner & Schlegelmilch, 2017). To explain this difference Simbrunner and Schlegelmilch (2017) suggest that “being a good person” is more important to North Americans than Western Europeans:

With questions of morality being perceived far more pivotal, North Americans presumably see a fluctuation in one’s own behavior with regard to moral dimensions as less acceptable than Western Europeans. Vice versa, Western Europeans are more likely to allow themselves to be immoral. Consequently, the value of a license, which enables an individual to behave immoral [*sic*], is smaller compared to the value a North American individual attributes to this license. (p. 219)

As Simbrunner and Schlegelmilch (2017) propose, research has found that North Americans and Western Europeans do indeed make different judgements about the morality or immorality of certain behaviours (Pew Research Centre, 2014). Though whether this is evidence of North Americans valuing morality more than Western Europeans is debatable given that the types of behaviour considered (e.g., extramarital affairs, gambling, homosexuality, abortion, premarital sex, alcohol consumption, divorce, and the use of contraceptives) could just as easily indicate higher levels of conservatism in North America. There are, as Simbrunner and Schlegelmilch also note, differences in the role of religion between both regions (with North America placing higher value on religious beliefs than Western Europe[[17]](#footnote-17)) (Pew Research Centre, 2015) and this might help to explain the observed difference in the size of the moral licensing effect. This seems a more plausible explanation than that North Americans care more than Western Europeans about being good people.

Also of note, was a reversed pattern of licensing in South-East Asians, whereby engaging in moral behaviour led to continued moral behaviour, i.e., moral *consistency* (Simbrunner & Schlegelmilch, 2017). Simbrunner and Schlegelmilch’s (2017) explanation is that because past experiences and future expectations are more relevant in guiding behaviour for South-East Asians than for Americans (see Rojas-Mendez et al., 2002) the need for consistency in one’s moral self-image is important for them. Thus, leading South-East Asians to behave in morally consistent ways.

What both these meta-analyses highlight is the importance of conducting well-powered studies on the moral licensing effect. Post hoc power analyses by Blanken et al. (2015) (on all the studies included in their meta-analysis) using their effect size estimate found that, on average, studies only had 28% power (Blanken et al., 2015[[18]](#footnote-18)). In short, these two meta-analyses found evidence for a modest and unreliable moral licensing effect that is moderated to a large extent by publication bias (Blanken et al., 2015; Simbrunner & Schlegelmilch, 2017).

#### 1.3.1. Pre-registered direct replications.

Recently Urban, Bahník and Kohlová (2017) conducted three direct replications of Mazar and Zhong (2010)[[19]](#footnote-19)(which were not included in the above meta-analyses), two of which were pre-registered. This is a particularly welcome endeavour, as the above meta-analyses did not include pre-registered studies. Across these three, well-powered studies, Urban et al. (2017) found engagement in pro-environmental behaviour had a negligible observed effect on subsequent immoral behaviour (-0.16 < *d* < 0.07). In other words, there was no evidence for a moral licensing effect. In fact, two of their studies found effects running in the opposite direction than that implied by moral licensing.

Similarly, Moery and Calin-Jageman (2016) conducted two direct (and one online conceptual) pre-registered replications of Eskine (2013). The results of Eskine (2013) suggest that exposure to organic foods (*vs.* comfort and control foods) triggers moral licensing effects (e.g., decreased likelihood of helping a stranger and harsher subsequent moral judgements). However, across three studies Moery and Calin-Jageman (2016) found, contra Eskine (2013), that exposure to organic food had a negligible effect on both moral behaviour (*d* = 0.06, 95% confidence interval [CI] [−0.14, 0.26], *N* = 377) and prosocial behavior (*d* = 0.03, 95% CI [−0.17, 0.23], *N* = 377). Urban et al. (2017) and Moery and Calin-Jageman (2016) are among the first to move moral licensing research in this new direction by conducting pre-registered, direct replication attempts. There are, at the time of writing, no successful pre-registered direct replications of the moral licensing effect.

### **1.4. Moral Cleansing.**

The mirror image of moral licensing is moral cleansing. This term encompasses behaviours that aim to restore a moral self that has been jeopardised by engaging in unethical activity (Jordan et al, 2011; Sachdeva et al, 2009; West & Zhong, 2015; Zhong & Liljenquist, 2006). A detailed overview of moral cleansing by West and Zhong (2015) proposes three separate categories of cleansing behaviour:

* **Restitution cleansing.** This is the most direct way to “cleanse” oneself following a transgression, simply by correcting the wrong. West and Zhong (2015) write that: “restitution cleansing is directly aimed at eliminating the provoking moral threat itself” (p. 221). For example, by reaffirming sacred values after having been asked to imagine violating them (Tetlock, et al, 2000) or donating more money to a black confederate after being given false feedback about one’s level of racial prejudice (Dutton & Lake, 1973).
* **Behavioural cleansing.** These are cleansing responses that are less direct than restitution cleansing. Zhong and West (2015) write:

**“**Restoring moral self-worth can involve counter-balancing across several disparate dimensions of moral character in ways that are usually opaque to outside observers” (p. 222). Behavioural cleansing consists of interchangeable prosocial behaviours. That is, threats to a particular aspect of moral self-worth prompt compensatory behaviours in an *unrelated* prosocial domain. For example, recalling past unsafe sexual activity leading to increased donations to a homeless shelter (Stone, Wiegand, Cooper & Aronson, 1997) or an increase in desire to engage in community service following false feedback on reckless driving (Steele, 1988). Zhong, Liljenquist and Cain (2009) suggest that these behavioural cleansing effects are a type of balancing between *“debits and credits”* which extends across moral domains (Zhong, Liljenquist & Cain, 2009).

* **Symbolic cleansing.** These are cleansing behaviours that extend beyond sub-domains of morality and are metaphorical or figurative in nature (West & Zhong, 2015). For example, giving participants the opportunity to wipe their hands following the recall of a past transgression led to a decrease in subsequent moral behaviour (Zhong & Liljenquist, 2006). Inexperienced video game players who played games involving violence against humans (compared to violence against objects) went on to show a preference for hygiene products in a selection task (Gollwitzer & Melzer, 2012). This suggests that physical cleansing can alleviate the emotional consequences of immoral behaviour (i.e., the “Macbeth Effect”; Zhong & Liljenquist, 2006)[[20]](#footnote-20).

### **1.5. Meta-Analytic Overview of Moral Cleansing.**

Several attempts to replicate certain types of moral cleansing effects (e.g., the “Macbeth effect”) have been unsuccessful (e.g., Fayard, Bassi, Bernstain & Roberts, 2009; Gámez, Díaz, & Marrero 2011; Earp, Everett, Madva & Hamlin, 2014[[21]](#footnote-21)). Therefore, recent meta-analyses have proven a useful first step in estimating the size of the moral cleansing effect. Blanken et al. (2014) performed a meta-analysis of the size of the moral cleansing effect across five studies (including Sachdeva et al., 2009, Studies 1 & 3; and three replication attempts) and found a mean effect size of *d* = 0.04, 95% CI [-0.11, 0.20]. The confidence intervals here (which bracket zero) indicate a non-significant effect, (*z* = 0.53, *p* = .593).

More recently Körner and Strack (in press) conducted a small meta-analysis to measure the strength of the “clean-slate” effect. This is the act of physically cleansing to remove the effects of a past experience on present behaviour (Lee & Schwartz, 2010, 2011, 2016) (e.g., washing one’s hands after a transgression to remove guilt – the “Macbeth effect” – or physical cleansing after failure to remove associated negative emotions; see Kasper, 2012). The results of this meta-analysis revealed a small to medium effect size, *d* = 0.39, 95% CI = [0.16, 0.62].

Meta-analyses of moral cleansing effects are fewer than those of moral licensing due to a noticeable paucity of moral cleansing studies compared to those on moral licensing[[22]](#footnote-22). Perhaps this reflects a certain disinterest in investigating moral cleansing- given how intuitive a notion it is to us. After all, the concept of “cleansing” (both figurative and literal) is a prevalent feature of our cultural environment as evidenced by the a) variety of absolution and ablution rituals found in most major world religions and b) the wealth of morality-purity metaphors which abound in everyday natural language (Lee & Schwarz, 2010; West & Zhong, 2015; see Appendix B).

Moreover, people tend to believe that the shift from immoral to moral behaviour is driven by the emergence of an individual’s virtuous *true* self as compared with similar changes from moral to immoral behaviour (or morally neutral changes) (see De Freitas, Tobia, Newman & Knobe, 2016; De Freitas, Cikara, Grossmann & Schlegel, 2017). Therefore, moral licensing might intrigue us more than moral cleansing because it violates intuitions about the essential nature of the self as virtuous.

### **1.6. Limitations of the Moral Compensation Model.**

A key conjecture of this thesis is that the moral compensation model does not adequately account for the flexibility of our real-world moral behaviours. This suggests they are subject to a more complex system than one that simply regulates between licensing and cleansing. This aside, there are further methodological and conceptual issues with the moral compensation model.

#### 1.6.1. Insufficient evidence to support a balancing mechanism.

According to Sachdeva at al. (2009), moral licensing and cleansing work together as part of a broader process of moral self-regulation. This is demonstrated, they claim, by showing licensing and cleansing effects within the same paradigm (Sachdeva et al., 2009). However, as previously mentioned, a recent study by Blanken et al. (2014) failed to replicate Sachdeva et al’s (2009) results. Blanken et al. concluded there is insufficient evidence at present to support the notion of a general “balancing” mechanism underpinning moral licensing and cleansing (Blanken et al., 2014, though see Simbrunner & Schlegelmilch, 2017).

#### 1.6.2. Moral Consistency.

Also at odds with the moral compensation model, is the relatively large area of research that details a “moral consistency” effect. This is when an individual with a salient self-concept as a moral person displays *more* moral behaviour than does an individual without a salient moral self-concept (or those who have a salient *im*moral self-concept) (e.g., Aquino, Freeman, Reed, Lim & Felps, 2009; Blasi, 1983; Conway & Peetz, 2012; Fu, Heyman, Quian, Guo & Lee, 2015; Joosten, van Dijke, Van Hiel & De Cremer, 2014; Reed, Aquino, & Levy, 2007). In addition, moral consistency effects violate an assumption of the moral compensation model, which is that people’s “normal state of being” (Sachdeva et al, 2009, p. 527) is one of moral equilibrium; neither too moral nor too immoral. Sachdeva et al. (2009) write:

We suggest that people aspire to maintain a comfortable moral self-image. Deviation from this level in either direction, either when the view of the self as a moral individual is threatened or when this comfortable level is surpassed, results in compensatory behaviors. (p. 527)

Are people motivated to offload a surplus of moral credentials because they no longer feel a state of moral neutrality?

Moreover, if we desire nothing more than a comfortable moral self-image then it seems counterproductive to a) have a tendency to see ourselves as more moral than others (Alicke, Vredenburg, Hiatt & Govorun, 2001; Tappin & McKay, 2017; Van Lange & Sedikes, 1998) and b) place so much value on possessing and projecting a moral self-image (Aronson, Cohen, & Nail, 1999; Greenwald 1980; Griffin & Ross 1991; Sanitioso, Kunda & Fong 1990; Schlenker, 1980; Steele, 1988; Rodriguez Mosquera, Manstead & Fischer, 2002).

Attempts to integrate moral consistency and compensation have suggested that the divergent effects are driven by the perspective that participants take (Joosten et al., 2014). Moral consistency effects may be driven by long-term concerns about ensuring a good reputation whereas moral compensation effects are on the fly “damage control” responses to reputation concerns (Joosten et al., 2014). I turn now to consider reputation effects in more detail, in particular how reputation concerns influence our moral behaviours.

## **2. Reputation and Surveillance**.

*"Our reputation is more important than the last hundred million dollars.”*

~ Rupert Murdoch

A positive social reputation carries significant benefits such as greater opportunities for future cooperation and increased help from others (Nowak & Sigmund, 1998a, 1998b, 2005; Leimar & Hammerstein, 2001; Sienen & Schram, 2006; Sigmund, 2012). In short, we help individuals with good reputations (Nowak & Sigmund, 2005) and avoid those with bad ones (Kurzban & Leary, 2001). It is trivially true that our reputation depends on other people’s perception of it (Engelmann & Zeller, 2017; Sperber & Baumard, 2012) therefore it is not surprising that humans are motivated to present themselves in a good light and strategically adjust their behaviour in front of others (Engelmann & Fischbacher, 2009; Kurzban, 2001; Milinski, Semman & Krambeck, 2002; Reis & Gruzen, 1976; Satow, 1975; van Rompay, Vonk, & Fransen, 2009; Van Vugt & Hardy, 2009).

### **2.1. The Effects of Being Watched.**

Studies have shown that when people know their behaviour is observable by others, they are more generous (Reis & Gruzen, 1976; Milinski, Semmann & Krambeck, 2002; van Vugt & Hardy, 2010); less likely to look at provocative stimuli (Risko & Kingstone, 2011); more likely to engage in moralistic punishment (Kurzban, DeScioli, & O'Brien, 2007; Piazza & Bering, 2008) and more altruistic in public goods games (Hardy & van Vugt, 2006). Moreover, moral licensing effects are muted when the subsequent licensing behaviour is public rather than private (Greene & Low, 2014).

These reputation management strategies are effective: altruistic group members are more highly regarded and more frequently chosen for future cooperative interactions (Hardy & van Vugt, 2006). They are so effective that people often engage in conspicuous displays of altruism to boost prestige rather than out of any real desire to see the benefit of their contributions (Hardy & van Vugt, 2006). For example, van Vugt & Hardy (2009) found that participants conspicuously or “wastefully” (in the sense that they knew the impact of their donations would be ineffective in attaining a public good) donated approximately 40-50% of their money when they were being observed (van Vugt & Hardy, 2009).

Children as young as five-years-old have been shown to strategically manage their behaviour when they are being watched, for example cheating less and being more generous (Engelmann, Herrmann & Tomasello, 2012; Engelmann, Over, Herrmann & Tomasello, 2013; Fujii, Takagishi, Koizumi & Okada, 2015; Leimgruber, Shaw, Santos & Olson, 2012; Takagishi et al., 2015). Even being told they are in the presence of an invisible spirit (“Princess Alice”) has a similar effect (Piazza, Bering & Ingram, 2011). Moreover, children demonstrate audience sensitivity – they are aware of exactly *who* is watching- and are more generous when being observed by a future collaborator or an ingroup member (as opposed to an outgroup member) (Engelmann et al., 2013).

### **2.2. Effects of Artificial Surveillance Cues on Behaviour.**

Some researchers suggest that even exposure to artificial cues of surveillance (e.g., stylised eye-spots) can affect behaviour and prompt changes in prosociality (Burnham & Hare, 2007; Oda, Niwa, Honma & Hiraishi, 2011). This effect has been documented in a variety of cooperative domains such as charitable giving (Fathi, Bateson & Nettle, 2014; Bateson, Nettle & Roberts, 2006; but see Manesi, Van Lange, Van Doesum & Pollet, in press; Northover, Pedersen, Cohen & Andrews, 2017a); generosity (Baillon, Selim, & Van Dolder, 2013; Haley & Fessler, 2005; Keller & Pfattheicher, 2011; Rigdon, Ishii, Watabe & Kitayama, 2009; but see Matsugasaki, Tsukamot & Ohtsubo, 2015); increased condemnation of moral transgressions (Bourrat, Baumard & McKay, 2011, Sparks & Barclay, 2015; but see Northover, Pedersen, Cohen & Andrews, 2017b) and environmentalism (Bateson et al., 2015; Manesi, Van Lange, & Pollet, 2015).

Recent studies have found that children as young as 3-years-old are sensitive to artificial cues of surveillance. Children who explicitly interacted with images of eyes (compared with cloth flowers) went on to show increased prosociality in a subsequent task (Kelsey, Grossmann & Vaish, 2018). However, there was no prosocial effect when the children were simply exposed to the surveillance cue, i.e., without any explicit mention or interaction with it (Kelsey, et al., 2018).

### **2.3. Meta-Analytical Overview of Artificial Surveillance Cues Effect.**

The literature contains many inconsistencies regarding the effects of artificial surveillance cues on prosocial behaviour and a significant number of studies have failed to replicate earlier findings (e.g., Carbon & Hesslinger, 2011; Fehr & Schneider, 2009; Lamba & Mace, 2010; Matsugasaki, Tsukamoto, & Ohtsubo, 2015; Northover et al., 2017b; Raihani & Bshary, 2012; Saunders, Taylor & Atkinson, 2016). As an example, Raihani and Bshary (2012) found no effect of eye images on cooperation in a large cross-cultural online sample (*N* = 776). Rather they found, unexpectedly, a *positive* effect when participants were exposed to the control image (images of flowers; Raihani & Bshary, 2012). A re-examination of the literature by Nettle et al. (2013), however, found that eye images increased the odds of donating *something* rather than nothing. More recently though, in a well-powered MTurk study (*N* = 1535), Saunders et al. (2016) failed to find evidence of increased donation amount or likelihood of donating in the presence of a range of artificial surveillance stimuli (e.g., images of faces *vs.* eyes[[23]](#footnote-23), abstract *vs.* photographic; the number of eyes/faces presented and the emotion portrayed in the eyes/faces; either angry, sad or neutral).

Northover et al. (2017a) conducted two meta-analyses to estimate the strength of the effect of artificial surveillance cues on measures of generosity. Their first meta-analysis used the amount given by participants as the measure of generosity and included data from 26 studies (*N* = 2732); the second used the proportion of participants who gave something rather than nothing (27 Studies; *N* = 19512) (Northover et al., 2017a). The first meta-analysis (amount given by participants) revealed a small mean effect size, *d* = .03, 95% CI [-0.08, 0.13] which was not significantly different from zero. The second (proportion who gave) revealed a small mean effect size of *d* = 0.16, 95% CI [-0.04, 0.35] and not significantly different from zero. Taken together the results of these meta-analyses provide no evidence that artificial surveillance cues increase a) how generous people are or b) the likelihood of people being generous at all (Northover et al., 2017a).

However, Northover et al. (2017a) only investigated the effect of artificial surveillance cues on *generosity*, excluding other cooperative behaviours that have been widely documented in the literature (e.g., littering, dishonesty, moral judgements). In a separate paper, Northover et al. (2017b) reported a further four studies (three of which were replications of Bourrat et al., 2011) in an attempt to ascertain whether surveillance cues had any effect on moral judgements. They found no supporting evidence for this effect (Northover et al., 2017b). A further two small meta-analyses investigating the effect of artificial surveillance cues on moral judgement (*N* = 6 for each) yielded mixed results.

Despite the inconsistencies in the literature, what *is* clear is that the effects of surveillance cues on prosocial behaviour are more complex than initial studies suggested. Furthermore they appear to be moderated by a number of situational variables such as the number of people nearby (Bateson, Nettle & Roberts, 2006; Ernest-Jones, Nettle & Bateson, 2011; Powell, Roberts & Nettle, 2012); the direction of gaze of the eye-spots (Manesi, Van Lange & Pollet, 2015); length of exposure to cues (Sparks & Barclay, 2013; *cf* Sparks & Barclay, 2015) and participant characteristics such as gender (see Rigdon, Ishii, Watabe & Kitayama, 2009), levels of chronic public self-awareness and prevention focussed self-regulation (Pfattheicher & Keller, 2011).

To be able to strategically manage behaviour to maintain or secure reputational benefits one must be aware that they are being watched. A wide body of research suggests that we have evolved capacities for the detection of agents. I now turn to this literature and discuss how, in particular, these capacities have been proposed as a major factor in the formation of religious beliefs.

## 3. Agency Detection Capacities.

*“Our agency detection device suffers from some hyperactivity, making it prone to find agents around us, including supernatural ones”*

~ Justin Barrett, *Why would anyone believe in God? (2004)*

For any animal, the difference between correctly identifying whether or not a particular stimulus is an intentional agent can be the difference between life and death. Mistaking a predator for something innocuous is the obvious example but conversely, mistaking your *prey* for something inanimate can be an equally costly error. In other realms too, such as mating, being able to correctly detect an agent (or not) could be the difference between passing on your genes, or not.

An influential theory of cognitive biases, Error Management Theory (EMT), posits that a general rule underlies biologically evolved systems of decision-making and judgement; given the choice between two errors, natural selection will favour behaviours that suppress the rate at which individuals commit the more costly error (Haselton, 2007; Haselton, Bryant, Wilke, Frederick & Galperin, 2009; Haselton & Buss, 2000; Haselton & Nettle, 2006; though see McKay & Efferson, 2010).

Proponents of EMT suggest that mistakenly inferring the presence of another agent (i.e., a false positive) carries only limited costs. A false *negative* (i.e., falsely assuming no agent is present), however, carries potentially catastrophic costs. According to EMT a system that errs on the side of caution and *over-*infers the presence of agency will, on balance, avoid more false negatives (Geary & Huffman, 2002[[24]](#footnote-24)). Thus, some researchers argue the evolutionary conditions were ripe for the development of a hyperactiveagency detection system or “device” (hereafter HADD[[25]](#footnote-25)) (Atran & Norenzayan, 2004; Barrett, 2000, 2004, 2012; Bertolotti & Magnani, 2010; Guthrie, 1993, 2001; Guthrie et al., 1980). In the face of ambiguous information about the presence of other agents the “hair-trigger” nature of this device will over-activate (Barrett, 2000; Barrett & Burdett, 2011) meaning that individuals will infer agency where it does not in fact (or is unlikely to) exist (Barrett, 2004). It is the hypersensitivity of this device that purportedly leads to an excess of false positives regarding agency attributions in some individuals, which in turn is thought to make attributions of unseen agents plausible among these individuals (Petrican & Burris, 2012).

### **3.1. Illusory Agency Perception.**

The HADD tradition developed around experimental and empirical evidence that reveals an intuitive tendency to interpret the presence of agents (McCauley, 2011). A classic, mid-20th century psychology study demonstrated that even when shown abstract moving stimuli (e.g., two-dimensional geometric shapes) people interpret the movements in agentic and intentional terms (Heider & Simmel, 1944). This agency interpretation bias has since been reliably found in both adults and children (Abel, Happe & Frith, 2000; Bloom & Veres, 1999; Csibra, Gergely, Bíró, Koos & Brockbank, 1999; Premack & Premack, 1995). The phenomenon of perceiving faces in unusual objects (e.g., faces in clouds and more recently, the face of Donald Trump on a cyst in a dog’s ear, see figure. 1), or “*Pareidolia*” (Liu et al., 2014) is well documented.

Studies on illusory agency perception have demonstrated that face detection is triggered by minimal face-like patterns even when these features are not centred in visual focus. Moreover, humans evince a pronounced tendency to interpret meaningful patterns in random noise, with face detection rates for pure white noise stimuli as high as 30-40% (Gosselin & Schyns, 2003; Hansen, Thompson, Hess, & Ellemberg, 2010), though this tendency has significant variation between individuals. From birth human infants demonstrate a preference for attending to face-like stimuli (Fantz, 1963; Johnson & Morton, 1991) and recent studies have found evidence to suggest this preference obtains *in utero* (Reid, Dunn, & Young, 2017).



*Figure 1.* An example of pareidolia: Donald Trump's face in a cyst found in a dog's ear.

### **3.2. Boundary Conditions of the HADD.**

An issue pertaining to the HADD is determining the conditions under which it is triggered (Maij, van Schie & van Elk, 2017). For example, the HADD might be activated by *any* uncertainty (regarding the presence of an agent or otherwise). Although the costs associated with a false positive (i.e., incorrectly assuming the presence of an agent) are considered relatively small (i.e., you waste energy) they are not insignificant. If the HADD is activated in the face of *all* uncertainty, the relatively small costs associated with a false positive add up. Furthermore, under this scenario you would simply be *wrong* a large proportion of the time. Therefore, taken to the extreme, a HADD that systematically activates in the face of any uncertainty regarding the presence of an agent would be unable to correctly calibrate intuitions with objective reality. This means individuals would make a high number of erroneous inferences or would be simply rendered delusional.

A second possibility put forward by researchers, is that *threatening*, ambiguous information regarding the possible presence of agents are necessary to elicit agency intuitions (Barrett, 2000; Guthrie, 1993; Lisdorf, 2007). There is a degree of evidence to support an association between negative and threatening stimuli and agency detection (see Maij et al., 2017; Nieuwboer, Van Schie & Wigboldus, 2014). For example, musically induced fear renders participants more likely to see an ambiguous stimulus as a dangerous animate object (i.e., a snake) (Prinz & Seidel, 2012). Nieuwboer, Van Schie and Wigboldus (2014) found that images of threatening natural phenomena elicited stronger attributions of intentionality and agency (Nieuwboer et al., 2014). Similarly, Valdesolo and Graham (2014) showed that aversion to uncertainty mediated the effect of awe on the propensity to perceive human agency in random events. Moreover, people seem more likely to attribute events to intentional agents when these events are negative. For example, Morewedge (2009) showed that participants playing the dictator game were more likely to believe that their opponent was a human agent, rather than a computer, when they experienced an unusually negative outcome (i.e., an unfair allocation) rather than an unusually positive one (Morewedge, 2009). In another series of studies Hoskin, Hunter and Woodruff (2014) found that exposing anxious participants to threatening images increased the rate of false alarms on a speech detection task. Barrett and Johnson (2003) found evidence to suggest lack of control led participants to make greater attributions of intentional agency to inanimate objects than participants in a *“*perceived control*”* condition (Barrett & Johnson, 2003).

However, a recent set of studies by Maij et al. (2017) revealed that exposing participants to weak and moderate threat manipulations did not increase false detections of agency or intentionality. In one of their experiments (Study 4) participants were exposed to one of two virtual reality scenarios, either walking around a dark basement (contextual threat scenario) or walking around a desert (neutral scenario). Participants in the contextual threat scenario did not show higher levels of agency detection (operationalized as the perception of human voices) than those in the neutral scenario. On the contrary, neutral scenarioparticipants showed higher levels of agency detection than those in a virtual threat scenario. One possible explanation proposed by the researchers is that feelings of loneliness elicited by the desert condition might have motivated individuals to seek sources of social connection, thus leading to an increase in agency detection (Maij et al., 2017).

### **3.3. HADD and the Formation of Religious Beliefs.**

The HADD (in its varying incarnations) has been implicated in the acquisition and transmission of religious beliefs (Barrett, 2000; Barrett & Lanman, 2008; Guthrie, 1980, 1993, 2001). According to HADD theorists (Barrett, 2000; Barrett & Lanman, 2008), ambiguous agency intuitions often lead to the feeling an *invisible* agent is present: therefore, allowing individuals to form beliefs in supernatural agents (Atran, 2004; Atran & Norenzayan, 2004; Barrett, 2000). Atran and Norenzayan (2004) state:

[S]upernatural agents are readily conjured up because natural selection has trip-wired cognitive schema for agency detection in the face of uncertainty. Uncertainty is omnipresent; so, too, is the hair triggering of an agency-detection mechanism that readily promotes supernatural interpretation and is susceptible to various forms of cultural manipulation. Cultural manipulation of this modular mechanism and priming facilitate and direct the process. Because the phenomena created readily activate intuitively given modular processes, they are more likely to survive transmission from mind to mind under a wide range of different environments and learning conditions than entities and information that are harder to process... As a result, they are more likely to become enduring aspects of human cultures, such as belief in the supernatural. (p. 720)

Yet the causal step from unseen agents to beliefs in Supernatural Agents remains poorly articulated in the HADD tradition (see Van Leeuwen & van Elk, 2018 for a review). A hypersensitive agency detection device may well cause anxiety about the possible presence of lurking agents but why these agents should be *supernatural* (rather than just human intruders,) is unclear, unless one already has beliefs about the supernatural.

Barrett and Lanman (2008) propose that it is not HADD experiences *per se* that cause the formation of supernatural beliefs but rather their interaction with other mentalising abilities which allow us to ascribe mental states and intentionality to these experiences (Barrett & Lanman, 2008). Therefore, the HADD experience is reinforced by pre-existing beliefs in supernatural agents and associated cultural notions, indeed these three factors are mutually reinforcing (Barrett & Lanman, 2008). It may well be, as Barrett and Lanman (2008) suggest, that HADD experiences, supernatural agent beliefs and certain cultural notions are mutually reinforcing but their reasoning seems circular. Under their account the formation of supernatural agent beliefs requires such beliefs to already exist. The question of how supernatural beliefs were formed in the first place remains unanswered.

In addition, the heavy emphasis that HADD theories place on agency detection intuitions as a major cause of religious beliefs seems misguided given that existing evidence has failed to provide support for this notion (Barnes & Gibson, 2013; Saler, 2010; van Elk, Rutjens, Van der Pligt & Van Harreveld, 2016; van Elk, 2013; Willard & Norenzayan, 2013). If agency detection intuitions do play a causal role in the genesis of religious beliefs, then it may be to a much lesser extent than many HADD theories have claimed. Cultural and social factors such as religious upbringing (Lanman & Burhmester, 2017) have been shown to play a substantial role in supernatural beliefs (Farias et al, 2017) and have generally been overlooked in the HADD tradition.

### **3.4. An Alternative to the HADD: The Interactive Religious Experience Model.**

According to Van Leeuwen and van Elk’s (2018) Interactive Religious Experience Model (hereafter IREM) one must distinguish between *general* religious beliefs and *personal* religious beliefs. General beliefs about supernatural agents (e.g., *God exists*) are not specifically about the person holding the belief and are formed primarily through cultural learning. In contrast, personal religious beliefs are beliefs about supernatural agents that contain indexical constituents[[26]](#footnote-26) (e.g., *I, me, myself, here, now*) and therefore exclusively refer to the holder of the belief (e.g., *Jesus visited* ***me*** *in hospital,* Van Leeuwen & van Elk, 2018). In addition, personal religious beliefs cannot be derived from general religious beliefs by inference alone (Van Leeuwen & van Elk, 2018, though see McKay, Jong & O’Lone, 2018).

So how then does one acquire personal religious beliefs? The IREM proposes that with general religious beliefs in place, believers may seek out situations that elicit low-level intuitions, such as agency intuitions (though not exclusively), which allow them to develop *personal* religious beliefs.

#### 3.4.1. Fleshing out the formation of personal (religious) beliefs.

The IREM articulates how believers develop personal religious beliefs (*God visited me*) from general religious beliefs (*God exists*) with reference to a particular type of information conveyed in indexical beliefs; *self-attached information* (Perry, 1979; 1990)*.* This is experiential information that agents derive from their own body or environment. This information reflexively binds that environment or sensation to the agent, i.e., it is *self*-relevant (Van Leeuwen & van Elk, 2018) and provides the foundation for indexical beliefs (e.g., *I saw a boat; I heard a noise*). In this way self-attached information is different to *objective* information (e.g., *2 + 2 = 4; Rome is the capital of Italy*) that consists of propositions that are true or false independently of whether the belief-holder believes them to be true, i.e., the truth value of objective information is not reflexively-bound.

Crucially, the self-attached information conveyed in certain experiences (i.e., agency intuitions) can link with objective information to create new *personal* religious beliefs. For example, a believer has a low-level intuition (e.g., sensing the presence of an invisible agent) that indicates *something* is happening to *me here now* (i.e., it conveys self-attached information). The self-attached information conveyed by the agency intuition links to objective information about supernatural agents depicted in culturally reinforced general religious beliefs. The result is a personal belief with the following content: *a particular supernatural agent appeared to me at a certain time.*

Low-level intuitions conveying self-attached information occur only rarely without prompting and the IREM suggests that people participate in religious practices to deliberately trigger the relevant intuitions that provide the content for personal beliefs. Such practices therefore serve to facilitate the formation of personal beliefs;moreover, once an individual has a personal religious belief, repeated participation in the relevant practice is necessary to sustain it (Van Leeuwen & van Elk, 2018).

#### 3.4.2. The Distinction between General and Personal Religious Beliefs.

Van Leeuwen and van Elk (2018) define a personal religious belief as a belief containing an indexical constituent (therefore referring to the person who holds the belief) and being about an individual’s own particular place within a wider religious narrative. Yet the distinction between personal and religious beliefs is often not clear. For example, they cite the following as an example of the content of a *personal* religious belief: *a witch caused my cousin’s illness.* Arguably this is not about “me” but *“my cousin”;* therefore, it is a belief about a supernatural agent causing someone *else* harm. Using Van Leeuwen and van Elk’s criteria it is more akin to a *general* religious belief (i.e., “General religious beliefs are beliefs about supernatural agents that *are not directly about the person holding the beliefs*” [emphasis added], p. 12).

To clarify, in the IREM it is not necessarily the case that a belief containing an indexical is automatically *personal* in the intended sense (if that were so then the example they give of a general religious belief “Jesus died for my sins” would be a *personal* religious belief by virtue of containing an indexical constituent). As Van Leeuwen and van Elk state, “knowing that a religious belief contains an indexical is not enough to secure its status as a personal belief, since if it was derived by logical reasoning from general beliefs it is not personal (p. 48).

According to the IREM one crucial feature of a personal religious belief is that it cannot be derived by inference alone from general religious beliefs. One might counter this claim. For example, take the general religious belief “God loves everyone”. Here it *does* seem that inference is sufficient to derive the personal religious belief “God loves me”. The first general belief entails the second personal belief (i.e., “me” is included within the scope of the quantifier “everyone”). Van Leeuwen and van Elk cite “God cured my cancer” as another example of a personal religious belief that could not have been derived by logic alone. Yet again it seems only a simple inferential step is needed from the general religious belief “God heals the sick” to derive the accompanying personal religious belief, “God cured my cancer” (see McKay, Jong & O’Lone, 2018).

For Van Leeuwen and van Elk though, when deciding whether a belief is personal or general one must look at more than the content of the belief:

It matters not that – in point of logical principle – there might have been some more general belief from which an indexical belief with the same content could have been derived. It matters, for us, only whether or not this is what actually happened. If one’s belief that Jesus visited me in the hospital did not in fact come from some logical reasoning on some general belief that Jesus visits everyone who is in the hospital (which seems a bizarre posit in any case), then the former belief counts as personal (and not otherwise). (p. 48)

#### 3.4.3. The role of transcendent experiences in religious belief.

Critcher and Lee (2018) also propose an account of the formation of religious beliefs that emphasises experiential origins. They suggest that when there is no direct evidence to substantiate a claim, individuals are influenced, and ultimately persuaded, by intuitions, feelings or experiences that are consistent with the claim’s truth (Critcher & Lee, 2018).

They further argue that the feeling of inspiration encourages belief in God because it produces a transcendent experience where one feels connected to something greater than the self (Piff, Dietze, Feinberg, Stancato & Keltner, 2015; van Elk, Karinen, Specker, Stamkou & Baas, 2016); “Such feelings are phenomenologically consistent with God’s existence and revelation; thus, they offer experiential evidence that God exists.” (Critcher & Lee, 2018, p. 2). Why should the experience of inspiration or transcendence encourage a belief about an Abrahamic God for example, rather than Mother Nature or some other cosmic life force? It is not clear on Critcher and Lee’s account (2018) why transcendental experiences would lead to *religious* beliefs rather than *spiritual* ones.[[27]](#footnote-27) They offer no mechanism for how to bridge the gap between “experiences” and the formation of religious beliefs that are recognisable within a framework of a specific doctrine.

In the IREM, however, this is nicely articulated. Believers deliberately seek out situations that elicit transcendent experiences or intuitions and interpret these according to a body of culturally prevalent religious beliefs. Thus, it is possible for an individual to have a relevant experience and to go on to form the personal *religious* belief ([The Christian] God visited me), if indeed they are Christian, rather than an experience that leads them to believe they encountered a supernatural agent that exists outside of the conceptual repertoire of their religious community.

Once in place, religious beliefs are proposed to have been an important factor in driving large-scale cooperation and social complexity. Perhaps more importantly for the purposes of this thesis is the intuitive idea that religious beliefs influence our moral behaviour. The following section provides a brief overview of the substantial body of research investigating the influence of religious beliefs on prosocial behaviour[[28]](#footnote-28).

## 4. Religion and Prosociality.

*“morality is everywhere the same for all men, because it comes from God”*

*~* Voltaire

Voltaire’s assertion that God determines morality has been pronounced repeatedly from antiquity to the present day[[29]](#footnote-29). From this we derive the long-standing assumption that without God we would not be moral; “If God does not exist, everything is permitted” (Dostoevsky, *The Brothers Karamazov*). This assumption is still widespread and recent studies have shown that a perceived lack of morality underpins much hostility towards atheists (Cook, Cottrell & Webster, 2015; Gervais, 2014; Gervais et al, 2017; Simpson & Rios, 2017). These lay intuitions about God and morality are given some credence by research that suggests a positive association between religious beliefs and prosociality (see Galen, 2012; McKay & Whitehouse, 2015 for a review).

For example religious beliefs and concepts have been linked with increased generosity and cooperation in experimental and public goods games (Ahmed & Hammarstedt, 2011; Ahmed & Salas, 2011; Benjamin, Choi & Fisher, 2010; Duhaime, 2015; Shariff & Norenzayan, 2007; cf. Benjamin, Choi, & Fisher, 2010; Xygalatas, 2013); increased willingness to help others (Malhotra, 2010; Pichon, Boccato, & Saroglou, 2007; Pichon & Saroglou, 2009; Van Cappellen, Saroglou & Toth-Gautier, 2016); increased forgiveness (Nieuwboer, Van Schie, Karremans & Wigboldus, 2015); and a decrease in self-interested and cheating behaviour (Aveyard, 2014; Bering, McLeod, & Shackelford, 2005; Carpenter & Marshall, 2009; Friese & Wänke, 2014; Laurin, Kay, & Fitzsimons, 2012; Mazar, Amir, & Ariely, 2008; Randolph-Seng & Nielsen, 2007).

Yet there are also many inconsistencies (see Day, 2017; Galen, 2012) such as failures to replicate previous findings (Gomes & McCullough, 2015; Miyatake & Higuchi, 2017; see Chapter 1, section 3.2) gender-specific effects (Pazhoohi, Pinho & Arantes, 2017; Willard, 2018) and evidence suggesting an association between religious belief and *anti*social behaviour (Blogowska, Lambert & Saroglou, 2013; Bushman, Ridge, Das, Key & Busath, 2007; Ginges, Hansen & Norenzayan, 2009; Johnson, Rowatt & LaBouff, 2010; LaBouff, Rowatt, Johnson, & Finkle, 2012; Saroglou, Corneille & Van Cappellen, 2009; Van Pachterbeke, Freyer & Saroglou, 2011).

### **4.1. Religion and Antisocial Effects.**

Although there is evidence to suggest a positive association between religion and prosociality, there are many examples in the literature that suggest the contrary, i.e., a link between religion and *anti*social behaviours and attitudes. For example, Blogowska, Saroglou & Lambert (2013) found that self-reported religiosity related to increased prejudice towards a moral out-group member (e.g., a homosexual individual). The results of a meta-analysis by Whitley (2009) found that negative attitudes towards homosexuals are associated with several measures of religiosity (Whitley, 2009). LaBouff, Rowatt, Johnson & Finkle, (2012) found that participants who were recruited as they passed by a religious building reported more negative attitudes to non-Christians.

Similarly, Johnson, Rowatt and LaBouff (2010) found that Christian religious primes in a diverse sample increased covert racial prejudice towards African-Americans. A meta-analysis by Hall, Matz and Wood (2010) found that across 51 studies (*N* = 22,075) greater religious identification (*r* = 0.10, 95% CI = [0.08, 0.12], extrinsic religiosity (*r* = 0.17, 95% CI = [0.14, 0.20]) and religious fundamentalism (*r* = 0.13, 95% CI = [0.10, 0.16]) all positively predicted prejudice towards racial outgroups (Hall, Matz & Wood, 2010).

Bushman, Ridge, Das, Key & Busath (2007) found that participants exposed to a biblical passage about aggressive retaliation sanctioned by God were subsequently more aggressive than participants who had read the same passage but without the reference to divinely sanctioned retaliation. Johnson, Li, Cohen and Okun (2013) also found that reminding participants of an authoritarian God increased reported intentions to engage in aggressive behaviour. Saroglou, Corneille and Van Cappellen (2009) found that religiously primed participants were, when encouraged by the experimenter, more vengeful to someone who criticised them than control participants. Religiously primed authoritarians were also more in favour of supporting societal norms despite the fact that doing so would cause harm (Van Pachterbeke, Freyer & Saroglou, 2011).

### **4.2. Literal Prosocial Behaviour: Punishment as a Case in Point.**

Many of the inconsistencies in the religious prosociality literature could well be due to a Westernised and liberal conceptualisation of morality (see McKay & Whitehouse, 2015) - one that views morality as consisting of “nice” behaviours. As McKay and Whitehouse write, “To identify morality with “niceness” is thus to ignore a plethora of moral concerns, motivations, and behaviors” (McKay & Whitehouse, 2015, p.4; see also Fiske and Rai, 2014). Ample evidence suggests a strong *positive* association between religiosity and the desire to punish (Applegate, Cullen, Fisher & Vander Ven, 2000; Batson, Schoenrade & Ventis, 1993; Miller & Hayward, 2008); in some cases support for extreme forms of punishment (Grasmick, Burskil & Blackwell, 1993, Grasmick, Davenport, Chamlin & Bursik, 1992; Leiber & Woodrick, 1997; Miller, 2006; Miller & Bornstein, 2006; Young, 1992). Is this evidence of religion promoting distinctly *anti*social effects? This depends, McKay and Whitehouse argue (2015), on how one construes the concept of prosociality (McKay & Whitehouse, 2015).

A literal view of prosociality encompasses all behaviours that further the interests of a group (Fehr & Gächter, 2002). Many of these behaviours are not “nice” (e.g., punishment) yet they play an important role in reinforcing cooperation within the group (Fehr & Gächter, 2000, 2002) and are therefore prosocial to the extent that they benefit the group as a whole. Similarly, increased hostility or aggression towards a member of an outgroup is literally prosocial to the extent that it protects and reinforces ingroup moral norms, which ultimately benefits the group (McKay & Whitehouse, 2015).

## 5. Supernatural Surveillance, Punishment and Morality.

*“Thus I will punish the world for its evil; and the wicked for their iniquity”*

*~* (Isaiah, 13:11)

In section 2 of this chapter, I discussed research suggesting we are deeply attuned to the presence of other agents and are, in particular, sensitive to being watched. However, it is not only conspecifics that we believe are watching us – most human societies have beliefs in Gods or divine agents who monitor them to varying degrees (e.g., Atran & Norenzayan, 2004; Bering, 2006; Boyer, 2007; Norenzayan & Shariff, 2008). The Supernatural Monitoring Hypothesis (SMH) proposes that beliefs in watchful supernatural agents positively influence moral behaviour because people *feel watched* (Norenzayan, Henrich & Slingerland, 2013; Shariff & Norenzayan, 2011). That is, beliefs in watchful and omniscient Gods entail a state of blanket surveillance, which could influence our moral behaviour via the same psychological mechanisms as normal cues of social surveillance. Baumeister, Bauer & Lloyd (2010) suggest “the belief that one’s actions are constantly and inescapably being observed by a divine being may be a strong stimulus and reminder to be aware of one’s actions” (p. 76).

One of the major findings in the Cognitive Science of Religion (CSR) has been to reveal the extent that supernatural agent beliefs are underpinned by mechanisms used in ordinary social cognition (e.g., Barrett, 2000; Bering, 2011; Boyer, 2007, 2003; Gray & Wegner, 2010). Thus, particular beliefs about supernatural agents (e.g., watchful Gods) may trigger specific responses involved in general social cognition. For example, in a series of studies Gervais and Norenzayan (2012a) found that participants explicitly primed with God concepts and participants primed to think about how others view them (a “people” prime) showed significantly higher attentiveness to the aspects of the self that are presented to others (i.e., *public self-awareness*)*,* than those in a control condition. But there was no significant difference in public self-awareness between the *God* prime and *people* prime condition, suggesting that thinking of God affects public self-awareness *in a way that is consistent with thinking about social evaluation by others.*

Even when presented with implicit God primes participants showed higher public self-awareness relative to controls. This latter finding suggests the effects seen in the explicit priming study were not caused by demand characteristics and further demonstrates that thoughts of God influence factors sensitive to perceived social surveillance (i.e., public self-awareness) and that they do so below conscious awareness.

### **5.1. Watchful Gods and the Fear of Punishment.**

The Supernatural Monitoring Hypothesisplaces key emphasis on the watchfulness of supernatural deities, i.e., they can see us when no one else can and therefore we adjust our behaviour accordingly (Norenzayan, 2013) as we would when being watched by conspecifics. This emphasis on the surveillance capacities of supernatural agents implies that, just as when other people watch us, the increase in prosocial behaviour is driven by a concern for reputation.

However, as Johnson says, “in this framing, Gods are more perverts than police” (Johnson, 2016, p. 25), and indeed if Gods were just voyeurs then what would it matter if they watch us? It matters because being observed by Gods entails consequences; deities are (to varying degrees) morally concerned agents (Roes & Raymond, 2003). When we transgress, they can intervene to administer rewards or punishments in this life, or the next (Johnson, 2016).

### **5.2. The Supernatural Punishment Hypothesis.**

A prominent body of research has stressed the importance of punishment in the evolution of human cooperation (Boyd, Gintis, Bowles & Richerson, 2003; Fehr & Gächter, 2000; Ostrom, Walker & Gardner, 1992; Sigmund, Hauert & Nowak, 2001; Yamagishi 1986). Without punishment, or at the least the credible threat of punishment, there is no sanctioning mechanism to stop those individuals who exploit public goods but do not contribute to the costs (i.e., “free riders”). This leads to the breakdown of cooperation (Kagel & Roth 1995; Olson 1965; Ostrom 1990). Punishment provides a deterrent to cheaters and pushes up the costs associated with free riding (Johnson, 2005).

However, punishment is costly and requires significant collective resources and effort. It is possible that individuals might contribute to the original public goods yet shirk their contributions to the cost of punishment (what is termed “second-order free riding”), leading once again to the breakdown of cooperation. The Supernatural Punishment Hypothesis (SPH) proposes that threat of divine punishment is an effective mechanism for upholding cooperation and that this threat has been a critical force in the evolution of large-scale cooperation and complex societies (Bering & Johnson, 2005; Johnson, 2005; Johnson & Bering, 2006; Johnson & Kruger, 2004; Johnson, 2015; Norenzayan, 2013; Norenzayan et al, 2016; Norenzayan & Shariff, 2008; Shariff, Norenzayan & Henrich, 2011; Schloss & Muray, 2011).

It has been suggested that moral codes and social norms can be maintained by beliefs in supernatural agents who are perceived to have the capacity to sanction those who transgress, whether in this life or the afterlife (Johnson & Kruger, 2004; Johnson, 2005). Moreover, once beliefs in punitive or moralising supernatural agents are well established then people can “offload” some of the costs of punishment thus providing an elegant solution to the second-order free rider problem (Johnson, 2005).

#### 5.2.1. Evidence for the Supernatural Punishment Hypothesis.

The main contention of the Supernatural Punishment Hypothesis is that *the threat of supernatural punishment* is the mechanism by which cheating is deterred and cooperation sustained (Johnson, 2016; Shariff & Norenzayan, 2011). There is a degree of support for this in the cross-cultural data. For example, in a study that looked at 186 human societies, beliefs in a punitive God were associated with several markers of social complexity such as size, sophisticated economies and policing and a decrease in selfish behavior and an increase in cooperation (Johnson, 2005).

Similarly, in a sample of 96 Austronesian societies supernatural punishment emerged as a driver of social complexity (Watts et al., 2014). Analyses of cross-cultural data sets (e.g., 583 global societies) have revealed that larger and more politically complex societies were associated with a tendency to have more beliefs in powerful deities with capacities for supernatural punishment (Botero et al., 2014). Purzycki et al. (2016) found that across eight diverse communities around the globe, increased impartial behaviour to fellow co-religionists was triggered by beliefs in powerful and punitive Gods. The more punitive people believed their deities to be, the greater the allocation amount given to geographically distant co-religionists (as compared to both themselves and other more local co-religionists) (Purzycki et al., 2016). Other researchers have suggested that increased trust among strangers may well be driven by shared beliefs in moralising, punitive deities (Norenzayan, 2013; Ruffle & Sosis, 2010; Shariff, 2011; Sosis, 2005; Tan & Vogel, 2008).

Prior to the formulation of the SPH, the “Hellfire Hypothesis” suggested that beliefs in supernatural sanctions deter individuals from criminal behaviour (Burkett & White, 1974; Higgin & Albrecht, 1977; Hirschi & Stark, 1969; Stark, 1996). More recent evidence suggests the same; the national crime rate in the US was found to negatively correlate with a belief in supernatural punishment (but positively correlate with a belief in heaven) (Shariff & Rhemtulla, 2012). Moreover, a meta-analysis including data from 60 published studies of the effect of religion on crime (Baier & Wright, 2001) revealed that religious beliefs (including beliefs in supernatural sanctions) and behaviours exert a moderate deterrent effect on criminal behaviour, *r* = -.12 (though see Jackson & Gray, in press).

Shariff and Norenzayan (2011) found that participants who viewed God as more punitive were less likely to cheat on an experimental task. In contrast, they found that participants who viewed God as more forgiving showed a greater tendency to cheat on the same task. Similarly, Yilmaz and Bahçekapili (2016) found that after being both explicitly and implicitly primed with punishing religious andsecular authorities (as opposed to non-punishing and neutral primes) individuals showed stronger pro-social intentions.

## 6. Powerful Gods and Individual Moral Responsibility.

*“There is one lawgiver, who is able to save and to destroy: who art thou that judgest another?”*

*~* (James, 4:12)

We can also calibrate our level of moral responsibility against the moral authority of powerful Gods, which then influences our behaviour. For example, the “Do not play God” principle (Sunstein, 2005) stipulates that one should refrain from engaging in morally relevant actions that disrupt the “natural order” of the universe. At its most extreme any deviation from divine destiny is viewed as a violation of the will of God[[30]](#footnote-30). A recent study found that this principle influences people’s moral preferences in the context of the “trolley dilemma” (Gangemi & Mancini, 2013). In this widely used paradigm participants are presented with the scenario of a runaway trolley headed towards five people on a track who will be killed if it is not stopped. They are then asked to choose between one of two morally undesirable outcomes. They can either choose to pull a switch that diverts the trolley onto another track where it will kill one person instead of five (the action option) or they can choose not to act at all meaning that all five are killed (the omission option). Almost all participants (96%) who chose the omission option justified their decision according to the “Do not play God” principle (i.e., *“I cannot decide who lives, and who dies”*). Similarly, Barak-Corren and Bazerman (2017) found that belief in a God who controls events in the universe reduced people’s sense of individual responsibility in ameliorating human tragedy[[31]](#footnote-31). Nieuwboer, Van Schie, Karremans and Wigboldus (2015) have conducted research demonstrating that priming participants with God concepts reduces perceptions of the moral responsibility of offenders (Nieuwboer et al., 2015). This suggests that powerful Gods provide a mechanism to outsource personal moral responsibility as well as the perception of the moral responsibility of others. In a series of studies, Laurin, Shariff, Henrich & Kay (2012) found that beliefs in powerful and intervening Gods, when salient, reduced people’s motivation to engage in altruistic punishment. This effect, they claimed, was due to people outsourcing responsibility for punishment to God, as a type of “social loafing” rather than fear of interfering in God’s will.

### **6.1. Divine Moral Licensing.**

Beliefs in a powerful God might well encourage individuals to offload moral responsibility (Laurin et al., 2012) but would they free people up to actively engage in immoral behaviours; a type of *divine* moral licensing? Beliefs in a God who intervenes in human affairs might encourage immoral behaviours by leading us to perceive temptation as a heavenly sanctioned sign, i.e., *it was meant to be*. Thus, a God “who is good” could lead people to view their misbehaviour as less immoral (Jackson & Gray, in press). A recent series of studies by Jackson and Gray (in press) provide compelling evidence for this notion. For example, participants who made divine attributions for immoral acts viewed them as more permissible. Across three of the studies (Studies 1-3), divine attribution, as opposed to general religious beliefs, emerged as a significant predictor of moral permissibility (Jackson & Gray, in press). Moreover, they found that divine attributions for immorality were most common in instances where human agency was ambiguous, as these led people to infer God’s agency (Studies 4-7). Finally, religious believers were more likely to engage in divine moral licensing than non-believers (Studies 9-13).

### **6.2. Religiosity and Punishment.**

The notion of God as the supreme lawgiver and judge is reflected throughout much religious doctrine, in particular Christian scripture[[32]](#footnote-32). In Islam too, the judgement of Allah, *hokm*, is distinct from human judgement. It is made explicit which one takes precedence: *"And Allah decides; there is no adjuster of His decision. And He is swift in account"* (Qu’ran, 13:41).This is potentially problematic; if punishment is the sole jurisdiction of God then conceivably one is interfering in his work by administering it oneself. If so, and no one contributes to punishment because of the fear of divine reprimand then cooperation is jeopardised (unless, of course, fear of divine reprimands is sufficient to curb first-order free-riding). Could punishing Gods then, as exemplars of divine order and authority, demotivate individuals from punishment?

A body of literature seems to suggest otherwise. Research has revealed a strong positive association between religiosity and the desire to punish (Miller & Hayward, 2008). For example, religiousness has been linked to increased punitive attitudes towards interpersonal harm; an increased probability of supporting harsh forms of punishment (Applegate, Cullen, Fisher & Vander Ven, 2000; Batson, Schoenrade & Ventis, 1993) and more stringent judgements of moral transgressions (Atkinson & Bourrat, 2011).

Religious fundamentalism and biblical literalism have also been found to be associated with greater punitiveness and support for the death penalty (Grasmick et al., 1993, Grasmick et al., 1992; Leiber & Woodrick, 1997; Miller, 2006; Young, 1992). For example, higher rates of religious fundamentalism in the US positively correlated with higher levels of imprisonment and capital punishment (Jacobs & Carmichael, 2001, 2004; Jacobs, Carmichael & Kent, 2005). Within the legal system a fundamentalist orientation among judges was associated with more frequent sentencing of African-Americans and less serious offenders (Myers, 1988).

This positive association between religion and punishment is problematic for the SPH’s proposal that beliefs in punitive supernatural Gods provide a way for people to “offload” some of the costs of punishment (thus solving the second-order free rider problem) (Johnson, 2005). The evidence above suggests that, in fact, religious adherents have *increased* punitive tendencies. What is relevant for this thesis however, are the specific conditions under which people displace moral responsibilities and judgements onto God. Of course, the strength of one’s affiliation to the religious group, or indeed any social group, plays a significant role in shaping moral behaviours and judgements. I turn now to a discussion of the influence of group affiliation on morality.

## 7. Group Affiliation and Moral Transference.

*“But I am an American and just in virtue of being one I can be tainted by the sins, if not of my father, then of my country.”*

~ Ken Taylor, Henry Waldgrave Stuart Professor of Philosophy, StanfordUniversity.

Research indicates that our moral behaviour is often influenced by other people, in particular those with whom we share a degree of psychological closeness (Gino, Ayal & Ariely, 2009; Gino & Galinsky, 2012; Gino, Gu & Zhong, 2009; Goldstein & Cialdini, 2007; Kouchaki, 2011; Moore & Gino, 2013). There is a long research tradition in social psychology that examines the influence of other people on our self-concept (Bem, 1967) and this work has been instrumental in better understanding the role that group affiliation has on our identities and subsequent behaviour. The prominent Social Identity Theory (Tajfel, 1981, 1982), for example, proposes that we derive our social identities primarily from the groups we belong to. Furthermore, we strive to maintain a positive social identity, and this can be done by making (overly) favourable ingroup comparisons with outgroups (Tajfel & Turner, 1979, 1986; see also Brown, 2000, for a review). One domain which plays a particularly important role in positive ingroup evaluation and group level self-concept, is morality (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Leach, Ellemers & Barreto, 2007; Levine & Campbell, 1972).

### **7.1. The Importance of Morality to Ingroup Identification.**

Cross-cultural research on ethno-centrism suggests that morality, above all other domains, is most important to positive ingroup evaluation. For example, in a global sample of societies morality was the only trait that ingroups continually ascribed to themselves more than to outgroups (Levine & Campbell, 1972). Although ingroups varied on the degree to which they perceived other attributes (sociability, strength, prestige, competence) relative to outgroups, this was not the case with morality (Leach et al., 2007). This is echoed in laboratory studies, which have shown that the more participants identify with an experimentally assigned ingroup the more they ascribe morality to this group. Even in a competence-orientated context where morality is not salient it still emerges as the most important factor in determining positive group evaluation (Ellemers, Pagliaro, Barreto & Leach, 2008; Leach et al., 2007).

In short, people who identify with their group have a bias to perceive that group as good and moral (e.g., Janis, 1983; Staub, 1997). These strands of research converge to suggest that morality is the most important dimension along which we assess the value of our group and that we are strongly inclined to perceive our group as morally superior to others (Roccas, Klar & Liviatan, 2006).

### **7.2. Vicarious Self-Perception.**

Research on vicarious self-perception demonstrates that people infer their own attributes by observing the behaviour of other group members, *as if* they had observed themselves performing the action (Galinsky, Wang & Ku, 2008; Goldstein & Cialdini, 2007). This claim is supported by evidence demonstrating that when perceiving the actions and emotions of similar others, the same neural mechanisms are employed as when we produce those actions and emotions ourselves (Conson, Salzano, Frolli & Mazzarella, 2017; Decety & Grezes, 2006).

In addition, experimentally inducing a sense of merged identity with a target by getting them to write a narrative from the target’s perspective has been shown to lead participants to perceive themselves as possessing the attributes of the target. They then incorporate these attributes into their own self-concept (Goldstein & Cialdini, 2007). Moreover, studies that expose participants to shared multi-sensory experiences have been found to alter their mental representations of their identity and to increase the perception of similarity with others (Paladino, Mazzurega, Pavani & Schubert, 2010; Tajadura-Jiménez, Grehl & Tsakiris, 2012). Similar effects have been documented in domains such as vicarious pain (Giummarra, Tracy, Young & Fitzgibbons, 2016; Jackson, Brunet, Meltzoff & Decety, 2006; Young, Gandevia & Giummarra, 2017) vicarious self-control (Ackerman, Goldstein, Shapiro & Bargh, 2009; Walter, Demetriades & Murphy, 2017) and vicarious cognitive dissonance (Blackman, Keller & Cooper, 2016; Cooper & Hogg, 2004; Norton, Monin, Cooper & Hogg, 2003).

### **7.3. Vicarious Moral Licensing.**

Unsurprisingly, given the primacy of moral information with regards to deriving a positive group identity, people readily incorporate *moral* information about group members into their own self-concept (Kouchaki, 2011). People not only use the past moral behaviour of similar others to inform their own self-concept but use these credentials to license subsequent immoral behaviour, a type of *vicarious* moral licensing (Kouchaki, 2011; Meijers, Noorewier, Verlegh, Zebregs & Smit, 2018; Newman & Brucks, 2017).

For example, Kouchaki (2011) asked white participants to indicate whether they thought an available job position was better suited for a particular race. Those who were told of the moral superiority of their group in a prior task were more likely to indicate that a job position was better suited for a white candidate than an African-American candidate (Kouchaki, 2011, Study 1). This suggests the past moral behaviour of the ingroup enhances one’s moral self-concept, which then licenses discriminatory attitudes on a subsequent task. An “intelligence” and “competitive” condition were also included to assess whether being told that one’s group is superior in other positive attributes (i.e., intelligence, competitiveness) produce the same licensing effects or whether it was specifically superiority in the moral domain that triggered vicarious licensing. The same vicarious licensing effect emerged when participants were given information about the superiority of their group’s intelligence, but not their competitiveness. It seems that in some circumstances even an unrelated domain can lead to vicarious moral licensing if it causes individuals to perceive themselves in a generally positive light (Kouchaki, 2011)[[33]](#footnote-33).

The degree to which a person is influenced by the behaviour of other group members depends on the strength of their identification with the group (Goldstein & Cialdini, 2007; Tajfel, 1982; Tajfel & Turner, 1979). Therefore, past ingroup behaviour can trigger vicarious moral licensing only if identification with the group is strong enough (Kouchaki, 2011; Newman & Brucks, 2017). For example, Kouchaki found that participants who were told of their group’s past non-discriminatory decisions expressed more prejudiced attitudes towards a Hispanic applicant on a subsequent job hiring task when they identified *more* with the ingroup (Kouchaki, 2011, Study 3).

Similarly, participants who strongly identified with a corporate brand (i.e., *Nike*) used information about that brand’s socially responsible behaviour to vicariously license subsequent selfish behaviour in the context of an economic game. Participants indicating lowers levels of “self-brand overlap” however, did not exhibit licensing behaviours (Newman & Brucks, 2017, Study 1). Meijers et al. (2018) conducted a series of studies to assess the vicarious moral licensing effect in dyadic relationships. They found that participants who both read about and imagined a close other (e.g., friend or partner) performing an environmentally-friendly action subsequently went on to behave in a less environmentally friendly way (compared to participants who read about or imagined their close other performing a neutral action) (Meijers et al., 2018). In line with both Kouchaki (2011) and Newman and Brucks (2017), Meijers et al. (2018) found that the degree of identification with the target moderated any vicarious licensing effects. Less environmentally friendly behaviours were only performed after participants read or thought about a close other’s ethical behaviour, rather than a non-close other. Meijers et al. (2018) also performed a meta-analysis on two of their studies, which revealed a significant vicarious moral licensing effect and yielded a small to moderate mean effect size.

### **7.4. Vicarious Moral Cleansing.**

Insofar as licensing and cleansing effects are in a hydraulic relationship, we would also expect to see evidence of vicarious moral *cleansing*. There is preliminary evidence for such an effect. Newman and Brucks (2017) assigned participants to either a *high self-brand overlap* condition (where they were asked to write how they were similar to a target brand, i.e., Nike) or a *low self-brand overlap* condition (participants were asked to write how they were different to Nike). They were then asked to read an article outlining either Nike’s positive socially responsible behaviour, Nike’s socially irresponsible behaviour or a control article. All participants then decided how many out of an allocation of 10 prize-draw entries they would give to a recipient. High self-brand overlap participants exposed to negative information about Nike’s socially irresponsible behaviour gave significantly more of their allocation to participants in both the positive information and control conditions thus indicating a vicarious moral cleansing effect. Thus, Newman and Brucks (2017) claim to have found evidence of a) vicarious moral cleansing and b) a superordinate moral balancing effect, i.e., where both vicarious licensing and cleansing operate together as part of a broader system of moral self-regulation in a symmetrical and systematic way (Sachdeva et al., 2009).

However, Meijer et al. (2018) found no evidence of a vicarious moral cleansing effect. Participants who were given information about a close other’s immoral behaviour did not show an increase in subsequent cleansing behaviours. The reasons why a vicarious moral licensing effect appears to exist, but a vicarious cleansing effect does not, will be discussed in detail in Chapter 7 of this thesis.

### **7.5. Collective Guilt.**

Vicarious moral cleansing effects can also be seen in individual responses to historical ingroup wrongdoing: *collective guilt.* Here people experience guilt for the past actions of their ingroup, often events in which they could not possibly have had any individual involvement (Doosje, Branscombe, Spears & Manstead, 1998). This prompts engagement in guilt-reducing and reparative behaviours, such as publicly apologising to victims and the giving of reparations[[34]](#footnote-34) (Branscombe, Spears, Ellemers & Doosje, 2002; Doosje et al, 1998; Hewston et al., 2004; Iyer, Leach & Crosby, 2003; Lickel, Schmader, & Barquissau, 2004; McGarty et al., 2005).

For example, a seminal study found that Dutch participants showed higher levels of collective guilt and reparative behaviours when confronted with negative narratives of their nation’s history of colonisation in Indonesia than participants presented with favourable narratives (Doosje et al., 1998). A series of longitudinal studies found that collective guilt in non-indigenous Chileans over mistreatment of the indigenous Mapuche predicted reparative attitudes (Brown, Gonzalez, Zagefka, Manzi & Čehajić, 2008).

Similarly, “white guilt” has been found to predict reparative attitudes to African Americans and endorsement of compensatory affirmative action policies (Swim & Miller, 1999; Harvey & Oswald, 2000; Iyer et al., 2003). Explicitly and implicitly reminding participants about their ingroup privilege increases levels of collective guilt (Greenaway, Fisk & Branscombe, 2017; Harth, Kessler, & Leach, 2008; Iyer et al., 2003; Knowles & Peng, 2005; Leach, Iyer, & Pedersen, 2006), especially when the reminders come from an ingroup source (Doosje, Branscombe, Spears & Manstead, 2006).

The experience of collective or group-based guilt has also been found to enhance willingness to engage in collaborative strategies with a target out-group (Solomatina & Austers, 2014), to motivate compensation (Čehajić-Clancy, Effron, Halperin, Liberman & Ross, 2011) and to promote intergroup forgiveness by encouraging the act of apologising (Hewston et al, 2004; Leonard, Yung & Cairns, 2015).

#### 7.5.1. Group Identification and Collective Guilt: A Paradox.

The mechanisms underlying collective guilt remain unclear and, if anything, research has revealed a paradox. If, as Social Identity Theory would suggest, people construct part of their self-identity on their group identification then one would expect those who strongly identify with their ingroup to experience greater group-based emotions than those who identify less with the ingroup. For example, a study of spectators at a university basketball game found that participants who strongly identified with the target team experienced greater enjoyment of the game compared to participants who identified with the team to a lesser degree (Wann & Branscombe, 1990).

A study with Israeli participants looking at the effect of collective guilt on the propensity to engage in collaborative behaviour (i.e., with Palestinians) found that the more participants felt collective guilt, the greater their readiness to cooperate. However, the stronger their *identification with the target ingroup* the less their readiness to collaborate with the outgroup (Solomatina & Austers, 2014). The results of Doosje et al’s study (1998) revealed that when Dutch participants were presented with explicitly unfavourable information about the history of Dutch colonisation, those who showed *low* levels of national identification *and* those who showed *high* levels of national identification felt more group-based guilt than when the information presented to them was explicitly favourable. Yet when presented with *ambiguous* information regarding Dutch colonisation, low identifiers were more willing to accept group-based guilt than high identifiers. This resulted in the curious scenario where individuals who identified *less* with their ingroup experienced greater levels of group-based guilt compared to individuals who strongly identified with the group.

Later studies too, found that under conditions of ambiguous information individuals with low identification felt more collective guilt than those with high identification (Schmitt, Branscombe, & Brehm, 2004). Research has indicated that highly identified individuals are the most motivated to protect their cherished group identity, particularly in the face of moral threat (Branscombe et al., 2002; Roccas, Klar & Liviatan, 2006). When presented with negative information concerning the ingroup they have a tendency to reinterpret this information so that they can still derive a positive social identity from membership to the group (Mummendey, Klink, Mielke, Wenzel & Blanz, 1999; Tajfel & Turner, 1979). Roccas et al. (2006) describe the paradox at the heart of group-based guilt:

Being identified with one’s group should be associated with experiencing stronger group-based emotions and thus should be associated with feeling stronger group-based guilt. But being identified with the group should also be associated with legitimization (*sic*) of the group’s wrongdoings and hence feeling little or no guilt. (p.699)

#### 7.5.2. Asymmetrical effects of (im)moral information on vicarious moral cleansing.

Moral and immoral information regarding the ingroup are not processed equal. It appears that while people readily attend to *moral* information about the ingroup they have tendencies to “morally disengage” with information that threatens ingroup morality (Bandura, 2016, 1999; Branscombe & Miron, 2004; Leidner & Castano, 2012; Mummendey et al., 1999; Tajfel & Turner, 1979) via a process of infrahumanisation (see Haslam & Loughnan, 2014, for a review) or outgroup derogation (Castano & Giner-Sorolla, 2006; Leidner & Castano, 2012; Leidner, Castano, Zaiser & Giner-Sorolla, 2010). One can attribute blame to an outgroup to alleviate collective responsibility (Rothschild, Landau, Sullivan & Keefer, 2012) or justify negative historical ingroup actions by reference to external circumstances (more so than for similar actions committed by other groups) (Doosje & Branscombe, 2003). These strategies are effective it is claimed because they alter the way that the victim or event is perceived.

Leidner and Castano (2012) propose a qualitatively different mechanism: *Moral shifting*. This process causes the individual to alter the moral frame of an event so that the transgression is seen as not immoral to start with, or even as an essentially moral act. Thus, group moral identity is protected not through exonerating cognitions but by assessing the morality of an event against different combinations of moral foundations (see Leidner & Castano, 2012)*.* According to the influential moral foundations theory (Haidt, 2012, 2001; Haidt & Graham, 2007; Haidt & Joseph, 2004; McKay & Whitehouse, 2015; Simpson & Rios, 2017) there are (at least) five psychological systems that provide the foundations upon which our multiple human moralities are constructed. Not all moralities converge on the same assessment of a particular event; one episode can be judged as utterly abhorrent when assessed in line with the foundation of Harm/Fairness for example but as a plausibly moral, and even laudable, act when assessed in line with different moral foundations, (e.g., Loyalty/Authority)[[35]](#footnote-35). Leidner and Castano write:

when people detect violations of internalized moral norms committed by fellow ingroup members, the accessibility and importance of moral foundations shift: loyalty and authority come to the foreground, whereas harm and fairness recede to the background. As a consequence, the ingroup’s behavior is less likely to be perceived as immoral (or it may even be perceived as moral) and the objective of protecting one’s social identity is met. (p. 42)

Leidner and Castano (2012) found some support for their claim. In a series of studies, they found that when confronted with atrocities committed by the ingroup (as compared to by an outgroup) participants subsequently used more Loyalty/Authority words than Harm/Fairness words. They also recognised Loyalty/Authority words more quickly and guessed an unknown Chinese character as having a Loyalty/Authority meaning more frequently after reading about ingroup committed atrocities (compared with outgroup committed atrocities).

What this suggests is that whereas moral information about the ingroup is readily incorporated into the self-image, the same is not true with regards to morally threatening ingroup information. Moreover, it provides further evidence that we draw upon more than just past individual moral behaviours to guide future moral actions suggesting once again that a simple compensatory mechanism is insufficient to account for people’s day-to-day moral behaviour.

## 8. Literature Synthesis.

*“So far, about morals, I know only that what is moral is what you feel good after and what is immoral is what you feel bad after.”*

~ Ernest Hemingway, Death in the Afternoon.

The literature reviewed in this chapter may seem disparate and disconnected. While it is true that I have reviewed several bodies of literature I deemed this necessary in the service of fully articulating the notion of moral transference. This is the cohesive strand that weaves together the seemingly heterogeneous literature described in this chapter. Moral Transference is, as the name might suggest, not limited to a particular body of research. It is, by its very nature, a mobile phenomenon. In this section, I will attempt to synthesise the theoretical perspectives outlined above to provide a clear through line to guide the reader through this thesis.

There are three main aspects, or *themes*, to moral transference that are central to this thesis and the various bodies of literature outlined above pertain to at least one of these themes. First, is the notion that the consequences of (im)moral behaviour are not subject to a strict chronology. For example, a transgression does not always entail immediate redress. To use the banking analogy again, one does not commit an unethical act and then pay for it immediately. The literature I have reviewed strongly suggests that people’s intuitions about (im)morality take place within an indeterminate timeframe. For example, the research on moral compensation shows that people can deal with the consequences of their past unethicality in the present, i.e., by engaging in moral cleansing. Similarly, the ‘value’ of a past moral act does not diminish over time; it is not valuable only at the time the moral act took place. People can “bank” a past moral behaviour and later recall it to licence immoral behaviour, i.e., moral licensing. And indeed, we can recall the past moral behaviours of group members and transfer the value to the present, leading us to engage in vicarious moral licensing. The literature on collective guilt shows that the consequences of a historical group wrongdoing can be transferred into the present. It is possible that this wrongdoing did not occur in your lifetime, yet the consequences are played out over a historical timeframe. Beliefs in powerful supernatural agents allow us to conceive of the consequences of immorality as being played out over a cosmic timescale. The concept of an all-seeing, punitive God for example, means that the consequences of an immoral deed are never-ending. If one is not punished on earth, then one has all of eternity to face the consequences of one’s misdeeds.

The second theme is that self-regulation alone does not determine our day to day moral behaviour. The wealth of research on surveillance and prosociality show that being watched profoundly influences our moral decision making and behaviour, even from a very early age. The need to safeguard and present a good reputation can lead us into sometimes extravagant acts of altruism. Similarly, beliefs in supernatural-watchers influence our moral behaviours, if not to appear good then to avoid punishment. The religious priming literature too indicates that certain religious concepts and beliefs can alter our everyday moral behaviours, decisions and judgements. We are a social species and our conspecifics also exert a strong influence over our day to day moral behaviour. We can use information about the morality of our group or individual members to inform our own moral self-concept, freeing us up to behave unethically.

The final theme is that certain moral concerns and responsibilities can be displaced onto others. Some of the literature reviewed in this chapter indicates that the consequences of an immoral act are not yoked to a particular individual. Again, as the collective guilt literature suggests, we can often take on the consequences of the wrongdoings of others, whether they be ingroup members or similar others. Beliefs in morally interested supernatural agents also offer a mechanism for people to relinquish moral responsibility.

Below is a synthesis matrix which further highlights the integration of the theoretical perspectives discussed in this chapter (Table 2).

Table 2. Synthesis matrix of the literature

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Moral compensation | Reputation and Surveillance | Agency detection | Powerful Gods and individual moral responsibility | Religion and prosociality | Supernatural surveillance and punishment | Group Morality |
| **Theme 1**  Moral consequences exist across an indefinite timescale | Past good deeds can motivate unethical behaviours (and vice versa). |  |  | The consequences of an immoral act can be delayed until they are addressed by a divine agent- across a comic timescale. |  | The consequences of a misdeed carry on to the afterlife; they exist across an eternal timescale. | Past historical wrongdoings of the group can be accounted for and addressed in the present. |
| **Theme 2**  moral behaviour is driven by more than simply balancing good and bad deeds |  | Reputation concerns can profoundly influence our moral behaviour and perception of our moral self. | The presence, or possible presence, of agents can shape our moral decision making and behaviour. | Beliefs in the moral capacities of divine agents can influence our moral decisions | Religious primes can influence our moral behaviour, (and indeed antisocial behaviour). | The presence of divine surveillance influences our moral behaviour, particularly when the supernatural agents can punish. | Information about the morality of one’s ingroup can determine subsequent moral decisions and actions. |
| **Theme 3**  Moral responsibilities and concerns can be displaced onto others |  |  |  | Beliefs in powerful Gods influence our perception of our moral responsibilities, e.g., we can ‘outsource’ some to Him. |  |  | Previous ingroup wrongdoings can lead to individual moral concerns and attempts to make ammends. |

The following five empirical chapters detail the eleven studies that form the body of this thesis. All have the over-arching goal of exploring instances of moral transference and are informed by and yoked to the literature described in this chapter. For example, in the next chapter (Chapter 3), I investigate whether the performance of a false confession can alleviate guilt for a past, unrelated transgression. This is based within the moral compensation literature, as indeed are the two studies outlined in Chapter 7 where I investigate whether information about ingroup immorality can trigger vicarious moral cleansing. The literature on Surveillance and Reputation forms the theoretical backbone to both Chapters 4 and 5. These chapters seek to determine whether the recollection of a past transgression increases surveillance perception (Chapter 4) and whether exposure to artificial surveillance cues triggers voluntary confession (Chapter 5). Chapter 6 draws upon the supernatural punishment literature to determine whether divergent God concepts have differential effects on support for state-sanctioned punishment. Chapter 7, as well as being informed by the Moral compensation literature, also draws on the group affiliation and morality literature. Here I seek to determine whether the type of group affiliation one has to an ingroup (i.e., identity fusion) can trigger vicarious moral cleansing. With that said, I turn now to the first empirical chapter.

# III. STUDY 1: DOES A FALSE CONFESSION ALLEVIATE GUILT FOR A PAST TRANSGRESSION?

*“It is the confession, not the priest, that gives us absolution.”*

~ Oscar Wilde, *The Picture of Dorian Gray.*

## 1. Introduction.

In 1947, the severely mutilated body of a young woman was found in a Los Angeles park. She had been sliced in half at the waist. This was the corpse of the victim the press dubbed “Black Dahlia”. Despite massive press coverage at the time and subsequent decades of interest, her killer has never been found. At the time, up to 60 people (mostly men) from across the United States came forward and confessed to her murder despite in some cases not even being in the same state at the time of the crime. Of these, only 25 were considered viable suspects by the LAPD[[36]](#footnote-36). In the years since the murder more than 500 people have confessed, many of who were not even born in 1947 (Corwin, 1996).

The sheer number of people willing to falsely claim responsibility for such a crime is striking and from a social psychological point of view, downright puzzling. False claims of responsibility for transgressions seem almost perverse because they violate a particularly robust false belief: No one would ever confess to a crime they didn’t commit[[37]](#footnote-37) (Douglas & Handrich, 2012; Kassin, 2017a). Moreover, given that we are deeply invested in projecting and maintaining a positive social reputation (e.g., to increase opportunities for being chosen as future co-operative partners) (Nowak & Sigimud, 1998a) it seems counterintuitive that some individuals would willingly admit to transgressions they did not commit, knowing that in doing so, they face probable retribution or punishment. In fact, seen in this light, even *genuine* confession is puzzling.

### **1.1. Voluntary False Confession.**

The reasons why people make false confessions are manifold and complex. Research over the past thirty years has investigated various psychological and situational factors that elicit both voluntary and coerced false confessions (Gudjonsson, 2003; Horgan, Russano, Meissner, & Evans, 2012; Kassin, 2017b; Kassin & Gudjonsson, 2004; Kassin & Kiechel, 1996; Meissner & Russano, 2003). Of these two types, a voluntary false confession seems the most bizarre. Why would an individual voluntarily accept responsibility for something they did not do? (Kassin, 2017a) The (non-exhaustive list) below outlines various factors that have been implicated in the genesis of voluntary false confessions (Kassin, 2017a; Kassin, 2017b; Kassin & Wrightsman, 1985):

* A scenario where others’ belief in your guilt is so strong that a false confession is made to lessen ensuing and inescapable punishment (Bering & Shackelford, 2005; Kassin, 2005; Shackelford, Liddle, Bering & Shalkoski, 2014).
* An inability to distinguish reality from fantasy. Breakdown in reality monitoring is normally associated with major psychiatric disorders and the occurrence of anomalous experiences and false memories (Johnson & Raye, 1981; McKay & Dennett, 2009; Radaelli, Benedetti, Cavallaro, Colombo & Smeraldi, 2013). However, some models of psychosis (Bentall, 2003; DeRosse & Karlsgodt, 2015) assume psychotic symptoms, such as deficits in reality monitoring, exist along a continuum and are present in the general population albeit in a milder form (Cohen, Eysenck & Levoi, 1986; Garrison, Bond, Gibbard, Johnson & Simons, 2017; Hicks, Marsh & Ritschel, 2002; Johns, 2005; Linscott & Van Os, 2013; Peters, 2010; Van Os, Linscott, Myin-Germeys, Delespaul, & Krabbendam, 2009).
* A pathological need for attention
* A condition called memory distrust syndrome where people develop distrust of their memory and have a tendency to rely on external cues and suggestions from others (Gudjonsson & MacKeith, 1982; Gudjonsson, 2003; Gudjonsson, 2017; Porter & Baker, 2015). For example, in an interrogation situation, a suspect’s memory of their alibi and innocence can be undermined by police, which leads them to potentially believe they have committed the offense but just cannot remember doing so[[38]](#footnote-38). This type of confabulation in the context of false confession is not typically due to neurological impairment but rather subtle psychological processes in situations of high emotional intensity (see Gudjonsson, Sigurdsson, Sigurdardottir, Steinthorsson & Sigurdardottir, 2014).

In this chapter, however, I explore a moral transference explanation where voluntary false confessions are an attempt to remove residual guilt, either from an unrelated transgression or because the person is congenitally high in trait guilt.

### **1.2. Moral Transference.**

The moral compensation literature suggests that for certain types of moral cleansing (i.e., behavioural cleansing) there is no intrinsic link between the transgression and the subsequent behaviour that one engages in to restore one’s moral self-concept (Jordan, Mullen & Murnighan, 2011; West & Zhong, 2015). This raises the possibility that via behavioural cleansing mechanisms, a false confession can provide a way to restore a moral self-concept damaged by an unrelated transgression. Similarly, previous research in forensic psychiatry has suggested that an unconscious need to remove guilt over past transgressions is a plausible factor in the genesis of voluntary false confessions (Gudjonsson, 2003; Kassin and Wrightsman, 1985). The feeling of guilt does not necessarily have to be associated with a previous identifiable transgression (Gudjonsson, 2003). For example, a study of “secondary psychopaths”[[39]](#footnote-39) (Gudjonsson & Roberts, 1983) found that the subjects’ poor self-concept and high trait anxiety were reflected in a constant feeling of guilt *regardless* of whether they were reporting an unethical act or not. In comparison, normal subjects only rated themselves as feeling guilty after they had committed an unethical act that violated a specific norm of behaviour. For Gudjonsson and Roberts, this finding implies that some individuals have a high level of generalised guilt which is unrelated to specific transgressions and this may influence their behaviour, including the need to volunteer a false confession (Gudjonsson & Roberts, 1983; Gudjonsson, 2003). Furthermore, this residual guilt can be related to imaginary acts, not exclusively to real ones (Gudjonsson, 2003). This highlights the potential for the transferenceof guilt, whereby, uncertain as to the source of the emotion, individuals falsely confess in an attempt to relieve it.

### **1.3. Incidental Guilt.**

Researchers in emotion processing have long identified that a type of transference occurs in the form of “incidental” emotions[[40]](#footnote-40) (e.g., Cohen, Pham, & Andrade, 2008; Loewenstein & Lerner, 2003)**.** For example, studies have found that positive affect induced by incidental events (e.g., a free lunch, good weather, receiving sweets) affects a range of cognitive processes and social behaviour (Isen, 1987; Schwarz & Clore, 1988; Razran, 1940). Recently researchers have turned their attention to incidental guilt (Jordan, Flynn & Cohen, 2015), where the source of guilt is unrelated to subsequent judgements (see Agrawal & Duhacheck, 2010; Hofmann & Fisher, 2012). In a series of studies, Jordan, Flynn and Cohen (2015) investigated whether feelings of guilt associated with one transgression could trigger forgiveness for another guilty individual. They found that trait and state-based guilt led transgressors to forgive unrelated third-party transgressors and this was mediated by their sense of identification with the transgressors. This demonstrates that guilt can be subject to transference i.e., the source of the guilt is unrelated to the subsequent behaviour (e.g., either the forgiveness of another or the false acknowledgement of responsibility for an unrelated transgression).

### **1.4. The Current Study.**

If voluntary false confessions are an attempt to remove residual guilt, then one might expect that having participants perform a false confession would lessen unconscious guilt.Therefore, the primary prediction of this study was that individuals who recalled a past transgression would go on to show lower levels of implicit guilt (here operationalised by performance on an Implicit Association Test, IAT; Greenwald, McGhee & Schwartz, 1998) if they subsequently performed a false confession than if they subsequently performed a control task.

An alternate possibility that I acknowledge is that, due to intuitive reasoning about the transmission of immorality (Eskine, Novreske & Richards, 2013; Rozin, Markwith & McCauley, 1994), having participants perform a false confession might increase levels of guilt. For example, individuals might feel morally tainted and contaminated by the performance of a false confession, leading them to feel guilty via a process of moral contagion (Eskine et al., 2013).

Given that removal of guilt enhances positive emotions (McKay, Herold & Whitehouse, 2013; Peer et al., 2014) and even physical wellbeing (La Barre, 1964; Major & Gramzow, 1999; Pennebaker, 1989, 1990, 1997; Smyth, 1998), I further predicted that false confessors would show higher levels of positive affect, and lower levels of negative affect, compared to controls. Finally, because confession is a feature of certain religions, I predicted that any effects would be more pronounced for religious individuals[[41]](#footnote-41).

## 2. Study 1.

### **2.1. Methods.**

#### 2.1.1. Participants.

A power analysis was performed for sample size estimation using GPower (Faul, Erdfelder, Buchner & Lang, 2009). With input parameters of alpha = 0.05, power = 0.8, *d =* 0.5, one-tailed, the projected sample size needed was *N* = 102. A total of 120 participants were recruited to ensure sufficient power despite exclusions due to the high error rate on the IAT (any participants with an error rate > 20% were excluded). Following IAT exclusions, 100 participants from Royal Holloway, University of London (37 males, 62 females. 1 who identified as ‘Other’; *Mage=* 24.11 years, *SD*= 5.85) were included in the final analyses. All participants received a £3 show-up fee at the end of the experiment.

#### 2.1.2. Measures.

*Implicit Guilt:*

Implicit guilt was operationalised by response times on a Single Category Implicit Association Task (SC-IAT; Karpinski & Steinman, 2006). The SC-IAT is a modification of the standard IAT, which measures the strength of evaluative associations (i.e., *guilty* and *not guilty*) with a *single* category or attitude object (i.e., *self*). Over two stages, just as in a traditional IAT, target words associated with the attitude object (i.e., *self*) and an evaluative dimension (*guilty* and *not guilty*) are randomly presented to participants. During the first stage, “good” words (i.e., *not guilty* words) and attitude object words (i.e., *self* words) are categorised on one response key, and “bad” words (i.e., *guilty* words) are categorised on another key. During the second stage bad words and attitude object words are categorised on one key and good words on a different key.

In this study each participant completed a total of 192 trials over two blocks. In one block participants were instructed to press the left “Ctrl” key if the target word that appeared on the screen belonged to the category "GUILTY" *or* "SELF” and to press the right “Ctrl” key if the target word on the screen belonged to the category "NOT GUILTY". In the other block they were instructed to press the left 'Ctrl' key if the target word belonged to the category "GUILTY” and to press the right “Ctrl” key if the target word belonged to the category "NOT GUILTY" or to the category "SELF". For all trials a red cross was displayed on the screen if participants responded incorrectly and a green tick appeared if they gave the correct response. The blocks were counterbalanced across all participants to avoid ordering effects. Individuals who closely associate the concepts “GUILTY” and “SELF” should make faster responses when those concepts share a response key. Because implicit measures are not consciously directed, they avoid the pitfalls of explicit measures (e.g., social desirability bias); I chose to use the SC-IAT for precisely this reason.

*Selection of final word sets for SC-IAT:*

For *guilty* and *not guilty*, two sets of candidate word lists, each comprised of five items were matched on several lexical variables obtained from the N-watch psycholinguistics programme (Davies, 2005). This programme computes (among other things) word frequencies based on the *Celex* database (Baayen, Piepenbrock & van Rijn, 1995). The two words groups did not significantly differ from one another on orthographic structure length *t*(9)= .79, *p*= .450; spoken *Celex* frequency *t*(9)= .13, *p*= .897; written *Celex* frequency *t*(9)= -.49, *p*=.633 or total *Celex* frequency *t*(9)= -.46, *p*= .657. The *self* words were taken from Karpinski and Steinman (2006, Study 2). The final word sets for the IAT were as follows:

* “guilty” words: *criminal, sinner, blameworthy, perpetrator, guilty*
* “not guilty” words: *innocent, pure, legitimate, virtuous, blameless,*
* “self” words: *participant’s first name, participant’s last name, me, I and myself.*

*Self-reported religious affiliation:*

Participants were asked to choose their religious affiliation from a list of twelve items (Appendix C).

*Positive and Negative Affect Scale:*

The positive and negative affect Scale (PANAS) is a 20-item self-report measure comprised of two mood scales: one that measures positive affect and one that measures negative affect (Watson, Clark & Tellegen, 1988). Participants are asked to rate on a scale of 1-5 (1=*Very slightly or not at all*, to 5= *Extremely*) items corresponding to positive affect (e.g., *Proud, Strong, Inspired)* and negative affect (e.g., *Jittery, Afraid, Irritable*). The PANAS has good internal consistency, with Cronbach’s alpha of 0.88 for positive items, and α = .87 for negative items (Crawford & Henry, 2004; Watson et al, 1988). In this study the wording instructed particip nants to base their responses on how they were feeling at the present moment (Appendix D). Positive affect and negative affect scores are computed by totalling the responses to each respective mood scale.

*Strength of belief in God:*

Participants were administered the Strength of Belief in God scale (Gervais, 2014). This asked them to indicate on a scale of 0 - 100 the extent to which they believe in God (or gods). They were told that if they felt certain that God (or gods) did *not* exist, to report this as “0” and if certain that God (or gods) *did* exist, then to rate this as “100”. Any participant who were absolutely uncertain either way was instructed to put “50”.

#### 2.1.3. Design and Procedure.

Following a battery of demographics questions all participants completed the religious affiliation and Belief in God measure. They were then asked to quietly recall a time they had behaved unethically (as in McKay, Herold & Whitehouse, 2013). Participants were informed they did not need to record or report any details of the unethical episode they recalled. They were then randomly assigned to a false confessionor controlcondition. Participants in the false confession condition were provided with a written narrative outlining a fictitious episode where they stole £40 (Appendix E). They were asked to read their narrative several times until they became familiar with the gist of the story and felt able to relay it (in the 1st person) to another participant, without having to read it off the sheet. They were instructed to include four key information points in their recall (see Appendix F). Participants in the control condition were given a neutral narrative to read (Appendix G) and instructed to include four key information points in their recall (Appendix H).

When participants indicated to the experimenter that they felt ready to repeat the narrative, a confederate was brought into the lab posing as a participant from a parallel study. Participants were given the cover story that this narrative recall task formed the main experimental condition for a study that was being run in an adjacent lab. All participants then completed the SC-IAT and the PANAS in isolation. These two tasks were counterbalanced across all participants. Finally, all participants completed a funnel debrief questionnaire to probe the strength of our cover story, which all participants indicated they had believed.

## 3. Results.

To compute an implicit guilt score for each participant I followed the procedure prescribed by Karpinski and Steinman (2006). For each participant the data from the practice trials (the first 24 trials in each block) were discarded, as were any non-responses and any responses less than 350ms. For the remaining participant scores, error responses were replaced with the block mean plus an error penalty of 400ms. The average response times of block 1 were then subtracted from the average response times of block 2. Finally, the result of this subtraction was divided by the standard deviation of all correct response times within blocks 1 and 2, yielding an IAT score where higher values indicated higher implicit guilt (Karpinski & Steinman, 2006).

A one-sample t-test was conducted to test whether implicit guilt scores were significantly different to zero. The results of the one sample t-test revealed that implicit guilt scores (*M* =-.202, *SD* =.287) were significantly lower than zero, t(99) = -7.029, p = .001. This points to a significant response advantage for the “self/not-guilty” pairing, compared to the “self/guilty” pairing.

IAT scores were non-normally distributed, as assessed by Shapiro-Wilks’s test (*p* < .05), therefore a Mann-Whitney *U* test was conducted to compare implicit guilt scores between conditions. Distributions of the IAT scores for the false confession and controlconditions were similar, as assessed by visual inspection. There was no statistically significant difference between false confession *(Mdn* = -.168) and control *(Mdn* = *-.*174)conditions, U = 1133, z = -.777, p = .437.

An independent *t*-test was conducted to check for differences in positive affect scores between the false confession (*M =* 24.698, *SD =* 8.238) and control (*M =* 25.702, *SD =* 7.896)conditions. The results were not significant, *t*(98) = .-.620, *p =* .537.

A Mann-Whitney U test was run (the distribution of negative affect scores significantly deviated from normal) to determine if there were differences in negative affect scores between conditions. Distributions of negative affect scores for the control and false confession condition were not similar, as assessed by visual inspection. Negative affect mean rank was statistically significantly higher in the false confession condition (59.70) than in the control condition (38.83), U = 705, z = -3.614, p = .001. This suggests false confession does effect negative affect, but not in the predicted direction. Rather than making participants feel better, the recital of a false confession made them feel worse.

Given the extent to which confession is institutionalised in many religious traditions, a moderated multiple regression was conducted to check for any effects of self-reported religious affiliation on implicit guilt scores. In the first block I entered a dummy variable for self-reported religious affiliation. I coded Atheist, Agnostic and None as 0 and the remaining responses to the affiliation question (Catholic, Protestant, Christian (other), Hindu, Buddhist, Muslim, Jewish and Sikh) as 1. I checked the written responses of participants who had indicated ‘Other’ and re-coded them accordingly. In the second block I simultaneously entered a dummy variable for participation in the false confession condition (participants in the false confession condition were coded as 1, all others coded as 0) and the interaction of this dummy with the self-reported religious affiliation variable. In Model 1, self-reported religious affiliation emerged as a (negative) significant predictor of implicit guilt scores[[42]](#footnote-42). The full model was not significant (see Table. 3).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.033 |  |  | 4.38 | 0.039 | 0.008 |  |  | 1.716 | 0.169 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | -0.153 | 0.037 |  | -4.151 | 0.001 | -0.178 | 0.051 |  | -3.491 | 0.001 |
| self-reported religious affiliation | -0.12 | 0.057 | -0.207 | -2.093 | 0.039 | -0.125 | 0.088 | -0.214 | -1.424 | 0.158 |
| false confession dummy |  |  |  |  |  | 0.054 | 0.074 | 0.094 | 0.727 | -0.469 |
| false confession dummy x self-reported religious affiliation | |  |  |  |  | -0.005 | 0.117 | -0.007 | -0.04 | 0.968 |

Table 3. Summary of multiple regression (predicting implicit guilt from self-reported religious affiliation, false confession dummy and their interaction)

Another moderated multiple regression was conducted to check whether strength of belief in God moderated any effects of the experimental condition on implicit guilt. In the first block I entered strength of belief in God (mean-centred). In the second block, I simultaneously entered a dummy variable for participation in the false confession condition and the interaction of this dummy with the strength of belief in God (mean-centred) variable. Neither regression model was significant (see Table 4).

Table 4. Multiple hierarchical regression (predicting implicit guilt from strength of beliefs in God, false confession dummy and their interaction).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | -0.004 |  |  | 0.574 | 0.451 | 0.006 |  |  | 0.381 | 0.767 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | -0.202 | 0.029 |  | -7.014 | 0.001 | -0.225 | 0.043 |  | -5.205 | 0.001 |
| Strength of belief in God | -0.001 | 0.001 | -0.076 | -758 | 0.451 | -0.001 | 0.001 | -0.075 | -0.464 | 0.644 |
| false confession dummy |  |  |  |  |  | 0.044 | 0.059 | 0.077 | 0.744 | 0.458 |
| false confession dummy x strength of belief in God |  |  |  |  |  | 0.00 | 0.002 | -0.019 | -0.122 | 0.903 |

## 4. Discussion.

I predicted that having recalled a past transgression, individuals who made a false confession would exhibit lower levels of implicit guilt relative to those who had not falsely confessed. However, there was no significant difference in implicit guilt scores across conditions. I will discuss possible explanations for these null findings as well as the other main findings of this study.

### **4.1. The Ambiguity of The Source of Guilt.**

In the present study participants were explicitly asked to recall a specific time when they behaved unethically, therefore the source of guilt (assuming the recall paradigm did induce guilt, see limitations section) was evident. It is possible that for a false confession to effectively alleviate residual feelings guilt in the hypothesized manner there must be some ambiguity regarding the source of the emotion. For example, in the forensic literature an individual’s habitual levels of unspecified guilt is a proposed factor in the genesis of false confessions (Gudjonsson, 2003). In the non-clinical population, the feeling of guilt is generally in response to the violation of a particular norm (Gudjonsson & Roberts, 1983) therefore a false confession for an unrelated transgression will not be effective as it is not rooted to the source of the guilt. However, for patients with clinical depression or OCD for example who both demonstrate heightened levels of generalized guilt (i.e., not rooted to a specific cause) (Berrios et al., 1992; Gudjonsson, 2003; Shafran, Watkins & Charman, 1996), a false confession may be undertaken in an attempt to exploit the cleansing function of genuine confession. Whether this would be successful or not, is a different matter.

### **4.2. Disingenuous Confession.**

Contrary to my prediction, levels of negative affect were significantly higher in the false confessioncondition. This might be related to the extent that one acknowledges responsibility for what one has *actually* done. For example, in a recent series of studies Pe’er et al. (2014) explored the emotional consequences associated with confessing to only *part* of a transgression. Participants who only partially confessed experienced an increase in negative emotions compared both to participants who a) chose not to confess at all and b) gave full confessions. Although dishonest participants correctly predicted that a partial confession would make them look credible to others, they failed to anticipate the negative self-directed emotional consequences associated with revealing only a part of their dishonesty.

A partial confession is not, in the strictest sense a false confession, given that one admits a degree of responsibility for a transgression they committed (rather than admitting to one that they did not commit). However, in neither instance does one confess to what one has actually done; both are “disingenuous” types of confession in that sense. Therefore, the findings that partial confessors experienced the highest level of negative affect compared to full confessors and non-confessors (Pe’er et al., 2014) and that false confessors experienced more pronounced negative affect than controls (Study 5) suggests that confession might well be an “all or nothing” affair. That is, for a confession to successfully alleviate guilt it has to be genuine, i.e., it must be to the full extent and it must pertain specifically to the transgression that caused the guilt.

### **4.3. Moral Indignation as a Distancing Response.**

To offset ethical dissonance[[43]](#footnote-43) following a transgression people employ a series of distancing strategies aimed at countering the threat to their moral selves. For example, after recalling a past misdeed individuals use an overcompensating ethical code, judge others more harshly and present themselves as more moral (Barkan, Ayal, Gino & Ariely, 2012). One could postulate that participants in the false confession condition employed a similar type of response. Participants may have experienced a type of moral “indignation” at being asked to confess to something they did *not* do as a way to distance themselves from their recalled unethical behaviour. This would explain the significant difference in negative affect between conditions which is due to indignation at being asked to admit to something they didn’t do.

### **4.4. Evolved Reputational Concerns Trigger Aversive Emotions Response.**

Another explanation for the finding that negative affect was higher in the false confession condition relates to emotional responses associated with reputational threat. We are, as a species, deeply invested in projecting and maintaining a positive social reputation to increase opportunities to be chosen as future co-operative partners (Milinski, Semmann, & Krambeck, 2002; Nowak & Sigimud, 1998; Pizarro & Tannenbaum, 2011; Van Vugt & Hardy, 2009). Thus, from an evolutionary perspective it would be in an individual’s interest to have an intuitive emotional response to reputational threat, such as in a situation where you are being adversely judged. During debriefing many of the participants in the *false confession* condition expressed their unease at thinking the confederate thought their confessions were genuine. One participant, for example, wrote, *“I didn’t like that I could be judged”*.

### **4.5. An Illusion of Innocence.**

The other main finding was the overall response advantage for the “self/not-guilty” pairing in the SC-IAT, compared to the “self/guilty” pairing. This result suggests that, in general, people have an implicit sense of innocence. This is in line with research on positive illusions that claims people have systematically positive misrepresentations of their personal qualities (Taylor & Brown, 1994; McKay & Dennett, 2009). Such illusions span a range of dimensions and include unrealistically positive self-evaluations, misperceptions of personal control and optimistic beliefs about the future. In general people tend to view themselves as “better than average” on most subjective and socially desirable dimensions (Myers, 2002), particularly in the moral domain (Tappin & McKay, 2017).

However, the interpretation that people feel implicitly innocent comes with a caveat. Our results in this respect could be an artefact of the particular words we chose; in particular the fact that our set of guilty words included *“criminal*”. If the words in the guilty set had included even more emotive words (e.g., *“rapist”* and *“murderer”*) then we would presumably have found an even more pronounced effect. It could be argued that although people may well not feel like a “criminal” (or on the stronger end of the semantic spectrum “rapists” or “murderers”), that does not entail that they view themselves as fundamentally innocent.

### **4.6. Religious Affiliation and Guilt.**

An unexpected finding was that religiously affiliated participants demonstrated lower levels of implicit guilt than non-affiliated participants. The autobiographical recall task involved the internalised recollection of a past transgression. This process is a salient feature of many religious practises and in Christianity especially, adherents are encouraged to sit and reflect upon their sins. The similarities between the recall task and the act of private prayer for example, are obvious. Therefore, the lower levels of implicit guilt seen in religiously affiliated participants might be caused by the recall task eliciting the feeling of being forgiven by God. In other words, their guilt had been divinely absolved before they performed the IAT. There is some evidence to suggest that the recollection of past transgressions elicits a feeling of forgiveness from God (van Oyen-Witvliet, Ludwig and Bauer, 2002). van Oyen-Witvliet, Ludwig and Bauer (2002) for example found that when people recalled a harmful, interpersonal transgression they reported more forgiveness from God (but less self-forgiveness, and less forgiveness from their victims) than when they thought about seeking forgiveness directly from the victim. This suggests that, for religious people, thinking or ruminating about one’s harmful behaviour leads to a sense of divine forgiveness.

## 5. Limitations.

Although conceptually novel, there were several limitations to this study. First, due to ethical constraints participants were not asked to report any aspect of the past transgression they recalled. Therefore, there was no way of assessing the homogeneity of the severity of the transgressions. Had all participants recalled a minor indiscretion rather than a severe misdemeanour (which, given that the sample included a large proportion of young undergraduate students, seems likely) then clearly feelings of guilt would not be that pronounced, potentially resulting in floor effects.

A second and related limitation was the absence of a manipulation check. Therefore, there was no way to validate whether or not the recall task did in fact induce feelings of guilt. Nor was it possible to rule out whether the recall task may have induced any unintended effects, such as feelings of shame[[44]](#footnote-44) for example, or anxiety. However, the administration of a manipulation check before the IAT may well have caused unwanted demand effects (Parrot & Hertel, 1999). The lack of a manipulation check is rendered less problematic in my view by the fact that autobiographical recall tasks are widely used methodology to effectively induce feelings of guilt (see Rebega, Apostal, Benga & Miclea, 2013, for a review).

## 6. Conclusion.

The main prediction of this study was not supported by the results, which yielded no difference in levels of implicit guilt between conditions. As discussed, one reason for this null result may be due to the ‘disingenuous’ nature of the false confession. Quite simply, perhaps for a confession to be effective it must be about the transgression that caused the feeling of guilt. The finding that negative affect was more pronounced in the false confession condition and the findings of Pe’er et al. (2014) suggest a necessary condition for the surge of positive emotions associated with confession is full and genuine disclosure.

The other finding of interest was the indication of an implicit innocence bias. Although as discussed this may be due to participants not associating with the emotive words chosen on the IAT. Therefore, exploring the potential existence of an innocence bias is a fruitful line of further investigation especially within the context of moral compensation.

In the next chapter, I again focus on the effects of recalling past misdeeds. However, there I explore whether the recollection of a past transgression makes us feel watched. This was, in part, motivated by the somewhat unexpected finding of this study that religiously affiliated participants demonstrated lower implicit guilt than non-affiliated participants. I have suggested that given its similarity to prayer rituals, recalling a past transgression may elicit a sense of divine forgiveness in religiously affiliated individuals. If so, this is predicated on the belief that God has access to one’s innermost thoughts; one is subject to divine surveillance. Yet despite religious prescriptions of God’s omniscience, research suggests we often conceive of God, particularly his mind, as being like our own (Barret & Keil, 1996; Jackson, Hester & Gray, 2018). Therefore, as it is for us, it is impossible for Him to attend to everything at once. We intuit that when something grabs His attention (such as recalling a misdeed) He becomes focussed on us and we are placed under divine observation. This would imply that, particularly for religious adherents, recalling a misdeed leads to heightened sensations of being watched. I explore this and the broader question of whether recalling a misdeed leads to heightened surveillance perception in the following chapter.

# IV. STUDIES 2 - 4: DO RECOLLECTIONS OF PAST TRANSGRESSION TRIGGER INCREASED SURVEILLANCE PERCEPTION?

*“The eyes of the Lord are in every place, Watching the evil and the good.”*

~ (Proverbs, 15:3)

*“Suspicion always haunts the guilty mind; The thief doth fear each bush an officer”*

~ William Shakespeare, Henry VI, part III.

## 1. Introduction.

The uncomfortable sensation that we are being watched while alone is an unsettling, yet familiar experience. As a psychological phenomenon, it has been of interest to researchers since the 19th century (Titchener, 1898; Coover, 1913). More recently, Johnson (2016) made the suggestive remark that these sensations are exacerbated in certain situations, specifically *after we transgress*, “Strikingly, the feeling [of being watched] is intensified precisely when we do not want to be watched, such as when we are doing something selfish, self-indulgent, or wrong” (Johnson, 2016, p.6.). Given the experimental evidence that suggests we possess a suite of sophisticated cognitive and behavioural responses to perceived reputational threat (see Chapter 2, section 2) I wondered whether recollections of past transgressions would trigger hypersensitivity to perceived social surveillance, in particular judgements of direct eye gaze.

In this Chapter I report three studies. The first (Study 2) reports the development of a novel measure of surveillance perception; a “portrait task” that consists of a selection of 46 portraits with varying degrees of eye gaze (i.e., averted, direct, ambiguous). Although the findings revealed no evidence that people exposed to a surveillance cue showed higher ratings of direct eye gaze on this measure, this wasthe case for a subset of participants (i.e., those who were religiously affiliated). In the second study (Study 3) I conducted a correlational analysis to assess whether there was any relationship between scores on the portrait task and a well-established measure of public self-awareness. The results revealed a modest positive correlation between the two measures as well as an association between guilt scores and public self-awareness. In the final study (Study 4) I investigated whether bringing to mind a past transgression led to heightened judgements of direct eye gaze in the portrait task. The results yielded supporting evidence for this prediction as well as revealing that religiously affiliated individuals made stronger judgements of direct eye gaze (although affiliation did not interact with theexperimental manipulation).

### **1.1. Heightened Social Awareness as a Response to Reputation Threat.**

Social reputation is important (Nowak & Sigmund, 1998, 2005; Leimar & Hammerstein, 2001; Sienen & Schram, 2006; Sigmund, 2012); particularly in large-scale cooperation where humans cooperate with genetically unrelated individuals, often in one-shot interactions (see Johnson & Krüger, 2004). We value those with good reputations and choose these individuals for cooperative interactions (Hardy & van Vugt, 2006; Nowak & Sigimund, 2005; Panchanathan & Boyd, 2004). Damage to one’s reputation entails significant costs (Fehr, 2004) and those with negative reputations are shunned or ostracised (Miliniski, Semmann & Kambeck, 2002; Over & Uskul, 2016). We are aware of these potential consequences and are motivated to engage in strategic behaviours to manage our reputations in front of potential collaborators from a very early age (Engelmann, Over, Herrmann & Tomasello, 2013; Engelmann & Rapp, 2018). Given the severe consequences associated with damage to reputation it seems likely that during episodes where they are under threat, such as when we transgress, we would pay particular attention to whether others are watching. This may well be the case even when privately recalling wrongdoing; we are often unaware of the social cues that motivate reputation-enhancing behaviours (Bateson, Nettle & Roberts, 2006) suggesting that strategic reputation management occurs below the level of consciousness.

### **1.2. Intuitive Direct Eye Gaze Detection.**

Eye gaze is a rich source of social information with *direct* eye gaze cues in particular triggering the process of mentalisation, (i.e., the decision to impute minds to others) (Khalid, Deska & Hugenberg, 2016). Evidence suggests that we tend to pay particular attention to cues of direct eye gaze (Conty, Tijus, Hugueville, Coelho, & George, 2006; Farroni, Csibra, Simion, & Johnson, 2002; Grossman, Johnson, Farroni, & Csibra, 2007; Senju & Hasegawa, 2005; Senju, Kikuchi, Hasegawa, Tojo, & Osanai, 2008; Senju, Yaguchi, Tojo, & Hasegawa, 2003; Wallace, Coleman, Pascalis, & Bailey, 2006; Yokoyama, Sakai, Noguchi, & Kita, 2014). This renders direct eye gaze detection a particularly salient aspect of agency detection and moreover one that is relatively easy to manipulate in the laboratory.

Infants not only preferentially attend to direct rather than averted eye gaze but also demonstrate enhanced neural processing for these cues (Farroni, Csibra, Simion, & Johnson, 2002; Grossman, Johnson, Farroni, & Csibra, 2007). In adulthood too, the brain preferentially processes cues of direct rather than averted eye gaze (Yokoyama, Sakai, Noguchi, & Kita, 2014). Several studies have found subjects to be quicker at detecting direct rather than averted eye gaze (Conty, Tijus, Hugueville, Coelho, & George, 2006; Senju & Hasegawa, 2005; Senju, Kikuchi, Hasegawa, Tojo, & Osanai, 2008; Senju, Yaguchi, Tojo, & Hasegawa, 2003; Wallace, Coleman, Pascalis, & Bailey, 2006) and cues of direct gaze reach conscious awareness faster than those of averted gaze (Stein & Sterzer, 2011). Remarkably, even in the absence of a functioning primary visual cortex (i.e., individuals who are cortically blind) there are still areas in the brain that continue to respond to cues of direct gaze (Burra et al., 2013).

### **1.3. Direct Eye gaze Judgements in the Face of Uncertainty.**

Studies have shown that in episodes of uncertainty people tend to assume that another person’s gaze is directed at them (Mareschal, Calder, & Clifford, 2013; Mareschal, Otsuka & Clifford, 2014). This tendency is also found when dealing with threatening stimuli, such as when looking at angry faces (Ewbank, Jennings & Calder, 2009; Harbort, Witthöff, Spiegel, Nick & Flecht, 2013) or during episodes of heightened social anxiety (Gamer, Hecht, Seipp & Hiller, 2011; Jun, Mareschal, Clifford & Dadds, 2013; Schulze, Lobmaier, Arnold & Renneberg, 2013). Therefore, in similarly threatening or stressful social contexts, such as when our reputation is threatened or when we fear exposure, we may be prone to making systematic over-attributions of direct eye gaze. Given that engaging in unethical behaviour carries risks of punishment by both conspecifics and deities, it seems plausible that when we *do* transgress, our cognitive system activates precautionary mechanisms of reputation management. One possibility is the activation of perceptions of social surveillance, specifically judgements of direct eye gaze.

### **1.4. Eye gaze Aversion and Guilt.**

We assume that people tend to *avoid* others’ eye gaze when they feel guilty. In everyday reasoning about human behaviour “not being able to look somebody in the eye” is taken as a sign of guilt. This is supported by recent research on the psychophysiology of guilt, which revealed a tendency to avoid looking at the eye region of a victim in episodes of interpersonal harm (Yu, Duan & Zhou, 2017). For example, Yu, Duan and Zhou (2017) asked participants to play several rounds of a dot–estimation task with two partners (in reality two confederates). In one condition participants were told that any of their failures on the task would cause electric shocks to one of the partners (*participant*-caused harm). In the other condition participants were told that any of one of the partner’s errors would cause the other partner to receive electric shocks (*partner*-caused harm). While watching video clips of the victim experiencing pain from the (simulated) electrical shock, participants who believed they had caused the pain (and were presumably high in guilt) fixated less on the victim’s eye region than those in the partner-caused harm condition (Yu et al., 2017, Study 1). In a second study, participants were again made to watch videos of a victim experiencing pain as a result of the participants’ previous errors; some were asked to focus on the eye region of the victim and others on the nose region. Those participants who were made to look at the eye region showed higher skin physiological responses to guilt than those who were made to look at the victim’s nose region (Yu et al., 2017).

A tendency for transgressors to avoid direct eye gaze with a victim does not rule out the hypothesis that recalling a past transgression will *increase* judgements of direct eye gaze. It is perhaps, precisely *because* we feel watched that we avoid the gaze of others. Avoiding someone’s gaze does not mean that you think they can’t see you. It can mean however, that you don’t want to see *them*. In fact, guilty people’s avoidance of others’ gaze might well be moderated by the extent that they judge others to be looking at them.

### **1.5. The Current Studies.**

The first two studies in this chapter (Studies 2 & 3) aimed to develop and validate a novel measure of surveillance perception. This was designed in response to previous research, which has relied on measures that do not quantify social surveillance perception directly (e.g., public self-awareness). The measure consisted of a set of portraits that varied in eye gaze direction. In this study participants completed the portrait task either with or without a surveillance cue (i.e., a video camera). I predicted that participants in the surveillance condition would make higher judgements of direct eye gaze across the series of portraits relative to controls.

The second study (Study 3) aimed to assess whether there was any association between responses on the portrait task and those on an established measure of public self-awareness i.e., the 3-item public self-awareness component of the Situational Self-Awareness Scale (SSAS) as used by Gervais & Norenzayan (Studies 1 & 2, 2012a). The final study (Study 4) aimed to test whether participants asked to recall a previous transgression would make stronger judgements of direct eye gaze in a set of ambiguous portraits relative to control participants.

## 2. Study 2.

In this study participants were shown a selection of European and Korean portraits on a computer screen and asked to rate the extent to which they thought the portraits were looking at them. The portraits had been used in a previous study on the cognitive appeal of direct eye gaze in portraiture (Morin, 2013; see Appendix I). Both types of portrait (i.e., European and Korean) had been previously classified (by two independent coders) as falling in one of three eye gaze categories: averted, direct or ambiguous (European Cohen's Kappa = 0.896, *S.E*. = 0.018; Korean Kappa = 0.689, *S.E*. = 0.096, Morin 2013). The main prediction of the current study was that participants completing the portrait task in the presence of a video camera[[45]](#footnote-45) would make stronger direct eye gaze ratings compared to controls. I also included a measure of religious affiliation to assess whether this might moderate any effects of participation in the surveillance condition on judgements of direct eye gaze.

### **2.1. Methods**

#### 2.1.1. Participants.

A total of 52 undergraduate students from Royal Holloway, University of London (41 males, 11 females; *Mage =* 20.02 years, *SD* = 3.11) were recruited. This sample size was in line with similar previous studies (e.g., Ellett & Chadwick, 2007, *N* = 60). Each participant was reimbursed either with partial course credit or with £2.

#### 2.1.2. Measures.

*Portrait task:*

Only male portraits were used. Each portrait was standardised to include only the head and torso and any extraneous background details were removed. This left a final set of 46 images consisting of three categories for each racial type (*Korean averted, Korean direct, Korean ambiguous, European averted, European direct, European ambiguous*). There were 26 ambiguous items in the total set of portraits.

*Portrait task structure and presentation:*

The portrait task was programmed and presented using the stimulus presentation software *Inquisit*. Each participant completed a single block of 46 trials. On each trial a portrait was displayed onscreen and participants were instructed to rate the extent to which they thought the portrait was looking at them (using a six-point scale anchored by 1 =*"Looking away from me"*; and 6 =*"Looking directly at me"*). The order of portraits was randomised for each participant. An average response score was calculated for each participant.

*Religiosity and religious affiliation:*

At the outset of the study participants were asked to indicate their religious affiliation from a list of 12 options (the same as in Study 1, see Appendix C) and to complete the Strength of Belief in God Scale (Gervais, 2014).

#### 2.1.3. Design and Procedure.

Participants were randomly assigned to one of two conditions: a surveillance(*N* = 26) or control condition (*N* = 26)*.* In the surveillance condition, participants performed the portrait task in the presence of a video camera that was placed in the middle of the room and mounted on a tripod. This was directed at the desk where participants completed the portrait task. When participants sat at the computer the experimenter went over to the camera and pretended to press record. The experimenter then informed participants it would be recording for the duration of the task. In reality, the camera was on but not recording[[46]](#footnote-46). Participants in the controlcondition completed the portrait task alone in the lab with no video camera present while the experimenter waited outside. All participants then completed a funnel debrief which was used to assess whether any had guessed the true purpose of the study, which none had.

### **2.2. Results and Discussion.**

An independent samples t-test was performed to check for differences in average response scorebetween conditions. There was no significant difference between the surveillance(*M* = 3.41, *SE =* .66) and controlconditions (*M* = 3.27, *SE* = .59), *t*(50) = .82, *p* = .415.

To explore whether religious affiliation might moderate any effects of surveillance condition on judgements of direct eye gaze I conducted a multiple regression analysis with average response score as the criterion variable. First, I computed a surveillancedummy variable (by coding all participants in the surveillanceconditionas 1, and all controls as 0) and a religious affiliation variable. This second variable was computed by coding all participants who listed a religious affiliation as 1, and those who selected “*Atheist”, “Agnostic”* or “*None”* as 0 (any participants who responded *“Other”* were asked to clarify their affiliative stance and then re-coded accordingly). I then entered the interaction of the surveillance dummy with the religious affiliation variable. I regressed average response scoreon religious affiliation, the surveillance dummy and their interaction (see Table 5 for means).

Table 5. Means for interaction in Study 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | **95% Confidence Interval** | |
|  | **Mean (average response score)** | **S. E** | **Lower Bound** | **Upper Bound** |
| Surveillance/religiously affiliated | 3.66 | .19 | 3.29 | 4.03 |
| Surveillance/non-affilaited | 3.23 | .16 | 2.91 | 3.55 |
| Control/religiously affiliated | 3.14 | .16 | 2.81 | 3.47 |
| Control/ non-affiliated | 3.42 | .18 | 3.06 | 3.77 |

There was no main effect of surveillance dummy on average response score (*β* = -.15, *p* = .435). However, the interaction between the dummy surveillance variable and religious affiliation was significant (*β* = .46, *p* = .046) (see Table 6). This indicates that religiously affiliated individuals made stronger judgements of direct eye gaze in the presence of a surveillance cue than religiously affiliated participants in the control condition and non-affiliated individuals in both the control and surveillance conditions.

In summary, the presence of a surveillance cue did not increase judgements of direct eye gaze in the portrait task overall (there was no main effect of surveillance dummy on average response scores). However, for religiously affiliated participants this effect *did* take place, i.e., the presence of a surveillance cue increased judgements of direct eye gaze in ambiguous portraits.

Table 6. Multiple regression (predicting average response score from religious affiliation, surveillance dummy and their interaction).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** |
|  | .039 |  |  | 1.70 | .180 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| intercept | 3.49 | .18 |  | 19.32 | 0.001 |
| religious affiliation | -.28 | .24 | -.23 | -1.16 | .253 |
| surveillance dummy | -.19 | .24 | -.15 | -.79 | .435 |
| religious affiliation x surveillance dummy | .70 | .34 | .46 | 2.05 | .046 |

One possible explanation for this null t-test result was that the sample size (*N* = 52) did not generate enough statistical power and was simply too small to detect an effect. One other possible explanation relates to the tacit acceptance of technological surveillance in modern society (Pinto & Nemorin, 2015). One upshot of this acceptance could be that any effects resulting from exposure to technical surveillance cues are diluted. The mean age of the sample was 20.13 years, and most participants will have grown up in a society of high surveillance, so differences that may have been pronounced 40 years ago may now be muted. For example, after being asked whether they thought the video camera was recording, several participants responded that they hadn’t even thought about it, thus indicating it was not consciously a salient feature for them.

The main finding however was that for a subset of participants (i.e., those who were religiously affiliated), the presence of a surveillance cue did have the intended effect. Religiously affiliated participants showed increased judgements of direct eye gaze in the presence of a surveillance cue. An obvious explanation for this finding is that because such individuals have beliefs in watchful and moralising deities, they are rendered particularly sensitive to cues of surveillance. This finding is in line with previous studies that have found highly religious people are more susceptible to false-positives in agency detection (Petrican & Burris, 2012; Riekki, Libdeman, Alenoff, Halme & Nuortimo, 2012; though see van Elk, 2013[[47]](#footnote-47)).

## 3. Study 3.

The aim of this study was to assess further the validity of the portrait task as a measure of direct surveillance perception. I investigated whether there was any relationship between an established measure previously used in the literature to quantify the related construct of public self-awareness and the portrait task. Following Gervais and Norenzayan (2012a, Study 1 & 2) I used the 3-item public self-awareness component of the situational self-awareness Scale (SSAS) as the established measure of social surveillance perception. Previous studies investigating factors contributing to perceived social surveillance have relied on measures like the public self-awareness scale that quantify *related* psychological constructs rather than measuring it directly. In addition, the public self-awareness measure is only a component of a scale designed to measure the super-ordinate category of situational self-awareness (Govern & Marsch, 2001). Therefore, validating the portrait task as a *direct* measure of surveillance perception would make a timely contribution and could potentially be of use to researchers who wish to measure surveillance perception directly in the future. I also included a measure of guilt to check for any association between guilt and judgements of direct eye gaze as assessed by the portrait task.

### **3.1. Methods.**

#### 3.1.1. Participants.

A power analysis with G\*Power (Faul et al., 2009) recommended an *N* of 319 to attain 95% power (*α* = .05, two-tailed) to detect a small effect of *r* = .20. I added 12% to account for potential attrition rates on MTurk. 358 participants were recruited online through Amazon Mechanical Turk and paid $1 each for their time. Thirty-four participants did not complete the study and three participants failed an attention check. These thirty-seven participants were excluded from subsequent analyses leaving a remaining *N* = 321 (169 males, 152 females; *Mage =* 35.08 years, *SD* = 11.74).

#### 3.1.2. Measures.

*Public self-awareness component of the situational self-awareness scale:*

The situational self-awareness scale (SSAS) is a 9-item state measure designed to quantify levels of public and private self-awareness and awareness of surroundings (Govern & Marsch, 2001). It consists of a private self-awareness component (*α* = 0.70) and a public self-awareness component (*α* = 0.82, see Govern & Marsch, 2001). In line with Gervais and Norenzayan (2012a, Studies 1 & 2), I used only the public self-awareness component to assess sensitivity to cues of social surveillance. Public self-awareness is a transient state in which people are aware of the impression they are making on others (Prentice-Dunn & Rogers, 1982). Not only are they aware they are the object of external evaluation they are also concerned that others view them positively (Solomon & Schopler, 1982). The public self-awareness component consists of 3 items: *“Right now I am self-conscious about the way I look.” “Right now I am concerned about what other people think of me.” “Right now, I am concerned about the way I present myself”*.For each item respondents are asked to indicate the number that best corresponds to their answer using a 7-point Likert-scale anchored by *(1) Strongly Disagree* through to *(7) Strongly Agree.* Public self-awareness scores were calculated by totalling the responses to these three items.

*The portrait task:*

The portrait task was the same as in Study 1. *Average response score* was calculated as before.

*The guilt inventory:*

The guilt inventory is a 45-item questionnaire designed to measure the multi-faceted construct of guilt (Jones, Schratter & Kugler, 2000). Respondents are instructed to indicate the number that best corresponds to their answer using a 5-point Likert-scale (anchored by *(1) Strongly Agree, (2) Agree, (3) Undecided, (4) Disagree* to *(5) Strongly Disagree*). The guilt inventory has good internal consistency, *α* = .79 (Jones et al, 2000; Kugler & Jones, 1992). It assesses guilt across three domains: trait guilt (e.g., *“Guilt and remorse have been a part of my life for as long as I can remember”*), state guilt(e.g., *“I have recently done something that I deeply regret”*) and moral standards (example item: *“I believe in a strict interpretation of right and wrong”*). Sub-scale and full-scale guilt scores are calculated by totalling responses to relevant items. Lower scores on the guilt inventory reflect *higher* levels of guilt.

#### 3.1.3. Design and procedure.

Participants first completed a series of demographics questions, which included a measure of their religious affiliation, before completing the SSAS, the portrait task, and the guilt inventory. The ordering of these tasks was randomised for each participant. After the first task all participants completed a two-stage attention check before completing the remaining two measures. Finally, all participants completed a funnel debrief and were thanked for their time.

### **3.2. Results and Discussion**.

The results revealed a significant positive correlation between public self-awareness and average response score. There was no significant association between guilt scores and average response score. There was however, a significant negative correlation between total guiltscores and public self-awareness, meaning that lower guilt scores (indicating *higher* levels of guilt) were associated with higher levels of public self-awareness scores. Correlations and descriptive statistics are reported in Table 7. Spearman’s rho was used for all correlations as all variables were measured as Likert-scale responses.

To assess whether religious affiliation moderated the association between average response score and PSA I conducted a regression analysis. First, I computed a religious affiliationvariable by coding all participants who listed a religious affiliation as 1, and those who selected “*Atheist”, “Agnostic”* or “*None”* as 0 (any participants who responded *“Other”* to the religious affiliation prompt were asked to clarify their affiliative stance and then re-coded accordingly). I then regressed average response score on religious affiliation, PSA and the interaction between these two variables. The regression model was significant *F*(3, 317) = 3.46, *p* = .018, with an *R2* = .032. However, religious affiliation, PSA and their interaction did not emerge as significant predictors of average response score.

Table 7. Correlations and descriptive statistics for Study 2 (*N* = 321).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | 1 | 2 | 3 | Mean | *SD* |
| 2. average portrait score | .983\*\* | 1.00 |  | 3.37 | .64 |
| 3. guilt total | .002 | -.004 | 1.00 | 122.36 | 18.23 |
| 4. public self-awareness | .121\* | .120\* | -.332\*\* | 10.39 | 4.68 |

*\*p* <.05 (two-tailed), \*\**p* <.01 (two-tailed).

One of the main findings was that as people felt guiltier (as indicated by *lower* scores on the guilt inventory) they experienced increased levels of public self-awareness. Although this was a correlational study and therefore it precludes any claims about causality, the direction of this association is one that compliments the main hypothesis of this study. The more we transgress, and presumably feel guilty, the more we feel ourselves to be the object of others’ attention and aim to act in a way that restores our reputation. Studies have found that public self-awareness can promote helping behaviour in a way that is motivated by reputational concerns, e.g., reducing the bystander effect (Van Bommel, Van Prooijen, Elffers & Van Lange, 2012). In a similar way I suggest that heightened feelings of being watched following a transgression activate impression management tools geared towards limiting any reputational damage.

The other finding of interest was the positive correlation between the responses to the portrait task and an established measure of public self-awareness. This provides some evidence to support the validation of the portrait task as a measure of surveillance perception. With these findings in mind I aimed to explore whether recalling a past transgression would heighten judgements of direct eye gaze on the portrait task.

## 4. Study 4.

In this study I examined whether recalling a past transgression increases perceptions of social surveillance, in particular judgements of direct eye gaze. Participants were asked to recall either, a past transgression or, a routinised daily activity. My primary prediction was that participants in the transgression recall condition would show higher ratings of direct eye gaze in a series of ambiguous portraits than controls. I also included measures of self-reported religious affiliation and strength of belief in God to a) explore the moderating effects of these variables and b) to further investigate the association between religiosity and heightened feelings of being watched found in previous research (Gervais & Norenzayan, 2012; Petrican & Burris, 2012; Riekki et al., 2013 and Study 2 of this thesis).

Research has found that showing people faces of visibly different out-group members, such as other races, is associated with intuitive threat responses (Phelps et al*.,*2000; Cunningham et al*.,*2004; Eberhardt, 2005; Lieberman, Hariri, Jarecho, Eisenberger & Bookheimer, 2005) suggesting that intuitive and unconscious indicators of potential threat are associated with visibly different out-group members (Ronquillo et al., 2007). This could lead people to make more judgements of direct eye gaze in racially dissimilar faces. As a precaution I included a measure of ethnicity to control for any possible effects of racial perception biases on judgements of direct eye gaze.

### **4.1. Methods.**

#### 4.1.1 Participants.

A power analysis recommended an *N* of 398 to attain 80% power (*α* = .05, one-tailed) to detect a medium effect of *d* = .25. I added 15% to account for potential attrition rates on MTurk. In total 457 participants were recruited online through Amazon Mechanical Turk and paid $1 for their time. Participants were excluded from analyses on the following conditions:

* They did not complete the study (138 participants[[48]](#footnote-48)).
* They guessed the hypothesis (18 participants)
* They indicated that they had not thought of a suitable incident for the recall task (16 participants)
* They indicated they had responded at random or not according to their true beliefs (3 participants)

After exclusions this left a remaining *N* = 282 (162 males, 117 females, 3 = Other; *M=* 33.75 years, *SD* = 9.89).

#### 4.1.2. Measures.

*Portrait task:*

In previous studies participants had repeatedly raised concerns about the length of the full 46-item task. Therefore, in this study a truncated version of the portrait task was used that included only the ambiguous portraits. A total of 26 items were randomly presented to each participant (16 European ambiguous, 10 Korean ambiguous, see Appendix I).

*Transgression recall:*

Participants in the transgression recall condition were instructed to recall a time in their life when they felt that they had behaved unethically (see Appendix J). All participants were assured that they would not be asked to disclose any information about the nature of the episode. They were asked to indicate whether they had a suitable episode to think about (“yes/no”). If they responded “no” they were taken to a screen that asked them to think carefully and were given several prompts (“it could be something said in an argument, or something from your childhood”). They were instructed to proceed when they had a suitable episode in mind. To encourage effortful recall a series of four yes-no questions were asked about the episode (*Think of the incident- picture it carefully, was there more than one person present? Did it happen during the day? Did it happen indoors? Was there a table present?).*

Participants in the neutral condition were asked to recall a neutral and routine activity they perform every day (see Appendix K). They were provided with the same prompt questions. Following McKay, Herold & Whitehouse (2013) participants were then asked to indicate on a scale of 1-6 how pleasant or unpleasant it was to recall the episode; how ashamed or unashamed they felt and how regretful or un-regretful they felt, with the higher scores being indicative of negative valence. There were significant differences between conditions in all three attributes. Participants in the transgression recall condition found the recall more unpleasant than controls*, U* = 4758.00, *z* = -7.698. *p* = .001; felt more ashamed than controls, *U* = 3506.00, *z* = -9.865, *p* = .001; and felt more regret about the episode than controls, *U* = 3086.00, *z* = -10.433, *p* = .001.

#### 4.1.3. Design and procedure.

Participants completed a series of demographics questions (including a measure of religious affiliation, ethnicity and political beliefs). All participants first completed a two-stage attention check (as discussed in Chapter 1, see Appendix A) where they were presented with a paragraph of information about research on affective judgement and context. They were asked to select from a list of emotions two that they had felt that day. However, in the main body of text the following instruction was given *“To show that you have read the instructions, please ignore the question below about how you are feeling and instead check only the "none of the above" option as your answer.”* To correctly pass the attention check participants had to select the “none of the above” option from the set of emotions. For those that failed first time, they were directed to a prompt screen that informed them they had failed an attention check and to please read each question carefully. They were then re-directed back to the manipulation check. No participants failed this a second time. Participants were then randomly assigned to the transgression recall or the neutral condition in which both completed a truncated 26-item version of the Portrait Task (i.e., ambiguous portraits only).

Finally, all participants completed a measure of their strength of belief in God (Gervais, 2014) and a funnel debrief during which they were asked if they had genuinely recalled an appropriate transgression when prompted. They were assured that they would not be penalised if they had not done so. Sixteen participants indicated they had not and were subsequently excluded from analyses. Finally, participants were informed of the true purpose of the study and thanked for their time.

### **4.2. Results and Discussion**.

Ratings of direct eye gaze for each level of condition were non-normally distributed, as assessed by Shapiro-Wilk's test (p < .05). Therefore, a Mann-Whitney *U* test was run to determine if there was a difference in average response scores between the transgression recall condition (*N* = 140) and the neutral condition (*N* = 142). The decision to use a Mann- Whitney U test was further determined by the fact that the DV was ordinal as well as non-normally distributed. The distribution of scores for both groups of the IV was identical, as assessed by visual inspection. Therefore, median values (rather than mean ranks) were reported (mean values are given in footnote below[[49]](#footnote-49)). There was a statistically significant difference in average ambiguous portrait scores between the transgression recall (*Mdn* = 98) and neutral (*Mdn* = 95) conditions, *U =* 8588, *z* = -1.975, *p* = .048, *r* = 0.12.

To check for any effects of self-reported religious affiliation on judgements of direct eye gaze I conducted a multiple regression. First a religious affiliationvariable was computed by coding all participants who listed a religious affiliation as 1, and those who selected “*Atheist”, “Agnostic”* or “*None”* as 0 (any participants who responded *“Other”* to the religious affiliation prompt were asked to clarify their affiliative stance and re-coded accordingly). Second, I computed a transgression dummy variable (participants in the transgression recall condition were coded as 1, all others coded as 0) and then the interaction of this dummy with the religious affiliation variable. I regressed average response score on religious affiliation, the transgression dummy and their interaction. Religious affiliationemerged as a significant predictor of judgements of direct eye gaze, (*p* = .033), and the transgression dummy emerged as marginally (negatively) significant (*p* = .057). The interaction between these two variables was not significant (see Table 8).

Table 8. Multiple regression analysis (predicting direct eye gaze from religious affiliation, transgression dummy and their interaction).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** |
|  | 0.019 |  |  | 2.80 | 0.040 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| intercept | 3.98 | .181 |  | 21.95 | .001 |
| religious affiliation | .251 | .117 | .179 | 2.15 | .033 |
| transgression dummy | -.216 | .113 | -.155 | -1.91 | .057 |
| transgression dummy x religious affiliation -.160 | | .165 | -.098 | -.967 | .335 |

## 5. General Discussion.

This chapter had two aims: first to develop and provisionally validate a novel measure of surveillance perception (the portrait task); and second to use this measure to investigate whether recalling a past transgression heightens judgements of surveillance perception. Across three studies I found evidence to support both these aims. The results of a correlational analysis (Study 3) revealed a significant correlation between responses on the portrait task and a well-established measure of public self-awareness. The main finding however (albeit only just at the level of significance) was that participants who recalled a past transgression made stronger judgements of direct eye gaze in a series of ambiguous portraits compared to those who recalled a neutral event (Study 4). Furthermore, both studies 2 and 4 yielded interesting results regarding religiosity and surveillance perception. First, religiously affiliated participants demonstrated stronger direct eye gaze judgements in the presence of a surveillance cue than non-affiliated participants (Study 2). Second, religious affiliation was a significant predictor of strength of direct eye gaze judgements in a set of ambiguous portraits (Study 4). In this section I will discuss these findings and address the implications for current research.

### **5.1. “God is watching”: Religion and Surveillance.**

In both Study 2 and 4, religious affiliation had interesting effects on surveillance perception. In Study 2 religiously affiliated participants in the surveillance condition made higher judgements of direct eye gaze. And in Study 4 religious affiliation emerged as a significant predictor of judgements of direct eye gaze. A ready explanation is that religiously affiliated individuals feel more watched *in general* as a result of pervasive beliefs in an all-seeing God(s). Previous studies have shown religious believers are more prone to illusory face perception (Riekki et al., 2012) and making false positive errors in face perception (Petrican & Burris, 2013; though see van Elk, 2013). The fact they should be particularly sensitive to cues of surveillance is perhaps explained by prolonged engagement in religion, which is replete with cues of supernatural surveillance that, furthermore, carry the reminder of punishment.

The interaction between religious affiliation and participation in the experimental recall did not emerge as a significant predictor of judgements of direct eye gaze, meaning that that the effect of recalling the transgression on surveillance perception was not stronger for religiously affiliated individuals. This is perhaps surprising and contradicts the suggestion at the end of the previous chapter (pg. 128). However, one upshot of believing in a pervasive and all-seeing God is that every transgression (even down to one’s immoral thoughts) is monitored. As religiously affiliated individuals’ feelings of being watched are high anyway, due to beliefs in a powerful watchful God(s), they may not necessarily feel any more watched than normal upon recalling past immoral activity because God already knows they have transgressed. However, for non-religious participants who perhaps do not have beliefs in an all-seeing God (or at least to the same degree) it might be that recalling a past transgression triggers heightened feelings of being watched because the fear of exposure is more salient.

### **5.2. Watchful Gods and Social Expansion.**

Beliefs in the existence of watchful Gods are cultural universals (Atran & Norenzayan, 2004; Barrett, 2000). Moreover, it has been proposed that such beliefs played a vital role in the expansion of human societies (Norenzayan et al., 2016; Preston & Ritter, 2013) by promoting moral behaviour while reducing the need for costly second-order policing (Bering & Johnson, 2005; Shariff, Norenzayan & Henrich, 2011). The findings from Studies 2 and 4 complement this notion. Early societies governed by religious institutions that were centred on beliefs in powerful watchful deities can be conceived of as proto “surveillance states”. Through repeated participation in religious rituals and exposure to cultural motifs of supernatural surveillance, group members were continually reminded they were being watched. A society policed by a watchful moralising deity who has the capacity for punishment entails a wider scope of jurisdiction than one policed merely by con-specifics allowing for the spread of large-scale societies (Norenzayan & Shariff, 2008; Norenzayan et al., 2016; Shariff et al., 2011; though see Baumard & Boyer, 2013).

The claim that moralising religions with beliefs in powerful watchful gods supported and perhaps encouraged the evolution of large-scale cooperation is controversial. Baumard and Boyer (2013) in particular contest this line of reasoning issue by highlighting the amoral character of many ancient Gods:

the gods of antiquity were generally not construed as being interested in people’s moral or prosocial behaviors. People did think the gods watched them, but that was to monitor the appropriate performance of rituals and sacrifices... As a result, there seems to be no reason to assume…that believers in such gods would have been more cooperative, or that this increased cooperation would have made their societies more successful.

(p. 276)

They cite the pantheons of ancient Sumer, Greece, Rome and Meso-America as possessing non-moral Gods; who were watchful but concerned with the appropriate performance of rituals and sacrifice rather than the transgressions of mortals (Baumard & Boyer, 2010). Moreover, large-scale and complex societies appeared millennia *before* the emergence of moralising religions. In the absence of expert knowledge of ancient religions, the extent to which the Gods of antiquity were moralising is a matter of interpretation and debate. Beliefs in moralising Gods *were* prevalent in ancient Mesopotamia in the third millennium B.C.E. for example. Furthermore, these beliefs in supernatural punishment were already well established having been inherited from the Sumerians (Prochownik, in press). The Gods of Ancient Greece are typically seen as amoral and capricious. Yet there were moral undertones to the interaction between mortals and ancient Gods. For example, rituals of “scapegoating” were common in Ancient Greece (Frazer, 1890/2009). Here a designated individual was sacrificed on behalf of the community so that the Gods may take away the collective sins of the community. Arguably these Gods were not disinterested in morality they were just not explicitly, or intrusively, moral (like a Christian God). Seen in this light, the focus shifts from explaining the ancient Gods’ laissez-faire attitude towards morality to explaining why certain deities (e.g., an Abrahamic God) are intensely concerned, almost obsessively so, with human morality.

## 6. Limitations.

One potential limitation was the high attrition rate documented in Study 4 i.e., MTurk workers who accepted the study but returned it unfinished (30.2%, 138 participants)[[50]](#footnote-50). Zhou and Fishbach (2016) tested MTurk dropout rates across a selection of six paradigms commonly used in published online studies (see Appendix L). They specifically included paradigms that are cognitively taxing (e.g., ego-depleting)[[51]](#footnote-51) or those involving memory recall (e.g., recall a low *vs*. high power episode). They concluded that there is cause for concern if dropout rates are 20% or above in studies using these common paradigms. Internal validity can potentially be jeopardised because of selective attrition (i.e., when participants self-select to drop out of a study due to the issues with the condition they were assigned to) and this increases as dropout rates gets bigger. As Zhou and Fishbach (2016) write:

The internal validity of an experiment is predicated on successful random assignment, which allows for unambiguous causal inference by statistically removing myriads of confounds. Whenever participants drop out of different conditions of an experiment for different reasons—that is, condition-dependent or selective dropout— confounds of the experimental manipulations are likely to be introduced, which would compromise the experiment’s internal validity and expose the experimenter to the risk of drawing false conclusions about the causal effect. Evidently, a high dropout rate opens the door for condition-dependent dropout to creep in. (p. 2)

The percentage of dropouts in study 4 (30.2%) is above the safety threshold set by Zhou and Fishbach (2016). Moreover, the recall task used in Study 4 was similar to the well-established power prime (Study 5, May & Monga, 2014) included in Zhou and Fishbach’s attrition rate replication study (2016). This task asks participants to recall either a time in their lives when someone else had power over them (low-power) or when they had power over someone else (high-power). They are then asked to describe the relevant situation in detail, including how they felt and what happened (see Appendix M for the text presented to participants in Study 5, May & Monga, 2014). Zhou and Fishbach (2016) found that out of a total of 156 MTurk workers there was a dropout rate of 33.8% for the low-power prime and one of 34.1% for the high-power prime, leaving 103 participants to complete the study. The similarities in both conceptual design and phrasing between the power prime and the guilt recall task in Study 4 are evident. The findings of Zhou and Fishbach (2016), in particular the high attrition rate documented with using the power prime, have immediate implications for the findings of Study 4.

Some researchers have recommended safeguards to reduce selective attrition when conducting online research[[52]](#footnote-52) (Horton, Rand & Zeckhauser, 2011; Reips, 2000) (e.g., pre-warning participants, appeals to conscience and stating commitment to the study). I put several of the recommended safeguards in place at the beginning of Study 4 on the information and consent page (see Appendix N). Here I explicitly a) informed participants what the study would involve, b) appealed to their conscience (i.e., telling participants that if large numbers of people quit this could affect the quality of data and c) got some assurance of commitment (i.e., getting participants to write that they do not mind performing memory recall tasks). These safeguards, in my opinion, render the attrition concerns of Study 4 less worrisome. As Zhou and Fishbach (2016) note attrition, in particular selective attrition, is an unfortunate feature of Internet research (Zhou & Fishbach, 2016). As such, Zhou and Fishbach (2016) argue that researchers should not only implement drop out reduction strategies (such as the ones I used in Study 4) but also openly discuss and report attrition rates to help manage threats to internal validity[[53]](#footnote-53).

## 7. Conclusion.

Across three studies I found preliminary support to validate a novel measure of surveillance perception I had designed. Scores on this measure correlated with scores on a well-established measure of public self-awareness. Religiously affiliated participants who completed this measure in the presence of a surveillance cue made stronger judgements of direct eye gaze than non-affiliated ones. This suggests that these participants are particularly sensitive to these cues by virtue of their beliefs in punitive supernatural deities. This finding complements a large body of research that proposes such beliefs were instrumental in the expansion of human societies.

In the final study, I found that individuals who recalled a past transgression made significantly higher direct eye-gaze ratings (of ambiguous portraits) than controls this result provides preliminary evidence for my primary hypothesis and for Johnson’s claim, stated at the outset of this chapter. Although there were limitations to this study (i.e., in particular the attrition rate on MTurk) this finding provides some justification for further research into the effects of wrongdoing on perceptions of social surveillance.

If recalling a past misdeed increases surveillance perception because we are aware that our reputation is under threat, then what would be the behavioural consequences? The obvious suggestion is that heightened feelings of being watched after transgressing would trigger reputation-salvaging behaviours, such as apologising or confessing. Therefore, in the following chapter I explore whether exposing participants to a surveillance cue after they have cheated in a task increases the likelihood that they voluntarily confess to the misdeed.

# V. STUDY 5: DO SURVEILLANCE CUES INCREASE THE LIKELIHOOD OF VOLUNTARY FALSE CONFESSION?

## 1.Overview.

Evolutionary accounts of voluntary confession have suggested that it evolved as a pre-emptive strategy to lessen probable retributive punishment (Bering & Shackelford, 2005; Shackelford, Liddle, Bering, & Shalkoski, 2014). As an advantageous social strategy, heuristic processes might in certain contexts trigger confession as an automatic behaviour thus making the urge to confess very hard to suppress (Rand et al., 2014). This may be the case in circumstances where one fears probable social exposure of wrongdoing. One gauge of this is the extent to which others witnessed the transgression. The study outlined in this chapter aims to complement this body of research by investigating whether surveillance cues increase the likelihood of voluntary confession among participants who cheat on an incentivised coin-tossing task.

### **1.1. The Puzzle of Confession.**

Confession is puzzling; given the potentially severe consequences of being identified as a transgressor, one might assume that behaviour that implicates oneself in the commission of wrongdoing would be avoided at all costs (Sznycer, Schniter, Tooby & Cosmides, 2015). Yet the urge to confess is real (Kassin & Gudjonsson, 2004; Weiner, Graham, Peter & Zmuidinas, 1991)[[54]](#footnote-54). One prominent account conceives of confession as an evolved pre-emptive strategy that served to lessen retributive punishment in the ancestral environment (Bering & Shackelford, 2005). Even today punishment entails fitness costs such as restriction to vital resources (i.e., food and mates), group ostracism and even death (Shackelford et al., 2014; Williams, 2007; Williams & Nida, 2011). Given that research has suggested people internalise socially advantageous strategies (Rand et al., 2014), Shackelford et al. (2014) argue that unbidden, or voluntary, confession is not under conscious control thus making it intuitive and, in certain situations, difficult to suppress; particularly in scenarios where one fears social exposure.

### **1.2. Confession as Strategic Reputation Management.**

Acknowledging responsibility for a transgression can have negative consequences (Shackelford et al., 2014); this is something we are aware of even from a young age (Rosenberg, Betts, Eisner & Ribeaud, 2011). However, experimental evidence suggests that confession can, in fact, be an effective reputation management strategy. For example, compared to individuals accused of a transgression who do not confess those who *do* elicit more feelings of forgiveness from their victims; are perceived as more credible and trustworthy and judged to be less likely to re-offend (Darby & Schlenker, 1982; Kirby & Johnson, 2005; Pe’er, Acquisti & Shalvi, 2014; Rumsey, 1976; Sternglanz, 2010; Weine et al., 1991). Similar effects have been noted following an apology, even in children (Darby & Schlenker 1982), but it is confession that is the most effective method for altering the negative opinions of others (Benoît & Drew, 1997; Holtgraves, 1989).

Pe’er, Acquisti and Shalvi (2014), for example, found that individuals who read a vignette about a protagonist who offered a confession (full or partial) rated him as more credible and honest than protagonists who did not confess at all (i.e., despite the fact they could be potentially innocent) (Pe’er et al., 2014). Even children are sensitive to the positive aspects of confession; studies have demonstrated that 7-9-year-olds anticipate parental approval following a confession (Smith & Rizzo, 2017). The benefits of confession are also culturally reinforced as Weiner et al. (1991) note, by a lay association between confession and forgiveness (e.g., reflected in sayings such as “a fault confessed is half forgiven”) that implies the act of confession entails absolution[[55]](#footnote-55).

### **1.3. Confession and Punishment.**

With regards to the effect of confession on punishment there is evidence to suggest that a voluntary confession can reduce ensuing punishment. For example, Utikal (2012) found that individuals who confessed to a harmful failure were significantly less likely to be punished by a retributive partner than those whose failure was randomly detected. In the law courts too, confession can play a role in mitigating punishment. For example, in a sample of 1,009 Chinese criminal court cases confession was associated with a lower risk of imprisonment and when punishment *was* administered, confession elicited shorter sentences (Lu & Miethe, 2003).

In 2004, sentencing guidelines for judges in the UK encouraged reduced penalties for criminals who confessed (Verkaik, 2004). The guidelines further stipulate that the point at which the criminal confesses determines the sentencing reduction. For example, following a murder charge, if the defendant confesses at the first reasonable point they are entitled to a sentencing discount of as much as a third (e.g., A 15-year murder sentence is reduced by 5 years). However, a confession given *after* the jury has retired results in no reduction to sentencing. These lines of evidence suggest, as Bering and Shackelford propose (2005), that a strategic confession can effectively reduce ensuing punishment.

### **1.4. Voluntary Confession in Response to Imminent Social Exposure.**

According to Shackelford et al. (2014), the more one fears social exposure following a transgression, the more likely it is that one will voluntarily confess. One measure of probable social exposure is whether there were witnesses to a crime*.* Interrogators in modern settings frequently exploit this to secure confessions and routinely present suspects with false evidence (such as the supposed existence of eye-witnesses) (Henkel, Coffman, & Dailey, 2008; Kassin et al., 2007; Leo, 2008). This is an effective way to generate true confessions (Inbau et al., 2001; Jayne & Buckley, 1999; Leo & Liu, 2009) and not surprisingly is widely used by interrogators (see Forrest, 2012; Kassin et al., 2007). There is also evidence to suggest this is an effective strategy to elicit confessions in children. A recent study found that children aged 4-7-years-old who had broken a toy were 1.6 times more likely to voluntarily confess if their parent told them another adult had seen them break it (and wanted them to tell the truth), than if their parent did not provide this information about a witness (Rush, Stolzenberg, Quas & Lyon, 2017).

Furthermore, in a study conducted with prisoners in an Arkansas state prison, detainees explicitly reasoned about their decision to confess in terms of probable threat of social exposure. Their decision to confess was associated with stronger beliefs in exposure and anticipatory anxiety about being caught (Shackelford et al., 2014). Admittedly these findings were based on prisoners’ post-hoc rationalisations for their confessions, but studies in the lab have revealed that informing participants of the existence of witnesses or surveillance increases the likelihood of false confession (e.g., Horselenberg et al, 2006; Rassin & Israëls, 2014; Kassin & Kiechel, 1996; Redlich & Goodman, 2003). For example, in studies using a computer crash paradigm (where participants are told not to press the ALT key, as the system will shut down; unbeknownst to participants the computer is rigged to crash, Kassin & Kiechel, 1996) telling participants that another individual in the room saw them press a forbidden computer key led to a false confession rate of 60.6 -69 % (Perillo & Kassin, 2010 (Study 1) and Kassin & Kiechel, 1996 respectively). Similarly, 50 % of honest participants falsely confessed to colluding with another student (in reality a confederate) on an experimental task when they were told that there was a hidden camera in the room (Perillo & Kassin, 2011, Study 3). Real world examples abound in the forensic literature (see Kassin & Gudjonsson, 2004) and given the high frequency with which many interrogators use false evidence ploys[[56]](#footnote-56) there have been many calls for reform[[57]](#footnote-57).

### **1.5. Moral Outrage as a Response to Reputational Threat.**

Given the importance of our social reputation (Wu, Balliet & Van Lange, 2016; see Chapter 2, Section 2) and the negative consequences associated with a bad one (e.g., Feinberg, Willer & Schultz, 2014; Kurzban & Leary, 2001) when it is under threat it is emotionally aversive. This is particularly when the threat is unjustified. For example, a report of individuals falsely accused of child abuse report documented the presentation of a range of severe psychological consequences, such as anxiety, depression, suicidal ideation and symptoms of PTSD (Burnett, Hoyle & Speechley, 2017). Similarly, a study of individuals accused of workplace bullying documented a similar host of negative psychological outcomes following their accusation (Jenkins, Winefield & Sarris, 2011). What was also documented in these individuals was a strong sense of injustice. In a study on the emotional effects of full *vs.* partial confession, Pe’er et al. (2014) found that asking honest participants if they had cheated on an online coin-toss task elicited a similar sense of injustice; whereby participants asked if they had cheated expressed feelings of moral ‘outrage’.

### **1.6. The Current Study.**

It has been suggested that the fear of social exposure can trigger confessional behaviour (both true and false) as a pre-emptive strategy to lessen punishment. In this study I explore whether exposure to artificial surveillance cues following cheating behaviour would pose the relevant threat and therefore trigger voluntary confession. Cheating behaviour was measured with an incentivized online coin-toss task (taken from Pe’er et al., 2014). Participants were instructed to “toss” the coin 10 times and to report the number of heads they got. They were told they would receive a financial reward for every head they reported thus incentivizing the over-reporting of heads. In reality this coin-toss task was programmed to only show 2 heads so any reported number greater than 2 was indicative of cheating.

It was crucial that participants believed the coin-toss task was genuinely random and given that levels of non-naiveté have been documented in MTurk workers (Pe’er, Samat, Brandimarte & Acquisti, 2017) I decided against using this research platform and instead used Prolific Academic. Studies have revealed that users of Prolific report less familiarity with common experimental tasks and higher levels of honesty and naivety than users of MTurk (Pe’er et al., 2017). As the current study involved an element of deception (i.e., it relied on participants believing the coin-toss was both genuine and random) Prolific Academic seemed preferable[[58]](#footnote-58).

The main prediction of this study was that participants who cheated on the coin toss task would be more likely to confess when exposed to a surveillance cue than to participants who saw no such cue. Therefore, the key DV was confession; this was a binary variable computed as follows: any cheating participant who reported the number of heads they received as 2 (the true amount) when asked to *honestly* report the number of heads they got was coded as having confessed. All others were coded as non-confessors (i.e., 0).

Threats to one’s social reputation are emotionally aversive particularly when they are unjustified. Therefore, I further predicted, in line with Pe’er et al. (2014), that honest participants would express higher levels of moral outrage compared with those who cheated. Moreover, given that the presence of surveillance cues enhances reputational concerns (Engelmann & Fischbacher, 2009; Engelmann, Herrmann & Tomasello, 2012; Milinski et al, 2002), I further hypothesised that this moral outrage effect would be most pronounced for honest participants in the surveillance condition.

## 2. Study 5.

### **2.1. Methods.**

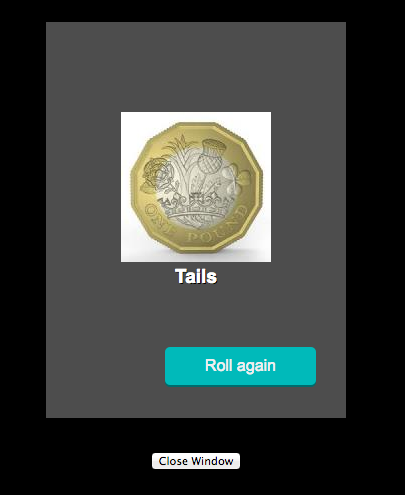
#### 2.1.1. Participants.

Following the findings of Pe’er et al. (2014)[[59]](#footnote-59), to ensure there were at least 50 cheaters in each condition I decided on 100 participants per condition and added an additional 20% of the sample size to account for potential attrition on Prolific Academic (*N* = 240). After excluding participants who did not complete the study (29 participants), or who indicated they had not understood the coin-toss task instructions (16 participants), 195 participants were included in the final analyses. All participants were paid a flat fee of £0.50 with the possibility of an additional payoff (up to a maximum of £1.50, including show-up fee) depending on the number of heads they reported. All participants were told that the study was designed to investigate the effects of enhancing people’s predictive abilities and actual accuracy on subsequent forecasting tasks.

#### 2.1.2. Measures.

*Coin Toss Task*

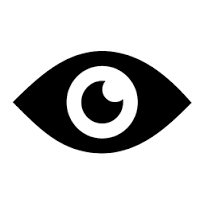
The cheating measure was a coin-tossing task previously used in Pe’er et al. (2014). Participants were told to click on a link to open an online random coin generator. It was programmed in Qualtrics so that it appeared in a separate window on the participant’s computer (see fig. 2). Each participant was told that they would receive an extra £0.10 for each head they reported in addition to the standard participation fee of £0.50. Participants were then instructed to toss the coin 10 times and report the number of heads they got. The task had been programmed so that only 2 heads appeared within the first 10 tosses; therefore, any participant who reported more than 2 heads was coded as having cheated. The task was also set so that participants could toss the coin more than 10 times, so they could check it was legitimate. The 11th toss was rigged to show a head, thus incentivising participants to include this in their reported number of heads. I also computed the extent to which they had over-reported by subtracting two from the number of reported heads, yielding an index of cheating behaviour where any positive number indicates cheating.



***Figure 2.* Screenshot of coin toss task presented to participants in Study 5.**

*Confession Probe*

Following the coin-tossing task, participants were informed (via onscreen text) that in previous trials other participants had cheated by over-reporting the number of correct guesses they’d got (this could be determined by comparing aggregate reports of heads to the aggregate number one would expect due to chance). Participants were then asked to state how many heads they had *actually* got (Appendix O). They were assured that they would be paid in accordance with their original report and would still be approved on Prolific whether or not they revealed that the original report was false. In the surveillance condition a stylised eye was placed at the top of the screen at this point in the procedure (*fig. 3*). In the neutral condition participants were presented only with the text of the confession probe and no surveillance cue.



*Figure 3.* Surveillance cue presented to participants in the surveillance condition in Study 5.

*Confession*

Any cheating participant who subsequently reported 2 (as their actual number of heads received) at the confession probe were scored as having confessed. All confessing participants were coded as 1, and non-confessors as 0.

*Moral Outrage*

This consisted of three items to gauge participants’ level of moral outrage. The three items (*Outraged, Insulted, Offended*) were randomly intermixed with items from the PANAS. Participants were instructed to indicate on a scale of 1- 5 (1 = *Very slightly or not at all*, to 5 = *Extremely*) the extent to which they were currently experiencing each feeling. A moral outrage score was computed by totalling the responses of these three items.

*PANAS:*

The PANAS was included as a cover for the three moral outrage items, which were embedded within it. The PANAS has good internal consistency, with Cronbach’s alpha of 0.88 for positive items, and α = .87 for negative items (Crawford & Henry, 2004; Watson et al, 1988).

#### 2.1.3. Design and Procedure.

Participants were randomly allocated to either a surveillance condition or a neutral condition. They then completed a short survey that asked questions about their experience of uncertain events and the belief in Good Luck scale (Darke & Freedman, 1997). These were included as filler items to avoid demand effects, as they were consistent with the cover story presented to participants. All participants were then told that they would be re-directed to a random coin-tossing website and were instructed to report the number of heads they got from a series of ten tosses. This task was designed so that for every head reported participants would earn a financial bonus; thus, cheating was incentivised. All participants were then given a confession probe, which asked them if they had reported honestly (see Appendix O). They then completed the PANAS (Watson et al., 1988), which included an additional three items to gauge levels of moral outrage. Finally, participants completed some demographics questions and were fully debriefed; this included an item asking whether they thought the coin-toss task was genuinely random. Before completion participants were asked whether they had read and understood the debriefing; they were not allowed to finish the study until they indicated in the affirmative. They were then thanked for their time and paid according to their original reported outcome response (as well as the initial £0.50).

## 3. Results.

Strikingly, only four participants (2.05% of the sample) confessed to over-reporting, and only one of these was in the surveillance condition. As a point of comparison, Pe’er et al. (2014) used the same paradigm and reported a confession rate of 18.79% - 24.60% (Studies 1 & 3, respectively). A closer inspection of the reported number of heads revealed this negligible confession rate was not due to increased levels of honesty on the coin toss task; 45.13% of the sample over-reported the number of heads they received. Furthermore, the mean number of reported heads was 2.93, meaning that, on average, participants exaggerated by nearly 50% the number of heads they had actually obtained. The unexpected negligible confession rate precluded any investigation of the main hypothesis.

I conducted an independent samples t-test to check whether there were any differences in levels of moral outrage between honest participants (*N =* 107) and dishonest participants (those who over-reported on the coin-toss task, *N* = 88). There was no significant difference in moral outrage scores for honest (*M* = 4.15, *SD* = 2.05) and dishonest (*M* = 4.57, *SD* = 2.69) participants, *t*(193) = 1.24, *p* = .218. Therefore, contrary to my hypothesis, there was no evidence to suggest that honest participants who felt their moral integrity was questioned demonstrated more moral outrage.

I conducted a moderated multiple regression to check whether surveillance cues moderated any effects of honesty on moral outrage. I regressed level of moral outrage onto a binary honesty variable (all those who did not over report were coded as 1, all others were coded as 0), a dummy variable for the surveillance condition and the interaction between the surveillance dummy and the honesty variable. The regression model was not significant, F(5, 229) = 1.76, p = .12, adj. R2= .04 and the interaction was not a significant predictor of moral outrage. Regression coefficients and standard errors can be found below in Table 9.

Table 9. Moderated multiple regression (predicting moral outrage from honesty, surveillance dummy and their interaction).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.002 |  |  | 1.33 | 0.251 | 0.004 |  |  | 0.72 | 0.541 |
| **Predictor** | **B** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | **B** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 4.55 | 0.25 |  | 18.21 | 0.001 | 4.32 | 0.36 |  | 12.12 | 0.001 |
| honesty | -0.39 | 0.34 | -0.08 | -1.15 | 0.251 | -0.17 | 0.48 | -0.04 | -0.35 | 0.729 |
| dummy surveillance |  |  |  |  |  | 0.46 | 0.50 | 0.10 | 0.92 | 0.360 |
| dummy surveillance x honesty |  |  |  |  |  | -0.44 | 0.68 | -0.08 | -0.65 | 0.518 |

As an exploratory analysis to check for any association between cheating and moral outrage I correlated degree of cheating with level of moral outrage. The results revealed a significant positive correlation, *r* = .248, *p* = .001, suggesting that, contrary to my initial prediction, the more people cheated the greater the level of moral outrage they expressed after being asked if they had behaved dishonestly.

## 4. Discussion.

The most unexpected result of this study was the negligible number of participants who admitted to over-reporting on the coin-toss task. Only four participants confessed, which is in stark contrast to Pe’er et al. (2014) who reported a confession rate an order of magnitude higher (18.79% and 24.60%, Studies 1 & 3 respectively). Given that 45.13% (*N* = 88) of the final sample cheated on the coin-toss task, the negligible confession rate was not a result of increased honesty in the sample (due to using Prolific Academic rather than MTurk, for example). In fact, at 45.13% the rate of over-reporting in Study 5 is *higher* than other studies, which have found the rate ranges from 20% to about 35% (Hilbig & Hessler, 2012; Fischbacher & Föllni-Heusi, 2013; Shalvi, Dana, Handgraaf & De Dreu, 2011).

Given that the present study and Pe’er et al. (2014, Studies 1 & 3) used similar methodology the obvious question is why did cheating participants in this study not confess at the same rate as in Pe’er et al. (2014)? I now provide possible explanations for the negligible confession rate in Study 5 and will also discuss the other findings of interest.

### **4.1. Time Constraints and Intuitive Processing.**

A commonly used method to induce individuals to rely on heuristic processing in the lab is the application of time pressure (Evan & Curtis-Holmes, 2005; Rand et al, 2014; Roberts & Newton, 2001; Suter & Hertwig, 2011; Wright, 1974). Forcing people to make quick decisions reduces time for deliberation, causing reliance on quicker, faster heuristics processing (Rand et al., 2014). As Rand et al. (2014) point out, applying time constraints does not however provide a failsafe way of inducing intuitive processing; rather, it can be an effective way of revealing “the dominant directions of the effects of intuitive versus reflective processing” (Rand et al., 2014, p.2). If confession, as an adaptive behaviour, relies on intuitive processing then it should be particularly prevalent in situations where people are under time pressure or considerable stress. In real life cases of suspects’ false confessions these factors play an important role. Suspects are often worn down and threatened with punishment[[60]](#footnote-60). As such they are potentially more prone to default to intuitive processing, resulting in a high percentage of false confessions. In this study, no such factors influenced participants’ decisions to confess; there was no time pressure or stress applied at the point of the confession probe, nor were there any threat of sanctions. Therefore, one could speculate that reliance on intuitive processing was over-ridden by more deliberative processes thus leading to a negligible confession rate. However, this explanation does not account for why participants confessed more in Pe’er et al’s studies than in Study 5.

### **4.2. Justifiable Dishonesty.**

One possible explanation for the discrepancy in confession rates between Study 5 and Pe’er et al. (2014) relates to the degree that people engage in dishonest behaviour. There is an emotional cost to transgressing and people appear to restrict the extent of their unethical activities in a way that lets them retain an honest self-concept (Mazar, Amir & Ariely, 2008) For example, according to previous research people engage in an *intermediate* degree of dishonesty (Hilbig & Hessler, 2013; Fischbacher & Föllni-Heusi, 2013; Shalvi, Handgraaf & De Dreu, 2011). That is, people do not generally cheat to the maximum extent but to a lesser degree (Hilbig & Hessler, 2013; Shalvi, Handgraaf & De Dreu, 2011; Fischbacher & Föllni- Heusi, 2013; Schweitzer, Ordóñez & Douma 2004; Schweitzer & Hsee, 2012) and this does not typically vary according to how much reward there is (Ariely, 2012). Moreover, people’s engagement in dishonesty appears to be a function of the degree of departure from the truth. That is, the willingness to cheat decreases when it involves a greater departure from the truth (Hilbig & Hessler, 2013). Retaining an element of plausibility while lying allows people to benefit from dishonesty in a way that preserves their moral self-image (Ariely, 2012; Bazerman & Tenbrunsel, 2011; Mazar, et al, 2008). One implication is that the closer one’s dishonesty is to the truth the less one considers it to be immoral and the easier it is to justify. The results of Study 5 provide some evidence to support this.

For example, the mean number of reported heads in Study 5 was 2.93. This was only slightly higher[[61]](#footnote-61) than the true amount (2 heads). It is possible that this slight departure from the truth allowed participants to justify their over-reporting and therefore they felt no need to confess. One implication of this line of reasoning is that there might be a minimum threshold of distance from the truth that must be met before one feels sufficiently compelled to confess. As a point of comparison in Pe’er et al’s Study 1 (2014) participants who cheated on the coin-toss task over-reported by a mean degree of 3.10. In their Study 3, a similar magnitude of cheating was observed; cheating participants added on average 3.47 reported correct guesses. Therefore, the higher magnitude of cheating found in cheaters in Pe’er et al (2014) than in Study 5 provides some support for the notion that the further one’s lies are from the truth, the harder they are to justify. Having over-reported by a sufficient degree (e.g., > 2.93) participants in Pe’er et al. (2014) were unable to justify their cheating. In response, they confessed to over-reporting on the coin-toss task to restore their self-image. Pe’er et al (2014) also noted that among cheaters larger lies were associated with larger confessions. That is, the greater the extent of cheaters’ over-reporting the more likely it was they would confess to a larger extent of their dishonesty (rather than to a lesser extent). It appears that the more one’s dishonesty strays from a point where one can justify it, not only is one more likely to confess, one is more likely to confess to a greater degree.

### **4.3. Moral Outrage as a Distancing Strategy.**

One finding of interest was the result that the more participants cheated on the coin-toss task, the more they expressed moral outrage after the confession probe. In other words, contrary to my hypothesis, the more an individual cheated the more outraged they were at being accused of the very thing they were guilty of. One explanation for this is that the hypocritical expression of moral outrage was a distancing mechanism that served to reduce the ethical dissonance associated with their cheating.

Previous research has demonstrated that individuals can employ a variety of distancing strategies to help reduce the threat to their self-concept caused by unethical behaviour (Barkan et al., 2012). For example, following the recall of a past transgression individuals were found to be more likely to reject an ethically questionable candidate on an admissions task. They recollection of a past misdeed also led participants to present themselves as ‘ultra-honest’ by demonstrating harsher ratings of an unethical behaviour, reporting less likelihood of engaging in the unethical behaviour themselves and prescribing more moral advice to others. This expression of moral hypocrisy is even more pronounced when judging out-group members (Gino, Ayal & Ariely, 2009). Therefore the ‘moral outrage’ that cheating participants expressed in Study 5 could be explained as an attempt to distance themselves from their dishonesty and to minimise the threat to their self-image.

## 5. Conclusion.

The results of this study did not allow me to investigate my initial prediction that surveillance cues trigger voluntary confessions. What was remarkable and unexpected about the results was that a negligible number of participants confessed. One possibility is that people’s engagement in strategic dishonesty allowed them a certain degree of flexibility in justifying their over-reporting. Thus, by over-reporting to a small degree people could justify their cheating and not feel the need to confess. The results also produced evidence in favour of a “moral outrage effect” or moral hypocrisy, a distancing response to wrongdoing, whereby guilty people express more outrage at being accused of the very thing they are guilty of. One documented upshot of this moral hypocrisy is a stronger desire for the punishment of others. In the next chapter I explore people’s intuitive punitive judgements. Not in relation to their own misdeeds however – instead I look at whether specific God beliefs influence their punitive decision making. If a god is forgiving and all sinners are pardoned, then is the moral responsibility for ensuring punishment placed upon our shoulders? Does belief in a forgiving God therefore lead to stronger support for punishment?

# VI. STUDIES 6-9: DIVINE FORGIVENESS AND MORAL SUPPORT FOR STATE-SANCTIONED PUNISHMENT

*‘I believe that forgiving them is God’s function. Our job is simply to arrange the meeting.’*

*~* General Norman Schwarzkopf, when asked about forgiving those who abetted and harboured the 9/11 perpetrators.

*‘To err is human; To forgive, divine.’*

*~* Alexander Pope.

## 1. Introduction.

Recent studies have found that beliefs in powerful, intervening Gods (both in general and when made salient) reducepeople’s engagement in, and endorsement of, punishment (Laurin, Shariff, Henrich & Kay, 2012). Those who believe in a powerful, morally interested God may view punishment as God’s remit and ‘outsource’ it to Him. In the present studies I investigate whether the specific manner in which God responds to moral infractions would affect people’s willingness to endorse state-sanctioned punishment. Reminding people of God’s punitive nature might indeed weaken endorsement of state-sanctioned punishment; but would being reminded of a God who responds to moral infractions with forgiveness *increase* support for punishment?

### **1.1. The Role of Punishment in Large-scale Cooperation.**

Humans are unique among species in their engagement in large-scale cooperation. Previous biological theories of cooperation (e.g., kin selection, reciprocal altruism, indirect reciprocity) have failed to fully explain this type of cooperation because it exists in circumstances where these theories predict it should not, i.e., between non-genetic relatives often in one shot interactions and in the absence of material or reputational gain (see Johnson & Krüger, 2004). What has become apparent is the critical role of punishment (Barclay, 2006; Boyd, Gintis, Bowles & Richerson, 2003; de Quervin et al., 2004; Fehr & Gächter, 2000, 2002; Johnson, 2005; Ostrom*,* Walker & Gardner, 1992; Sober & Wilson, 1998) in large-scale cooperation, in particular *altruistic punishment* (Fehr & Gächter, 2000, 2002, 2003). However, the cost of punishment very quickly escalates (Dreber*,* Rand, Fudenberg & Nowak, 2008; Herrmann, Thöni & Gächter, 2008; Nikiforakis, 2008; Nikiforakis & Engelmann, 2011). As Johnson and Krüger point out ‘while punishment appears to be necessary, there is no incentive for anyone to do it’ (Johnson & Krüger, 2004, p. 162). If altruistic punishment is necessary for human cooperation, then how is *that* sustained?

### **1.2. Supernatural Punishers.**

One solution invokes the notion of culturally evolved (Norenzayan & Shariff, 2008; Norenzayan et al., 2016) or elaborated concepts of supernatural deities who punish non-cooperation (Bering, 2006, 2011; Johnson & Bering, 2006; Johnson, 2009, 2011, 2016; Johnson & Krüger, 2004; Schloss & Murray, 2011). An inevitable consequence of living in large groups is that individuals cannot monitor others, all of the time. This allows for transgressions to go both undetected and unpunished therefore increasing the incentive to engage in norm-violating behaviors. Beliefs in a powerful and morally concerned God, however, provide a convenient ‘eye in the sky’ (Gervais & Norenzayan, 2012a). Such a deity can observe all actions, at all times. Moreover, such a God has the power to administer punishment in this life *or* the next, meaning that punishment of defectors is inevitable (Johnson, 2016; Johnson & Krüger, 2004).

### **1.3. Supernatural Punishment and Cooperation.**

At the societal level there is some evidence to suggest that positive effects of religion on prosocial behaviour are driven by supernatural punishment. The national crime rate in the U.S., for example, is negatively correlated with a belief in hell but positively correlated with a belief in heaven (Shariff & Rhemtulla, 2012). Belief in a punitive God was found to be associated with both a decrease in selfish behaviour and an increase in cooperation in a sample of 186 cultures (Johnson, 2005, Atkinson & Bourrat, 2011). Furthermore, Shariff and Norenzayan (2011) found that people who thought of God as a primarily punitive and wrathful agent tended to cheat less on an experimental academic task, while those who thought of God as essentially forgiving tended to cheat moreon the same task. Similarly, Yilmaz and Bahçekapili (2016) found that after being both explicitly and implicitly primed with punishing religious andsecular authorities (as opposed to non-punishing and neutral primes), individuals showed stronger prosocial intentions. Taken together, these findings suggest that beliefs in supernatural agents can enhance prosociality to the extent that those agents have the capacity for punishment*.* Moreover, as discussed in Chapter 2 (section 5.2), beliefs in a punitive God could provide a way for individuals to offload individual responsibility for punishment to divine hands without compromising cooperation (Laurin et al., 2012).

### **1.4. Punitive and Forgiving God Concepts.**

Laurin et al.(2012) reasoned that belief in morally concerned Gods with the power to dispense punishments and rewards might undermine incentives to engage in, or support, earthly punishment. Across a series of five studies, these authors found evidence that beliefs in a powerful God (both in general and when made salient) reduce people’s support for both altruistic and state-sanctioned punishment. The idea that belief in God undermines earthly punishment seems, however, to apply only to belief in a powerful and *punitive* God. This highlights a systematic tendency in the literature to conflate a powerful and moralising God with an exclusively punitive one (Johnson et al., 2017; Johnson & Cohen, 2016).[[62]](#footnote-62) The concept of an omnipotent, powerful God as *both* wrathful and benevolent is a feature of many world religions (Johnson et al*.*,2017; Johnson, Li, Cohen & Okun, 2013; Johnson & Cohen, 2016; McCullough, Bono & Root, 2005; Nieuwboer et al., 2015; Rye et al., 2000). The God of the Old Testament, for example, is mercurial and at least as likely to respond to human moral infractions with forgiveness as with punishment and can even convey both natures within the same dictate (see Exodus, 34:6-8). In Islam too, the concept of a powerful God as both punitive and forgiving is a central part of the religious experience, ‘Despair not of the God of Mercy: verily, God forgives all sins. Truly He is Oft-forgiving, Most Merciful’ (Qu’ran, 39:53). The Hindu pantheon includes at least two forgiving Gods who are further distinguished by the type of forgiving nature they possess[[63]](#footnote-63).

Moreover it has been demonstrated that people’s cognitive representations of God do not necessarily overlap with doctrine or dogma (Barrett, 1999; Gorsuch, 1968) and the roles of motivation and cognitive biases, such as egocentrism, jointly influence how people conceptualise God, his mind in particular (Epley, Converse, Delbosc, Monteleone & Cacioppo, 2009; Jackson, Hester & Gray, 2018). Thus the notion of a monolithic “powerful and punitive God” concept is not supported by evidence suggesting that people hold more nuanced and flexible representations of God that are influenced by a host of situational and motivational factors (e.g., socialisation (Chou & Uata, 2012; Dickie, Ajega, Kobylak & Nixon, 2006); compensatory control (Kay, Gaucher, McGregor & Nash, 2010); intergroup conflict (Caluori, Jackson, Grat & Gelfand, 2018) and cognitive biases such as ego-centrism (e.g., Epley et al., 2009; Ross, Lelkes & Russell, 2012).

### **1.5. Divergent Effects of Forgiving and Punitive God Concepts.**

Across a variety of domains and in particular with regards to prosociality,[[64]](#footnote-64) concepts of God as forgiving or punitive appear to be associated with divergent patterns of behavior, social attitudes, well being and even disease progression (Ironson et al., 2011; Johnson et al.,2017; Johnson et al., 2013; Sherman et al., 2015). Despite evidence suggesting that punishing Gods promote prosociality (Shariff & Norenzayan, 2011; Yilmaz & Bahçekapili, 2016), there is evidence suggesting punitive God concepts provoke distinctly antagonistic behavior (Bushman, Ridge, Das, Key & Busath, 2007; Johnson et al., 2013). Positive religious primes (as opposed to neutral religious primes) however, have been found to positively affect prosociality by increasing charitable giving (Pichon, Boccato & Saroglou,2007). Similarly, Harrell (2012) found that priming the reward related aspects of religious and secular authority led to an increase in prosociality.

### **1.6. Priming Methodologies: Delineating the Dimensions of Religious Cognition.**

The use of priming methodologies has allowed researchers to experimentally manipulate religious concepts and make headway in revealing the underlying causal effects of religious cognition on prosociality (see Ritter & Preston, 2013). Yet there are many inconsistencies (see Chapter 1, section 3.2). For example, religious primes have been found to increase honesty and prosociality and curtail selfish impulses (Ahmed & Salas, 2011; Shariff & Norenzayan, 2007; Randolph-Seng & Nielsen, 2007), yet at the same time religious primes have also been found to *decrease* prosocial attitudes and behavior in some contexts (e.g., Ginges, Hansen & Norenzayan, 2009; Johnson, Rowatt & LaBouff 2010; Saroglou, Corneille & Van Cappellen 2009). A recent meta-analysis by Shariff et al., (2016) found that religious primes in general only work to the extent that the participants are religious themselves. However other studies observe similar effects of religious priming on non-believers and believers alike (e.g., Laurin, Kay, & Fitzsimons, 2012) while others show mixed effects across different studies (e.g., Gervais & Norenzayan, 2012b; Shariff & Norenzayan, 2007).

These seemingly contradictory findings may be due to priming multiple religious concepts at once, using the erroneous assumption that all these primes activate a monolithic religious prosociality construct (Preston & Ritter, 2013). Research suggests that different religious concepts can activate different prosocial goals (see Preston & Ritter, 2013). For example, Laurin et al. (2012) consistently found that being religious, distinct from holding beliefs in a powerful God, was a *positive* predictor of costly punishment. Preston and Ritter (2013) found that *religion* primes enhance parochial altruism (that is, prosocial behavior preferentially directed to the ingroup), whereas *God* primes directed prosocial impulses towards religious outsiders (these effects were seen even in participants who reported no belief in God or religious affiliation). Therefore, if *religion* primes trigger concerns related to ingroup prosociality, it is not surprising that they would also strengthen support for state-sanctioned punishment, in so far as punishment benefits the group at large and is thereby prosocial (McKay & Whitehouse, 2015; 2016). What is clear is that people hold nuanced conceptualisations of the divine and this extends to the nature of God (Jackson et al., 2018).

### **1.7. God as a Behavioural Model.**

Rather than outsourcing punitive responsibility to God one possibility is that instead, he provides a behavioral model (Lipsey, Pogge, Shepperd & Miller, 2016; Preston, Ritter & Hernandez, 2010; Unnever, Cullen & Bartkowski, 2006). In this case, forgiving God concepts would *decrease* support for state-sanctioned punishment because God models forgiveness and people follow his divine example. There is some evidence to support this; previous research has found consistent associations between images of God as wrathful and judgemental and support for capital punishment and with stronger punitive attitudes in general (Bader, Desmond, Mencken & Johnson, 2010; Evans & Adams, 2003; Unnever & Cullen, 2006). In contrast, images of God as gracious and loving have been found to be negatively associated with these variables (Unnever et al*.*, 2006). Similarly, participants who reflected upon their religious beliefs demonstrated stronger levels of forgiveness towards a violent perpetrator than participants who had not reflected upon these beliefs (Nieuwboer et al., 2015).

However, the findings of Laurin et al. (2012) suggest that beliefs in a powerful God, both in general and when made salient, *reduce* people’s support for state-sanctioned punishment. Rather than beliefs in a powerful God promoting punitiveness as an appropriate response to wrongdoing, Laurin et al. argue that people view punishment as the remit of God, as is made explicit in much Christian scripture.[[65]](#footnote-65)

### **1.8. Current Research.**

Given the findings of Laurin et al. (2012), I sought to examine whether forgivingGod concepts and (salient) beliefs would affect people’s endorsement of state-sanctioned punishment. The rationale for Laurin et al.’s (2012) studies implies that people will only ‘outsource’ to the extent that the God they believe in is punitive. A God who ultimately responds to human wrongdoing with forgiveness, on the other hand,may make human responsibility for restoring fairness salient, thus *increasing* support for state-sanctioned punishment. I tested this prediction across four studies. In Studies 6 and 7, passages adapted from the parable of Lazarus (Luke 16: 19-31) were used to explicitly prime either the forgiving or punitive nature of God. Together these two studies investigated the effects of the experimental manipulation on support for state-sanctioned punishment in response to a corporate crime (Study 6) and a victim-directed transgression (Studies 7). In Studies 8 and 9, I manipulated the salience of forgiving God beliefs and examined the effects of this manipulation on support for state-sanctioned punishment. I hypothesised that participants explicitly primed with the forgiving nature of God would show increased support for state-sanctioned punishment compared to participants administered punitive God primes. Second, I predicted that, when made salient, forgiving God beliefs would increase individuals’ support for state-sanctioned punishment.

## 2. Study 6.

In this study I examined the effect of explicit forgiving (*cf*. punitive) God primes on endorsement of state-sanctioned punishment. I predicted that participants exposed to a forgiving God prime would show more support for state-sanctioned punishment than participants exposed to a punitive God prime. I also included measures of self-reported religious affiliation and strength of belief in God in the analyses to explore the moderating effects of these variables and to further investigate the positive religiosity-punishment association found in previous research (de Quervain et al*.*, 2004; Grasmick*,* Davenport, Chamlin & Bursik, 1992; Grasmick, Kinsey & Cochran, 1991; see Henrich, Heine & Norenzayan, 2010).

### **2.1. Method.**

#### 2.1.1. Participants.

Participants were 403 Amazon Mechanical Turk workers who took part in exchange for $1 (203 females, 199 males and 1 who identified as Other; *Mage =* 35.12 years, *SD* = 11.79). A power analysis recommended an *N* of 398 to attain 80% power (*α* = .05, one-tailed) to detect a small effect of *d* = .25[[66]](#footnote-66) (see Thalheimer and Cook, 2002). I added an additional 15% (59) to account for potential attrition on MTurk, thus 457 participants were recruited in total. Of these, 29 participants did not complete the study, 21 failed the comprehension check, and four were removed for providing illegitimate responses on the religious affiliation measure (e.g., Satanist, Wiccan, Witch). None of the participants guessed the hypothesis.

#### 2.1.2. Measures.

*Self-reported religious affiliation*

Participants were asked to choose their religious affiliation from a list of twelve items (Appendix C). Any participants who selected “Other” were prompted to describe their affiliation in more detail and re-coded accordingly.

*Strength of belief in God*

Participants were administered the Strength of Belief in God scale (Gervais, 2014). This asked them to indicate on a scale of 0-100 the extent to which they believe in God (or gods). They were told that if they felt certain that God (or gods) did *not* exist, to report this as ‘0’ and if certain that God (or gods) *did* exist, then to rate this as ‘100’. Any participant who was absolutely uncertain either way was instructed to put ‘50’.

*Forgiving and punishing God primes*

The priming methodology was based on that of Yilmaz and Bahçekapili (2016, Study 2) who adapted extracts of religious texts that made clear either the punishing or forgiving nature of God. We created two short passages adapted from the parable of Lazarus (Luke 16: 19-31) that outlined the story of a recently deceased rich and immoral man. As his coffin is paraded through the streets, a child observing the event asks his father what will become of the man’s soul? The two vignettes were identical except for the last sentence; the child’s father refers to either God’s forgiving nature (*‘Our Lord is merciful and forgiving, my son*’) (Appendix P) or his punishing nature *(‘God will punish the world for its evil, and the wicked for their sins, my son’)* (Appendix Q).

*Endorsement of state-sanctioned punishment*

Participants read a short vignette about a protagonist who steals from his company in order to fund his gambling habit (adapted from Carlsmith, Darley & Robinson, 2002) (Appendix R). Participants were then asked to indicate, using a seven-point scale (*1 = ‘the lowest possible’* and *7 = ‘the highest possible’*) how many of their tax dollars they would like to see put towards the state (a) *catching* and (b) *punishing* the transgressor. Responses to these items were then averaged to yield a *punishment index* for each participant, ranging from 1-7.

#### 2.1.3. Design and procedure.

Participants completed a series of demographics questions including a measure of self-reported religious affiliation and strength of belief in God (taken from Gervais, 2014). All participants were then randomly assigned to one of two conditions; *forgiving God prime* (*N* = 195) or *punishing God prime* (*N* = 208). All participants then underwent a comprehension check before completing a two-item punishment task (adapted from Carlsmith et al., 2002 and used in Laurin et al., 2012, see Appendix R). They then completed a funnel debrief that included a suspicion probe which allowed us to assess whether participants had guessed the true purpose of the study.

### **2.2. Results and Discussion.**

I predicted that participants exposed to the *forgiving God* prime would show stronger support for state-sanctioned punishment compared to those exposed to the *punishing God* prime. Due to issues of non-normality, I conducted a non-parametric Mann-Whitney *U* test to check for differences in support for state-sanctioned punishment between conditions. Distributions of support for state-sanctioned punishment for the *forgiving God* prime and the *punishing God* prime conditions were similar, as assessed by visual inspection, therefore Medians are reported. There was no significant difference in punishment indices between the *forgiving God prime* (*Mdn* = 3) and *punishing God prime* (*Mdn* = 2.50) conditions, *U =* 18744, *z* = -1.32, *p* = .185.

To check for any effects of self-reported religious affiliation on punishment indices, a hierarchical multiple regression was conducted. In the first block I entered a dummy variable for *self-reported religious affiliation.* I did this by transforming *self-reported religious affiliation* into a binary variable. I coded Atheist, Agnostic and None as 0 *(non-religious affiliation*) and the remainder (Catholic, Protestant, Christian (other), Hindu, Buddhist, Muslim, Jewish and Sikh) as 1 *(self-reported religious affiliation*). I checked the written responses of participants who had indicated ‘Other’ and re-coded them accordingly. In the second block, I simultaneously entered a dummy variable for participation in the *forgiving God prime* condition (participants in the *forgiving God prime* condition were coded as 1, all others coded as 0) and the interaction of this dummy with the self-reported religious affiliation variable. Model 1, with self-reported religious affiliation as the only predictor, was statistically significant, *p* = .029, (self-reported religious affiliation positively predicted punishment indices). The full model was not significant (*p* = .051) though religious affiliation emerged as a significant predictor of punishment indices (*p* = .019) (see Table 10). This is in line with previous research that has found evidence of a positive association between general religiosity and punishment (de Quervain et al., 2004; Grasmick et al., 1992; Grasmick, Kinsey & Cochran, 1991).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.009 |  |  | 4.80 | 0.029 | 0.007 |  |  | 2.62 | 0.051 |
| **Predictor** | **B** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | **B** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 2.78 | 0.12 |  | 23.95 | 0.001 | 2.63 | 0.15 |  | 17.91 | 0.001 |
| Self-reported religious affiliation | 0.35 | 0.16 | 0.11 | 2.19 | 0.029 | 0.42 | 0.18 | 0.13 | 2.35 | 0.019 |
| Dummy forgiveness |  |  |  |  |  | 0.25 | 0.16 | 0.08 | 1.55 | 0.123 |
| Dummy x self-reported religious affiliation |  |  |  |  |  | -0.004 | 0.003 | -0.05 | -0.85 | 0.396 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

Table 10. Multiple hierarchical regression (predicting endorsement of state-sanctioned punishment from self-reported religious affiliation, dummy forgiveness and their interaction).

To check for effects of strength of belief in God on punishment indices, a second hierarchical multiple regression was conducted. In the first block, I entered *strength of belief in God (mean-centred).* In the second block, I simultaneously entered a dummy variable for participation in the *forgiving God prime* condition and the interaction of this dummy with the *strength of belief in God (mean-centred)* variable. Neither regression model was significant (see Table 11).

Table 11. Hierarchical multiple regression (predicting endorsement of state-sanctioned punishment from strength of belief in God, dummy forgiveness and their interaction).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.001 |  |  | 1.44 | 0.231 | 0.006 |  |  | 1.38 | 0.25 |
| **Predictor** | **B** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | **B** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 2.97 | 0.08 |  | 37.05 | 0.001 | 2.86 | 0.11 |  | 25.63 | 0.001 |
| Strength of belief in God (mean-centered) | 0.002 | 0.002 | 0.06 | 1.20 | 0.231 | 0.004 | 0.003 | 0.09 | 1.36 | 0.176 |
| Dummy forgiving prime |  |  |  |  |  | 0.23 | 0.16 | 0.07 | 1.46 | 0.146 |
| Dummy x strength of belief in God (mean-centered) |  |  |  |  |  | -0.003 | 0.004 | -0.05 | -0.76 | 0.449 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

## 3. Study 7.

The type of crime outlined in the punishment scenario in Study 6 was a corporate crime, detailing a perpetrator who steals from a large company to fund his gambling habit. One potential limitation of this vignette is that the indirect, ‘white-collar’ nature of the crime means it is not the type of wrongdoing that our cognitive system evolved to have intuitive responses to (Buss, 1999; Cosmides & Tooby, 1987) (notwithstanding that Laurin et al. found significant effects with the original vignette). In Study 7, I addressed this issue by altering the type of wrongdoing in the vignette so that it involved direct harm to specified vulnerable victims (see Appendix S).

### **3.1. Methods.**

#### 3.1.1. Participants.

Participants were 396 MTurk workers who participated in exchange for $1 (145 females, 251 males; *Mage =* 35.95 years, *SD* = 11.49). An a priori power analysis indicated a sample size of 398, (parameters set at *d* = 0.25 *α* =. 05, power = 0.8, one-tailed). An additional 15% of the recommended N was added (59) to account for potential attrition on MTurk, thus 457 were recruited in total. Participants who failed the comprehension check (43) or did not complete the study (18) were removed from subsequent analyses. The study hypotheses, design, data collection, and analyses plan were pre-registered with AsPredicted (ref no: 3510; see https://aspredicted.org/wa6q6.pdf).

#### 3.1.2. Measures.

*Religiosity measure*

This was a single item asking participants to rate how religious they were on a scale of 1-5 (1 = *Not at all religious* to 5 = *Very religious*) (taken from Laurin et al., 2012, study 1).

*Endorsement of state-sanctioned punishment*

In this version, participants were presented with the story of a trusted, local handyman (John) who steals from his elderly clients once inside their homes (Appendix S). Participants were again asked to rate how many of their tax dollars they would like to see put towards the state a) catching and b) punishing John on a scale of 1-7. These scores were then averaged to yield a punishment index for each participant, ranging from 1-7.

#### 3.1.3. Design and procedure.

Participants completed a general religiosity measure (taken from Laurin et al., 2012) and a measure of political affiliation (to reduce demand characteristics by making religion less salient). Participantswere then randomly assigned to one of two conditions; a *forgiving God prime (N* = 192) and a *punishing God prime* (*N* = 204) condition where they read the same respective vignettes as in Study 6. Participants completed a comprehension check then a victim-related version of the punishment task from Study 6. Finally, participants completed a funnel debrief that included a suspicion probe to assess whether they had guessed the true purpose of the experiment.

### **3.2. Results and Discussion.**

I conducted a confirmatory analysis to check for differences in punishment indices between the two priming conditions. Due to issues of non-normality, I conducted a non-parametric Mann-Whitney *U* test to check for differences in support for state-sanctioned punishment between conditions. Distributions of support for state-sanctioned punishment for the *forgiving God* prime and the *punishing God* prime conditions were similar, as assessed by visual inspection, therefore Medians are reported. There was no significant difference in levels of endorsement of state-sanctioned punishment between *forgiving God prime* (*Mdn* = 4) and *punishing God prime* (*Mdn* = 4) conditions, *U =* 18760.50, *z* = -.726, *p* = .468.

I ran an exploratory hierarchical multiple regression to check for effects of self-reported religiosity on punishment indices. In the first block I entered the religiosity variable (mean-centred), adding a dummy variable for participation in the *forgiving God* condition and the interaction of this variable with self-reported religiosity (mean-centred) in the second block. The full model was not statistically significant, though *self-reported religiosity* emerged as a significant predictor of punishment indices (*p* = .036) (See Table 12).

I did find that endorsement levels of state-sanctioned punishment were higher in this study *(forgiving God M* = 4.09; *punishing God M* = 3.94) than in Study 6 (*forgiving God M* = 3.09; *punishing God M* = 2.85). This lends preliminary support to the suggestion that punitive responses may be more pronounced for victim-directed transgressions rather than for corporate crimes.

Table 12. Multiple regression (predicting endorsement of state-sanctioned punishment from self-reported religiosity, dummy forgiveness and their interaction).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.014 |  |  | 6.69 | 0.010 | 0.003 |  |  | 2.48 | 0.060 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 8.03 | 0.18 |  | 45.95 | 0.001 | 7.89 | 0.24 |  | 32.37 | 0.001 |
| religiosity (mean-centered) | 0.3 | 0.11 | 0.13 | 2.59 | 0.010 | 0.34 | 0.16 | 0.15 | 2.1 | 0.036 |
| Dummy forgiving prime |  |  |  |  |  | 0.28 | 0.35 | 0.04 | 0.809 | 0.419 |
| Dummy x religiosity (mean-centered) |  |  |  |  |  | -0.08 | 0.23 | -0.03 | -0.37 | 0.715 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

## 4. Studies 8 and 9.

Laurin et al*.* (2012) found that, when activated, beliefs in a powerful interventionist God reduced people’s endorsement of state-sanctioned punishment. In Studies 6 and 7, I attempted to build on this finding by priming participants with the notion of a God with specific characteristics (either forgiving or punitive) and measuring endorsement of state-sanctioned punishment. The experimental manipulation had no effect. One possibility is that whereas I primed the notion of a certain type of God, Laurin et al. (2012) measured existing beliefs in God and manipulated when this measure was taken (either before the punishment measure, so as to render the beliefs salient when punishment was measured; or after the measure). There is both a methodological and conceptual difference between priming the notion of *x* and activating an existing belief in *x*. So as to achieve a closer replication of Laurin et al.’s methodology, in Studies 8 and 9, I measured existing beliefs in a forgiving (*vs.* punitive) God and manipulated whether this measure was taken before or after the punishment measure (the victim-directed transgression in Study 8; the corporate crime in Study 9). I predicted that participants whose forgiving God beliefs were made salient (measured before administration of the punishment measure) would show higher levels of endorsement of state-sanctioned punishment compared to those whose beliefs were not made salient.

### **4.1. Method (Study 8).**

#### 4.1.1. Participants.

Participants were197 MTurk (82 females, 114 males, 1 who identified as Other; *Mage=* 34.21 years, *SD*= 11.46). In the salient condition (*N* = 95) the gender breakdown was as follows: Female, 34; Male, 60; Other, 1. In the non-salient condition (*N* = 102) the gender breakdown was: Female, 54; Male, 48. The hypotheses, design, data collection, and analysis plan for both studies were pre-registered with AsPredicted (ref no: 3583; see https://aspredicted.org/a8pi5.pdf).

#### 4.1.2. Measures.

*Forgiving God Scale*

This was adapted from the 12-item ‘View of God Inventory’ (Ironson *et al.* 2011) (Appendix T). The final scale was comprised of eight items.[[67]](#footnote-67) Four of these items assessed forgiving God beliefs (e.g., *‘I believe God is all forgiving’*), and the other four assessed punishing God beliefs (e.g., *‘ I think God will punish people for what they have done’*). Participants were asked to rate, on a 5-point Likert scale, how much they agreed with each statement (*1 = Strongly disagree,* to *5 = Strongly agree*). The forgiving and punishing items were averaged to create, respectively, a *forgiving God average* and *punishing God average*. The *punishing God average* was then subtracted from the *forgiving God average* to yield an overall *forgiving God Score*, with larger numbers indicating stronger forgiving God beliefs

#### 4.1.3. Design and procedure.

Participants were randomly assigned to one of two conditions: *salient* or *non-salient*. In the *salient* condition participants were asked to complete the forgiving God scale before completing the victim-directed transgression scenario (the same as in Study 7, see Appendix S). In the *non-salient* condition, participants completed the forgiving God scale *after* the DV.

### **4.2. Results and Discussion.**

As a confirmatory analysis, a hierarchical multiple regression was conducted predicting participants’ endorsement of state-sanctioned punishment. In the first block, I entered *forgiving God score (mean-centred)*. In the second block, I simultaneously entered a dummy variable for experimental condition (coded as *0 = not-salient, 1 = salient*) and the interaction between the dummy and *forgiving God score (mean-centred)*.

Model 1, with *forgiving God score* as the only predictor, was not statistically significant. The full model was not significant either, however *forgiving God score (mean-centred)* emerged as a significant predictor of punishment indices (*p* = .049) (see Table 13).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.01 |  |  | 2.9 | 0.09 | 0.006 |  |  | 2.1 | 0.11 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 4.09 | 0.12 |  | 35.24 | 0.001 | 4.22 | 0.16 |  | 26.10 | 0.001 |
| Forgiving God score (mean-centered) | -0.14 | 0.08 | -0.12 | -1.7 | 0.090 | -0.23 | 0.12 | -0.20 | -2.00 | 0.049 |
| Dummy salience |  |  |  |  |  | -0.3 | 0.23 | -0.10 | -1.3 | 0.194 |
| Dummy x forgiving God score (mean-centered) |  |  |  |  |  | 0.21 | 0.17 | 0.13 | 1.26 | 0.210 |
|  |  |  |  |  |  |  |  |  |  |  |

Table 13. Hierarchical multiple regression (predicting state-sanctioned punishment from forgiving God score, dummy salience and their interaction).

I conducted two further exploratory analyses to check for any individual effects of a) *punishing god average* and b) *forgiving god average* on punishment indices. These were calculated by averaging responses to a) punishing God items and b) forgiving God items. First, I conducted a hierarchical multiple regression predicting punishment indices from *punishing god average (mean-centred),* the dummy salience variable and their interaction. Both models were highly significant and *punishing god average* emerged as a significant predictor of punishment indices, (*p =* .001 in both models; see Table 14).

Second, I conducted a hierarchical multiple regression predicting punishment indices from *forgiving god average (mean-centred)*, the dummy salience variable and their interaction. Model 1 with *forgiving God average* as the only predictor was not significant. The full model was not significant, however *forgiving god average* emerged as a significant predictor of punishment indices (*p =* .046) (see Table 12).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
| **Punishing god average** |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.073 |  |  | 16.51 | 0.001 | 0.009 |  |  | 6.11 | 0.001 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***P*** |
| Intercept | 4.10 | 0.11 |  | 36.44 | 0.001 | 4.13 | 0.16 |  | 25.77 | 0.001 |
| Punishing god average (mean-centered) | 0.40 | 0.10 | 0.28 | 4.06 | 0.001 | 0.51 | 0.15 | 0.36 | 3.52 | 0.001 |
| Dummy salience |  |  |  |  |  | -0.15 | 0.23 | -0.04 | -0.63 | 0.530 |
| Dummy x punishing god average (mean-centered) |  |  |  |  |  | -0.25 | 0.20 | -0.12 | -1.20 | 0.230 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Forgiving god average** |  |  |  |  |  |  |  |  |  |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.012 |  |  | 3.33 | 0.07 | 0.015 |  |  | 2.10 | 0.102 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 4.09 | 0.12 |  | 35.28 | 0.001 | 4.24 | 0.16 |  | 26.29 | 0.001 |
| Forgiving god average (mean-centered) | 0.19 | 0.1 | 0.13 | 1.83 | 0.070 | 0.23 | 0.11 | 0.16 | 2.01 | 0.046 |
| Dummy salience |  |  |  |  |  | -0.28 | 0.24 | -0.09 | -1.18 | 0.240 |
| Dummy x forgiving god average (mean-centered) |  |  |  |  |  | -0.14 | 0.13 | -0.10 | -1.10 | 0.284 |

Table 14. Hierarchical multiple regression (predicting state-sanctioned punishment from punishing and forgiving God average, salience and their interactions).

### **4.3. Methods (Study 9).**

#### 4.3.1. Participants.

Participants were 211 MTurk workers (70 females, 140 males, 1 who identified as Other; *Mage=* 34.31 years, *SD*= 10.05). In the salient condition (*N* = 98), the gender breakdown was as follows: Female, 34; Male, 63; Other, 1. In the non-salient condition (*N* = 113), the gender breakdown was: Female, 36; Male, 77.

#### 4.3.2. Measures.

*Forgiving God Scale*

This was the same scale used in Study 8. As before the forgiving and punishing items were averaged to create, respectively, a *Forgiving God average* and *Punishing God average*. The *Punishing God average* was then subtracted from the *Forgiving God average* to yield an overall *Forgiving God Score*, with larger numbers indicating stronger forgiving God beliefs.

#### 4.3.3. Design and procedure.

As in Study 8, participants were randomly assigned to one of two conditions: *salient* or *non-salient*. In the *salient* condition, participants were asked to complete the forgiving God scale before completing the corporate crime punishment scenario (the same as in Study 6). In the *non-salient* condition, participants completed the forgiving God scale *after* the DV.

### **4.4. Results and Discussion.**

A hierarchical multiple regression was conducted predicting participants’ endorsement of state-sanctioned punishment. In the first block, I entered the *forgiving God score (mean-centred)*. In the second block, I simultaneously entered a dummy variable for experimental condition (*0 = not-salient, 1 = salient*), and the interaction between the dummy and the *forgiving God score (mean-centred)*. Neither model was significant (see Table 15).

Table 15. Hierarchical multiple regression (predicting state-sanctioned punihsment from forgiving God score, dummy condition and their interaction).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STUDY 3b)** |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | -0.001 |  |  | 0.7 | 0.40 | 0.01 |  |  | 0.93 | 0.43 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 2.90 | 0.54 |  | 5.42 | 0.001 | 3.42 | 0.73 |  | 4.72 | 0.001 |
| Forgiving God score (mean-centered) | -0.07 | 0.09 | -0.06 | -0.84 | 0.403 | -0.01 | 0.12 | -0.01 | -0.06 | 0.953 |
| Dummy salience |  |  |  |  |  | -1.17 | 1.08 | -0.34 | -1.09 | 0.278 |
| Dummy x forgiving God score (mean-centered) |  |  |  |  |  | -0.15 | 0.18 | -0.28 | -0.86 | 0.393 |

I conducted exploratory analyses to check for effects of both the *punishing god average* and the *forgiving god average* on punishment indices. In each case I conducted a hierarchical multiple regression predicting punishment indices from the relevant average score (punishing or forgiving), the dummy salience variable and their interaction. Neither model was significant in either case (see Table 16).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
| **Punishing god average** |  |  |  | **Model 1.** |  |  |  |  | **Model 2.** |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | 0.011 |  |  | 3.41 | 0.066 | 0.005 |  |  | 1.48 | 0.22 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 3.34 | 0.12 |  | 28.71 | 0.001 | 3.45 | 0.16 |  | 21.60 | 0.001 |
| Punishing god average (mean-centered) | 0.20 | 0.11 | 0.13 | 1.85 | 0.066 | 0.21 | 0.15 | 0.13 | 1.35 | 0.178 |
| Dummy salience |  |  |  |  |  | -0.24 | 0.23 | -0.07 | -1.01 | 0.313 |
| Dummy x punishing god average (mean-centered) |  |  |  |  |  | -0.02 | 0.22 | -0.01 | -0.10 | 0.919 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Forgiving god average** |  |  |  |  |  |  |  |  |  |  |
|  | **Adj *R2*** |  |  | ***F*** | ***p*** | ***R2* change** |  |  | ***F*** | ***p*** |
|  | -0.002 |  |  | 0.56 | 0.454 | 0.013 |  |  | 1.13 | 0.339 |
| **Predictor** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** | ***B*** | ***s.e.*** | ***Beta*** | ***t*** | ***p*** |
| Intercept | 3.34 | 0.12 |  | 28.52 | 0.001 | 3.44 | 0.16 |  | 21.4 | 0.001 |
| Forgiving god average (mean- centered) | 0.08 | 0.11 | 0.05 | 0.75 | 0.454 | 0.24 | 0.17 | 0.15 | 1.41 | 0.160 |
| Dummy salience |  |  |  |  |  | -0.25 | 0.24 | -0.07 | -1.05 | 0.297 |
| Dummy x forgiving god average mean-centered) |  |  |  |  |  | -0.29 | 0.22 | -0.14 | -1.03 | 0.195 |

Table 16. Hierarchical multiple regression (predicting state-sanctioned punishment from punishing and forgiving God average, dummy salience and their interaction).

Of note was the unexpected significant *positive* correlation between punishing God average and forgiving God average in both Study 8 and 9: Study 8 *rs*(195) = .178, *p* = .012; Study 9 *rs*(209) = .136, *p* = .049. It is of course possible that people who believe God is punitive also believe he is forgiving. The two are not mutually exclusive; people hold nuanced conceptions of the divine, therefore the nature of God can be multifaceted.

Arguably however, the measure consisted of two separate scales; one that measured forgiving God beliefs and the other that measured punitive God beliefs, each consisting of four items. It is possible therefore that, due to framing effects, both sets of beliefs were made salient to participants. In light of this, I conducted an exploratory correlational investigation between forgiving God beliefs, punitive God beliefs and support for state-sanctioned punishment (see Table 17 below.)

Table 17. Correlations and descriptive statistics for Study 9.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | 1 | 2 | 3 | Mean | *SD* |
| 1. forgiving God average | 1.00 |  |  | 3.40 | 1.09 |
| 2. punishing God average | 0.217\*\* | 1.00 |  | 2.95 | 1.06 |
| 3. punishment index | 0.052 | 0.127 | 1.00 | 3.30 | 1.70 |

\*\* *p* < .001 level (two-tailed)

## 5. General Discussion.

I predicted that participants exposed to a forgiving (*vs*. punitive) God prime (Studies 6 & 7), or whose forgiving God beliefs were made salient (*vs.* were not made salient; Studies 8 & 9) would make stronger endorsements of state-sanctioned punishment. However, there was no significant difference between conditions in any of these four well-powered studies. First, I consider two reasons why forgiving God concepts and beliefs might not lead to increased support for punishment. Second, I discuss conceptual issues related to the distinction between divine and state-sanctioned punishment. Third, I suggest that the emotional framing of the punitive decision may be an important factor in forgiving God concepts reducing support for state-sanctioned punishment in the way we hypothesised at the outset. Fourth, I discuss the role of forgiving God beliefs and concepts in relation to prosociality and human cooperation.

### **5.1. God as a Behavioural Model.**

If God provides a behavioural model (Lipsey et al*.*, 2016; Preston et al., 2010; Unnever et al., 2006), then forgiving God concepts and beliefs would decrease support for state-sanctioned punishment because God models forgiveness and people follow his divine example. There is evidence to support this, for example, people who stated they had a personal relationship with a loving God showed decreased support for capital punishment (Unnever et al., 2006). However, if God provides a behavioral model, then we would expect punitive God concepts and beliefs to (contra Laurin et al.) *increase* support for punishment. In other words, one would still expect a difference in endorsement of state-sanctioned punishment between conditions, just not in the direction I predicted based on Laurin et al. (2012).

### **5.2. Second Order Free-rider Problem and the Supernatural Punishment Hypothesis.**

Second, if the Supernatural Punishment Hypothesis (SPH) is construed as a solution not just to first-order free riding but also *second-order* free riding (Johnson, 2005, 2016), then punitive God beliefs and concepts should (contra Laurin et al., 2012) *increase*support for punishment because people fear divine sanction themselves for not supporting punishment. But again, under this scenario, one would still expect a difference in levels of endorsement of state-sanctioned punishment between conditions, just not in the predicted direction.

### **5.3. A Mutualistic Moral Framework.**

I dismiss both of these explanations, because not only are they contradicted by the findings of Laurin et al*.* (2012), but they also predict a difference between conditions, which was not supported by the findings of the studies outlined in this chapter.

The null findings may be more understandable if interpreted within a mutualistic moral framework (Baumard, André & Sperber, 2013) where morality is about demonstrating and enforcing fairness (Baumard, 2011; Baumard & Sheskin, 2015). A transgression creates an imbalance between the perpetrator and victim. Accordingly, people will act to restore the balance of interests either by a) harming the criminal or b) compensating the victim (Baumard & Sheskin, 2015). In both punishment scenarios in Studies 6 and 7, a supernatural mediator is available to redress the imbalance caused by the wrongdoing; either via punishment *or* forgiveness i.e., God can punish John, or he can forgive him. What matters, is that the inequality caused by the transgression has been addressed, thus satiating the intuitive desire for redress. In a scenario where people’s intuitions about fairness are not satisfied or left ambiguous, support for state-sanctioned punishment may increase. Uncertain to God’s response to the wrongdoing, individuals may take it upon themselves to restore equilibrium and thus demonstrate more motivation to invest in state-sanctioned mechanisms for punishment. Future studies may seek to investigate whether priming the *unknowable or mysterious* nature of God (e.g., *‘God moves in mysterious ways, my son’)* increases support for state-sanctioned punishment of transgressors.

### **5.4. The Divine Origins of State-sanctioned Punishment.**

One of the goals of this research was to follow the methodology and rationale of Laurin et al. (2012) as closely as possible. However, while Laurin et al.(2012) reasoned that belief in punitive Gods might demotivate earthly punishment, one cannot discount the historical evidence that suggests state-sanctioned punishment was originally construed as *the same* as God’s judgement (see Exodus, 24:3, Deuteronomy, 16-18). In the Abrahamic religions, for example, it is made explicit that God himself is the supreme lawgiver, and earthly bodies of law are not only sanctioned by him but also administered in his name (e.g., Mosaic law and Beth Din courts in Judaism, Sharia Law in Islam). Even today in many courtrooms across the globe, oaths are sworn on sacred texts and punishment is often administered *in the name of God*. This apparent conflation of earthly and divine punishment belies the potential of *outsourcing* punishment to God.

### **5.5. Complex Systems of Law as a Product of Large-scale Societies.**

One might also contest the sequencing of events implied by Laurin et al.’s argument. Their reasoning implies that large-scale societies with moralising High Gods would have less incentive to invest in and develop mechanisms of state-sanctioned punishment. Therefore, one might reasonably expect that such societies would have *fewer* sophisticated or complex systems of state-sanctioned punishment than smaller, non ‘high god’ societies. But this does not seem to be the case; complex systems of law and punishment seem to be a *product* of such societies.

### **5.6. Impersonal versus Personal Punitive Decision-making.**

There is some evidence in mock juror studies that God concepts do play some role in punitive decision-making in the way I hypothesised. For example, in Miller and Bornstein’s (2006) mock jury study of death penalty sentencing participants were the most punitive towards a defendant whose attorney quoted Biblical scripture prescribing mercy (i.e., a forgiving God prime). Although other interpretations of these results are possible (jurors may have found this a cynical legal ploy, for instance), this finding raises the possibility that when the stakes are high, or *when an individual plays some role in the punitive outcome*, then forgiving God concepts *do* increase support for earthly punishment in the way we hypothesised.There is a considerable difference between having to take a personal decision to determine the severity of John’s punishment and deciding how many of your tax dollars you would like put toward an institutional 3rd party (i.e., the state) punishing John. Indeed, previous research has revealed that similar distinctions in moral judgement are underpinned by distinct neural activity (Greene, Sommerville, Nystrom, Darley & Cohen, 2001).

### **5.7. Forgiving Gods and Cooperation.**

Given that a) there was no significant difference in endorsement of state-sanctioned punishment between forgiving and punishing conditions, and b) the body of research that has highlighted the profound effects of forgiveness on prosociality (Karremans*,* Van Lange & Holland, 2005; Karremans & Van Lange, 2004; McCullough, Fincham & Tsang, 2003; Pargament, McCullough & Thoresen, 2000), one possibility is that *both* forgiving and punishing God concepts promote co-operation. There was some support for this notion in the results as *both* punishing and forgiving God beliefs (when salient) were significant predictors of support for state-sanctioned punishment in response to a victim-directed transgression (Study 8). Therefore, I am sympathetic to researchers who have claimed that extant theories of the evolution of human co-operation have focussed on supernatural punishment to the detriment of the role of divine forgiveness (Johnson et al.,2017; Johnson & Cohen, 2016). A multitude of God representations are available in the historical and cross-cultural record, and as Johnson and Cohen (2016) argue, an authoritarian ‘Big’ God is only part of the story. Although punishing Gods may well have been a factor in driving large-scale cooperation, they do not, single-handedly, sustain it (Johnson et al., 2017; Johnson & Cohen, 2016).

The threat of supernatural punishment may curtail unethical behaviour but does not prevent it altogether. Given that people will invariably transgress, a mechanism that reconciles those who have transgressed with the group becomes necessary (Wilson, 2002). If not, the cost of punishment and subsequent exclusion rate would be so high that the stability of the group would be jeopardised. Beliefs in forgiving Gods could be one such mechanism. Fully exploring the cohesive role of forgiving God beliefs, for example, their re-integrative function (McCullough et al*.*, 2005) is an important endeavour. For less heinous wrongdoings, it is more advantageous to rehabilitate and reconcile such individuals back into the group. This reconciliation often includes a renewed sense of commitment to the group and shared norms (see McCullough et al., 2005).

## 6. Limitations.

There were several limitations in this series of studies. First, none of the studies included a manipulation check. Therefore, it is not possible to know whether the null results in Studies 6 and 7, for example, were due to the experimental manipulation having no effect or whether the prime itself failed. This is closely related to the second limitation: the passages used for the explicit primes in Studies 6 and 7 were subtle and perhaps too weak to have the intended effect. Moreover, both passages contained references to death that could have inadvertently primed terror management concerns (i.e., mortality salience has been shown to have an effect of increasing prosociality, especially towards the ingroup) (Jonas, Schimel, Greenberg & Pyszczynski, 2002; Zaleskiewicz, Gasiorowska & Kesebir, 2015). On reflection, the use of previously used priming methods that have been shown to activate the relevant aspects of God (e.g., Harrel, 2012; Yilmaz and Bahçekapili, 2016) may have been preferable. Future research should take this into account when using priming methodologies to explore the effects of different God concepts on prosociality.

A third potential limitation for (all) studies was the lack of a neutral condition. The initial aim was simply to establish whether there was a difference in support for state–sanctioned punishment between the forgiving God and punishing God primes. The results revealed no significant difference between the two conditions in both Study 6 and 7. Had there been such a difference it would then have been appropriate to run a follow-up with a neutral prime to clarify this. The fact there was not a difference between the two prime conditions, in my view, makes the absence of a neutral condition less problematic than it otherwise would have been.

One final point of note was the gender disparity seen in Study 9 (and in the Salient condition in Study 8). At an approximate ratio of 2:1 (males to females), our sample is atypical of the MTurk population. The normal gender breakdown on MTurk is roughly equal (51% female; 49% male), though significant deviations from the average have been noted across countries (see Difallah, Filatova & Ipeirotis 2018). I acknowledge this as a potential limitation given the evidence suggesting reliable gender differences in religiosity (e.g., De Bono & Kuschpel 2014; Roth & Kroll 2007; Stark 2002; Sullins 2006; Thompson 1991; Walter & Davie 1998).

## 7. Conclusion.

Across four well-powered studies, there were no significant differences between punitive or forgiving God representations on the endorsement of state-sanctioned punishment. The results suggest that in response to a victim-directed transgression, both punishing *and* forgiving God beliefs (when salient) are significant predictors of endorsement of state-sanctioned punishment. Therefore, future research should investigate in more detail the role of divine forgiveness in human prosociality and cooperation. Extant theories (e.g., the Supernatural Punishment Hypothesis) might have systematically over-emphasised the role of divine punishment to the detriment of divine forgiveness.

Punishment may not be the only mechanism that allows for the restoration of fairness and, as I have suggested, it is possible that people perceive that God can redress the imbalance caused by a wrongdoing through the administration of forgiveness. In the next chapter I consider other (earthly) mechanisms for the restoration of fairness, such as through restorative behaviours, i.e., moral cleansing. In particular, I look at restorative responses to transgressions committed by members of our ingroup. Are the moral consequences of past ingroup misdeeds transferred onto to us? That is, do we feel the need to redress the imbalance caused by the wrongdoings of others?

# VII. STUDIES 10-11: DOES INGROUP IMMORALITY TRIGGER VICARIOUS MORAL CLEANSING?

*“We are like chameleons, we take our hue and the colour of our moral character, from those who are around us” ~* John Locke.

## 1. Introduction.

In 1838, two Jesuit priests at the prestigious Georgetown University in Washington sold 272 slaves to pay off a series of debts. This sale secured the survival of the Institution. When this was brought to light in 2016 it triggered a series of student protests as well as an internal investigation at the University (Swarns, 2016). A series of compensatory actions were taken as a result, one of which was the offer of priority admission to any young descendants of the 272 slaves.

The Georgetown story suggests a vicarious aspect to moral cleansing. In the same way that recalling past immoral actions renders individuals more likely to engage in subsequent moral behaviours to restore a damaged moral self-concept (Sachdeva, Iliev & Medin, 2009; West & Zhong, 2015; Zhong & Liljenquist, 2006), information about the past misdeeds of one’singroup may also trigger compensatory behaviours (Lickel, Schmader, Curtis, Scarnier & Ames, 2005). In the two studies in this chapter, I aimed to demonstrate the existence of a vicarious moral cleansing effect and to elucidate the mechanisms by which people incorporate information about the (im)moral behaviour of similar others into their individual self-image and how this affects subsequent behaviour.

### **1.1. Theoretical Background.**

People readily incorporate moralinformation about similar others into their own self-concept (Goldstein & Cialdini, 2007) and research has shown that individuals can use the moral credentials of the ingroup to license subsequent immoral behaviour (Kouchaki, 2011), i.e., *vicarious* moral licensing (see Chapter 2, section 7.3). For example, participants told their group was more moral than similar others were more likely to express a preference for hiring a white candidate (over a black candidate) in a hypothetical job selection task (Kouchaki, 2011, Study 1a). Similarly, when given information about a group member’s previous non-discriminatory behaviour (i.e., having chosen a Hispanic applicant in a prior task), participants went on to make discriminatory ratings against the Hispanic for a job position that was stereotypically associated with whites (Kouchaki, 2011, Study 1b).

Given that moral licensing and moral cleansing are theorised to work together in a broader system of moral regulation (Brañas-Garza et al., 2013; Sachdeva et al., 2009) one may reasonably hypothesise that a vicarious moral *cleansing* effect also exists.

### **1.2. Vicarious Guilt.**

A prominent conceptualisation of guilt is as a self*-*conscious emotion (Baumeister, Stillwell & Heatherton, 1994; Tangney & Fischer, 1995) that involves a perception of personal responsibility for a negative event (Manstead & Tetlock, 1989; Smith & Ellsworth, 1985). More recent research however has investigated the vicarious aspect of guilt. Here the person who experiences the emotion is not the agent of the wrongdoing. For example, people report feeling guilty when recalling a close other’s misdeed (Chen, Wei, Shang, Wang & Zhang, 2016; Lickel et al., 2005; Scarnier, Schmader, & Lickel, 2009; Welten, Zeelenberg, & Breugelmans, 2012); being reminded of the historical wrongdoing of one’s country (Doosje et al, 1998; Iyer et al., 2003); thinking about the immoral activity of an organisation that one belongs to (Bernhard, 2016) or when thinking about the wrongdoings of one’s children (Scarnier et al., 2009). Even indirect contact with a moral transgressor has been found to elicit feelings of vicarious guilt (Eskine, Novreshe & Richards, 2013).

Taken together these findings strongly suggest that the emotional consequences of immoral activity can be transferred on to individuals other than the proximal agent of the wrongdoing; guilt can be *transferred*. However, does the experience of vicarious guilt trigger engagement in compensatory behaviours? That is, do the misdeeds of similar others produce a desire for vicarious moral cleansing?

### **1.3. Vicarious Moral Cleansing.**

According to self-completion theory (Gollwitzer & Kirchhof, 1998; Wicklund & Gollwitzer, 1982), when one’s individual or a valued group identity is threatened, one engages in compensatory behaviours as a way of re-affirming the identity in question (Gollwitzer, Wicklund, & Hilton, 1982). Therefore, vicarious moral cleansing may occur when one’s group identity is threatened by the immorality of its members and one seeks to find ways to reassert the identity in question.

There is a degree of evidence to support this. For example, Bernhard (2016) found that German bankers who strongly identified with their profession were not only more likely to experience higher levels of vicarious guilt (for misconduct and unethicality in their profession) but also to support the implementation of strict regulations in their profession; a form of vicarious moral cleansing.

### **1.4. The Moderating Effect of Identification.**

The degree to which a person is influenced by the behaviour of ingroup members depends on the strength of their identification with the group (Goldstein & Cialdini, 2007; Tajfel, 1981; Tajfel & Turner, 1979). Therefore, the strength of a shared identity with a group or individual should be an important factor in vicarious moral compensation. There is compelling evidence to support this. For example, Kouchaki (2011, Study 4) found that the feeling of shared identity or connectedness to an ingroup member who had previously demonstrated non-prejudiced credentials moderated vicarious moral licensing effects (i.e., the effects were stronger when participants were more strongly identified with the group). Similarly, the results of a Bernhard (2016) revealed that bankers who supported the implementation of compensatory regulations in the banking industry had indicated a high degree of professional identification. These lines of evidence suggest that vicarious licensing (and perhaps cleansing) depend on the degree of identification with a group and furthermore, the strength of identification must pass a certain level for vicarious effects to occur.

### **1.5. Identity Fusion.**

In both studies in this chapter, a specific type of group alignment, i.e., “identity fusion”, was measured to assess the strength of participants’ bond to the group. Here individuals experience an intense sense of oneness with a particular group (Fredman, Bastian & Swann, 2017; Swann, Jetten, Gómez, Whitehouse & Bastian, 2012) and research suggests it is a strong predictor of extreme pro-group behaviour (Swann et al., 2014; Swann, Gómez, Buhrmester, López-Rodríguez, Jiménez & Vásquez, 2014). Unlike in social identity theory (Tajfel, 1981) where the salience of the social self reduces the salience of the personal self and vice versa, both selves are theorised to remain salient in identity fusion[[68]](#footnote-68). Moreover, the boundary between the personal and social self becomes porous allowing aspects of both to mix with one another leading the individual to become fused. A fused individual feels not only profoundly connected to the group but also to its individual members and any threat to the group may well be interpreted as a threat to their personal self also (Swann, Burhmester et al., 2014; Swann, Gómez et al., 2014).

### **1.6. Overview of The Current Studies.**

The studies in this chapter had two aims. First, both sought to conceptually replicate the vicarious moral licensing effect documented by Kouchaki (2011). Second, both studies aimed to extend this finding by building on the suggestive evidence of Bernhard (2016) and provide confirmatory evidence for the existence of a vicarious moral cleansing effect.

In Study 10, American participants were exposed to positive, negative or neutral information about the morality of fellow Americans. In line with Kouchaki (2011) I predicted that participants exposed to positive information about ingroup morality would demonstrate more prejudicial attitudes on a hypothetical job-hiring scenario, i.e., vicarious moral licensing. Conversely, I hypothesised that Americans who were exposed to negative information about ingroup morality would demonstrate less prejudicial attitudes on the same task, i.e., vicarious moral cleansing. Following Kouchaki’s (2011) finding that group identification moderates vicarious moral licensing effects I made the further prediction that identity fusion would underpin any observed effects. Study 10 was preregistered with AsPredicted (https://aspredicted.org/43pe7.pdf).

Study 11 again aimed to re-examine Kouchaki’s finding (2011) and to establish whether any vicarious licensing effects are due to shared membership with the target. It differed from Study 10 as it included information about the relative (im)morality of a salient outgroup (i.e., Iran) as well as information about ingroup (im)morality. Study 11 was also pre-registered on AsPredicted (https://aspredicted.org/5mz6z.pdf).

## 2. Study 10.

### **2.1. Methods.**

#### 2.1.1. Participants.

A power analysis was performed for sample size estimation using the software package G\*Power (Faul et al., 2009). With input parameters of *d* = 0.2 (a small effect), *alpha* = 0.05, *power* = 0.9 and 3 groups, the recommended sample size was *N* = 321. I added an extra 15% of the recommended total to account for attrition rates on MTurk giving a final *N* = 369. All participants were recruited through MTurk and paid $1.85 for their time (equivalent to an hourly rate of $7.25). The study was preregistered with AsPredicted (https://aspredicted.org/43pe7.pdf). Participants first had to respond to a series of four questions: 1) highest level of education, 2) marital status, 3) age and 4) nationality. Unbeknownst to participants, anyone who indicated a nationality other than American was not able to take part. Sixty-nine participants were excluded for either failing a comprehension check after the experimental manipulation or failing an attention check, which left a remaining *N* = 300. The pre-registration stated the exclusion of any non-white participants leaving a final sample of 235 participants.

#### 2.1.2. Measures.

*Moral compensation.*

This was a job-hiring scenario used by Kouchaki (2011) and taken originally from Monin and Miller (2001) (see Appendix U). Participants were asked to imagine themselves as the Chief of Police in a rural US town with historic problems of racial hostility in the force. They were then asked to indicate whether they thought the job was better suited for any one ethnicity on a 7-point scale: (1) = *“Yes, much better for a black candidate”*, (7) = *“Yes, much better for a white candidate”* and a mid-point of (4) = *“No, I do not feel this way at all”*). In Kouchaki’s study (2011, Study 1a) this job selection task was used as a measure of vicarious moral licensing, whereby higher scores on the scale indicated higher levels of discrimination and thus higher vicarious moral licensing. In this study it was used to measure bothvicarious moral licensing (high scores) and vicarious moral cleansing (low scores) as it suggests that participants either felt licensed to make a discriminatory response (by selecting a stronger preference for a white candidate) or refrained from making a discriminatory response in an effort to cleanse their moral identity (i.e., stronger preference for a black candidate). Therefore, higher scores were taken to be indicative of vicarious moral licensing and lower scores of vicarious moral cleansing.

*Identity fusion.*

I used the7-item verbal fusion measure of Gómez, Brooks, Buhrmester, Vázquez, Jetten and Swann (2011, see Appendix V). Example items are “I am one with my country”and “I am strong because of my country” (α ranging from .70 to .93).Items are scored on a scale of (0) = *totally disagree* to (7) = *totally agree.* Higher scores indicate stronger levels of identity fusion with a target group (i.e., America).

#### 2.1.3. Design and procedure.

Participants were told that the study was designed to understand how people construct their national identities. After giving their consent, participants first had to respond to a series of four questions: 1) highest level of education, 2) marital status, 3) age and 4) nationality. If they responded they were not US citizens, they were excluded at this stage of the study. Participants then completed the verbal measure of identity fusion to assess the strength of their affiliation with the US. This was followed by a 23-item measure of essentialist thinking about social categories (Bastian and Haslam, 2006), included as a filler questionnaire (this scale was thematically compatible with the cover story given to participants). The questionnaire also included an attention check (see Appendix A). Any participants who failed this check twice were excluded from the analyses.

All participants were then randomly assigned to one of three conditions: immoral, moral and neutral. All participants read a message that contained information about recent cross-cultural research investigating moral behaviours and attitudes. In this message they were told that several studies had revealed their ingroup (fellow Americans) to be either *less* moral than citizens from other countries (immoral condition), *more* moral than citizens from other countries (moral condition), or neither more nor less moral than others (neutral condition). All participants then completed the filler task (to reduce demand characteristics) before responding to the job-hiring scenario (Kouchaki, 2011; taken from Monin & Miller, 2001, see Appendix U). Finally, they completed a series of demographics questions (including an ethnicity item) and were fully debriefed.

### **2.2. Results.**

As stated in the pre-registration, I investigated whether there were any significant differences in job suitability ratings across conditions for white participants only (*N* = 235*)*. Suitability ratings for each condition were non-normally distributed, as assessed by Shapiro-Wilks’s test (*p* < .05). Moreover, the data were ordinal thus rendering a one-way ANOVA strictly inappropriate. Therefore, a Kruskal-Wallis H test was conducted to determine if there were significant differences in job suitability ratings between conditions: neutral (*N* = 89), immoral (*N* = 75) and moral (*N* = 71). Distributions of job suitability ratings were not similar for all groups (i.e., they had different variabilities) as assessed by visual inspection of a boxplot. The mean ranks of job suitability ratings were statistically significantly different between groups, *Χ2*(2) = 6.90, *p* = .036, *η2 =*.03 (a small effect on Cohen’s classification, 1988).

Subsequently pairwise comparisons were performed using Dunn’s (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted *p*-values are presented, and values are mean ranks unless otherwise stated. This post hoc analysis revealed a statistically significant difference in job suitability ratings between the moralcondition (131.49) and the neutralcondition (107.95), *p* = .026, indicating that participants in the moral condition made more discriminatory hiring decisions. No other inter-group differences were significant.

Kouchaki (2011, Study 1a) did not exclude non-white participants from her analysis. Although her sample (*N =* 127) was predominantly white (76% self-identified as white/european american), it was not exclusively so (10% self-identified as asian/asian american, 7% self-identified as latino/hispanic, and the remainder self-identified as other or unknown.) This is problematic as the logic of the measure is that preference for a non-white candidate is a signal of morality (non-discrimination), yet participants who are themselves non-white may prefer a non-white candidate for self-interested reasons.

Following Kouchaki’s original analysis (2011) I also checked for any differences in job suitability ratings between conditions using the total recruited sample (*N =* 300). The full ethnic breakdown of participants was as follows: 78% self-identified as white/caucasian, 4% self-identified as hispanic/latino, 8% self-identified as black/african-american, 15% self-identified as native american/alaskan native, 5% self-identified as asian, 2% self-identified as mixed race and <1% self-identified as other.

Job suitability ratings for each condition were non-normally distributed, as assessed by Shapiro-Wilks’s test (*p* < .05). As in the previous analysis the data were ordinal thus rendering a one-way ANOVA strictly inappropriate. Therefore, a Kruskal-Wallis H test was conducted to determine if there were significant differences in job suitability ratings between conditions: neutral (*N* = 108), immoral (*N* = 101) and moral (*N* = 91) (see Table 18 for cell means). Distributions of job suitability ratings were not similar for all groups (i.e., they had different variabilities) as assessed by visual inspection of a boxplot. The mean ranks of job suitability ratings were statistically significantly different between groups, *Χ2*(2) = 7.12, *p* = .028, *η2 =*.024.

Subsequent pairwise comparisons were performed using Dunn’s (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted *p*-values are presented, and values are mean ranks unless otherwise stated. This post hoc analysis revealed a statistically significant difference in job suitability ratings between the moralcondition (166.00) and the neutralcondition (138.58), *p* = .023, indicating that participants in the moral condition made more discriminatory decisions –suggesting vicarious moral licensing. No other inter-group differences were significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  | **95% Confidence Interval for the Mean** | |
|  | **N** | **Mean** | **SD** | **Lower bound** | **Upper bound** |
| **Control** | 89 | 4.08 | 1.13 | 3.84 | 4.31 |
| **Immoral** | 75 | 4.20 | 0.81 | 4.01 | 4.39 |
| **Moral** | 71 | 4.51 | 1.18 | 4.23 | 4.79 |
| **Total** | 235 | 4.25 | 1.07 | 4.11 | 4.38 |

Table 18. Cell means for the effect in Study 10

A moderated multiple regression was conducted to check whether identity fusion moderated any effects of the experimental condition on job suitability ratings. As stated in the pre-registration, this analysis was conducted on white participants only (*N =* 235). I regressed job suitability ratings onto identity fusion scores (mean-centred), a dummy variable for the moral condition; a dummy for the immoral condition; the interaction between the moral dummy and fusion and the interaction between the immoral dummy and fusion. The regression model was not significant, F(5, 229) = 1.76, p = .12, adj. R2= .04 and neither interaction was a significant predictor of job suitability ratings (though the coefficient for participation in the moral condition was significant (*p* = .026), indicating vicarious moral licensing). Regression coefficients and standard errors can be found below in Table 19.

Table 19. Results of moderated multiple regression analysis regressing job suitability ratings on fusion, a dummy variable for the moral condition, a dummy for the immoral condition, and the interaction of these dummy variables with fusion.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | ***B*** | ***SEb*** | ***Beta*** | ***t*** | ***p*** |
| **(Constant)** | 3.85 | .20 |  | 19.21 | .001 |
| **Fusion (mean-centred)** | .002 | .001 | 0.14 | 1.37 | .171 |
| **Dummy Moral** | .69 | .31 | 0.30 | 2.24 | .026\* |
| **Dummy moral x Fusion** | -.002 | .002 | -.13 | -.97 | .331 |
| **Dummy immoral** | .28 | .29 | .12 | .98 | .329 |
| **Dummy immoral x Fusion** | .000 | .002 | -.09 | -.66 | .510 |

*B* = unstandardized regression coefficient; *SEb =* Standard error of the coefficient; *Beta =* standardised coefficient; An asterisk indicates significance at the 0.05 level.

### **2.3. Discussion.**

This study replicated the vicarious moral licensing effect of Kouchaki (2011). However, there was no evidence of vicarious moral cleansing. Given that both moral cleansing and licensing are theorised to work together as part of a broader system of moral balancing, this apparent asymmetry is puzzling; if one exists, then so should the other. One potential explanation for this relates to a self-serving bias, which causes us to present information in a way that allows us to justify our behaviours and retain a positive self-image (Finch & Cialdini, 1989; Gino & Galinsky, 2012; Mazar, et al, 2008; Sanitioso, Kunda, & Fong, 1990). For example, when people feel conflicted about their environmental behaviours, recalling previous environmentally friendly behaviours can help rationalise their environmentally detrimental behaviours (Hope, Jones, Webb, Watson, & Kaklamanou, 2017). Therefore, a similar bias might lead us to discount or minimise negative information about the ingroup and present this information in a way that allows us to protect and retain a positive group image.

Vicarious moral licensing on the other hand can be readily be explained by research that suggests positive information is integrated more readily into people’s beliefs than negative information (i.e. an “optimism bias”), even when this positive news contradicts prior beliefs and leads to unrealistic optimism and overly positive self-perceptions (e.g., Eli & Rao, 2011; Sharot & Garrett, 2016; Sharot, Korn & Dolan, 2011; Tappin, McKay & van der Leer, 2017; though see Shah, Harris, Bird, Catmur & Hahn, 2016). This bias also has a vicarious aspect and occurs when learning about events affecting friends (Kappes, Crockett, Faber, Savulescu & Kahane, 2018).

In short, it is possible that people readily attend to positive information about the group, via the optimism bias, which they use to bolster their moral credentials, and subsequently engage in vicarious licensing. On the other hand, a self-serving bias allows people to minimise or discount negative information about the group thus explaining the absence of vicarious cleansing effects. To provide further evidence of a discrepancy between these two mechanisms (i.e., the presence of vicarious moral licensing, but not vicarious moral cleansing) I thought it prudent to probe this apparent asymmetry further.

### **3. Study 11.**

Given the findings of Study 10, I hypothesised that positive information about ingroup morality would trigger vicarious moral licensing and second, that negative information about ingroup morality would *not* trigger vicarious moral cleansing effects. I further predicted that vicarious moral licensing effects will be found only in response to positive information about *in*group morality, therefore I included negative/positive/neutral information about the morality of a salient outgroup (i.e., Iran).

### **3.1. Methods.**

#### 3.1.1. Participants.

A power analysis was run which indicated a sample size of *N =* 420, (parameters set at *d* = 0.2, *alpha* = 0.05, *power* = 0.9, 6 groups). An additional 15% of the recommended *N* was added to account for potential attrition on MTurk, giving a total *N* = 483. As in the previous study, participants were asked a series of preliminary questions at the outset to ensure that only white, US citizens participated. Of those who did take part, 109 participants were excluded: 14 correctly guessed the hypothesis, 27 did not fully complete the study, 1 failed an attention check embedded in a filler questionnaire and 67 failed a 4-item comprehension check following the experimental manipulation (which asked them to indicate the result of the reported study). This left a remaining *N* = 374 (182 males; 192 females).

#### 3.1.2. Measures.

*Moral compensation.*This was assessed with the same police job-hiring task as in Study 10. As before, higher scores were indicative of vicarious moral licensing and lower scores indicative of vicarious moral cleansing.

*Verbal fusion.* Again, the 7-item Gómez et al. (2011) verbal fusion measure was used. Higher scores indicate stronger levels of identity fusion with the target group (i.e., the US).

#### 3.1.3. Design and procedure.

Having given informed consent, all participants then completed the verbal fusion measure (Gómez et al., 2011) before being randomly assigned to one of six conditions: ingroup immoral information, ingroup moral information, ingroup neutral information, outgroup immoral information, outgroup moral information and outgroup neutral information. As in Study 10, participants in the ingroup immoral information condition were told that a recent study of citizens from multiple countries revealed their ingroup (i.e., fellow Americans) to be less moral than other countries. In the ingroup moral information condition they were told that the study revealed their ingroup was more moral than others and in the ingroup neutral information condition they were told that the study revealed their ingroup to be neither more nor less moral than citizens of other countries tested in the study (Appendix W). In the outgroup conditions, participants were told that a study of citizens from multiple countries found that Iranians were the least moral (outgroup immoral information condition), the most moral (outgroup moral information condition) or no more nor less moral than citizens of other countries (outgroup neutral information condition) (Appendix X).

All participants then completed a filler task (the same as in Study 10) to reduce demand characteristics before responding to the job-hiring scenario. Finally, they completed a series of demographics questions, were fully debriefed and paid $1.85 for their time.

### **3.2. Results.**

A 2 x 3 ANOVA was conducted to examine the influence of group type (ingroup *vs.* outgroup) and valence of information (moral, neutral and immoral) on job suitability ratings. All effects were non-significant at the 0.05 level. There was a marginally significant interaction between group type and valence of information (*p* = .077). With due caution, I probed this further. Pairwise comparisons revealed a statistically significant difference in mean job suitability ratings between ingroup (*M* = 4.10, *SD* = .13) and outgroup (*M* = 4.59, *SD* = .13) at the moral level, *p* = .01, but for no other combination (see fig. 4).

*Figure 4.* Plot of estimated marginal means for job suitability ratings, Study 11.

Information about the superior morality of an outgroup led to stronger job suitability ratings for a white candidate than information about the moral superiority of one’s ingroup. Moreover, for participants in the three ingroup conditions, those in the moral condition made the strongest job suitability ratings for a black candidate; a complete reversal of the results of Study 10.

### **3.3. Discussion.**

With the exception of the outgroup conditions, the materials and procedure in Studies 10 and 11 were identical. Therefore, the results of Study 11 are unexpected. Why would the vicarious moral licensing effect found in Study 10 have not just disappeared, but apparently reversed? This might have been due to moral consistency effects, i.e., acting *in line* with the valence of the presented moral information. Prior research suggests the behaviour of others can provide us with information about social norms and can lead us to behave in a similar manner as those around us (e.g., Cialdini, 1993; Gino, Ayal, & Ariely, 2009; Goldstein, Martin, & Cialdini, 2008). For example, studies have found that when exposed to an ingroup member’s unethical behaviour, participants often align with the behaviour and behave dishonestly themselves as the ingroup member’s behaviour provides information about what is normatively appropriate in the given context (Gino, Ayal, & Ariely, 2009; Gino, Gu, & Zhong, 2009). Similarly, the priming of a (positive) role model can lead people to behave in prime-consistent ways (Eibach, Libby & Ehrlinger, 2009; Fitzsimons & Bargh, 2003). Joosten et al. (2014) suggest that morally consistent behaviours are driven by long-term reputational concerns whereas moral licensing effects are on the fly responses to moral information.

But as the materials in both studies here were identical, it remains unclear why positive ingroup moral information would trigger a licensing effect in one study, and consistency effects in the other.

## 4. General Discussion.

The studies in this chapter had two aims: First, to replicate Kouchaki’s vicarious moral licensing effect (2011), and second, to provide evidence of vicarious moral cleansing. The results of Study 10 revealed that discriminatory attitudes were highest among participants presented with moral information about the ingroup; thus replicating the vicarious moral licensing effect (Kouchaki, 2011).

With regards to the second aim neither Study 10 nor Study 11 revealed evidence of a vicarious moral cleansing effect. Also of note in Study 11, was the surprising finding that vicarious moral licensing was *lower* among participants given moral information about their ingroup, a reversal of the findings from Study 10. I will discuss each of these findings in turn.

### **4.1. Vicarious Moral Licensing.**

In Study 10 participants presented with moral information about their ingroup demonstrated a preference for hiring a white candidate in a subsequent job-hiring scenario. In line with Kouchaki (2011), this suggests that participants used the moral credentials of the ingroup to bolster their moral self-image, which then freed them up to exhibit racially discriminatory judgements. However, although this replicates the vicarious moral licensing effect, there is a caveat. The exclusion of non-white participants brought the sample size in Study 10 down to 235 participants. This was far below the *N* recommended by an a priori power analysis (*N* = 321). An analysis of obtained power with a small effect size (*η2 =*.03), an alpha level of .036 and three groups revealed the power achieved was .02. Therefore, the analysis conducted on the white only sample was severely underpowered (but see Lakens, 2014. Following Kouchaki (2011), I also conducted analyses on a mixed-race sample (*N* = 300) and the vicarious moral licensing effect held. The effect did not hold however Study 11; conversely participants in the moral ingroup condition made the strongest job suitability ratings for a black candidate (i.e., vicarious moral licensing).

### **4.2. The Absence of Vicarious Moral Cleansing.**

After Studies 10 & 11 were conducted two publications appeared in the literature that were conceptually and methodologically very similar: Newman and Brucks (2017) and Meijer et al. (2018). Both sought to find evidence of vicarious moral licensing and cleansing effects and both involved exposing participants to positive/negative/neutral information about an ingroup’s (or close other’s) morality.

In line with Kouchaki and the results of Study 10, both Newman and Brucks (2017) and Meijer et al. (2018) found evidence of vicarious moral licensing (see Chapter 2, section 7.3 for a full discussion). With regards to the existence of vicarious moral cleansing however, the results were mixed.

#### 4.2.1. Overview of subsequent vicarious cleansing studies.

The set of studies on consumer behaviour by Newman & Brucks (2017), revealed that participants with a high degree of self-brand overlap with Nike were more generous in an economic game after being informed about Nike’s socially irresponsible (i.e., immoral) behaviour compared with neutral information (Newman & Brucks, 2017, Study 1). A further study found that for high self-brand overlap participants, the increase in moral behaviour following exposure to negative information about Nike was mediated by guilt (Newman & Brucks, 2017, Study 2). Newman and Brucks suggest that for those participants high in self-brand overlap, exposure to Nike’s immoral behaviour activated negative representations of the moral self and feelings of vicarious guilt which together triggered engagement in vicarious cleansing behaviours.

However, a subsequent set of studies by Meijer et al. (2018) tested for vicarious moral licensing and cleansing effects in the environmental domain. Across four studies they found evidence of vicarious moral licensing in response to the pro-environmental behaviour of a close other – but *not* the opposite. That is, participants exposed to information about a close other’s negative environmental behaviours did *not* express more intentions to engage in environmentally friendly behaviours, i.e., there was no evidence of vicarious moral cleansing (Meijer et al., 2018, Studies 3 & 4). Moreover a series of internal meta-analyses by Meijer et al., (2016) revealed a significant vicarious moral licensing effect with a small-to-moderate mean effect size (*d* = −0.43, *SE* = 0.19, CI, 95% [–0.80, –0.06], *p* = .022) but no support for a vicarious cleansing effect, (*d* = 0.24, *SE* = 0.20, CI, 95% [–0.15, 0.62], *p* = .230).

Given that the results of Meijer et al. (2018) and Study 10 point to the absence of vicarious moral cleansing the obvious question is why would it not exist if vicarious moral licensing does? I propose some possible explanations.

### **4.3. Boundary Conditions of Vicarious Moral Cleansing.**

One response to the documented absence of vicarious moral cleansing is that it may exist, but it does not work together with vicarious moral licensing in the way suggested by a moral balancing model. This may explain why we do not see evidence of both mechanisms within the same paradigm. If so, this undermines the contention that both mechanisms work together within a system of moral balancing, as they do in a compensation model (Sachdeva et al., 2009; Brañas-Garza et al., 2013).

It might be, for example, that vicarious cleansing does exist but only in response to a *specific* past ingroup wrong (The Georgetown example) rather than general information about the immorality of the group. A past wrongdoing (e.g., the sale of slaves to pay off university debt) is a concrete example of the ingroup’s immorality. However, simply being told that your ingroup is less moral than others may not be sufficiently compelling (and may be easier to marshal defensive resources against), particularly if one identifies strongly with the group.

### **4.4. Protecting the Group Moral Image.**

Research has found evidence to suggest that the affective consequences of negative information are stronger and longer lasting than those of positive information (Baumeister, Bratslavsy, Fikenauer & Vohs, 2001; Ikegami, 1993). When it comes to information about the self, people are motivated to avoid bad feedback to protect themselves from the negative impact (Baumeister & Cairns, 1992; Baumeister et al., 2001). Therefore, individuals might be motivated to reinterpret negative information about the morality of the ingroup to protect a valued group image, leading to an absence of vicarious cleansing. Meijer et al. (2018) converge upon a similar explanation and suggest people may seek alternative ways to resolve the need to compensate for the immorality of close others by employing cognitive dissonance reduction strategies. They write:

people seem to be able to construe information in such a way that is most beneficial for themselves, which may explain the absence of vicarious cleansing effects: Because people might construe the morally questionable behavior of close others as less immoral, people are less likely to feel the need to cleanse their close other’s morally questionable behaviors (p. 19).

Moral competence is a particularly important factor in determining positive group evaluation (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Ellemers, Pagliaro, Barreto & Leach, 2008) especially in relation to other groups (Leach, Ellemers & Barreto, 2007). We have a tendency to perceive our group as more moral than others (Leach et al, 2007; Levine & Campbell, 1972; Roccas, Klar & Liviatan, 2006). Therefore, when faced with information that threatens group morality, we may be motivated to process this information so as to interpret it in line with an illusion of ingroup moral superiority (Tappin & McKay, 2017). For example, in response to ingroup atrocities individuals have been shown to demonstrate a type of defensive strategy called ‘moral shifting’ (Leidner & Castano, 2012); when the immoral actions of ingroup members threaten the group image, the accessibility and salience of certain moral foundations shift[[69]](#footnote-69). Leidner and Castano (2012) propose that the moral foundations of harm and fairness become less accessible and those of loyalty and authority become more salient and assume more significance (Leidner & Castano, 2012). This shift protects the group identity as the behaviour of ingroup members is reinterpreted according to different moral foundations (Haidt & Graham, 2007). For example, the killing of outgroup members by ingroup soldiers might be less likely to be considered immoral when assessed in line with the moral foundations of loyalty and authority. In fact, due to this shift in moral foundations, it may even be considered a moral act (Leidner & Castano, 2012)[[70]](#footnote-70).

## 5. Limitations.

There were several conceptual and methodological limitations to the studies in this chapter. First, the framing of the job-hiring scenario potentially provides participants with a legitimate reason for favouring a white candidate over a black candidate. Second, although the job-hiring scenario has been previously used in the moral licensing literature (Monin & Miller, 2001; Kouchaki, 2011), there is a further issue of construct validity. And third, there is a tendency in the literature to conflate non-discriminatory behaviour with moral behaviour. I will now discuss each of these limitations in turn.

### **5.1. Deontological Considerations.**

In the job-hiring task, it was made explicit to participants that the working environment was hostile to a black candidate. The rationale for this was that rather than asking participants to make an explicitly racist choice, they would have a legitimate reason to favour a white candidate (Kouchaki, 2011). However, this is problematic, as there is no way to tell whether lower suitability ratings for a black candidate are due to the expression of prejudice or to deontological considerations. From the latter perspective it would be immoralto deliberately place a black candidate in a disadvantageous position[[71]](#footnote-71).

### **5.2. Problems with the Job-Hiring Scenario*.***

Participants’ responses to the job-hiring task are measured on a 7-point continuous scale ranging from 1 *(“yes, much better for a Ethnic Minority/Black candidate”*) to 7 *(“yes, much better for a White candidate”*), with a midpoint labelled 4 (*“No, I do not feel this way at all”*). When responding to the question, *“Do you feel that this specific position is better suited for any one ethnicity?”* participants only have to select the mid-point (4) to establish themselves as non-prejudiced. This has two consequences: first, the *mid-point* of the scale is the maximally egalitarian option and second, as a result, *any* deviation from this mid-point could be construed as indicative of discrimination, albeit negative *or* positive[[72]](#footnote-72).

### **5.3. The Conflation of Egalitarianism and Morality.**

As McKay and Whitehouse (2015) noted, many of the inconsistencies in the religious prosociality literature may be due to a Westernised and liberal conceptualisation of morality. Under this view morality consists of “nice” behaviours. Similarly, there is a tendency in the moral compensation literature to assume that the expression of liberal or egalitarian attitudes is inherently “moral”. As McKay and Whitehouse note, the liberal and westernised conception of morality that pervades academia does so at the expense of “a plethora of moral concerns, motivations, and behaviors” (McKay & Whitehouse, 2015, p.4). Drawing conclusions about universal human moral behaviour is problematic if one presupposes a particular conception of morality. For example, if it’s not clear what the “moral” signal is, in that it is ambiguous or culturally relative, then one’s measure will fail to be useful as a measure of moral licensing or cleansing.

## 6. Conclusion.

Study 10 replicated a previous finding in the literature; vicarious moral licensing (Kouchaki, 2011) but provided no evidence of a vicarious moral *cleansing* effect. This is potentially due to defensive processes geared towards protecting a valued ingroup identity. The results of Study 11 yielded a surprising pattern of results. Contrary to the initial prediction, the highest levels of vicarious moral licensing were seen in response to information about *out*group morality rather than ingroup morality. That is, being told an outgroup was more moral than the ingroup led to an increase in vicarious moral licensing. In addition, the lowest levels of vicarious moral licensing were found in participants who had been exposed to moral information about their ingroup, contrary to the finding of Study 10. The broader literature has also revealed inconclusive findings regarding a vicarious cleansing effect (Newman & Brucks, 2017, *cf* Meijers et al., 2018) therefore, to rectify this, the obvious next step for researchers should be to attempt rigorous replications of previous key findings.

# VIII. Conclusion.

*“Science, my lad, is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth.”*

~ Jules Verne, A Journey to the Centre of the Earth.

## 1. Revisiting Moral Transference.

Writing in 1890, James Frazer devoted an entire chapter in his cultural study of mythology (The Golden Bough) to ‘the transference of evil’. He detailed various rituals and practises from around the globe that all possess the same central notion[[73]](#footnote-73): the consequences of immorality can be displaced or transferred on to other people. With early twentieth century European arrogance, he wrote “the notion that we can transfer our guilt and sufferings to some other being who will bear them for us is familiar to the savage mind” (Frazer, 1890/2009, p. 557). The idea of moral transference is, I would argue, intuitive to all minds and is certainly as much a feature of Frazer’s Western Christianity as it is to the “savage” societies he documents.

In this thesis I have tried to explore some of the psychological and social factors that trigger and constrain moral transference. In this final chapter, I will first summarise the main findings of my research and how they relate to the existing literature. Second, I will provide a few suggestions for further studies. Third, I will discuss the relevance of the null findings, specifically within the context of the ‘replication crisis’. Fourth, I will address some of the limitations of this research, before concluding with a few thoughts about future directions for research on moral transference.

### **1.1. The “Direct Eye Gaze Effect”.**

In 1997 an American truck driver called Matthew Gibson battered a woman to death. He didn’t know her name and no one ever asked questions or came looking for her, so for 17 years he thought he had got away with it. In 2014, however, he received a series of voicemails and text messages from Walmart informing an Anita Townsend that her prescription was ready for collection. To Gibson, these seemingly innocuous messages meant two things. First, Anita Townsend must be the name of the woman he had murdered. And second, someone must know about his crime after all these years. Gibson’s quandary was exacerbated when he received an envelope with no return address containing only an advertisement from Walmart. This further convinced him that he was being watched. Having found religion a few years earlier, he was wracked with guilt (according to his attorney) and drove from North Carolina to Winslow, Arizona where he walked into the local police station, stating that he wanted to confess to a murder he had committed more than a decade previously (Washburn & Off, 2014).

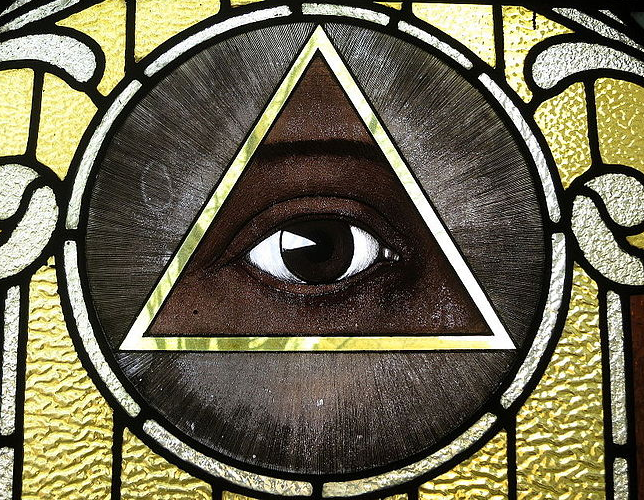
Matthew Gibson is obviously an extreme example - yet the results of Study 4 suggest that recalling a past transgression heightens sensitivity to social surveillance; a “direct eye gaze effect”. That is, participants who recalled a past transgression made stronger judgements of direct eye gaze in a set of ambiguous facial stimuli than those who recalled a neutral event. One fruitful line of enquiry could be to probe whether the “direct eye gaze effect” occurs in children? If so, at what point in development?

There is compelling evidence to suggest for example that young children are aware of their social reputations (Engelmann, Herrmann & Tomasello, 2012; Engelmann, Over, Herrmann & Tomasello, 2013; Fujii, Takagishi, Koizumi & Okada, 2015; Leimgruber, Shaw, Santos & Olson, 2012; Takagishi et al., 2015) and strategically adjust their behaviour when others are present (Engelmann, Over, Herrmann & Tomasello, 2013). Given that one of my proposed explanations for the “direct eye gaze effect” relates to the activation of reputation management concerns, the prediction one could make at this stage is that the “direct eye gaze effect” is yoked to children’s developing sense of their social reputation. For example, if given an incentive to do something naughty, would transgressing children check to see whether they had been seen by a caregiver, or indeed by an imaginary observer (see Piazza, Bering & Ingram, 2011)?

### **1.2. Religion and Surveillance.**

The other main findings of interest pertain to religion and surveillance. In Study 2 I found that religiously affiliated participants showed greater judgements of direct eye gaze in the presence of a surveillance cue than did non-affiliated participants. In Study 4, religious affiliation emerged as a significant predictor of judgements of direct eye gaze for ambiguous portraits[[74]](#footnote-74), i.e., religiously affiliated participants were more likely to perceive the portraits were looking directly at them than non-affiliated participants were.

First, why would religiously affiliated individuals have increased sensitivity to surveillance cues? One obvious explanation is that they (Christians in particular) are particularly attuned to such cues because they are a salient feature of their religious lives. Participation in religious activities can involve repeated exposure to the concept of a watchful God. For example, this concept is explicit in Christian scripture[[75]](#footnote-75) and is also reinforced through the use of religious imagery, such as the eye of providence (see fig. 5).

**

*Figure 5.* The Eye of Providence in a stained-glass window at St Raphael's Catholic church, Springfield, Ohio.

Second, why would religiously affiliated individuals have a tendency to think the portraits were looking at them? Through repeated exposure, surveillance cues might automatically trigger introspection and a focus on one’s misdeeds. This heightened focus on one’s moral foibles might lead religious adherents to believe that they are being looked at either because they are concerned their reputation is under threat or because they believe that God is looking at them via the surveillance stimuli. Answers to these questions are beyond scope of this thesis but they provide a compelling line of enquiry for researchers interested in exploring the association between religion and surveillance.

This association between religiosity and increased surveillance perception is to be expected under the Supernatural Punishment Hypothesis. This suggests that religious people feel more watched in general; they are erring on the side of caution. Johnson (2015) writes:

If false negative errors (assuming stealth and getting caught) are more costly than false positive errors (assuming detection and missing a reward), then only exaggerated estimates of the probability of detection – such as a belief that supernatural agents are observing your behavior all the time – will help you to avoid the worst of the two errors. (p. 169)

In other words, to avoid the more-costly errors associated with selfish behaviour, religious believers use the exaggerated estimate that they are under constant divine surveillance i.e., “The Lord sees everything you do. Wherever you go, he is watching” (Proverbs, 5:21) (though see McKay, Ross, O’Lone and Efferson, 2017).

Taken to the extreme, this is problematic. Yes, a religious believer who avoids all unethical behaviour due to his belief that God is constantly watching will never make the costly error of transgressing and being caught (McKay & Efferson, 2010). However, he will also never gain access to a host of benefits associated with unethical behaviour (see Schloss & Murray, 2010), which in some cases may be more than just ‘missed opportunities’. Consider a starving person who does not steal a loaf of bread, as he believes God is watching. In this scenario, the decision not to steal could lead to an error more-costly than being caught, i.e., he dies of starvation. So explaining the findings of Studies 2 and 4 as merely an upshot of religious believers feeling more watched in general, comes with associated theoretical problems (see McKay et al., 2017).

### **1.3. Forgiving Gods and Cooperation.**

The results of Study 8 revealed that both forgiving and punishing God beliefs, when salient, were significant predictors of endorsement of state-sanctioned punishment in response to a victim-directed transgression. One implication of this finding is that existing accounts of human prosociality and cooperation have underestimated the role of divine forgiveness in favour of a focus on divine punishment (Johnson & Cohen, 2016). At some point, virtually every individual is going to transgress, regardless of whether they believe in a watchful, punitive God or not. A punitive God who condemns everyone would ultimately undermine cooperation but a God who is more discriminate in his sanctions and forgives lesser transgressions provides a safety net. This way, non-serious transgressions are forgiven and through mechanisms of absolution (such as confession) ‘everyday sinners’ are let back into the fold with a renewed sense of commitment to the group.

Research has noted the prevalence of beliefs about the moral significance of thoughts in Christian (particularly Protestant) individuals, a type of cognitive bias termed moral thought-action fusion (e.g., Berman, Abramowitz, Pardue & Wheaton, 2010; Rassin & Koster, 2003). This is the belief that having a bad thought is as bad as performing the equivalent action. The concept of the moral equivalence between thoughts and actions is reflected in much Christian doctrine[[76]](#footnote-76). Religions vary with respect to the moral status accorded to thoughts. Protestant teaching views thoughts as morally equivalent to actions whereas Judaism does not and this is reflected in the degree to which Protestant and Jewish individuals display moral thought-action fusion (see Siev & Cohen, 2007). By widening the set of behaviours that constitute transgressing, such as including immoral thoughts, there is little chance an individual will avoid transgressing altogether. Perhaps this is the point. Given the unattainable moral standards imposed by certain religious prescriptions, we will always need mechanisms of forgiveness, which in turn strengthen our commitment to the group and its values thus ensuring cooperation.

### **1.4. Vicarious Moral Licensing.**

In Chapter 7, I conceptually replicated Kouchaki’s vicarious moral licensing effect (Study 10). The results revealed that US participants who were told that members of their country were more moral than others went on to show a preference for hiring a white participant in a job selection task. However, this result was not found in the follow up study (Study 11). In fact, Study 11 found the opposite: US participants told that their ingroup were more moral than others went on to show lower levels of vicarious moral licensing. These contradictory findings add to the burgeoning literature on vicarious moral compensation. During the final stages of the write up of this PhD two sets of studies were published on vicarious moral licensing (i.e., Meijers et al., 2018; Newman & Brucks, 2018). Until then the most prominent study on this effect was Kouchaki (2011). A promising course of action is two-fold. First, the vicarious moral licensing literature needs more direct, pre-registered replications. Neither of the two recent articles on vicarious moral licensing contained studies that were direct replications, nor were any pre-registered. This makes Studies 10 and 11 (of this thesis), at the time of writing, the only pre-registered studies on the vicarious moral licensing effect.

Second, as interest in vicarious moral licensing grows, so too does the need for a comprehensive meta-analysis, which includes pre-registered and direct replications. Although Meijer et al. (2018) are the first, to my knowledge, to conduct a meta-analysis on vicarious moral licensing effects, this was only an internal analysis of their four studies and these were not pre-registered. An obvious next step at this point would be to conduct a meta-analysis including: the five studies from Kouchaki (2011); the four from Newman and Brucks (2018); the four from Meijer et al. (2018) and the two unpublished studies (Studies 10 and 11) from this thesis. It is probable that there are other researchers working on vicarious moral licensing who may well have data they would be willing to share. Considering that Kouchaki (2011), Newman and Brucks (2018) and Meijer et al. (2018) were all published in a diverse range of publications[[77]](#footnote-77) there might well be researchers in other fields (consumer psychology, marketing research, organisational psychology etc.) working on vicarious moral licensing. Therefore, an interdisciplinary effort to pool collective resources and data could be a productive future course of action.

## 2. In Support of the Null Hypothesis.

*“The moral of this story is that the finding of statistical significance is perhaps the least important attribute of a good experiment; it is never a sufficient condition for concluding that a theory has been corroborated, that a useful empirical fact has been established with reasonable confidence—or that an experimental report ought to be published.”*

~ Lykken (1968, p.158)

The term ‘null result’ comes from the Latin nullus resultarum, which is literally translated as “no consequence”. In statistical hypothesis testing a null result is not significantly different from what is to be expected by the null hypothesis. Its probability does not exceed the threshold set for rejection of the null hypothesis. In psychology this is conventionally set at p-value 0.05 (but see Benjamin et al., 2018, for a recent alternative proposal, cf Lakens et al., in press). More loosely a study that yields a null result can mean one that is not publishable or, if the term is misconstrued, one that has ‘failed’. The majority of studies in this thesis had null, or non-significant results. However, this does not mean that the majority of these studies failed, or that the results were simply of no consequence. Null results are fundamental to science and the tendency to under-report them has had problematic consequences (Fiddler, Thorn, Barnett, Kambouris & Kruger, 2018; Laws, 2013).

As discussed at the outset, this thesis took place during a shift in the research zeitgeist. Previously the focus was on novel findings, which were quantified as anything with a p-value of < .05. This led to a pervasive publication bias and often the use of questionable research practises such as stopping data collection when finding significant results and hypothesising after the results are known, or HARKing (Ferguson & Heene, 2012; Fiedler & Schwarz, 2016; John, Loewenstein & Prelec, 2012; Motyl et al., 2017; Simmons, Nelson, Simonsohn, 2011). The reporting of null results was neglected in favour of the over-reporting of inflated or meretricious effect sizes (Kühberger, Fritz & Scherndl, 2014). Laws (2013) writes:

Negative findings and replications are science’s road signs telling us how to moderate our journey – we may like to ignore them or find them frustrating, but they are vital to progress - and contrary to existing trends, journals must allocate more space and importance to both null findings and replications. (p. 2).

As psychology moves towards encouraging large-scale replication and pre-registration it must also end its aversion to publishing null results. Some journals now dedicate sections to null results and replications indeed some are dedicated to publishing only null results (e.g., Journal of Articles in Support of the Null Hypothesis[[78]](#footnote-78)). The increased reporting of null results in combination with pre-registration and replication should be considered a mark of “gold standard” science.

## 3. Limitations.

There are always limitations to any PhD research; issues of time and funds are the most obvious. The first issue was certainly a factor in the laboratory studies I conducted (Studies 1 & 2). For example, in Study 1, an a priori power analysis recommended N = 102 (with input parameters of: alpha = 0.05 and power = 0.8, one-tailed, d = 0.5). Though it is possible that the estimate of the effect size (i.e., d = 0.5) was too optimistic. Therefore, despite the enormous effort required to test 120 participants, the sample size was potentially too small. Given that a recent meta-analysis examining the relationship between the Implicit Association Test and intergroup behaviour revealed that the studies included were chronically underpowered[[79]](#footnote-79), the small samples sizes in Study 1 could be perceived as problematic (see Kurdi et al., in press).

The value of laboratory studies is in no doubt, for me however the benefits of conducting studies online have profoundly shaped the way I think about research. The ease with which one can access very large populations, in some cases specific populations (e.g., US participants only), is in stark contrast to the laborious process of recruiting and testing students in a laboratory. Moreover, in terms of running studies that are sufficiently powered, this is much easier to do if they are conducted online.

One other limitation was that several studies lacked a manipulation check, thus precluding more concrete conclusions being drawn. For example, the lack of a manipulation check in Study 4 makes it impossible to make any precise statements about the role of guilt on surveillance perception. It is possible, for example, that an increase in generalised negative affect or shame led to increased judgements of direct eye-gaze rather than guilt specifically. Similarly, in Studies 6 and 7 the lack of a manipulation check means that it is not possible to say with any certainty whether the null effects were due to the independent variable having no effect on support for state-sanctioned punishment, or whether the manipulation (i.e., the forgiving vs. punitive God prime) failed. Moreover, the primes used in Studies 6 and 7 may well have made concepts of death and immortality salient, which could have had effects on participants in both conditions (see Chapter 6, section 6).

The inclusion of manipulations checks is favoured by high impact journals (see Hauser, Ellsworth & Gonzalez, 2017, Simons & Holcombe, 2014) and a popular view among social psychologists is that the inclusion of manipulation checks is necessary for well-designed psychological laboratory studies (see Fayant et al., 2017). They undoubtedly have benefits but the inclusion of a manipulation check for its own sake could be misguided (Hauser et al., 2018). The most obvious concern is that is that including a manipulation check may inadvertently inform the participant about the purpose of the study (Parrot & Hertel, 1999). For example, in the studies that involved a transgression recall, administering a manipulation check to assess guilt afterwards could well have exacerbated demand effects in participants. There are other potential concerns too (Hasuer et al. 2018). For example, giving participants a measure asking them to express how guilty or remorseful they feel after thinking about a past misdeed may well serve to alleviate the emotion before they get to the dependent variable (Hauser et al., 2018). Alternatively, getting participants to complete a guilt or remorse scale after the recall could ‘crystalise’ vague feelings of guilt this leading them to experience more guilt (see Hauser et al., 2018). Hauser et al (2018) write:

There are many possible ways in which a manipulation check might affect the participant. It could undo the effects of the manipulation; it could enhance the effects of the manipulation; or it could interact with the manipulation. There is no way to know which of these processes may be occurring in any particular experiment without empirically investigating it, but in any case, the assumption that the manipulation check is a neutral unobtrusive measure of the effects of the manipulation may be unjustified. (p.

This is not to suggest that the inclusion of manipulation checks is a bad thing or that the exclusion of manipulation checks is a good thing. Rather I wish to suggest that the absence of a manipulation check in a study included as a matter of course is not necessarily a limitation. The relative pros and cons of the inclusion of a manipulation check should be considered at the design stage, and the decision whether or not to include one should be based on what would better serve the overall purposes of the study. One solution that Hauser et al (2018) propose is to exclude a manipulation check until its efficacy has been demonstrated, in pilot studies for example. In an ideal world this seems a productive way of resolving the issue of including manipulation checks. However, the realities of PhD research meant that due to time constraints the piloting of manipulation checks was not a viable option for this research.

Another potential limitation ­ was that some of the studies were perhaps too “adventurous” and not close enough to existing studies[[80]](#footnote-80) thus making effect size calculations difficult and raising concerns about the validity of the measures (e.g., the portrait task). With hindsight I think the lure of novelty that was prevalent when I started this PhD, proved too strong. It is certainly easy to be seduced when major research funders, such as the European Research Council, prioritise “high-risk/high-gain” research[[81]](#footnote-81) (see Leijten, Roseboom & Hofer, 2010). The literature should be the starting point for any study design, both in terms of gauging the required effect size but also the efficacy of the materials. For example, the vignettes used in Studies 6 and 7 to prime the forgiving and punishing nature of God may have been too weak to have the intended effect on support for state-sanctioned punishment. Yet there is certainly no dearth of relevant primes in the literature, which have already had their usefulness demonstrated (e.g., Yilmaz & Bahçekapili, 2016). That said, the presentation or adaptation of extracts of religious text is a priming technique used previously in the literature (e.g. Carpenter & Marshall, 2009; Yilmaz & Bahçekapili, 2016, Study 2; Rand et al., 2014, Study 1).

Creativity belongs in science and but there is a cost to novelty (see Condon, Graham & Mroczek, 2017). It has been a contributing factor to the replication crisis, from which the field is reeling and attempting to repair itself. If there are robust and well-established paradigms and materials in the literature, they should act as a guide for future research. Especially within the context of the current research climate, that places a premium, quite rightly so, on the importance of replication (Zwaan et al., 2017). Yet there are concerns that an open science framework and the use of practices, such a pre-registration, would constrain scientific creativity (e.g., Goldin-Meadow, 2016; Gonzales & Cunningham, 2015; Scott, 2013). For example, the President of the American Psychological Association voiced her worry that pre-registration would stifle scientific discovery; “Science isn’t just about testing hypotheses—it’s also about discovering hypotheses…Aren’t we supposed to let the data guide us in our exploration? How can we make new discoveries if our studies need to be catalogued before they are run?” (Goldin-Meadow, 2016, para 4). As Frankenhuis & Nettle (2018) note, the status quo allows high researcher creativity but low reliability, and the understandable concern is that an open science framework entails a trade-off between creativity and rigour. For them, pre-registration can be liberating for researchers. For example, the process of specifying how you will collect, process and analyse data before having seen it removes some of the unease researchers face when making decisions about their data after having seen it. It also removes any suggestion of p- hacking. Second, the increase in journals that publish pre-registered studies relieves the researcher of the burden of having to produce “positive” results. Third, Frakenhuis and Nettle (2018) suggest that an open science framework encourages a more pluralistic approach to statistics (e.g., an increase in the use of Bayesian analyses), which allows the field to break free from the chains of the significant p-value. They write:

The new practices can foster creativity because they cultivate an open and flexible mind-set, create a more collaborative and constructive climate, and generate more accurate information and make it more accessible. We do not think that reliability versus creativity is a zero-sum game: It is possible for knowledge to become more reliable with researchers continuing to enjoy great creativity. This is because Open Science liberates researchers in some ways even as it constrains them. (p. 444)

## 4. Concluding Remarks.

In Chapter 2, I outlined three central themes that underpin the literature which provided the theoretical framework for this thesis. First, was the notion that the consequences of (im)moral actions are not subject to temporal constraints. That is, the emotional consequences of wrongdoing, or the need to compensate for a transgression, are not tied to the time of the wrongdoing itself. In many cases these can be delayed across a historical timescale or transferred to a cosmic one. As discussed in chapter 1 and 2, this is a recurrent feature of much research investigating aspects of our moral behaviour. For example, the moral compensation model allows us to use past moral actions to license unethical behaviour in the present. Similarly, the consequences of past historical wrongdoings can be displaced into the present and influence both the emotional states and behaviours of group members. Peoples’ beliefs in a powerful God, who can punish or reward, also influence perceptions of moral transference. The consequences of a sin or good deed continue after we die and can be addressed by God (i.e., rewarded or punished) in the afterlife. Some of the findings of this thesis suggest the same. For example, in Study 10, I found that past ingroup morality can be transferred to the present and influence individual moral judgements (i.e., vicarious moral licensing). Similarly, in Study 4 the recollection of a past misdeed influenced socio-cognitive processes in the present, that is judgements of surveillance perception.

Arguably, one could have concluded that the consequences of (im)moral behaviour are not subject to temporal constraints from the existing literature. However, the research in this thesis strengthens that conclusion and specifically places it within a moral transference framework. Given the current drive towards replicability in Psychology, the fact that this PhD supports existing findings and themes, rather than documenting a plethora of novel findings, is perhaps where its value lies.

Several key findings of this thesis also strongly point to the role of religious concerns and beliefs as factors that influences moral transference. For example, in Study 4, I found that religiously affiliated participants who recalled a past transgression made stronger judgements of direct eye gaze in a series of ambiguous portraits compared to those who recalled a neutral event. This effect occurred only for religiously affiliated participants, not for non-affiliated ones. From this we can conclude that religious beliefs and concepts have an (as yet undetermined) effect on moral transference.

In Study 8, forgiving God beliefs (when salient) were significant predictors of support for state-sanctioned punishment in response to a victim-directed transgression (as were punishing God beliefs). This allows us to draw the conclusion that forgiving God concepts and beliefs (when salient) influence our perceptions of the moral responsibility of punishment. Moreover, we can conclude that the role of forgiving God concepts in a variety of moral domains (i.e., endorsement of punishment) has been systematically overlooked. The results of Study 8 provide an interesting line of future enquiry. I would like this line of research to gain momentum so that we have a more comprehensive understanding of the influence of forgiving God concepts on a variety of aspects of human society, and in particular how such specific God concepts promote everyday moral behaviours.

The ultimate goal of course would be to formulate an overall theory of moral transference. That may be a long way off, but psychological research into morality is showing no signs of decline. Morality fascinates us all, from psychologist to lay person. As Kant (1788/2015) said, “two things awe me most: the starry sky above me and the moral law within me”.

# References.

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1-16.

Abraham, A. (2017, August 11). The strange case of the innocents who confessed to murders they didn’t commit. *The Vice.* Retrieved from https://www.vice.com/en\_uk/article/xww8nk/the-strange-case-of-the-innocents-who-confessed-to-murders-they-didnt-commit

Abramowitz, J. S., D. Huppert, J., Cohen, A. B., Tolin, D. F., & Cahill, S. P. (2002). Religious obsessions and compulsions in a non-clinical sample: The Penn Inventory of Scrupulosity (PIOS). *Behaviour Research and Therapy*, *40*(7), 825-838.

Ackerman, J. M., Goldstein, N. J., Shapiro, J. R., & Bargh, J. A. (2009). You wear me out. *Psychological Science*, *20*(3), 326-332.

Agrawal, N., & Duhachek, A. (2010). Emotional compatibility and the effectiveness of anti-drinking messages: A defensive processing perspective on shame and guilt. *Journal of Marketing Research, 47*(2), 263–273.

Ahmed, A., & Hammarstedt, M. (2011). The effect of subtle religious representations on cooperation. *International Journal of Social Economics, 38,* 900–910.

Ahmed, A., & Salas, O. (2011). Implicit influences of Christian religious representations on dictator and prisoner’s dilemma game decisions. *The Journal of Socio-Economics, 40*, 242–246.

Alden, L. E., Teschuk, M., & Tee, K. (1992). Public self-awareness and withdrawal from social interactions. *Cognitive Therapy and Research, 16*, 249–267.

Alicke, M. D., Vredenburg, D. S., Hiatt, M., & Govorun, O. (2001). The ‘‘better than myself effect’’*. Motivation and Emotion, 25,* 7–22.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews*, *32*(1), 175-191.

Anderson, R. G., & Kichkha, A. (2017). Replication, meta-analysis, and research synthesis in economics. *American Economic Association, 107*(5), 56-59.

Applegate, B. K., Cullen, F. T., Fisher, B. S., & Vander Ven, T. (2000). Forgiveness and fundamentalism: reconsidering the relationship between correctional attitudes and religion. *Criminology 38,* 719–754.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83,* 1423–1440.

Aquino, K., Freeman, D., Reed, A., Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of behaviour: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology, 97*(1), 123-141.

Ariely, D. (2012). *The (honest) truth about dishonesty*. New York, NY: HarperCollins.

Aronson, J., Cohen, G. L., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 127–148). Washington, DC: American Psychological Association.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. European Journal of Personality, 27(2), 108-119.

Atkinson, Q. D., & Bourrat, P. (2011). Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evolution and Human Behavior, 32*(1), 41-49.

Atran, S. (2004). *In Gods we trust: The evolutionary landscape of religion*. Oxford, UK: Oxford University Press.

Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences, 27,* 713–770.

Aveyard, M. E. (2014). A call to honesty: Extending religious priming of moral behavior to Middle Eastern Muslims. *PLoS ONE, 9*(7), e99447.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX* *Lexical Database. Release 2* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, *Philadelphia*.

Bader, C. D., Desmond, S. A., Carson Mencken, F., & Johnson, B. R. (2010). Divine iustice: The relationship between images of God and attitudes toward criminal punishment. *Criminal Justice Review*, *35*(1), 90-106.

Baier, C. J., & Wright, B. R. E. (2001). “If you love me keep my commandments”: A meta-analysis on the effect of religion on crime. *Journal of Research on Crime and Delinquency, 38*(1), 3-21.

Baillon, A., Selim, A., & Van Dolder, D. (2013). On the social nature of eyes: The effect of social cues in interaction and individual choice tasks. *Evolution and Human Behavior, 34,* 146–154.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452-454.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3,*193–209.

Bandura, A. (2016). *Moral disengagement: How people do harm and live with themselves*. New York, NY: Worth Publishers, Macmillan Learning.

Barak-Corren, N., & Bazerman, M. H. (2017). Is saving lives your task or God’s? Religiosity, belief in God, and moral judgement. *Judgment and Decision Making, 12*(3), 280–296.

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior, 27*(5), 325–344.

Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *Journal of Experimental Psychology (General), 141*(4), 757-773.

Barnes, K., & Gibson, N. J. (2013). Supernatural agency: Individual difference predictors and situational correlates. *International Journal for the Psychology of Religion, 23*(1), 42-62.

Barrett, J. L. (1999). Theological correctness: Cognitive constraint and the study of religion. *Method & Theory in the Study of Religion*, *11*(4), 325–39.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Science 4*(1), 29-34.

Barrett, J. L. (2004). *Why would anyone believe in God?* Walnut Creek, CA: AltaMira Press.

Barrett, J. L. (2012). *Born believers: The science of children’s religious belief*. New York, NY: Atria Books.

Barrett, J. L., & Burdett, E. R. (2011). The cognitive science of religion. *Psychologist, 24*(4), 252-255.

Barrett, J. L., & Johnson, A. H. (2003). The role of control in attributing intentional agency to inanimate objects. *Journal of Cognition and Culture*, *3*(3), 208–217.

Barret, J. L., & Keil, F. C. (1996). Conceptualising a non-natural entity: Anthropomorphism in God concepts. *Cognitive Psychology, 31*(3), 219-247.

Barrett, J. L., & Lanman, J. A. (2008). The science of religious beliefs. *Religion*, *38*(2), 109–124.

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology, 42*(2), 228-235.

Bastian, B., Jetten, J., & Fasoli, F. (2011). Cleansing the soul by hurting the flesh: The guilt-reducing effect of pain. *Psychological Science, 22*(3), 334-335.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2,* 412–414.

Bateson, M., Robinson, R., Abayomi-Cole, T., Greenlees, J., O’Connor, A., & Nettle, D. (2015). Watching eyes on potential litter can reduce littering: evidence from two field experiments. *PeerJ*, *3*, e1443.

Batson, C. D., Schoenrade, P., & Ventis, W. L. (1993). *Religion and the individual: A social-psychological perspective.* New York, NY: Oxford University Press.

Baumard, N. (2011). Punishment is not a group adaptation: Humans punish to restore fairness rather than to support group cooperation. *Mind and Society, 10*(1), 1-26.

Baumard, N., André, J. B., & Sperber, D. (2013). A Mutualistic Approach to Morality. *Behavioral and Brain Sciences*, *36*(1), 59-122.

Baumard, N., & Boyer, P. (2013). Explaining moral religions. *Trends in Cognitive Science, 17*(6), 272-280.

Baumard, N., & Sheskin, M. (2015).Partner choice and the evolution of a contractualist morality. In J. Decety & T. Wheatley (Eds.), *The Moral Brain* (pp.35-48). Boston, MA: M.I.T. Press.

Baumeister, R. F., Bauer, I. M., & Lloyd, S. A. (2010). Choice, free will, and religion. *Psychology of Religion and Spirituality, 2*(2), 67-82.

Baumeister, R., Bratslavsky, E., Finkenauer, C., & Vohs, K. (2001). Bad is stronger than good. *Review Of General Psychology*, *5*(4), 323-370.

Baumeister, R. F., & Cairns, K. J. (1992). Repression and self-presentation: When audiences interfere with self-deceptive strategies. *Journal of Personality and Social Psychology, 62,* 851-862.

Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin, 115*, 243–267.

Bazerman, M. H., & Tenbrunsel, A. E. (2011). *Blind spots: Why we fail to do what's right and what to do about It.*Princeton, NJ: Princeton University Press.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*(3), 183-200.

Benjamin, D. J., Choi, J. J., & Fisher, G. W. (2010)*. Religious identity and economic behavior* (NBER Working Paper No. 15925). Cambridge, MA: National Bureau of Economic Research.

Benjamin, D. J., et al. (2018). Redefine statistical significance. *Nature Human Behavior, 2,* 6-10.

Benoît, W. L., & Drew, S. (1997). Appropriateness and effectiveness of image repair strategies. *Communications Report, 10,* 153-163.

Bentall, R. P. (2003). *Madness explained: Psychosis and human nature*. London, England: Penguin.

Bering, J. M. (2011). *The Belief Instinct: The psychology of Souls, Destiny and the Meaning of life.* New York, NY: Norton.

Bering, J. M. (2006). The folk psychology of souls. *Behavioural and Brain Sciences, 29*(5), 453-62.

Bering, J. M., & Johnson, D. D. P. (2005). ‘‘O lord ... you perceive my thoughts from afar’’: Recursiveness and the evolution of supernatural agency*. Journal of Cognition and Culture Special Issue: On Psychological and Cognitive Foundations of Religiosity, 5,* 118–142.

Bering, J. M., McLeod, K., & Shackelford. T. K. (2005). Reasoning about dead agents reveals possible adaptive trends. *Human Nature, 16*(4), 360-81.

Bering, J. M., & Shackelford, T. (2005). Evolutionary psychology and false confession. *American Psychologist, 9*, 1037-1038.

Berinsky, A., Druckman, J. N., & Yamamoto, T. (in press). Why replications do not fix the file drawer problem: A model and evidence from a large-scale vignette experiment. *PNAS,* retrieved from https://asiapolmeth.princeton.edu/sites/default/files/polmeth/files/teppeiyamamoto.pdf

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com’s Mechanical Turk. *Political Analysis, 20,* 351–368.

Berman, N. C., Abramowitz, J. S., Pardue, C. M., & Wheaton M. G. (2010) The Relationship between religion and thought-action fusion: Use of an in vivo paradigm. *Behaviour Research and Therapy, 48*(7), 670-674.

Bernhard, F. (2016). Employees’ vicarious guilt and the support of stricter regulations. *Academy of Management Proceedings*, *2016*(1), 15621.

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience and Biobehavioral Reviews, 55*, 536-546.

Berrios, G. E., Bulbena, A., Bakshi, N., Dening, T. R., Jenaway, A., Markar, H., … Mitchell, S. L. (1992). Feelings of guilt in major depression. *British Journal of Psychiatry*, *160*(06), 781-787.

Bertolotti T., & Magnani, L. (2010). The role of agency detection in the invention of supernatural beings. In L. Magnani., W. Carnielli & C. Pizzi (Eds.), *Model-based reasoning in science and technology: Studies in computational intelligence, vol 314* (pp. 239-262). Berlin, Heidelberg: Springer-Verlag.

Billingsley, J., Gomes, C. M., & McCullough, M. E. (2018). Implicit and explicit influences of religious cognition. Preprint retrieved from https://psyarxiv.com/6amfe

Blackman, S. F., Keller, K. T., & Cooper, J. (2016). Egocentrism and vicarious dissonance. *Journal of Experimental Social Psychology, 62,* 1-6.

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin, 41*(4), 540–558.

Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. (2014). Three attempts to replicate the moral licensing effect. S*ocial Psychology, 45*(3), 232–238.

Blasi, A. (1983). Moral cognition and moral actions: A theoretical perspective. *Developmental Review, 3*(2), 178-210.

Blogowska, J., Saroglou, V., & Lambert, C. (2013). Religious prosociality and aggression: It's real. *Journal for the Scientific Study of Religion, 52*(3), 524-536.

Bloom, P., & Veres, C. (1999). The perceived intentionality of groups. *Cognition, 71*(1), B1-B9.

Bortolini, T., Newson, M., Natividade, T. C., Vázquez, A., & Gómez, A. (2018). Identity fusion predicts endorsement of pro-group behaviours targeting nationality, religion, or football in Brazilian samples. *British Journal of Social Psychology, 57*, 346-366.

Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *Proceedings of the National Academy of Sciences, 111,* 16784-16789.

Bourrat, P., Baumard, N., & Mckay, R. (2011). Surveillance cues enhance moral condemnation. *Evolutionary Psychology, 9,* 193–199.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, *100*(6), 3531-3535.

Boyer, P. (2003). Religious though and behavior as by-products of brain function. *Trends in Cognitive Science, 7*(3), 119-124.

Boyer, P. (2007). *Religion explained: The evolutionary origins of religious thought*. New York, NY: Basic Books.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41,* 135–143.

Brambilla, M., Sacchi, S., Rusconi, P., & Cherubini, P., & Yzerbyt, V.Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology, 51,*149-166*.*

Brañas-Garza, P., [Bucheli](https://philpapers.org/s/Marisa%20Bucheli), [M., Espinosa](https://philpapers.org/s/Mar%C3%ADa%20Paz%20Espinosa), M.. P,  & [García-Muñoz](https://philpapers.org/s/Teresa%20Garc%C3%ADa-Mu%C3%B1oz), T. (2013). Moral cleansing and moral licenses: Experimental evidence. Economics and Philosophy, 29(2), 199-212.

Brandt, M. J., Ijserman, H., Dijksterhuis, A., Farach, F., Geller, J., Giner-Sorola, R., …van’t Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50,* 217-224.

Branscombe, N. R., & Miron, A. M. (2004). Interpreting the Ingroup's Negative Actions Toward Another Group: Emotional Reactions to Appraised Harm. In L. Z. Tiedens & C. W. Leach (Eds.), *Studies in emotion and social interaction. The social life of emotions* (pp. 314-335). New York, NY: Cambridge University Press.

Branscombe, N. R., Spears, R., Ellemers, N., & Doosje, B. (2002). Intragroup and Intergroup Evaluation Effects on Group Behavior. *Personality and Social Psychology Bulletin, 28*(6), 744-753.

Brown, R. (2000). Social identity theory: past achievements, current problems and future challenges. *European Journal of Social Psychology, 30*(6), 745-778.

Brown, R., González, R., Zagefka, H., Manzi, J., & Čehajić, S. (2008). Nuestra Culpa: Collective guilt and shame as predictors of reparation for historical wrongdoing. *Journal of Personality and Social Psychology, 94,* 75-90.

Buhrmester, M. D., 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.

Buhrmester, M.D., Talaifar, S., & Gosling, S.D. (2018). An valuation of Amazon’s Mechanical Turk, Its rapid rise, and its effective use. *Perspectives on Psychological Science*, *13*(2) 149–154.

Burkett, S., & White, M. (1974). Hellfire and Deliquency among LDS Adolescents. *Journal for the Scientific Study of Religion, 13,* 455-62.

Burnett, R., Hoyle, C., & Speechley, N. E. (2017). The Context and Impact of Being Wrongly Accused of Abuse in Occupations of Trust. *The Howard Journal of Crime and Justice, 56*(2), 176-197.

Burnham, T. C., & Hare, B. (2007). Engineering human cooperation. *Human Nature, 18*(2), 88-108.

Burra, N., Hervais-Adelman, A., Kerzel, D., Tamietto, M., De Gelder, B., & Pegna, A. J. (2013). Amygdala activation for eye contact despite complete cortical blindness. *Journal of Neuroscience*, *33*(25), 10483-10489.

Bushman, B. J., Ridge, R. D., Das, E., Key, C. W., & Busath, G. L. (2007). When God sanctions killing: Effect of scriptural violence on aggression. *Psychological Science, 18*, 204-207.

Buss, D. M. (1999). *Evolutionary psychology.* London, UK: Allyn and Bacon.

Caluori, N. E., Jackson, J. C., Gray, K., & Gelfand, M. J. (2018). Warfare changes people’s view of God. *Manuscript in preparation*.

Carbon, C., & Hesslinger, V. M. (2011). Bateson et al.’s (2006) Cues-of-being-watched Paradigm Revisited. *Swiss Journal of Psychology*, *70*(4), 203-210.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just desserts as motives for punishment*. Journal of Personality and Social Psychology, 83*,284–299.

Carnes, N. C., & Winer, J. P. (2017). Do unto others: How cognitive fusion shapes the transmission of moral behavior. *Journal of Experimental Psychology: General, 146*(4), 472-484.

Carpenter, T. P., & Marshall, M. A. (2009). An examination of religious priming and intrinsic religious motivation in the moral hypocrisy paradigm. *Journal for the Scientific Study of Religion, 48,* 386-393.

Carter, E. C., & McCullough, M. E. (in press). Combining meta-analysis and replication. *Advances in Methods and Practices in Psychological Science,* *(in press).*

Carter, E. C., Kofler, L. M., Forster, D. E., McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General, 144,*796–815.

Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating doing good later allow you to be bad now? *Journal of Experimental Social Psychology*, *5,* 110-116.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*, 2156–2160.

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infrahumanization in response to collective responsibility for intergroup killing. Journal of Personality and Social Psychology, 90(5), 804-818.

Čehajić-Clancy, S., Effron, D. A., Halperin, E., Liberman, V., & Ross, L. D. (2011). Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures*. Journal of Personality and Social Psychology, 101,* 256-270.

Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578). New York, NY: Guilford Press.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice.*Princeton, NJ: Princeton University Press.

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex, 49,* 609-610.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46,* 112–130.

Chandler, J., & Paolacci, G. (2017). Lie for a dime: When most pre-screening responses are honest but most study participants are imposters. *Social Psychological & Personality Science, 8,* 500–508.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12,* 53-81.

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. (2015). Non-naïve participants can reduce effect sizes. *Psychological Science, 26*(7), 1131-1139.

Chen, J., Wei, J., Shang, P., Wang, X., & Zhang, J. (2016). Do we feel the same level of guilt about the wrongdoings of close ones and distant ones? The mediating role of behavioural control. *International Journal of Psychology*, *53*(4), 261-268.

Chivers, T. (2017, July 29). These people are trying to fix a huge problem in science. *Buzzfeed.* Retrieved from https://www.buzzfeed.com/tomchivers/signifying-nothing?utm\_term=.ifzqvnlXa#.pc61keY2n

Chou, H. T., Uata, D. (2012). The impact of parental discipline on the image of God. *Mental Health, Religion & Culture, 15*(7), 677–88.

Cialdini, R. B. (1993). *Influence: Science and practice.* New York, NY: Harper Collins College.

Clifford, S., & Jerit, J. (2016). Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions. *Public Opinion Quarterly, 80,* 858–887.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science, 1,* 120–131.

Cohen, G., Eysenck, M. W., & Levoi, M. E. (1986). *Memory: a cognitive approach*. Milton Keynes, UK: Open University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Cohen, J. B., Pham, M. T., & Andrade, E. B. (2008). The nature and role of affect in consumer behavior in C. Haugtvedt, P. Herr, & F. Kardes (Eds.), *Handbook of consumer psychology* (pp. 297-348). Mahwah, NJ: Lawrence Erlbaum.

Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia, 61,* 65-79.

Condon, D. M., Graham, E. K., & Mroczek, D. (2017). On replication research. Retrieved from https://doi.org/10.17605/OSF.10/2FN5X

Conson, M., Salzano, S., Frolli, A., & Mazzarella, E. (2017). The peer's point of view: Observing a peer performing an action enhances third-person perspective in adolescents. *Journal of Adolescence, 56,* 84-90.

Conty, L., Tijus, C., Hugueville, L., Coelho, E., & George, N. (2006). Searching for asymmetries in the detection of gaze contact versus averted gaze under different head views: A behavioural study. *Spatial Vision, 19,* 529–545.

Converse, P. D., & DeShon, R. P. (2009). A tale of two tasks: Reversing the self-regulatory resource depletion effect. *Journal of Applied Psychology, 94,*1318–1324.

Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual bbstraction moderates whether past moral deeds motivate consistency or compensatory ehavior. *Personality and Social Psychology Bulletin, 38*(7), 907-919.

Cook, C. L., Cottrell, C. A., & Webster, D. (2015). No good without God: Anti-atheist prejudice as a function of threats to morals and values. *Psychology of Religion and Spirituality, 7*(3), 217-226.

Cooper, J., & Hogg, M. A. (2004). Feeling the anguish of others: A theory of vicarious dissonance. *Advances in Experimental Social Psychology, 39,* 359-403.

Coover, J. (1913). "The feeling of being stared at." *The American Journal of Psychology,* *24*(4), 570-575.

Corwin, M. (1996, March 25). False confessions and tips still flow in Simpson case. *LA Times*. Retrieved from http://articles.latimes.com/1996-03-25/news/mn-51081\_1\_false-confession

Cosmides, L., & Tooby, J. (1987). From evolution to behavior: evolutionary psychology as the missing link. In J. Dupre (Ed.) *The latest on the best: essays on evolution and optimality* (pp. 277-306). Cambridge, MA: MIT Press.

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66,* 93-99.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*. *43*(3), 245–265.

Critcher, C. R., & Lee, C. J. (2018). Feeling is believing: Inspiration encourages belief in God. *Psychological Science*, *29*(5), 723-737.

Crump, J. C., McDonnel, J. V., Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioural research. *PLoS ONE, 8*, e1058.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of ‘pure reason’ in infancy. *Cognition, 72*(3), 237-267.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7-29.

Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American Psychological Association publication manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology, 64,* 138-146.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science,15*(12), 806–813.

Darby, B. W., & Schlenker, B. R. (1989). Children’s reactions to transgression: Effects of the actor’s apology, reputation and remorse. *British Journal of Social Psychology, 28,* 353–364.

Darby, B. W., & Schlenker, B. R. (1982). Children’s reactions to apologies. *Journal of* *Personality and Social Psychology, 43,* 742–753.

Darke, P. J., & Freedman, J. L. (1997). The belief in good luck scale. *Journal of Research in Personality, 31*(4), 486-511.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavioral Research Methods, 37*(1), 65-70.

Davies, M. F. (2005). Mirror and camera self-focusing effects on complexity of private and public aspects of identity. *Perceptual and Motor Skills, 100*(3), 895–898.

Day, J. M. (2017). Religion and human development in adulthood: Well-being, prosocial behavior, and religious and spiritual development. *Behavioral Development Bulletin, 22*(2), 298-313.

De Bono, K., & Kuschpel, A. (2014). Gender differences in religiosity: the role of self-monitoring. *North American Journal of Psychology, 16*(2), 415-425.

De Freitas, J., Cikara, M., Grossman, I., & Schlegel, R. (2017) Origins of the Belief in Good True Selves. *Trends in Cognitive Science, 21*(9), 634-636.

De Freitas, J., Tobia, K., Newman, G. E., & Knobe, J.(2016). Normative judgments and individual essence. *Cognitive Science, 41,* 1551-6709.

de Quervain, D., Fischbacher, U., Treye, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305,* 1254–1259.

Decety, J., & Grezes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain Research, 1079*(1), 4-14.

DeRosse, P., & Karlsgodt, K. H. (2015). Examining the psychosis continuum. *Current Behavioral Neuroscience Reports, 2,* 80–89.

Dickie, J. R., Ajega, L. V., Kobylak, J. R., & Nixon, K. M. (2006). Mother, father, and self: Sources of young adults’ God concepts. *Journal for the Scientific Study of Religion, 45*(1), 57–71.

Difallah, D,. Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of Mechanical Turk workers. In *Proceedings of WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey*, CA, USA, February 5–9, 2018.

Doosje, B., & Branscombe, N. R. (2003). Attributions for the negative historical actions of a group. *European Journal of Social Psychology, 33*(2), 235-248.

Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history*.* *Journal of Personality and Social Psychology, 75*(4), 872-886.

Doosje, B., Branscombe, N. R., Spears, R., & Manstead, S. R. (2006). Antecedents and consequences of group-based guilt: The effects of ingroup identification. *Group Processes and Intergroup Relations, 9*(3), 325-338.

Dostoevsky, F. (1880). *The Brothers Karamazov.* Moscow The Russian Messenger.

Douglas, L. K., & Handrich, R. (2012). “Only the guilty would confess to crimes” Understanding the mystery of false confessions. *The Jury expert. The Art and Science of Litigation Advocacy,* 24(6),1-18.

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It’s all in the mind, but whose mind? *PLoS ONE, 7*(1), e29081.

Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don’t punish. *Nature, 452*(7185), 348–351.

Drummond, C. (2009). Replicability is not reproducibility: Nor is it good science. *Proceedings of the Evaluation Methods for Machine Learning Workshop 26th ICML,* Montreal, Quebec, Canada.

Duhaime, E. P. (2015). Is the call to prayer a call to cooperate? A field experiment on the impact of religious salience on prosocial behavior. *Judgment and Decision Making, 10* (6), 593-596.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6,* 241-252.

Dunning, D. (2007). Self-image motives and consumer behavior: How sacrosanct self-beliefs sway preferences in the marketplace. *Journal of Consumer Psychology, 17,* 237–249.

Dutton, D. G., & Lake, R. A. (1973). Threat of own prejudice and reverse discrimination in interracial situations. *Journal of Personality and Social Psychology, 28*(1), 94-100.

Earp, B. D., Everett, J. A. C., Madva. E. N., & Hamlin, J. K .(2014). Out, damned spot: Can the “Macbeth effect” be replicated? *Basic and Applied Social Psychology, 36*(1), 91-98.

Eberhardt, J. L. (2005). Imaging race. *American Psychologist,* *60*(2), 181–90.

Eddy, M. (2012, November 17). *For 60th Year Germany Honors Duty to Pay Holocaust Victims.* Retrieved from http://www.nytimes.com/2012/11/18/world/europe/for-60th-year-germany-honors-duty-to-pay-holocaust-victims.html

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favouring whites. *Journal of Experimental Social Psychology, 45*(3), 590-593.

Effron, D. A., & Conway, P. (2015). When virtue leads to villany: Advances in research on moral self-licensing. *Current Opinion in Psychology, 6*, 32-35.

Effron, D. A., Monin, B., & Miller, D. T. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology, 103,* 916-932.

Effron, Daniel A. (2014). Beyond 'Being good frees us to be bad:' Moral self-licensing and the fabrication of moral credentials. In P. A. M. Van Lange & J. W. Van Prooijen (Eds.), *Cheating, corruption, and concealment: Roots of unethical behavior* (pp. 33-54). Cambridge, UK: Cambridge University Press.

Eibach, R. P., Libby, L. K., & Ehrlinger, J. (2009). Priming family values: How being a parent affects moral evaluations of harmless but offensive acts. *Journal of Experimental Social Psychology, 45*, 1160–1163.

Eil, D., & Rao, J. M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics, 3*(2), 114–38.

Ellemers, N., Pagliaro, S., Barreto, M., & Leach, C. W. (2008). Is it better to be moral than smart? The effects of morality and competence norms on the decision to work at group status improvement. *Journal of Personality and Social Psychology*, *95*(6), 1397-1410.

Ellett, L., & Chadwick, P. (2007). Paranoid cognitions, failure and focus of attention in college student Cognition and Emotion, *21,* 558-576.

Engelmann, D., & Fischbacher, U. (2009). Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior 67*, 399–407.

Engelmann, J. M., & Rapp, D.J. (2018). The influence of reputational concerns on children’s prosociality. *Current Opinion in Psychology, 20,* 92-95.

Engelmann, J. M., & Zeller, C. (2017). Doing the right thing for the wrong reason: Reputation and moral behavior. In J. Kiverstein (Ed.), *The Routledge Handbook of Philosophy of the Social Mind* (pp. 247-261). London, UK: Routledge.

**Engelmann, J. M.,** Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care more about their reputations with ingroup members and potential reciprocators. Developmental Science, *16*(6), 952-95.

Engelmann, J., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS ONE 7*(10), e48433.

Epley, N., Converse, B. A., Delbosc, A., Monteleone, G. A., & Cacioppo, J. T. (2009). Believers estimates of Gods beliefs are more egocentric than estimates of other people’s beliefs. *Proceedings of the National Academy of Sciences, 106*(51), 21533–8.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior, 32*, 172–178.

Eskine, K. J. (2013). Wholesome foods and wholesome morals?: Organic foods reduce prosocial behavior and harshen moral judgments. *Social Psychological and Personality Science, 4,* 251–254.

Eskine, K. J., Novreske, A., & Richards, M. (2013). Moral contagion effects in everyday interpersonal encounters. *Journal of experimental social psychology, 49,* 947-950.

European Research Area Board Annual Report. (2009). Preparing Europe for a new renaissance. A strategic view of the European research area. Retrieved from http://ec.europa.eu/research/erab/pdf/erab-first-annual-report-06102009\_en.pdf

Evans, T. D., & Adams, M. (2003). Salvation or damnation? Religion and correctional ideology. *American Journal of Criminal Justice,* *28*, 15-35.

Evans, J. S. B. & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Thinking and* *Reasoning 11,* 382–389.

Ewbank, M. P., Jennings, C., & Calder, A. J. (2009). Why are you angry with me? Facial expressions of threat influence perception of gaze direction. *Journal of Vision,* *9*(12), 16-16.

Fantz, R. L. (1963). Pattern vision in newborn infants. *Science,* *140*(3564), 296-297.

Farias, M., Van Mulukom, V., Kahane, G., Kreplin, U., Joyce, A., Soares, P., … Möttönen, R. (2017). Supernatural belief is not modulated by intuitive thinking style or cognitive inhibition. *Scientific Reports*, *7*(1), 15100.

Farroni, T., Csibra, G., Simion, F., and Johnson, M.H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Science USA, 99,* 9602–9605.

Fathi, M., Bateson, M., & Nettle, D. (2014). Effects of watching eyes and norm cues on charitable giving in a surreptitious behavioral experiment. Evolutionary Psychology,12(5), 878-887.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160

Fayant, M. P., Sigall, H., Lemonnier, A., Retsin, E., & Alexopoulos, T. (2017). On the limitations of manipulation checks: an obstacle toward cumulative science*. International Review of Social Psychology, 30,* 125–130.

Fayard, J. V., Bassi, A. K., Bernstein, D. M., & Roberts, B. W. (2009). Is cleanliness next to godliness? Dispelling old wives’ tales: Failure to replicate Zhong and Liljenquist (2006). *Journal of Articles in Support of the Null Hypothesis, 6,* 21-31.

Fehr, E. (2004). Don’t lose your reputation. *Nature, 432,* 449-450.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review,* *90*(4), 980-994.

Fehr, E., & Gächter. S. (2002). Altruistic punishment in humans. *Nature,* *415*(6868), 137-140.

Fehr, E., & Gächter, S. (2003). The puzzle of human cooperation- A reply. *Nature, 421*, 912.

Fehr, E., & Schneider, F. (2009). Eyes are on us, but nobody cares: are eye cues relevant for strong reciprocity? *Proceedings of the Royal Society B, 277,* 1315–1323.

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science*, 25(3), 656-664.

Ferguson, C. J., & Beresin, E. (2017). Social science's curious war with pop culture and how it was lost: The media violence debate and the risks it holds for social science. *Preventative Medicine, 99,* 69-76.

Ferguson, C. J., Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science’s aversion to the null. *Perspectives on Psychological Science, 7,* 555–561.

Fiddler, F., Thorn, F. S., Barnett, A., Kambouris, S., & Kruger, A. (2018). The Epistemic importance of establishing the absence of an effect. *Advances in Methods and Practices in Psychological Science, 1*(2), 237-244.

Fiedler, K., Schwarz, N. (2016). Questionable research practices revisited*. Social Psychological & Personality Science, 7,* 45–52.

Fielden, T. (2017, Feb 22). Most scientists ‘can’t replicate studies by their peers.’ *BBC News.* Retrieved from http://www.bbc.co.uk/news/science-environment-39054778

Finch, J. F., & Cialdini, R. B. (1989). Another indirect tactic of (self-) image management boosting. *Personality and Social Psychology Bulletin, 15,* 222-232.

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, A. P., & Rai, T. S. (2014). *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships*. Cambridge, UK: Cambridge University Press.

Fitzsimons, G. M., & Bargh, J. A. (2003). Thinking of you: Nonconscious pursuit of interpersonal goals associated with relationship partners*. Journal of Personality and Social Psychology, 84,* 148–164.

Forrest, K. D. (2012). False evidence ploys and interrogations: Mock jurors’ perceptions of false-evidence ploy type, deception, coercion, and justification. *Behavioral Sciences and the Law, 30,* 342-364.

Forsythe, R., Horowitz, J. L., Savin N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior 6*, 347-369.

Frankenhuis, W.E., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives on Psychological Science, 13*(4), 439-447.

Frazer, J. G. (1890/2009). *The Golden Bough*. Oxford, UK: Oxford University Press.

Frederick, S. (2005). On the ball: Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Fredman, L. A., Bastian, B., & Swann, W. B. (2017). God or Country? Fusion with Judaism predicts desire for retaliation following Palestinian stabbing intifada. *Social Psychological and Personality Science, 8*(8) 882-887.

Friese, J., & Wänke, M. (2014). Personal prayer buffers self-control depletion. *Journal of Experimental Social Pscyhology, 51,* 56-59.

Fu, G., Heyman, G. D., Quian, M., Guo, T., & Lee, K. (2015). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science, 19*(2), 275-283.

Fujii, T., Takagishi, H., Koizumi, M., & Okada, H. (2015). The effect of direct and indirect monitoring on generosity among preschoolers. *Scientific Reports, 5,* 9125.

Galen, L. W. (2012). Does religious belief promote prosociality? A critical examination. *Psychological Bulletin, 138*(5), 876-906.

Galinsky, A. D., Wang, C. S., & Ku, G. (2008). Perspective-takers behave more stereotypically. *Journal of Personality and Social Psychology, 95*(2), 404-419.

Gamer, M., Hecht, H., Seipp, N., Hiller, W., 2011. Who is looking at me? The cone of gaze widens in social phobia. *Cognition and Emotion 25,* 756–764.

Gámez, E., Diaz, J., Marrero, H. (2011). The uncertain universality of the Macbeth effect with a Spanish sample. *Spanish Journal of Psychology, 14,* 156-162.

Gangemi, A., & Mancini, F. (2013). Moral choices: the influence of the do not play god principle. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society, Cooperative Minds: Social Interaction and Group Dynamics* (pp. 2973-2977). Austin, TX: Cognitive Science Society.

Garrison, J. R., Bond, R., Gibbard, E., Johnson, M. K., & Simons, J. S*.* (2017). Monitoring what is real: the effects of modality and action on accuracy and type of reality monitoring error. *Cortex, 87,*108-117.

Garvey, A., & Bolton, L. (2017). The licensing effect revisited: How virtuous behavior heightens the pleasure derived from subsequent hedonic consumption. *Journal of Marketing Behavior*, *2,* 291-298.

Geary, D. C., & Huffman, K. J. (2002). Brain and cognitive evolution: Forms of modularity and functions of mind. *Psychological Bulletin, 128,* 667–698.

Gervais, W. M., Xygalatas, D., McKay, R. T., Van Elk, M., Buchtel, E. E., Aveyard, M., … Bulbulia, J. (2017). Global evidence of extreme intuitive moral prejudice against atheists. Nature Human Behaviour, 1(8), 0151.

Gervais, W. M. (2014). “Everything Is Permitted? People Intuitively Judge Immorality as Representative of Atheists.” *PLoS ONE 9*(4), e92302.

Gervais, W. M., & Norenzayan, A. (2012a). Like a camera in the sky? Thinking about God increases public self-awareness and socially desirable responding. *Journal of Experimental Social Psychology, 48*, 298–302.

Gervais, W. M., & Norenzayan, A. (2012b). Analytic thinking promotes religious disbelief. *Science,* *336*(6080), 493–496.

Ginges, J., Hansen, I., & Norenzayan, A. (2009). Religion and support for suicide attacks. *Psychological Science, 20,* 224-230.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel*. Psychological Science, 20*(3), 393-398.

Gino, F., & Galinsky, A. D. (2012). Vicarious Dishonesty: When psychological closeness creates distance from one’s moral compass. *Organizational Behavior and Human Decision Processes, 119*(1), 15-26.

Gino, F., Gu, J., & Zhong, C. B. (2009). Contagion or restitution? When bad apples can motivate ethical behavior*. Journal of Experimental Social Psychology, 45*(6), 1299–1302.

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking*. Journal of Applied Psychology, 93*(5), 1165-1173.

Giummarra, M. J., Tracy, L. M., Young, K. A., & Fitzgibbon, B. M. (2016). The social side of pain: What does it mean to feel another’s pain? In S. van Rysewyk (Ed.), *Meanings of Pain* (pp. 353-373). Basel, Switzerland: Springer International Publishing.

Goldin-Meadow, S. (2016). Why preregistration makes me nervous. *Observer.* Retrieved from http://www.psychologicalscience.org/observer/why-preregistration-makes-me-nervous

Goldstein, N. J., & Cialdini, R. B. (2007). The spyglass self: A model of vicarious self-perception. *Journal of Personality and Social Psychology, 92*(3), 402-417.

Goldstein, N. J., Martin, S. J., & Cialdini, R. B. (2008). *Yes! 50 scientifically proven ways to be persuasive*. New York, NY: Free Press.

Gollwitzer, P., & Kirchhof, O. (1998). The willful pursuit of identity. In J. Heckhausen, & C. S. Dweck (Eds.), *Life-span perspectives on motivation and control*(pp. 389-423). New York, NY: Cambridge University Press.

Gollwitzer, M., & Melzer, A. (2012). Macbeth and the Joystick: Evidence for moral cleansing after playing a violent video game. *Journal of Experimental Psychology, 48*(6), 1356-1360.

Gollwitzer, P. A., Wicklund, R. A., & Hilton, J. L. (1982). Admission of failure and symbolic self-completion: Extending Lewinian theory. *Journal of Personality and Social Psychology, 43*, 358–371.

Gomes, C. M., & McCullough, M. E. (2015). The effects of implicit religious primes on dictator game allocations: A preregistered replication experiment. *Journal of Experimental Psychology: General, 144*(6), e94-e104.

Gómez, Á, Brooks, M.L., Buhrmester, M. D., Vázquez, A., Jetten, J. & Swann, W. B., Jr. (2011). On the nature of identity fusion: Insights into the construct and a new measure. *Journal of Personality and Social Psychology, 100,* 918- 933.

Gonzales, J. E., & Cunningham, C. A. (2015). The promise of pre-registration in psychological research. *Psychological Science Agenda.* Retrieved from http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26,* 213–224.

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*(1), 38–44.

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.

Gorsuch, R. L. (1968). The conceptualization of God as seen in adjective ratings. *Journal for the Scientific Study of Religion, 7*(1), 56.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*(5), 505-509.

Govern, J. M., & Marsch, L. A. (2001). Development and validation of the situational self-awareness scale. *Consciousness and Cognition, 10,* 366–378.

Grasmick, H. G., Bursik Jr., R. J., & Blackwell, B. S. (1993). Religious beliefs and public support for the death penalty for juveniles and adults*. Journal of Crime and Justice, 16*, 59-86.

Grasmick, H. G., Davenport, E., Chamlin, M. B., & Bursik, R J. (1992). Protestant fundamentalism and the retributive doctrine of punishment. *Criminology*,*30*(1), 21-46.

Grasmick, H, G., Kinsey, K., & Cochran, J. K. (1991). Denomination, religiosity and compliance with the law: A study of adults. *Journal for the Scientific Study of Religion,* *30,* 99-107.

Gray, K., & Wegner, D. M. (2010). Blaming God for our pain: Human suffering and the divine mind. *Personality and Social Psychology Review, 14*(1), 7-16.

Greenaway, K. H., Fisk, K., & Branscombe, N. R., (2017). Context matters: Explicit and implicit reminders of ingroup privilege increase collective guilt among foreigners in a developing country. *Journal of Applied Social Psychology, 47*(12), 677-681.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgement. Science, *293*, 2105-2108.

Greene, M., & Low, K. (2014). Public integrity, private hypocrisy and the moral licensing effect. *Social Behavior and Personality: an international journal, 42*(3), 391-400.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603–618.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality & Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.

Griffin, D. W. and Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In M P. Zanna (Ed.), *Advances in Experimental Social Psychology*, *Vol. 24,* (pp. 319-359). New York, NY: Academic Press.

Grossman, T., Johnson, M. H., Farroni, T., & Csibra, G. (2007) Social perception in the infant brain: Gamma oscillatory activity in response to eye gaze. *Social Cognitive and Affective Neuroscience, 2,* 284–291.

Gudjonsson, G. H. (2017). Memory distrust syndrome, confabulation and false confession. *Cortex, 87*, 156-165.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. New York, NY: Wiley.

Gudjonsson, G.H., & MacKeith, J. A. C. (1982). False confessions. Psychological effects of interrogation. A discussion paper. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 253-269). Kluwer: Deventer.

Gudjonsson, G. H., & Roberts, J. C. (1983) Guilt and self-concept in "secondary psychopaths." *Personality and Individual Differences, 4*(1), 65-70.

Gudjonsson, G. H., Sigurdsson, J. F., Sigurdardottir, A. S., Steinthorsson, H., & Sigurdardottir, V. M. (2014). The role of the memory distrust in cases of internalized false confessions. *Applied Cognitive Psychology, 28*, 336e348.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555,* 175-182.

Guthrie, S. E. (1993). *Faces in the clouds*. Oxford, UK: Oxford University Press.

Guthrie, S. E. (2001). Why gods? A cognitive theory. In J. Andresen (Ed.), *Religion in mind: Cognitive perspectives on religious belief, ritual, and experience* (pp. 94-114).Cambridge, UK: Cambridge University Press.

Guthrie, S., Agassi, J., Andriolo, K.R., Buchdahl, D., Earhart, H.B., Greenberg, M., Jarvie, I., Saler, B., Saliba, J., Sharpe, K.J. & Tissot, G. (1980). A cognitive theory of religion [and comments and reply]. *Current Anthropology, 21*(2), 181-203.

Hagger, M., et al. (2016). A multilab preregistered replication of the ego-depletion Effect. *Perspectives on Psychological Science, 11*(4) 546–573.

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. (2012). *The righteous mind: How good people are divided by politics and religion*. London, UK: Penguin.

Haidt, J., & Graham, J. (2007). When morality opposes justice: conservatives have moral intuitions that Liberals may not recognize*. Social Justice Research, 20,* 98–116.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately preparely intuitions generate culturally variable virtues. *Daedalus, 133*, 55 – 66.

Haley, K. J., & Fessler, D. M. (2005). Nobody’s watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26,* 245-256.

Hall, D. L., Matz, D. C., & Wood, W. (2010). Why don’t we practise what we preach? A Meta-analytic review of religious racism. *Personality and Social Psychology Review, 14*(1), 126-139.

Hansen, B. C., Thompson, B., Hess, R.F., & Ellemberg, D. (2010). Extracting the internal representation of faces from human brain activity: An analogue to reverse correlation. *NeuroImage, 51,* 373-390.

Harbort, J., Witthöft, M., Spiegel, J., Nick, K., & Hecht, H. (2013). The widening of the gaze cone in patients with social anxiety disorder and its normalization after CBT. *Behaviour Research and Therapy 51*, 359–367.

Hardy, C. L., & van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin, 32*(10), 1402-1413.

Hare, R. D. (1975). Psychopathy. In P. H. Venables & M. J. Christie (Eds.), *Research in Psychophysiology* (pp. 325-348). New York, NY: Wiley.

Harrell, A. (2012). Do religious cognitions promote prosociality? *Rationality and Society, 24*(4), 463–482.

Harris, C. R., Coburn, N., Rohrer, D., & Pashlwe, H. (2013). Two failures to replicate high-performance goal priming effects. *PLos ONE, 8,* e72467.

Harth, N. S., Kessler, T., & Leach, C. W. (2008). Advantaged group's emotional reactions to intergroup inequality: The dynamics of pride, guilt, and sympathy. *Personality and Social Psychology Bulletin, 34*(1), 115-129.

Harvey, R. D., & Oswald, D. L. (2000). Collective guilt and shame as motivation for white support of black programs. *Journal of Applied Social Psychology, 30*(9), 1790-1811.

Haselton, M. G. (2007). Error management theory. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 311–12). Thousand Oaks, CA: Sage.

Haselton, M, G., Bryant, G. A., Wilke, A., Frederick, D., & Galperin, A. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition 27(5),* 733-763.

Haselton, M., & Buss, D. (2000). Error management theory: A new perspective on biases in cross-sex mind reading*. Journal of Personality and Social Psychology, 78,* 81–91.

Haselton, M. G. & Nettle, D. (2006) The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review* *10*(1), 47–66.

Haslam, N., & Loughnan, S. (2014). Dehumanization and Infrahumanization. *Annual Review of Psychology, 65,* 399-423.

Hauser, D. J, Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology, 9. https://*doi.org/10.3389/fpsyg.2018.00998

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400-407.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243-259.

Henkel, L. A., Coffman, K. A. J., & Dailey, E. M. (2008). A survey of people’s attitudes and beliefs about false confessions. *Behavioral Sciences and the Law, 26*, 555–584.

Henrich. J., Heine S. J., & Norenzayan A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–135.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319,* 1362–1367.

Hewstone, M., Cairns, E., Voci, A., McLernon, F., Niens, U., & Noor, M. (2004). Intergroup forgiveness and guilt in Northern Ireland: Social psychological dimensions of the “Troubles”. In N. R. Branscombe, & B. Doosje (Eds.), *Collective guilt: International perspectives* (pp. 193–215). New York, NY: Cambridge University Press.

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28*(3), 503e508.

Higgins, P., & Albrecht, G. (1977). Hellfire and delinquency revisited. *Social Forces, 55,* 952-58.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: how the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology, 49*, 263-266.

Hirschi, T., & Stark, R. (1969). Hellfire and delinquency. *Social Problems, 17,* 202-15.

Hofmann, W., & Fisher, R. R. (2012). How guilt and pride shape subsequent self-control. *Social Psychological and Personality Science, 3*(6), 682–690.

Holtgraves, T. (1989). The form and function of remedial moves: Reported use, psychological reality, and perceived effectiveness. *Journal of Language and Social Psychology*, *8*, 1–16.

Hope, A. L., Jones, C. R., Webb, T. L., Watson, M. T., & Kaklamanou, D. (2017). The role of compensatory beliefs in rationalizing environmentally detrimental behaviors. *Environment and Behavior, 50,* 401-425.

Horgan, A. J., Russano, M. B., Meissner, C. A., & Evans, J. R. (2012). Minimization and maximization techniques: assessing the perceived consequences of confessing and confession diagnosticity. *Psychology, Crime & Law, 18,* 65-78.

Horselenberg, R., Merckelbach, H., Smeets, T., Franssens, D., Gjalt-Jorn Ygram Peters, G-J. P., & Zeles, G. (2006). False confessions in the lab: Do plausibility and consequences matter? *Psychology, Crime & Law, 12*(1), 61-75.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14,* 399-425.

Hoskin, R., Hunter, M. D., & Woodruff, P. W. (2014). The effect of psychological stress and expectation on auditory perception: A signal detection analysis. *British Journal of Psychology*, *105*(4), 524–546.

Ikegami, T. (1993), Positive-negative asymmetry of priming effects on impression formation. *European Journal of Social Psychology, 23,* 1–16.

Inbar, Y., Pizarro, D. A., Gilovich, T., & Ariely, D. (2013). Moral masochism: On the connection between guilt and self-punishment. *Emotion, 13*(1), 14-18.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogations and confession* (4th ed.). Sudbury, MA: Jones and Bartlett.

Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *JAMA 294*(2), 218–228.

Ioannidis, J. P. (2014). How to make more published research true. *PLoS Medicine, 11,*e1001747.

Ironson, G., Stuetzle, R., Ironson, R. D., Balbin, E., Kremer, H., George, A., Schneiderman, N., & Fletcher, M. A. (2011). View of God as benevolent and forgiving or punishing and judgmental predicts HIV disease progression. *Journal of Behavioral Medicine,* *34*(6), 414-425.

Isen, A. M. (1987). Positive affect, cognitive processes, and social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 20* (pp. 203-253), San Diego, CA: Academic Press.

Isaiah Haastrup: ‘God should decide’ when baby dies. (2018, January 23). *BBC News.* Retreived from https://www.bbc.co.uk/news/uk-england-london-42787097

Iyer, A., Leach, C. W., & Crosby, F. J. (2003). White guilt and racial compensation: The benefits and limits of self-focus. *Personality and Social Psychology Bulletin, 29,* 117–129.

Jackson, J. C., & Gray, K. (in press). When a good God makes bad people: Testing a theory of religion and immorality. *Journal for Personality and Social Psychology.*

Jackson, J. C., Hester. N., & Gray, K. (2018). The faces of God in America: Revealing religious diversity across people and politics. *PLoS ONE 13*(6), e0198745.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*(5), 752-761.

Jacobs, D., & Carmichael, J. T. (2001). The Politics of Punishment across Time and Space: A Pooled Time-Series Analysis of Imprisonment Rates. *Social Forces,* *80*, 61-89.

Jacobs, D., & Carmichael, J. T. (2004). Ideology, Social Threat, and the Death Sentence: Capital Sentences across Time and Space. *Social Forces, 83*(1), 249-278.

Jacobs, D., Carmichael, J. T., & Kent, S. L. (2005). Vigilantism, Current Racial Threat, and Death Sentences. *American Sociological Review, 70*(4), 656-677.

Janis, I. (1983). *Groupthink* (2nd edition revised). Boston, MA: Houghton Mifflin.

Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and again, and again. *Science, 334,* 1225.

Jayne, B. C., & Buckley, J. P. (1999). *The investigator anthology*. Chicago, IL: Reid.

Jenkins, M., Winefield, H., & Sarris, A. (2011). Consequences of being accused of workplace bullying: an exploratory study. *International Journal of Workplace Health Management*,*4*(1), 33-47.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23,*524–532.

Johns, L. C. (2005). Hallucinations in the general population. *Current Psychiatry Reports, 7*, 162-167.

Johnson, D. D. P. (2005). God’s punishment and Public Goods: A test of the supernatural punishment hypothesis in 186 world countries. *Human Nature, 16*(4), 410–446.

Johnson, D. D.P. (2009). The error of God: Error management theory, religion and the evolution of cooperation. In S. A. Levin (Ed.), *Games, groups and the global good* (pp. 169-180). Berlin: Springer-Verlag.

Johnson, D. D.P. (2011). Why God is the best punisher*. Religion, Brain and Behavior, 1,* 77–84.

Johnson, D.D.P. (2015). *The god of wrath: Supernatural punishment and the evolution of cooperation*. New York, NY: Oxford University Press.

Johnson, D. (2016). *God is watching you: How the fear of God makes us human*. New York, NY: Oxford University Press.

Johnson, D. D. P., & Bering, J. M. (2006). Hand of God, mind of man: Punishment and cognition in the evolution of cooperation. *Evolutionary Psychology, 4*(1), 219-233.

Johnson, D. D. P., & Krüger, O. (2004). The God of wrath: Supernatural punishment and the evolution of cooperation. *Political Theology,* *5*(2), 159-176.

Johnson, K. A., & Cohen, A. B. (2016). Authoritarian and benevolent god representations and the two sides of prosociality. *Behavioral And Brain Sciences,* *39*. 32.

Johnson, K. A., Li, J. Y., Cohen, A. B., & Okun, M. A. (2013). Friends in high places: The influence of authoritarian and benevolent god-concepts on social attitudes and behaviors. *Psychology of Religion and Spirituality, 5*(1), 15–22.

Johnson, K. A., Liu, R., Minton, E. A., Bartholomew, D. E., Peterson, M., Cohen, A. B., & Kees, J. (2017). Citizens’ representations of God and support for sustainability policies. *Journal of Public Policy and Marketing 36*(2), 362- 378.

Johnson, M. H., & Morton, J. (1991)*. Biology and Cognitive Development: The Case of Face Recognition.* Oxford, UK: Blackwell.

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Johnson, M. K., Rowatt, W. C., & LaBouff, J. P. (2010). Priming Christian Religious Concepts Increases Racial Prejudice. *Social Psychology and Personality Science, 1*(2), 119-126.

Jonas, E., Schimel, J. E., Greenberg, J., & Pyszczynski, T. (2002). TheScrooge Effect: Evidence that mortality salience increases prosocial attitudes and behavior. *Personality and Social Psychology Bulletin, 28*, 1342-1353.

Jones, W. H., Schratter, A. K., & Kugler, K. (2000). The Guilt Inventory. *Psychological Reports, 87,* 1039-1042.

Joosten, A., van Dijke, M., Van Hiel, A., & De Cremer, D. (2014). Feel Good, Do-Good?! On Consistency and Compensation in Moral Self-Regulation. *Journal of Business Ethics, 123*(1), 71-84.

Jordan, J., Flynn, F., & Cohen, T. R. (2015). Forgive them for I have sinned: The relationship between guilt and forgiveness of others’ transgressions. *European Journal of Social Psychology*, *45*, 441-459.

Jordan, J., Mullen, E., & Murnighan, J. K (2011). Striving for the moral self: the effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin, 37,* 701-713.

Jun, Y. Y., Mareschal, I., Clifford, C. W., & Dadds, M. R. (2013). Cone of direct gaze as a marker of social anxiety in males. *Psychiatry Research, 210*(1),193-198.

Kagel, J. H., Roth, A. E. (1995). *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Kant, I., Gregor, M., & Reath, A. (2015). *Critique of Practical Reason*. Cambridge, UK: Cambridge University Press.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B, 361*(1476), 2109-2128.

Kappes, A., Crockett, M., Faber, N. S., Savulescu, J., & Kahane, G. (2018). Concern for others lads to vicarious optimism. *Psychological Science, 29*(3), 379-389.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1),16-32.

Karremans, J. C., & Van Lange, P. A. M. (2004). Back to caring after being hurt: The role of forgiveness. *European Journal of Social Psychology, 34*, 207-227.

Karremans, J. C., Van Lange, P. A. M., & Holland, R. W. (2005). Forgiveness and its association with prosocial thinking, feeling and doing beyond the relationship with the offender. *Personality and Social Psychology Bulletin, 31*(10),1315-1326.

Kasper, K. (2012). Washing one’s hands after failure enhances optimism but hampers future performance. *Social Psychological and Personality Science, 4*(1), 69–73.

Kassin, S. M. (2005). On the Psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60*(3), 215-228.

Kassin, S. M. (2017a). False confessions: How can psychology so basic be so counterintuitive? *American Psychologist, 72*(9), 951-964.

Kassin, S. M. (2017b). False confessions. *WIREs Cognitive Science, 8*(6), e1439.

Kassin, S. M., & Gudjonsson, G. (2004). The psychology of confession: A review of the literature and issues. *Psychological Science in the Public Interest, 5*(2),33-67.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization and confabulation. *Psychological Science, 7*(3), 125-128.

Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A. M., & La Fon, D. (2007). Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior, 31,* 381–400.

Kassin, S. M., & Neumann, K. (1997). On the power of confession evidence: An experimental test of the fundamental difference hypothesis. *Law and Human Behavior, 21*(5)*,*469-484.

Kassin, S. M., & Wrightsman, L. S. (1985). Confession evidence. In S. M. Kassin and L. S.Wrightsman (Eds.), *The psychology of evidence and trial procedure* (pp. 67-94). Beverly Hills, CA: Sage.

Kay, A. C., Gaucher, D., McGregor, I., & Nash, K. (2010). Religious belief as compensatory control. *Personality and Social Psychology Review, 14*(1), 37–48.

Keller, J., & Pfattheicher, S. (2011). Vigilant self-regulation, cues of being watched and cooperativeness. *European Journal of Personality, 25,* 363–372.

Kelsey, C., Grossmann, T., & Vaish, A. (2018). Early reputation management: Three-year-old children are more generous following exposure to eyes. *Frontiers in Pscyhology, 9,* article 698.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2,* 196-217.

Khalid, S., Deska, J. C., & Hugenberg, K. (2016). The eyes are the windows to the mind: Direct eye gaze triggers the ascription of others' minds. *Personality and Social Psychology Bulletin, 42(*12), 1666-1677.

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of* *Marketing Research, 43,* 259–266.

Kirby, J. A., & Johnson, P. (2005). Students’ reactions to cheating: An examination of the effects of confession on forgiveness and emotions. *Journal of College and Character, 6*(4), 32-41.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., & Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology, 45,* 142–152.

Knowles, E. D., & Peng, K. (2005). White selves: Conceptualizing and measuring a dominant-group identity. *Journal of Personality and Social Psychology, 89*(2), 223-241.

Kochanska, G., DeVet, K., Goldman, M., Murray, K., & Putnam, S. P. (1994). Maternal reports of conscience development and temperament in young children. *Child Development, 65*, 852-868.

Körner, A., & Strack. F. (in press). Conditions for the clean slate effect after success or failure, *The Journal of Social Psychology.*

Kouchaki, M. (2011). Vicarious moral licensing: The influence of others' past moral actions on moral behavior. *Journal of Personality and Social Psychology, 101*(4), 702–715.

Kouchaki, M., & Jami, A. (2018). The licensing effect of prosocial marketing messages. *Management Science, 64*(1), 102–111.

Krummenacher, P., Mohr, C., Haker, H., & Brugger, P. (2010). Dopamine, paranormal belief, and the detection of meaningful stimuli. *Journal of Cognitive Neuroscience, 22*, 1670–1681.

Kugler, K., & Jones, W. H. (1992). On conceptualizing and assessing guilt. *Journal of Personality and Social Psychology, 62,* 318-327.

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE, 9*(9), e105825.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (in press). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist.*

Kurzban, R. (2001). The social psychophysics of cooperation: Nonverbal communication in a public goods game. *Journal of Nonverbal Behavior, 25*(4), 241–259.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior, 28*(2), 75-84.

Kurzban. R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin 127,* 187–208.

La Barre, W. (1964). Confession as cathartic therapy in American Indian tribes. In A. Kiev (Ed.), *Magic, Faith and Healing: Studies in Primitive Psychiatry Today* (pp. 36-49)*.* Glencoe, IL: Free Press.

LaBouff, J. P., Rowatt, W. C., Johnson. M. K., & Finkle, C. (2012). Differences in attitudes toward outgroups in religious and nonreligious contexts in a multinational sample: A situational context priming study. *International Journal for the Psychology of Religion, 22*(1), 1-9.

Lakens, D. (2014, December 19). Observed power, and what to do if your editor asks for post-hoc power analyses. [blog post]. Retrieved from http://daniellakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html

Lakens, D., et al (in press). Justify you alpha. *Nature Human Behavior.* https://doi.org/10.31234/osf.io/9s3y6

Lamba, S., & Mace, R. (2010). People recognise when they are really anonymous in an economic game. *Evolution and Human Behavior, 31*(4), 271-278.

Lanman, J. A., & Burhmester, M. D. (2017). Religious actions speak louder than words: exposure to credibility-enhancing displays predicts theism. *Religion, Brain and Behavior, 7*(1), 3-16.

Last ‘sin-eater’ to be celebrated with church service. (2010, September 19). *BBC News.* Retrieved from http://www.bbc.co.uk/news/uk-england-shropshire-11360659

Laurin, K., Kay, A. C., & Fitzsimons, G. M. (2012). Divergent effects of activating thoughts of God on self-regulation. *Journal of Personality and Social Psychology, 102,* 4-21.

Laurin, K., Shariff, A. F., Henrich, J., & Kay, A. C. (2012). Outsourcing punishment to God: beliefs in divine control reduce earthly punishment. *Proceedings of the Royal Society B: Biological Sciences,* *279*(1741), 3272-3281.

Laws, K. R. (2013). Negativland- A home for all findings in psychology. *BMC Psychology*, 1:2.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*(2), 234-249.

Leach, C. W., Iyer, A., & Pedersen, A. (2006). Anger and guilt about ingroup advantage explain the willingness for political action. *Personality and Social Psychology Bulletin, 32*(9), 1232-1245.

Lee, S. W. S., & Schwarz, N. (2010). Dirty hands and dirty mouths embodiment of the moral-purity metaphor is specific to the motor modality involved in moral transgression. *Psychological Science*, *21*,1423–1425.

Lee, S. W. S., & Schwarz, N. (2011). Wiping the slate clean: Psychological consequences of physical cleansing. *Current Directions in Psychological Science, 20*(5), 307–311.

Lee, S. W. S., & Schwarz, N. (2016). Clean-moral effects and clean-slate effects: Physical cleansing as an embodied procedure of psychological separation. In R. Duschinsky, S.Schnall, & D.Weiss (Eds), *Purity and danger now: New perspectives* (pp.136-161). London, UK: Routledge.

Leiber, M. J., & Woodrick, A. C. (1997). Religious beliefs, attributiona styles, and adherence to correctional orientations. *Criminal Justice and Behavior, 24*(4), 495–511.

Leidner, B., & Castano, E. (2012). Morality shifting in the context of intergroup violence. *European Journal of Social Psychology, 42*(1), 82-91.

Leidner, B., Castano, E., Zaiser, E., & Giner-Sorolla, R. (2010). Ingroup glorification, moral disengagement, and justice in the context of collective violence. *Personality and Social Psychology Bulletin, 36*(8), 1115-1129.

Leijten, J., Roseboom, H., & Hofer, R. (2010). More frontier research for Europe. A venture approach for funding high risk-high gain research. *Joint Institute for Innovation Policy.* Retrieved from https://ec.europa.eu/research/erab/pdf/erab-study-high-risk-high-gain-2010\_en.pdf

Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London B: Biological Sciences, 268,* 745–753.

Leimgruber, K. L., Shaw, S., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PloS ONE, 7*(10), e48292.

Leo, R. A. (2008). *Police interrogation and American justice*. Cambridge, MA: Harvard University Press.

Leo, R. A., & Liu, B. (2009). What do potential jurors know about police interrogation techniques and false confessions? *Behavioral Sciences & the Law, 27,* 381–399.

Leonard, M. A., Yung, S, M., & Cairns, E. (2015). Predicting intergroup forgiveness from in-group identification and collective guilt in adolescent and adult affiliates of a Northern Irish cross-community organization. *Peace and Conflict: Journal of Peace Psychology, 21*(2), 155–167.

LeVine, R. A., & Campbell, D. T. (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior.* Oxford, UK: John Wiley & Sons.

Lickel, B., Schmader, T., & Barquissau, M. (2004). The evocation of moral emotions in intergroup contexts: The distinction between collective guilt and collective shame. In N. R. Branscombe & B. Doosje (Eds.), *Collective guilt: International perspectives* (pp. 35–55). New York, NY: Cambridge University Press.

Lickel, B., Schmader, T., Curtis, M., Scarnier, M., & Ames, D. R. (2005). Vicarious shame and guilt. *Group Processes and Intergroup Relations, 8*(2), 145-157.

Lieberman, M. D., Hariri, A., Jarecho, J.M., Eisenberger, N.I., Bookheimer, S.Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience, 8*(6), 720–2.

Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist, 67*, 111–129.

Lilienfeld, S. O. (2017). Psychology’s replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science, 12*(4), 660 – 664.

Lindeman, M., & Svedholm, A. M. (2012). What’s in a term? Paranormal, superstitious, magical and supernatural beliefs by any other name would mean the same. *Review of General Psychology, 16,* 241–255.

Lindsay, S. D. (2107). Preregistered direct replications in psychological science. *Psychological Science, 28*(9), 1191-1192.

Linscott, R. J., & Van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: On the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine, 43,* 1133–1149.

Lipsey, N., Pogge, G., Shepperd, J., & Miller, W. (2016). *Views of God and consequences for behavior.* Poster session presented at 17th Annual meeting of the Society for Personality and Social Psychology, San Diego, CA.

Lisdorf, A. (2007). What’s HIDD’n in the HADD? *The Journal of Cognition and Culture, 7*(3), 341-353.

Liu, J., Li. J., Feng, L., Li, L., Tian, J., & Le, K. (2014). Seeing Jesus in toast: neural and behavioural correlates of face pareidolia. *Cortex, 45,* 60-77.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision-making. In R. Davidson, H. Goldsmith, & K. Scherer (Eds.), *Handbook of affective science* (pp. 619–642). Oxford, UK: Oxford University Press.

Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100,* 91-108.

Lowry, P. B., D’Arcy, J., Hammer, B., & Moody, G. D. (2016). “Cargo Cult” science in traditional organization and information systems survey research: A case for using non-traditional methods of data collection, including Mechanical Turk and online panels. *The Journal of Strategic Information Systems, 25*, 232–240.

Lu, H., & Miethe, T. D. (2003). Confessions and case disposition in China. *Law and Society Review, 37*(3), 549-578.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151-9.

Maij, D. L., Van Schie, H. T., & Van Elk, M. (2017). The boundary conditions of the hypersensitive agency detection device: an empirical investigation of agency detection in threatening situations. *Religion, Brain & Behavior*, 1-29.

Major, B., & Gramzow, R. H. (1999). Abortion as stigma: Cognitive and emotional implications of concealment. *Journal of Personality and Social Psychology, 77,* 735-745.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537–542.

Malhotra, D. (2010). (When) Are religious people nicer? Religious salience and the “Sunday effect” on prosocial behaviour. *Judgment and Decision Making, 5,* 138–143.

Manesi, Z., Van Lange, P. A. M., & Pollet, T. V. (2015). Butterfly eyespots: Their potential influence on aesthetic preferences and conservation attitudes. *PloS ONE, 10*, e0141433.

Manesi, V., Van Lange, P. A. M., Van Doesum, N. J., & Pollet, T.V. (in press). What are the most powerful predictors of charitable giving to victims of typhoon Haiyan: Prosocial traits, socio-demographic variables, or eye cues? *Personality and Individual Differences.*

Mann, N. H., & Kawakami, K. (2012). The long, steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General, 141,* 187-197.

Manstead, A. S., & Tetlock, P. E. (1989). Cognitive appraisals and emotional experience: Further evidence. *Cognition & Emotion, 3,* 225–239.

Marder, J., & Fritz, M. (2015, Feb 11). The internet’s hidden science factory. *PBS News Hour.* Retrieved from https://www.pbs.org/newshour/science/inside-amazons-hidden-science-factory

Mareschal, I., Calder, A. J. & **Clifford, C. W. G.** (2013). Humans have an expectation that gaze is directed towards them. *Current Biology, 23,* 717-721.

Mareschal, I., Otsuka, Y., & **Clifford, C. W. G.** (2014). A generalized tendency toward direct gaze with uncertainty. *Journal of Vision 14*(12), *27*, 1–9.

Marriage, M. (2018, January 23).Men only: Inside the charity fundraiser where hostesses are put on show. *Financial Times.* Retrieved from https://www.ft.com/content/075d679e-0033-11e8-9650-9c0ad2d7c5b5

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavioral Research Methods, 44*, 1-23.

Mason, W., & Watts, D. J. (2009). Financial incentives and the performance of crowds. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 77–85). New York, NY: ACM.

Matsugasaki, K., Tsukamoto, W., & Ohtsubo, Y. (2015). Two Failed Replications of the Watching Eyes Effect. *Letters on Evolutionary Behavioral Science, 6*(2), 17-20.

May, F., & Monga, A. (2014). When time has a will of its own, the powerless don’t have the will to wait: Anthropomorphism of time can decrease oatience. *Journal of Consumer Research, 40*, 924 –942.

Mazar, N., & Zhong, C. B. (2010). Do green products make us better people? *Psychological Science, 21,* 494–498.

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.

McCauley, R. N. (2011). *Why religion is natural and science is not.* Oxford, UK: Oxford University Press.

McCullough, M, E., Bono, G., & Root, L. M. (2005). Religion and Forgiveness. In R. F Paloutzian & C. L. Park (Eds.), *Handbook of the Psychology of Religion and Spirituality* (pp. 394-411). New York, NY: Guilford.

McCullough, M. E., Fincham, F. D., & Tsang, J. (2003). Forgiveness, forbearance, and time: The temporal unfolding of transgression-related interpersonal motivations. *Journal of Personality and Social Psychology*, *3,* 540-557.

McGarty, C., Pedersen, A., Wayne Leach, C., Mansell, T., Waller, J., & Bliuc, A. (2005). Group-based guilt as a predictor of commitment to apology. British Journal of Social Psychology, 44(4), 659-680.

McKay, R. T., & Dennett, D. (2009). The evolution of misbelief. *Behavioral and Brain Sciences, 32,* 493-561.

McKay, R. T., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior, 31,* 301-319.

McKay, R. T., Herold, J., & Whitehouse, H. (2013). Catholic Guilt? Recall of confession promotes pro-sociality *Religion, Brain and Behavior, 3*(3),201-209.

McKay, R. T., Jong, J., & O’Lone, K. (2018). Idolizing the indexical: commentary on Van Leeuwen and van Elk. *Religion, Brain and Behavior,* https://doi.org./10.1080/2153599X.2018.1453533

McKay, R. T., Ross, R. M., O’Lone, K., & Efferson, C. (2017). The error of God revisited. *Religion, Brain and Behavior, 8*(3), 306-310.

McKay, R. T., & Whitehouse, H. (2015). Religion and Morality. *Psychological Bulletin, 141*(2), 447-473.

McKay, R. T., & Whitehouse, H. (2016). Religion promotes a love for thy neighbour: But how big is the neighbourhood. *Behavioral and Brain Sciences, 39*, 35.

Meijers, M. H., Noordewier, M. K., Verlegh, P. W., Zebregs, S., & Smit, E. G. (2018). Taking Close Others’ Environmental Behavior Into Account When Striking the Moral Balance? Evidence for Vicarious Licensing, Not for Vicarious Cleansing. *Environment and Behavior*, 001391651877314. https://doi.org/10.1177/0013916518773148

Meijers, M. H., Verlegh, P. W., Noordewier, M. K., & Smit, E. G. (2015). The dark side of donating: how donating may license environmentally unfriendly behavior. *Social Influence*, *10*(4), 250-263.

Meissner, C. A., & Russano, M. B. (2003). The Psychology of Interrogations and False Confessions: Research and Recommendations. *The Canadian Journal of Police and Security Services, 1*(1), 53-64.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4,* 344-357.

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K. K., Tuller, D. M., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology, 48*, 774-777.

Metzger, M. (1997). *History of the Liturgy: The Major Stages*. Collegeville, MN: Liturgical Press.

Milinski, M., Semman, D., & Krambeck, H. J. (2002). Reputation helps solve the ‘tragedy of the commons’. *Nature, 415,* 424-426.

Miller, G. F. (2007). Sexual selection for moral values. *The Quarterly Review of Biology, 82*(2), 97-125.

Miller, M.K (2006). *Religion in Criminal Justice*. New York, NY: LFB Scholarly Publishing.

Miller, M. K., & Bornstein, B. H. (2006). The use of religion in death penalty sentencing trials. *Law and Human Behavior, 30* (6), 675-684.

Miller, M. K., & Hayward, D. R. (2008). Religious characteristics and death penalty. *Law and Human Behavior, 32*(2), 113-123.

Miyatake, S., & Higuchi, M. (2017). Does religious priming increase the prosocial behaviour of a Japanese sample in an anonymous economic game? *Asian Journal of Social Psychology, 20*(1), 54-59.

Moery, E., & Calin-Jageman, R. J. (2016). Direct and conceptual replications of Eskine (2013). Organic food exposure has little to no effect on moral judgments and prosocial behavior. *Social Psychological and Personality Science, 7*(4), 312-319.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81,* 33–43.

Moore, C., & Gino, F. (2013). Ethically adrift: How others pull our moral compass from true North, and how we can fix it. *Research in Organizational Behavior, 33,* 53-77.

Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology: General*, *138*(4), 535–545.

Morey, M. R., & Yadav, S. (2018). Documentation of the file drawer problem in academic finance journals. *The Journal of Investing, 27*(1), 143-147.

Morin, O. (2013). How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution. *Evolution and Human Behavior, 34,* 222-229.

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., , . . . Skitka, L. J. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology, 113*, 34–58.

Mummendey, A., Klink, A., Mielke, R., Wenzel, M., & Blanz, M. (1999). Socio-structural characteristics of intergroup relations and identity management strategies: Results from a field study in East Germany. *European Journal of Social Psychology, 29,* 259-285.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1(1), 0021.

Myers, D. B. (2002). *Social psychology* (7th ed.). Boston, MA: McGraw Hill.

Myer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection task. *Judgment and Decision Making, 13*(3), 246-259.

Myers, M. A. (1988). Social background and the sentencing behavior of judges. *Criminology, 26*(4), 649-676.

Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby effect. *Emotion 9*(1), 118-22.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the Dictator Game: it's not how much you give, it's being seen to give something. Evolution and Human Behavior, 34(1), 35-40.

Newman, K.P., & Brucks, M. (2018). The inﬂuence of corporate social responsibility efforts on the moral behavior of high self-brand overlap consumers. *Journal of Consumer Psychology, 28*(2), 253- 271.

Nguyen, A. Q. (2018, January 25). Presidents’ club dinner: Why good deeds never justify bad actions. *The Conversation.* Retrieved from https://theconversation.com/presidents-club-dinner-why-good-deeds-never-justify-bad-actions-90678

Nieuwboer, W., Van Schie, H. T., Karremans, J., & Wigboldus, D. (2015). Supernatural agency and forgiveness. *Journal for the Cognitive Science of Religion, 3*(1), 85.

Nieuwboer, W., Van Schie, H. T., & Wigboldus, D. (2014). Priming with religion and supernatural agency enhances the attribution of intentionality to natural phenomena. *Journal for the Cognitive Science of Religion, 2*(2), 97-120.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics,* *92*(1-2), 91-112.

Nikiforakis, N., & Engelmann, D. (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior and Organization, 78*(3), 319-332.

Nilsson, A., Bergquist, M., & Schultz, W. P. (2016). Spillover effects in environmental behaviors, across time and context: a review and research agenda. *Environmental Education Research, 23*(4), 573-589.

Nisan, M., & Horenczyk, G. (1990). Moral balance: The effect of prior behavior on decision in moral conflict. *British Journal of Social Psychology, 29,* 29–42.

Noblet, C. L., & McCoy, S. K. (2017). Does one good turn deserve another? Evidence of domain-Specific licensing in energy behavior. *Environment and Behavior,* 1-25.

Norenzayan, A. (2013). *Big gods: how religion transformed cooperation and conflict.* Princeton, NJ: Princeton University Press.

Norenzayan. A., Henrich, J., & Slingerland, E. (2013). Religious prosociality: A synthesis. In P. J. Richerson & M. H. Christiansen (Eds.), *Cultural Evolution: Society, Technology, Language and Religion, vol 12* (pp. 378-379). Cambridge, MA: MIT Press.

Norenzayan A., & Shariff, A. (2008). The origin and evolution of religious prosociality. *Science 322*, 58 – 62.

Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E. & Henrich, J. (2016) The cultural evolution of prosocial religions. *Behavioral and Brain Sciences, 39*, 1-65.

Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017a). Artificial surveillance cues do not increase generosity: Two meta-analyses. *Evolution and Human Behavior, 38,*144-153.

Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017b). Effect of artificial surveillance cues on reported moral judgment: Experimental failures to replicate and two meta-analyses. *Evolution and Human Behavior, 38,*561-571.

Norton, M. I., Monin, B., Cooper, J., & Hogg, M. A. (2003). Vicarious dissonance: Attitude change from the inconsistency of others. *Journal of Personality and Social Psychology, 85*(1), 47-62.

Nosek, B. A., Spies, J. R., Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7,* 615–631.

Nosek, B. A., & Lakens, D. (2014). Registered reports: a method to increase the credibility of published results. *Social Psychology, 45,* 137–141.

Nosek, B., et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1428.

Nowak, M. A., & Sigmund, K. (1998a). Evolution of indirect reciprocity by image scoring*. Nature, 393,* 573–577.

Nowak, M. A., & Sigmund, K. (1998b). The dynamics of indirect reciprocity. *Journal of Theoretical Biology, 194*(4), 561-574.

Nowak, M. A., & Sigmund, K. (2005). The evolution of indirect reciprocity. *Nature, 437,* 1291-1298.

NWO makes 3 million available for replication studies pilot. (2016, July 19). Retrieved from https://www.nwo.nl/en/news-and-events/news/2016/nwo-makes-3-million-available-for-replication-studies-pilot.html

Oda, R., Niwa, Y., Honma, A., & Hiraishi, K. (2011). An eye-like painting enhances the expectation of a good reputation. *Evolution and Human Behaviour, 32,* 166- 171.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a wword: Self governance is possible. *American Political Science Review, 86*, 404-417.

Over, H., & Uskul, A. K. (2016). Culture moderates children’s responses to ostracism situations. *Journal of Personality and Social Psychology, 110*(5), 710-724.

Paladino, M., Mazzurega, M., Pavani, F., & Schubert, T. W. (2010). Synchronous Multisensory Stimulation Blurs Self-Other Boundaries. *Psychological Science*, *21*(9), 1202-1207.

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature 432*, 499–502.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23,* 184–188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Pargament, K. I., McCullough, M. E., & Thoresen, C. E. (2000). The frontier of forgiveness: Seven directions for psychological study and practice. In M. E. McCullough., K. Pargament & C. Thoresen (Eds.), *Forgiveness: Theory, research and practice* (pp. 299-320). New York, NY: Guilford.

Parrot, W. G., & Hertel, P. (1999). Research methods in cognition and emotion. In T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 61–81). New York, NY: John Wiley & Sons Ltd.

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., & Weiner, K. S. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *Journal of Neuroscience, 32*(43), 14915-14920.

Pazhoohi, F., Pinho, M., & Arantes, J. (2017). Effect of religious day on prosocial behavior: A field study. *The International Journal for Psychology of Religion, 27*(2), 116-123.

Pe’er, E., Acquisti, A., & Shalvi, S. (2014). ‘‘I cheated, but only a little’’- partial confessions to unethical behavior. *Journal of Personality and Social Psychology, 106*, 202-217.

Pe’er, E., Samat, S., Brandimarte, L., & Acquisti, A. (2017). Beyond the Turk: An empirical comparison of alternative platforms for online behavioural research*. Journal of Experimental Social Psychology, 70, 1*53-163.

Pe’er, E., Vosgerau, J., & Aquisti, S. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*(4), 1023-1031.

Pennebaker, J. M. (1989). Confession, Inhibition and Disease. *Advances in Experimental Social Psychology, 22,* 211-244.

Pennebaker, J. W. (1990). *Opening up: The healing power of confiding in others*. New York, NY: William Morrow.

Pennebaker, J. W. (1997). *Opening up: The Healing Power of Expressing Emotions*. New York, NY: Guilford Press.

Perillo, J., & Kassin, S. (2011). Inside interrogation: The lie, the bluff and false confession. *Law and Human Behavior, 35*(4), 327-337.

Perry, J. (1979). The problem of the essential indexical. *Nous, 13,* 3-21.

Perry, J. (1990). Self-notions. *Logos, 11,* 17-31.

Perry, J. (1998). Myself and I. In M. Stamm (Ed.), *Philosophie in Synthetischer Absicht,* (pp. 83-108). Stuttgart: KlettCotta.

Peters, E. (2010). Are delusions on a continuum? The case of religious and delusional beliefs. In I. Clarke (Ed.), *Psychosis and spirituality: Consolidating the new paradigm*, *2nd ed* (pp. 127–138). Chichester, UK: John Wiley.

Peters, E. N. (2008). *A Modern Guide to Indulgences: Rediscovering This Often Misinterpreted Teaching.* Chicago, IL: Hillenbrand Books.

Petrican, R., & Burris, C. T. (2012). Am I the stone? Overattribution of agency and religious orientation. *Psychology of Religion and Spirituality, 4,* 312-323.

Pew Research Center. (2014). *Global views on morality*. Retrieved from http://www.pewglobal.org/2014/ 04/15/global-morality/

Pew Research Center. (2015). *Global Attitudes Survey*. Retrieved from http://www.pewglobal.org/2015/06/23/spring-2015-survey/

Pew Research Center. (2018a, 25 April). *When Americans say they believe in God, what do they mean?* Retrieved from http://www.pewforum.org/2018/04/25/when-americans-say-they-believe-in-god-what-do-they-mean/

Pew Research Center. (2018b, 29 May). *Being Christian in Western Europe*. Retrieved from http://www.pewforum.org/2018/05/29/being-christian-in-western-europe/

Piff, P. K., Dietze, P., Feinberg, M., Stancato, D. M., & Keltner, D. (2015). Awe, the small self, and prosocial behavior. *Journal of Personality and Social Psychology*, *108*(6), 883-899.

Pfattheicher, S., & Keller, J. (2015). The watching eyes phenomenon: The role of a sense of being seen and public self-awareness. *European Journal of Social Psychology, 45,* 560–566.

Pfeiffer, T., Bertram, L., & Ioannidis, J. P. A. (2011). Quantifying Selective Reporting and the Proteus Phenomenon for Multiple Datasets with Similar Bias. *PLos One, 6*(3), e18362.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funagama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*(5), 729–38.

Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior*, *29*(3), 172-178.

Piazza, J., Bering, J. M., & Ingram, G. (2011). "Princess Alice is watching you": children's belief in an invisible person inhibits cheating. *Journal of Experimental Child Psychology, 109*(3), 311-320.

Pichon, I., Boccato, G., & Saroglou, V. (2007). Nonconscious influences of religion on prosociality: A priming study. *European Journal of Social Psychology, 37,* 1032–1045.

Pichon, I., & Saroglou, V. (2009). Religion and helping: Impact of target, thinking styles and just-world beliefs. *Archive for the Psychology of Religion, 31*, 215–236.

Pinto, L. E., & Nemorin, S. (2015). ‘Elf on the Shelf’ and the normalization of surveillance. *Our Schools and Our Selves, 24*(2),53-62.

Pizarro, D. A., & Tannenbaum, D. (2011). 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). Washington, DC: American Psychological Association.

Porter, S. B., & Baker, A. Y. (2015). CSI (Crime Scene Induction): Creating False Memories of Committing Crime. *Trends in Cognitive Sciences,* *9*(12), 716-718.

Powell, K. L., Roberts, G., & Nettle, D. (2012). Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket. *Ethology, 118,* 1096–1101.

Premack, D., & Premack, A. J. (1995). Origins of human social competence. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 205-218). Cambridge, MA: The MIT Press.

Prentice-Dunn, S., & Rogers, R. W. (1982). Effects of public and private self-awareness on deindividuation and aggression. *Journal of Personality and Social Psychology, 43*, 503- 513.

Preston, J. L., & Ritter, R. S. (2013). Different effects of religion and God on prosociality with the ingroup and outgroup. *Personality and Social Psychology Bulletin, 39*(11), 1471-1483.

Preston, J. L., Ritter R. S., & Hernandez J. I. (2010). Principles of religious prosociality: A review and reformulation. *Social and Personality Psychology Compass* *4*(8), 574-590.

Preston, J. L., Salomon, E., & Ritter, R. S. (2014). Religious prosociality: Personal, cognitive, and social factors. In V. Saroglou (Ed.), *Religion, personality, and social behavior* (pp. 149–169). New York, NY: Psychology Press.

Prinz, J., & Seidel, A. (2012). Alligator or squirrel: Musically induced fear reveals threat in ambiguous figures. *Perception*, *41*(12), 1535–1539.

Prochownik, K. (in press). Gods and goodness by the rivers of Babylon. A cognitive scientific approach to ancient Mesopotamian theology. In T. Oshima (Ed.), *Teaching morality in antiquity: wisdom texts, images, and oral traditions*. Tübingen: Mohr Siebeck.

Puckle, B. S. (1926). *Funeral Customs: their Origin and Development*. London, UK: T. Werner Laurie Ltd.

Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., … Henrich, J. (2016). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, *530*(7590), 327-330.

Radaelli, D., Benedetti, F., Cavallaro, R., Colombo, C., & Smeraldi, E. (2013). The reality monitoring deficit as a common europsychological correlate of schizophrenic and affective psychosis. *Behavioral Sciences, 3*(2), 244e252.

Raihani, N. J., & Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B,* 1-9.

Ramsey, S. R., Thompson, K. L., McKenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior, 58,* 354–360.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology, 299*, 172–179.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677.

Randolph-Seng, B., & Nielsen, M. E. (2007). Honesty: One effect of primed religious representations. *International Journal for the Psychology of Religion,* *17*, 303-315.

Rassin, E., & Israëls, H. (2014). False confessions in the lab: A review. *Erasmus Law Review, 4,* 219-224.

Rassin, E., & Koster, E. (2003). The correlation between thought–action fusion and religiosity in a normal sample. *Behaviour Research and Therapy, 41*(3), 361-368.

Razran, G. H. (1940). Conditioned response changes in rating and appraising socio-political slogans. *Psychological Bulletin, 37*(48).

Read, D., Loewenstein, G., & Kalyanaraman, S. (1999). Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic*. Journal of Behavioral Decision Making, 12*, 257- 273.

Rebega, O. L., Apostol, L., Benga, O., & Miclea, M. (2013). Inducing guilt: A literature review. *Procedia - Social and Behavioral Sciences*, *78*, 536-540.

Redlich, A. D., & Goodman, G. S. (2003). Taking responsibility for an act not committed: The influence of age and suggestibility. *Law and Human Behavior, 27*(2), 141-156.

Reed, A., Aquino, K., & Levy, E. (2007). Moral identity and judgments of charitable behaviors. *Journal of Marketing,* *71*(1), 178–193.

Reid, V. M., Dunn, K. J., & Young, R. J. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology*, *12*(27), 1825-1828.

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 89–117). San Diego, CA: Academic Press.

Reis, H. T., & Gruzen, J. (1976). On mediating equity, equality, and self-interest: the role of self-presentation in social exchange. *Journal of Experimental Social Psychology, 12,* 487-503.

Riekki, T., Lindeman, M., Aleneff, M., Halme, A., & Nuortimo, A. (2012). Paranormal and religious believers are more prone to illusory face perception than skeptics and non-believers. *Applied Cognitive Psychology, 27*(2), 150-155.

Rigdon, M., Ishii, K., Watabe, M., & Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology, 30*(3), 358–67.

Risko, E.F. & Kingstone, A. (2011). Eyes wide shut: Implied social presence, eye tracking and attention. *Attention Perception and Psychophyics, 73*(2),291-296.

Ritter, R. S., & and Preston, J. L. (2013). Representations of Religious Words: Insights for Religious Priming Research. *Journal for the Scientific Study of Religion,* *52*, 494-507.

Roberts, M. J. & Newton, E. J. (2001). Inspection times, the change task, and the rapid response selection task. *Quarterly Journal of Experimental Psychology A, 54*, 1031–1048.

Roccas, S., Klar, Y., & Liviatan, I. (2006). The paradox of group-based guilt: Modes of national identification, conflict vehemence, and reactions to the in-group's moral violations. *Journal of Personality and Social Psychology, 91*(4), 698-711.

Rodriguez Mosquera, P. M., Manstead, A. S. R., & Fischer, A. H. (2002). The role of honor concerns in emotional reactions to offenses. *Cognition and Emotion, 16*(1), 143-163.

Roes, F.L., & Raymond, M. (2003). Belief in Moralizing Gods. *Evolution and Human Behaviour, 24*(2), 126-135.

Rojas-Méndez, J. I., Davies, G., Omer, O., Chetthamrongchai, P., & Madran, C. (2002). A time attitude scale for cross-cultural research. *Journal of Global Marketing, 15*(3–4), 117–147.

Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z., Nandy, A., & Maddox, K. B. (2007). The effects of skin tone on race-related amygdala activity: an fMRI investigation. *Social Cognitive and Affective Neuroscience*, *2*(1), 39-44.

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology, 67*, 206–221.

Rosenberg, K. J., Betts, L. R., Eisner, M., & Ribeaud, D. (2011). Social antecedents of children’s trustworthiness. *Infant and Child Development, 21*(3), 310-322.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45,* 775–777.

Ross, L. D., Lelkes, Y., & Russell, A. G. (2012). How Christians reconcile their personal political views and the teachings of their faith: Projection as a means of dissonance reduction. *Proceedings of the National Academy of Sciences*, *109*(10), 3616–22.

Roth, L. M., & and Kroll, J. C. (2006). Risky business: Assessing risk-preference explanations for gender differences in religiosity. *American Sociological Review, 72*(2), 205-220.

Rothschild, Z. K., Landau, M. J., Sullivan, D., & Keefer, L. A. (2012). A dual-motive model of scapegoating: Displacing blame to reduce guilt or increase control. *Journal of Personality and Social Psychology, 102*(6), 1148-1163.

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication bias in meta-analysis: prevention, assessment and adjustments*. Chichester, UK: Wiley.

Rozin, P. (1999). The process of moralization. *Psychological Science, 10*(3), 218-221.

Rozin, P., Markwith, M., & McCauley, C. R. (1994). Sensitivity to indirect contacts with other persons: AIDS aversion as a composite of aversion to strangers, infection, moral taint and misfortune. *Journal of Abnormal Psychology, 103,* 495–504.

Ruffle, B. J., & Sosis, R. (2010). Do religious contexts elicit more trust and altruism? An experiment on facebook*. Unpublished manuscript.*

Rumsey, M. G. (1976). Effects of defendant background and remorse on sentencing judgments. *Journal of Applied Social Psychology, 6,* 64–68.

Rush, E. B., Stolzenberg, S. N., Quas, J. A., & Lyon, T. D. (2017). The effects of the putative confession and parent suggestion on children’s disclosure of a minor transgression. *Legal and Criminological Psychology, 22,* 60-73.

Rye, M. S., Pargament*,* K. P., Ali, M. A., Beck., G. L., Dorff, E.N., Hallisey, C., Narayanan, V., & Williams, J. G*.* (2000). Religious perspectives on forgiveness. In M. E. McCullough, K. Pargament & C. Thoresen (Eds.), *Forgiveness: Theory, research and practice* (pp. 17-40). New York: Guilford Press.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20,* 523–528.

Sagioglou, C., & Forstmann, M. (2013). Activating Christian religious concepts increases intolerance of ambiguity and judgment certainty. *Journal of Experimental Social Psychology, 49,* 933-939.

Saler, B. (2010). Theory and criticism: The cognitive science of religion. *Methods and Theory in the Study of Religion, 22*(4), 330-339.

Sample, I. (2015, August, 27). Study delivers bleak verdict on validity of psychology experiment results. *The Guardian.* Retrieved from https://www.theguardian.com/science/2015/aug/27/study-delivers-bleak-verdict-on-validity-of-psychology-experiment-results

Sánchez- Tójar, A., Nakagawa, S., Sanchez-Fortun, M., Martin, D. A., Ramani, S., Girndt, A.,…Schroeder, J. (2018). Meta-analysis challenges a textbook example of status signalling: Evidence for publication bias. Preprint retrieved from https://www.biorxiv.org/content/early/2018/03/16/283150

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated Recruitment of Autobiographical Memories. *Journal of Personality and Social Psychology, 59*(2), 229–41.

Saroglou, V., Corneille., O., & Van Cappellen, P. (2009). ‘Speak, Lord, your servant is listening’: Religious priming activates submissive thou*ghts and behaviors. International Journal for the Psychology of Religion, 19*, 143-154.

Satow, K. (1975). Social approval and helping. *Journal of Experimental Social Psychology, 11,* 501–509.

Saunders, T. J., Taylor, A. H., & Atkinson, Q. D. (2016). No evidence that a range of artificial monitoring cues influence online donations to charity in an MTurk sample. *Royal Society Open Science, 3*(10), 150710.

Scarnier, M., Schmader, T., & Lickel, B. (2009). Parental shame and guilt: Distinguishing emotional responses to a child’s wrongdoings. *Personal Relationships, 16*(2), 205–220.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17,* 551-566.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity and interpersonal relations*. Monterey, CA: Brooks/Cole.

Schlenker, B. R. (1982). Translating actions into attitudes: An identity-analytic approach to the explanation of social conduct. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 194-248). New York, NY: Academic Press.

Schloss, J. P., & Murray, M. T. (2011). Evolutionary accounts of belief in supernatural punishment: a critical review. *Religion, Brain and Behavior, 1*(1), 46-99.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*, 90–100.

Schmitt, M. T., Branscombe, N. R., & Brehm, J. W. (2004). Gender inequality and the intensity of men’s collective guilt. In N. R. Branscombe & B. Doosje (Eds.), *Collective guilt: International perspectives* (pp. 75-92). New York, NY: Cambridge University Press.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*(8), 299-309.

Schulze, L., Lobmaier, J.S., Arnold, M., & Renneberg, B. (2013). All eyes on me?! Social anxiety and self-directed perception of eye gaze. *Cognition and Emotion*, *27*(7), 1305-1313.

Schumann, K., McGregor, I., Nash, K. A., & Ross, M. (2014). Religious magnanimity: Reminding people of their religious belief system reduces hostility after threat. *Journal of Personality and Social Psychology, 107,* 432-453.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, *Vol. 25* (pp. 1–65). New York, NY: Academic Press.

Schwartz, S. H. & Bardi, A. (2001). Values and Behavior: Strength and Structure of Relations. *Personality and Social Psychology Bulletin*, *29*(10), 1207-1220.

Schwarz, N., & Clore, G. L. (1988). How do I feel about it? Informative functions of affective states. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior* (pp. 44 – 62). Toronto, Canada: Hogrefe.

Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: elastic justification and motivated communication of uncertain information*. Journal of Risk and Uncertainty*, *25*, 185-201.

Schweitzer, M. E., Ordóñez, L., & Douma, B. (2004). Goal setting as a motivator of unethical behavior. *Academy of Management Journal, 47*(3), 422-432.

Scott, S. (2013, July 25). Pre-registration would put science in chains. *Times Higher Education*. Retrieved from https://www.timeshighereducation.com/comment/opinion/pre-registration-would-put-science-in-chains/2005954.article

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition, 12,* 127–144.

Senju, A., Kikuchi, Y., Hasegawa, T., Tojo, Y., & Osanai, H. (2008). Is anyone looking at me? Direct gaze detection in children with and without autism. *Brain and Cognition, 67,* 127–139.

Senju. A., Yaguchi, K., Tojo, Y., & Hasegawa, T. (2003) Eye contact does not facilitate detection in children with autism. *Cognition, 89*, 43–51.

Shackelford, T. K., Liddle, J. R., Bering, J. M., & Shalkoski, G. (2014). Unbidden confession as an evolved pre-emptive strategy against punishment: A preliminary investigation with prisoners. *Personality and Individual Differences, 61-62,* 86-90.

Shafran, R., Watkins, E., & Charman, T. (1996). Guilt in obsessive-compulsive disorder. *Journal of Anxiety Disorders, 10*(6), 509-516.

Shah, P., Harris, A. J. L., Bird, G., Catmur, C., & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology, 90,* 71–127.

Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*, 181-190.

Shalvi, S., Handgraaf, M. J., & De Dreu, C. K. (2011). Ethical manoeuvring: why people avoid both major and minor lies. *British Journal of Management, 22*(s1),S16-S27.

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., & Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLoS ONE, 8*(4), e56515.

Shariff, A. F. (2011). Big gods were made for big groups. *Religion, Brain & Behavior, 1*(1), 89–93.

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming god concepts increases prosocial behavior in an anonymous economic game. *Psychological Science,* *18*, 803-809.

Shariff, A. F., & Norenzayan, A. (2011). Mean Gods Make Good People: Different Views of God Predict Cheating Behavior. *The International Journal for the Psychology of Religion, 21*(2), 85-96.

Shariff, A. F., Norenzayan, A., & Henrich, J. (2011). The birth of high gods: how the cultural evolution of supernatural policing influenced the emergence of complex, cooperative human societies, paving the way for civilization. In T. M. Schaller, A, Norenzayan, S. J. Heine, T. K. Yamagishi (Eds*.*)*, Evolution, culture, and the human mind* (pp. 119– 136). New York, NY: Psychology Press.

Shariff, A., & Rhemtulla, M. (2012). Divergent effects of beliefs in heaven and hell on national crime rates. *PloS One, 7*(6), e39048.

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2016). Religious Priming. A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review, 20*(1), 27-48.

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. Nature Neuroscience, 14(11), 1475-1479.

Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in Cognitive Sciences, 20*(1), 25–33.

Sherman, A. C., Merluzzi, T. V., Pustejovsky, J. E., Park, C. L., George, L., Fitchett, G., Jim, H. S. L., Munoz, A. R., Danhauer, S. C., Snyder, M. A ., & Salsman, J. M. (2015). A meta‐analytic review of religious or spiritual involvement and social health among cancer patients. *Cancer*,*121*(21), 3779-3788.

Sienen, I., & Schram, A. (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review, 50*(3), 581-602.

Siev, J., & Cohen, A. B. (2007). Is thought–action fusion related to religiosity? Differences between Christians and Jews. *Behaviour* *Research and Therapy, 45*, 829–837.

Sigmund, K. (2012). Moral assessment in indirect reciprocity. *Journal of Theoretical Biology, 299,* 25-30.

Sigmund, K., Hauert, C., & Nowak, M. (2001). Reward and punishment. *PNAS, 98*(19), 10757-10762.

Simbrunner, P., & Schlegelmich, B.B. (2017). Moral licensing: a culture-moderated meta-analysis. *Management Review Q, 67,* 201-225.

Simmons, J. P., Nelson, L. N., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359-1366.

Simons, D. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*(1), 76-80.

Simons, D. J., & Holcombe, A. O. (2014). Registered Replication Reports. *The Observer, 27*. Available at: https://www.psychologicalscience.org/observer/ registered-replication-reports

Simons, D. J., Holcombe, A. O., and Spellman, B. A. (2014). An introduction to registered replication reports at Perspectives on Psychological Science. *Perspectives on Psychological Science, 9,* 552–555.

Simonson, I. (1990). The effect of purchase quantity and timing on variety seeking behavior. *Journal of Marketing Research, 32*, 150-162.

Simpson, A., & Rios, K. (2017). The moral contents of anti‐atheist prejudice (and why atheists should care about it). *European Journal of Social Psychology, 47*(4), 501-508.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813–838.

Smith, C. E., & Rizzo, M. T. (2017). Children’s confession- and lying-related emotion expectancies: Developmental differences and connections to parent-reported confession behavior. *Journal of Experimental Child Psychology, 156,* 113-128.

Smyth, J. M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology, 66,* 174-178.

Sober, E., & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, MA: Harvard University Press.

Solomatina, I., & Austers, I. (2014). Collective Guilt Makes Conflicting Parties More Collaborative: Quasi-experimental Study of the Israeli-Palestinian Conflict. *Psychology Research, 4*(4), 271-284.

Solomon, M. R., & Schopler, J. (1982). Self-consciousness and clothing. *Personality and Social Psychology Bulletin, 8,* 508-514.

Sosis, R. (2005). Does religion promote trust? The role of signalling, reputation, and punishment. *Interdisciplinary Journal of Research on Religion, 1*, 1–30.

Sparks, A., & Barclay, P. (2013). Eye images increase generosity, but not for long: The limited effect of a false cue. *Evolution and Human Behavior, 34,* 317–322.

Sparks, A., & Barclay, P. (2015). No effect on condemnation of short or long exposure to eye images. *Letters on Evolutionary Behavioral Science, 6,* 13–16.

Spellman, B. (2015). A short (personal) future history of revolution 2.0*. Perspectives on Psychological Science, 10*(6) 886–899.

Sperber, D., & Baumard, N. (2012). Moral Reputation: An Evolutionary and Cognitive Perspective. *Mind and Language, 27*(5), 495-518.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd edition). Oxford, UK: Blackwell.

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable over time. *Judgment and Decision Making, 13*(3), 260-267.

Stark, R. (1996). Religion as context: Hellfire and delinquency one more time. *Sociology of Religion*, *57,* 163-73.

Stark, R. (2002). Physiology and faith: Addressing the “universal” gender difference in religious commitment. *Journal for the Scientific Study of Religion, 41*(3), 495-507.

Staub, E. (1997). Blind versus constructive patriotism: Moving from embeddedness in the group to critical loyalty and action. In D. Bar-Tal & E. Staub (Eds.), *Nelson-Hall series in psychology. Patriotism: In the lives of individuals and nations* (pp. 213-228). Chicago, IL: Nelson-Hall Publishers.

Steele, C. M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21,* 261-302.

Stein, T., & Sterzer, P. (2011). High-level face shape adaptation depends on visual awareness: Evidence from continuous flash suppression. *Journal of Vision, 11*(8), 1–14.

Sternglanz, W. R. (2010). Exoneration of Serious Wrongdoing via Confessing to a Lesser Offense. In M.S. McGlone & M.L. Knapp (Eds.), *The Interplay of Truth and Deception* (pp 165-193). New York, NY: Routledge.

Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10,* 479–491.

Stone, J., Wiegand, A. W., Cooper, J., & Aronson, E. (1997). When exemplification fails: hypocrisy and the motive for self-integrity. *Journal of Personality and Social Psychology, 72*(1), 54-65.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*, 59–71.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition, 131,*159–171.

Sullins, D. P. (2006). Gender and religion: Deconstructing universality, constructing complexity. *American Journal of Sociology,* *112*(3), 838-880.

Sunstein, C. (2005). Moral heuristic. *Behavioral and Brain Science, 28,* 531–573.

Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS One, 6*(3), e16836.

Suter, R. S. & Hertwig, R. (2011). Time and moral judgment. *Cognition 119,* 454–458.

Swann, W. B., Buhrmester, M., Gómez, A ., Jetten, J., Bastian, B., Vázquez, A., ... Zhang, A. (2014). What makes a group worth dying for? Identity fusion fosters perception of familial ties, promoting self-sacrifice. *Journal of Personality and Social Psychology, 106,* 912–926.

Swann, W. B., Gómez, A., Buhrmester, M. D., López-Rodríguez, L., Jiménez, J., & Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect, cognition, and moral decision-making. *Journal of Personality and Social Psychology, 106,* 713–727.

Swann, W. B., Gómez, A., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behaviour. *Journal of Personality and Social Psychology, 96*(5), 995-1011.

Swann, W. B., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review, 119*(3), 441-456.

Swanson, R. N. (2007). *Indulgences in Late Medieval England: passports to paradise?* Cambridge, UK: Cambridge University Press.

Swarns, R. L. (2016, April 16). *272 slaves were sold to save Georgetown. What does it owe their descendants?* New York Times. Retrieved from https://www.nytimes.com/2016/04/17/us/georgetown-university-search-for-slave-descendants.html?\_r=0)

Świątkowski, W., & Dompnier, B. (2017). Replicability crisis in social psychology: Looking at the past to find new pathways for the future. *International Review of Social Psychology, 30*(1), 111–124.

Swim, J. K., & Miller, D. L. (1999). White guilt: Its antecedents and consequences for attitudes toward affirmative action. *Personality and Social Psychology Bulletin, 25*(4), 500-514.

Sznycer, D., Schniter, E., Tooby, J., & Cosmides, L. (2015). Regulatory adaptations for delivering information: the case of confession. *Evolution and Human Behavior, 36*(1), 44-51.

Tajadura-Jiménez, A., Grehl, S., & Tsakiris, M. (2012). The other in me: Interpersonal multisensory stimulation changes the mental representation of the self. *PLoS ONE, 7*(7), e40682.

Tajfel, H. (1981). Social stereotypes and social groups. In J. C. Turner & H. Giles (Eds.), *Intergroup behavior* (pp. 144 –167). Oxford, UK: Blackwell.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*, 1–39.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Pacific Grove, CA: Brooks/Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In W. G. Autin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks- Cole.

Takagishi, H., Fujii, T., Koizumi, M., Schug, J., Nakamura, F., & Kameshima, S. (2015). The development of the effect of peer monitoring on generosity differs among elementary school-age boys and girls. *Frontiers in Psychology, 6,* 895.

Tan, J. H. W., & Vogel, C. (2008). Religion and trust: An experimental study. *Journal of Economic Psychology, 29,* 832– 848.

Tangney, J. P., & Fischer, K. W. (1995). *Self-conscious emotions*. New York, NY: Guilford.

Tappin, B. M., & McKay. R. T. (2017). The illusion of moral superiority. *Social Psychological and Personality Science, 8*(6), 623- 631.

Taylor, S. E., & Brown, J. D. (1994). Positive well-being and illusions revisited separating fact from fiction. *Psychological Bulletin, 116*(1),21-27.

Tennie, C., & Moore, R. (2018, June 13). A proposal for systematic replication in Psychology. [blog post]. Retrieved from https://zoneoflatentsolutions.wordpress.com/2018/06/13/a-proposal-for-systematic-replication-in-psychology/

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals.Journal of Personality and Social Psychology, 78(5), 853-870.

Thai, M. T., Hornsey, M. J., & Barlow, F. K. (2016). Friends with moral credentials: Minority friendships reduce attributions of racism for majority group members who make conceivably racist statements. *Social Psychology and Personality Science, 7*(3), 272-280.

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, 1-9.

Thompson, E, H., Jr. (1991). Beneath the status characteristic: Gender variations in religiousness. *Journal for the Scientific Study of Religion, 30*(4), 382-394.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis. *Journal of Clinical Epidemiology, 53*(2), 207-216.

Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioural energy conservation campaign. *Energy Policy, 57,* 160-171.

Titchener, E. (1898). The Feeling of Being Watched. *Science, 8*(208), 895-897.

Tuk, M. A., Zhang, K., & Sweldens, S. (2015). The propagation of self-control: Self-control in one domain simultaneously improves self-control in other domains. *Journal of Experimental Psychology: General, 144*,639–654.

Unnever, J. D., & Cullen, F. T. (2006). Christian fundamentalism and support for capital punishment. *Journal of Research in Crime and Delinquency*, *43*, 169-197.

Unnever, J. D., Cullen, F. T., & Bartkowski, J. P. (2006). Images of God and capital punishment: Does a close relationship with a loving God matter? *Criminology,* *44*(4), 835-866.

Urban, J., Bahník, Š., & Kohlová, M. B. (2017). Green consumption does not make people cheat: Three replications of a moral licensing experiment. [Preprint]. Retrieved from https.//doi.org/10/17605/OSF.10?WYNJB

Utikal, V. (2015). A fault confessed is half redressed- Confession and punishment. *Journal of Economic Behavior and Organization, 81*(1), 314-327.

Valdesolo, P., & Graham, J. (2014). Corrigendum. *Psychological Science*, *25*(3), 844–844.

Van Bavel, J. (2016, May 28). Why do so many studies fail to replicate? *The New York Times.* Retrieved from https://www.nytimes.com/2016/05/29/opinion/sunday/why-do-so-many-studies-fail-to-replicate.html

Van Bommel, M., Van Prooijen, J. W., Elffers, H., & Van Lange, P. A. M. (2012). Be aware to care: Public self-awareness leads to a reversal of the bystander effect. *Journal of Experimental Social Psychology, 48,* 926-930*.*

Van Cappellen, P., Saroglou, V., & Toth-Gautier, M. (2016). Religiosity and Prosocial Behavior Among Churchgoers: Exploring Underlying Mechanisms. *The International Journal for the Psychology of Religion, 26*(1), 19-30.

van Elk, M. (2013). Paranormal believers are more prone to illusory agency detection than skeptics. *Consciousness and Cognition, 22,* 1041–1046.

van Elk, M. (2017). The self-attribution bias and paranormal beliefs. *Consciousness and Cognition, 49*, 313-321.

Van Elk, M., & Aleman, A. (2017). Brain mechanisms in religion and spirituality: An integrative predictive processing framework. *Neuroscience & Biobehavioral Reviews*,*73*, 359-378.

Van Elk, M., Karinen, A., Specker, E., Stamkou, E., & Baas, M. (2016). ‘Standing in awe’: The effects of awe on body perception and the relation with absorption. *Collabra*, *2*(1), 4.

van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., & Wagenmakers, E. J. (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology, 6,* 1365.

van Elk, M., Rowatt, W., & Streib, H. (2017).Good dog, bad dog: Introducing open science badges. *International Journal for the Psychology of Religion, 28*(1), 1-2.

Van Elk, M., Rutjens, B. T., Van der Pligt, J., & Van Harreveld, F. (2014). Priming of supernatural agent concepts and agency detection. *Religion, Brain & Behavior*, *6*(1), 4-33.

Van Lange, P. A., & Sedikides, C. (1998). Being more honest but not necessarily more intelligent than others: Generality and explanations for the Muhammad Ali effect. *European Journal of Social Psychology, 28,* 675–680.

Van Leeuwen, N., & Van Elk, M. (2018). Seeking the supernatural: the Interactive Religious Experience Model. *Religion, Brain & Behavior*, 1-31.

Van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychological Medicine, 39,* 179–195.

van Oyen-Witvliet, C., Ludwig, T. E., & Bauer, D. J. (2002). Please forgive me: Transgressors’ emotions and physiology during imaginary of seeking forgiveness and victim responses. *Journal of Psychology and Christianity, 21,* 219- 233.

Van Pachterbeke, M., Freyer, C., & Saroglou, V. (2011). When authoritarianism meets religion: Sacrificing others in the name of abstract deontology. *European Journal of Social Psychology, 41*(7), 898-903.

van Rompay, T. J. L., Vonk, D. J., & Fransen, M. L. (2009). The eye of the camera: Effects of security cameras on prosocial behavior. *Environment and Behavior, 41*(1), 60–74.

Van Vugt, M., & Hardy, C. L. (2009). Cooperation for reputation: Wasteful contributions as costly signals in public goods. *Group Processes & Intergroup Relations, 13*, 101–111.

van’t Veer & Giner-Sorolla, R. (2016). Pre-registration in social psychology-A discussion and suggested template. *Journal of Experimental Social Psychology, 76,* 2-12.

Verkaik, R. (2004, Sept 21). *Criminals offered shorter sentences in return for guilty plea.* The Independent. Retrieved from https://www.independent.co.uk/news/uk/crime/criminals-offered-shorter-sentences-in-return-for-guilty-plea-547076.html

Vitello, P. (2009, February 10). For Catholics, a door to absolution is reopened*.* *New York Times.* Retrieved from https://archive.nytimes.com/query.nytimes.com/gst/fullpage-9D03E0D61F38F933A25751C0A96F9C8B63.html

Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2017). Death Before Dishonor: Incurring Costs to Protect Moral Reputation. *Social Psychological and Personality Science*, 1-10.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., Kievit, R., and van der Maas, H. L. J. (2015). A skeptical eye on psi. In E. May and S. Marwaha (Eds.), *Extrasensory Perception: Support, Skepticism, and Science*, Santa Barbara, CA: ABC-CLIO.

Wallace. S., Coleman, M., Pascalis, O., & Bailey, A. (2006). A study of impaired judgment of eye-gaze direction and related face-processing deficits in autism spectrum disorders. *Perception, 35,* 1651–1664.

Wallington, S. A. (1973). Consequences of transgression: Self-punishment and depression. *Journal of Personality and Social Psychology, 28*, 1–7.

Walter, T., & Davie, G. (1998). The religiosity of women in the modern west. *British Journal of Sociology, 49*(4), 640-660.

Walter, N., Demetriades, S. Z., & Murphy, S. T. (2017). Just a Spoonful of Sugar Helps the Messages Go Down: Using Stories and Vicarious Self-Affirmation to Reduce e-Cigarette Use. Health Communication, 1-9.

Wann, D. L., & Branscombe, N. R. (1990). Die-Hard and Fair-Weather Fans: Effects of Identification on BIRGing and CORFing Tendencies. *Journal of Sport and Social Issues, 14*(2), 103-117.

Washburn, M., & Off, G. (2014, September 19) *Feeling guilty, N.C. man admits to desert murder*. Charlotte Observer. Retrieved from http://www.charlotteobserver.com/news/local/article9193550.html#.VCMk3xbp-3w

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, 47*, 1063-1070.

Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., & Gray, R. D. (2014). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society B 282*, 20142556.

Weiner, B., Graham, S., Peter, O., & Zmuidinas, M. (1991). Public confession and forgiveness. *Journal of Personality, 59*, 281–312.

Welten, S. C., Zeelenberg, M., & Breugelmans, S. M. (2012). Vicarious shame. *Cognition and Emotion, 26*(5), 836–846.

Wertheim, E. H., & Schwartz, J. C. (1983). Depression, guilt, and self-management of pleasant and unpleasant events*. Journal of Personality and Social Psychology, 45,* 884-889.

West, C., & Zhong, C. B. (2015). Moral cleansing. *Current Opinion in Psychology, 6,* 221-225*.*

Whitley, B. E., Jr. (2009). Religiosity and attitudes toward lesbians and gay men: A Meta-analysis. *International Journal for the Psychology of Religion, 19*(1), 21-38.

Wicklund, R. A., & Gollwitzer, P. M. (1982). *Symbolic Self-completion.* Hillsdale, NJ: Erlbaum.

Willard, A. K. (2018). Religion and prosocial behavior among the Indo-Fijians. *Religion, Brain and Behavior, 8*(2), 227-242.

Willard, A. K. & Norenzayan, A. (2013) Cognitive biases explain religious belief, paranormal belief, and belief in life’s purpose. *Cognition 129*(2): 379–391.

Williams, C., & Bybee, J. (1994). What do children feel guilty about? Developmental and gender differences. *Developmental Psychology, 30*(5), 617-623.

Wilson, D. S. (2002). *Darwin’s cathedral: Evolution, religion and the nature of society.* Chicago, IL: University of Chicago Press.

Williams, K. D., & Nida, S. A. (2011). Ostracism: Consequences and coping. *Current Directions in Psychological Science, 29,* 71-75.

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58,* 425-452.

Wright, P. (1974). The harassed decision maker: time pressures, distractions, and the use of evidence. *Journal of Applied Psychology, 59,* 555.

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Reputation, gossip, and human cooperation. *Social and Personality Psychology Compass 10*(6), 350–364.

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P., Middleton, K. R., & Wing, R. R. (2014). Failure to replicate depletion of self-control*. PLoS ONE, 9,*e109950.

Xygalatas, D. (2013). Effects of religious setting on cooperative behavior: A case study from Mauritius. *Religion, Brain and Behavior, 3,* 91–102.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*(1), 110-116.

Yao, M. Z., & Flanagin, A. J. (2006). A self-awareness approach to computer mediated communication. *Computers in Human Behavior, 22*(3), 518–544.

Yilmaz, O., & Bahçekapili. H. G. (2016). Supernatural and secular monitors promote human cooperation only if they remind of punishment. *Evolution and Human Behavior,* 37(1), 79-84.

Yokoyama, T., Sakai, H., Noguchi, Y., & Kita, S. (2014). Perception of direct gaze does not require focus of attention. *Scientific Reports, 4*(3858), 1-7.

Young, K. A., Gandevia, S. C., & Giummarra, M. J. (2017). Vicarious pain responders and emotion: Evidence for distress rather than mimicry. *Psychophysiology, 54*(7), 1081-1095.

Young, R. L. (1992). Religious Orientation, Race and Support for the Death Penalty, Journal for the Scientific Study of Religion, 31(1), 76-87.

Yu, H., Duan, Y., & Zhou, X. (2017). Guilt in the eyes: Eye movement and physiological evidence for guilt-induced social avoidance. *Journal of Experimental Social Psychology, 71,* 128-137.

Zaleskiewicz, T., Gasiorowska, A., & Kesebir, P. (2015). The Scrooge effect revisited: Mortality salience increases the satisfaction derived from prosocial behavior. *Journal of Experimental Social Psychology, 59*, 67-76.

Zhong, C. B., Ku, G., Lount, R. B., & Murnighan, J. K. (2010). Compensatory ethics. *Journal of Business Ethics, 92,* 323–339.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*(8), 1451–1452.

Zhong, C. B., Liljenquist, K., & Cain, D. M. (2009). Moral self-regulation. Licensing and compensation. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 75-89). Charlotte, NC: Information Age Publishing.

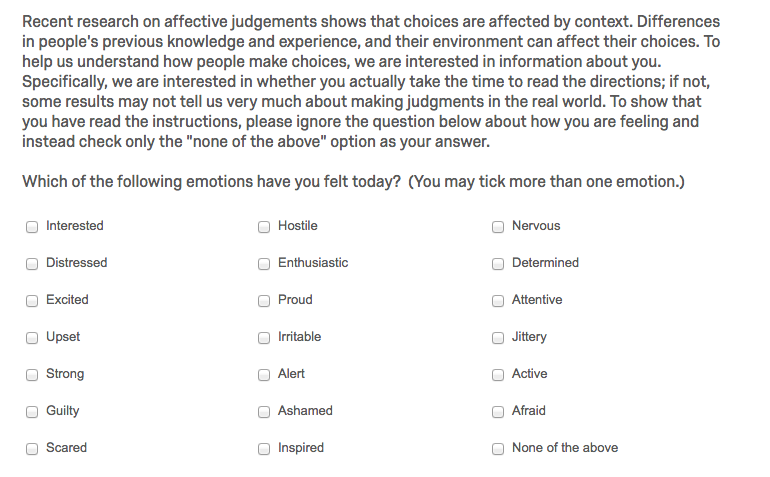
Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 493-504.

Zwaan, R., Etz, A., Lucas, R., & Donnellan, M. (2017). Making replication mainstream. *Behavioral and Brain Sciences.* 1-50. https://doi.org/10.1017/S0140525X17001972

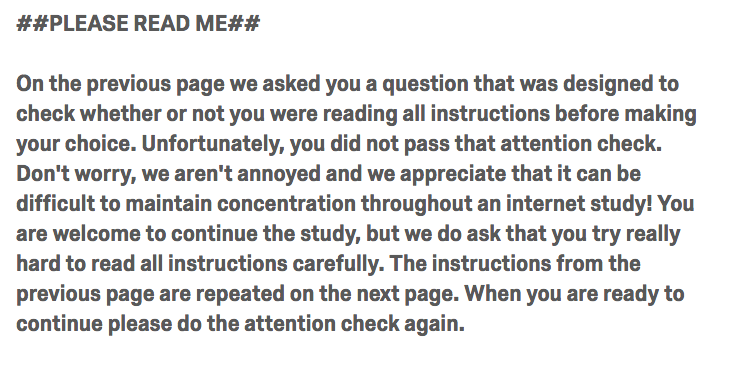
# Appendices.

## Appendix A: Recommended safeguards against participant inattention on MTurk.

**i)** Attention check used in online studies (Studies 3, 4, 6, 7, 8, 9, 10, and 11).

**

**ii)** Reminder presented to inattentive MTurk participants.

**

## Appendix B: Morality-purity metaphors.

The following list comprises of some examples of morality-purity metaphors in natural language.

**Zmazani lažov** (*Croatian*)

Literal translation: dirty liar

Meaning: someone who lies a lot

**Tamna mrlja u njegovoj proslosti** (*Croatian*)

Literal translation: a dark stain in his past

Meaning: a previous moral transgression

**Dreck am Stecken haben** (*German*)

Literal translation: having dirt on your (walking) cane

Meaning: somebody is hiding something but pretends to be good.

**Sporcarsi le mani** (*Italian*)

Literal translation: get your hands dirty

Meaning: something that has to be done but requires some kind of moral violation or compromise.

**Bianco immacolato** (*Italian*)

Literal translation: immaculate white

Meaning: extremely pure

**Nieczyste myśli** (*Polish*)

Literal translation: unclean thoughts

Meaning: dishonest/bad thoughts

**Zgnilizna moralna** (*Polish*)

Literal meaning: moral rot

Meaning: a state of immorality or moral decline

**Blanco igual papel** (*Portuguese*)

Literal translation: White as a sheet of paper

Meaning: honest/innocent

**Má špinu za ušami** (*Slovakian*)

Literal translation: he has dirt behind his ears

Meaning: he’s guilty

**Má čisty štít** (*Slovakian*)

Literal translation: his shield is clean

Meaning: he’s innocent/honourable

## Appendix C: Religious affiliation measure.

What is your religious affiliation?

* 1. Christian (Catholic)
  2. Christian (Church of England / Protestant)
  3. Christian (Other)
  4. Hindu
  5. Buddhist
  6. Muslim
  7. Jewish
  8. Sikh
  9. None
  10. Atheist
  11. Agnostic
  12. Other (Please specify): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_)

## Appendix D: Positive and negative affect scale (Watson et al., 1988).

This scale consists of a number of words that describe different feelings and emotions. Read each item and then list the number from the scale below next to each word.

***Indicate to what extent you feel this way right now, that is, at the present moment.***

**1                                    2                          3                           4                                5**

 Very slightly or   A little              Moderately         Quite a bit              Extremely

 Not at all

*\_\_\_\_\_\_\_\_\_\_*1. Interested                                       \_\_\_\_\_\_\_\_\_\_ 11. Irritable

\_\_\_\_\_\_\_\_\_\_ 2. Distressed                                       \_\_\_\_\_\_\_\_\_\_ 12. Alert

\_\_\_\_\_\_\_\_\_\_ 3. Excited                                           \_\_\_\_\_\_\_\_\_\_ 13. Ashamed

\_\_\_\_\_\_\_\_\_\_ 4. Upset                                             ­­­ \_\_\_\_\_\_\_\_\_\_14. Inspired

\_\_\_\_\_\_\_\_\_\_ 5. Strong                                            \_\_\_\_\_\_\_\_\_\_15. Nervous

 \_\_\_\_\_\_\_\_\_\_6. Guilty                                           \_\_\_\_\_\_\_\_\_\_16. Determined

\_\_\_\_\_\_\_\_\_\_ 7. Scared                                            \_\_\_\_\_\_\_\_\_\_ 17. Attentive

\_\_\_\_\_\_\_\_\_\_ 8. Hostile                                          \_\_\_\_\_\_\_\_\_\_ 18. Jittery

\_\_\_\_\_\_\_\_\_\_ 9. Enthusiastic                                   \_\_\_\_\_\_\_\_\_\_19. Active

\_\_\_\_\_\_\_\_\_\_10.Proud                                          \_\_\_\_\_\_\_\_\_\_ 20. Afraid

## Appendix E: False confession narrative presented to participants in Study 1.

*Two summers ago, I worked in a cloakroom for a local arts festival. It was only a six-week contract and was a good job to have over the summer break. I was responsible for checking in people’s coats and bags as they arrived at the venue. As the visitors handed over their belongings, I would give them a token with a printed number that matched the number on the voucher I then clipped on the item.*

*On my last day, my colleague asked to go on lunch early- leaving only me in the cloakroom. Normally there were always at least two people working in there. I remember hearing a thud and when I turned around I saw that one of the visitor’s bags had fallen off the shelf. I went to pick it up and noticed that the main section of the bag was unzipped. I could see a wallet. I opened it up and saw that there were two £20 notes in it. I took the £40 and stuffed the notes in my back pocket before putting the wallet back, zipping up the bag and replacing it on the shelf. My colleague returned ten minutes later and I finished my shift. I collected my bag and left, still with the stolen money in my pocket.*

## Appendix F: Recall instructions for participants in the false confession condition (Study 1).

Please read this narrative carefully and familiarise yourself with it. You will be asked to repeat this narrative out loud to another participant from memory. You do not have to memorise this narrative word for word but please ensure that you

include the following four points in your recall:

* **That you worked at in a cloakroom at a summer arts festival checking in people’s bags**
* **You were left alone in the cloakroom and saw an open bag with money inside (£40)**
* **You took the £40 and put it in your pocket**
* **You left without telling anyone what you’d done.**

## Appendix G: Control narrative presented to participants in Study 1.

*Two summers ago, I worked in a cloakroom for a local arts festival. It was only a six-week contract and was a good job to have over the summer break. I was responsible for checking in people’s coats and bags as they arrived at the venue. As the visitors handed over their belongings, I would give them a token with a printed number that matched the number on the voucher I then clipped on the item.*

*On my last day, my colleague asked to go on lunch early- leaving only me in the cloakroom. Normally there were always at least two people working in there. I remember hearing a thud and when I turned around I saw that one of the visitor’s bags had fallen off the shelf. I went to pick it up and then out it back in its correct position on shelf. My colleague returned ten minutes later and I finished my shift. I collected my bag and left.*

## Appendix H: Recall instructions for participants in the control condition (Study 1).

Please read this narrative carefully and familiarise yourself with it. You will be asked to repeat this narrative out loud to another participant from memory. You do not have to memorise this narrative word for word but please ensure that you include the following four points in your recall:

* **That you worked at in a cloakroom at a summer arts festival checking in people’s bags**
* **You were left alone in the cloakroom and saw a bag had fallen on the floor**
* **You picked up the bag and put it back on the shelf**
* **You continued as normal until your colleague returned**

## Appendix I: Examples of portraits presented to participants in Studies 2-4.

**i)** Examples of ambiguous eye gaze.

**

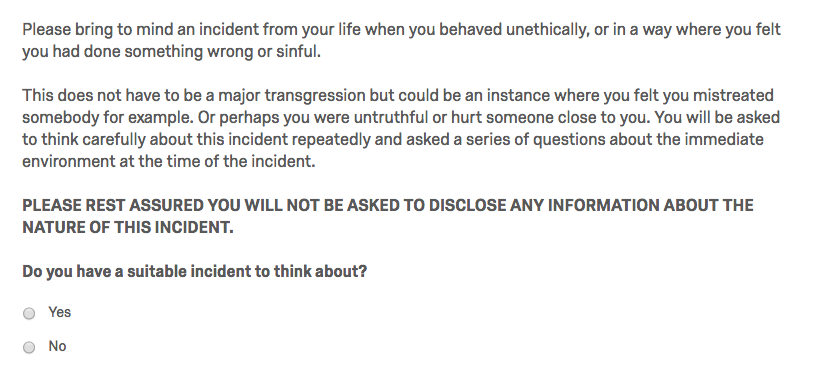
**ii)** Examples of direct eye gaze.



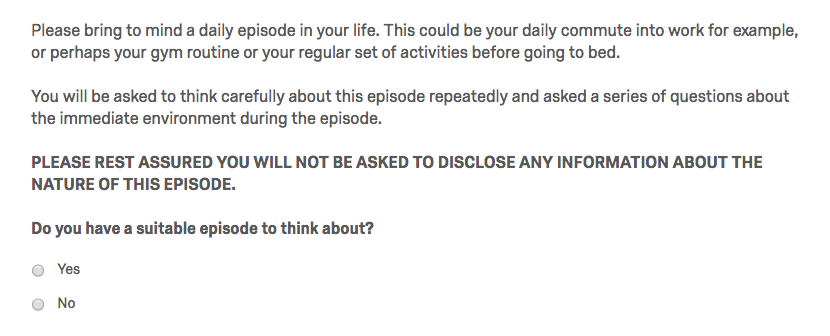
**iii)** Examples of averted eye gaze.

**

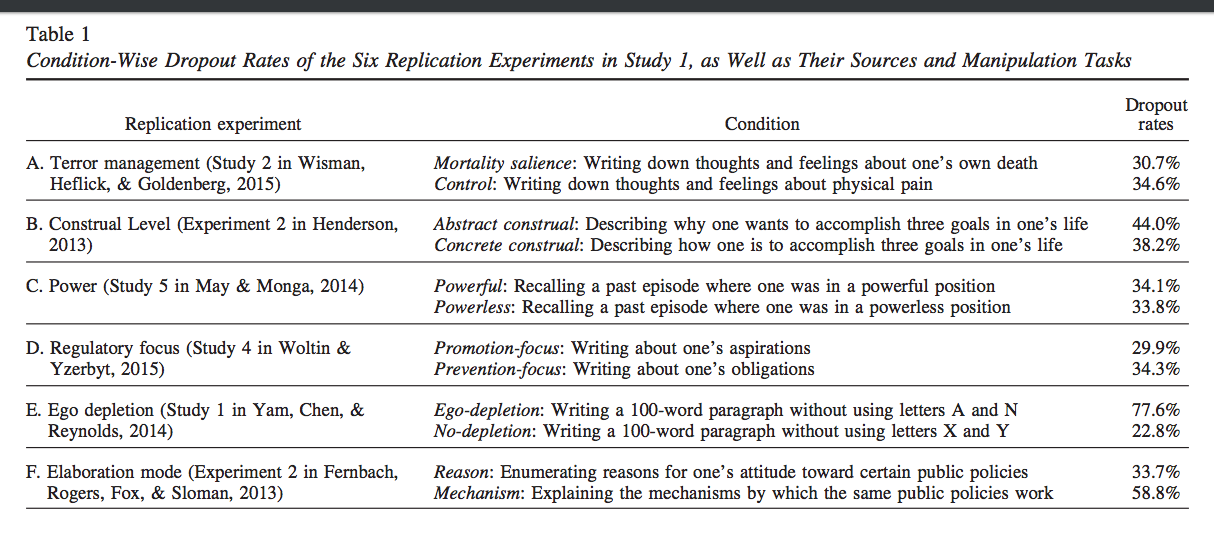
## Appendix J: Transgression recall instructions presented to participants in Study 4.



## Appendix K: Neutral recall instructions presented to participants in Study 4.



## Appendix L: The set of studies included in Zhou and Fishbach (2016).

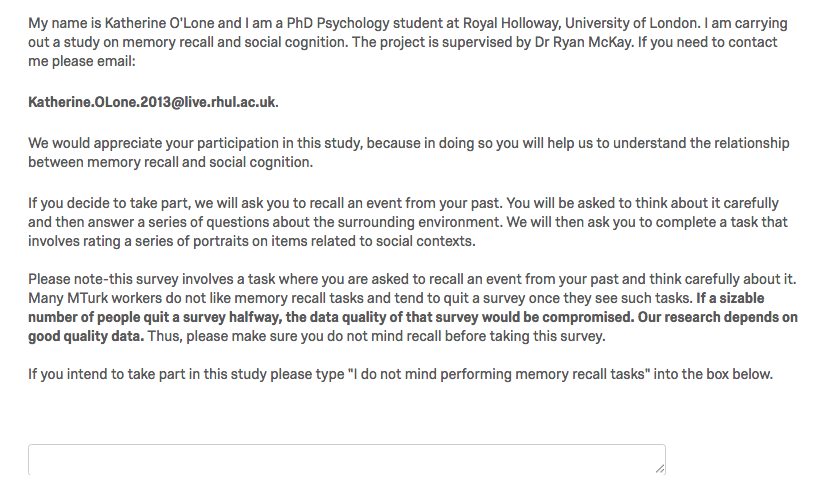
**

## Appendix M: (high/low) power prime (taken from Study 5, May & Monga, 2014).

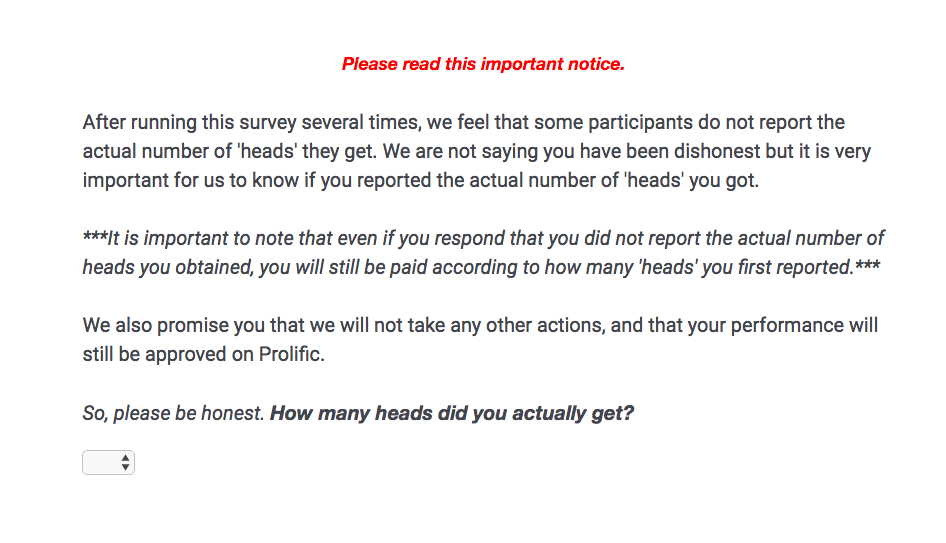
High-power: “Please recall a particular incident in which you had power over another individual or individuals. By power, we mean a situation in which you controlled the ability of another person or persons to get something they wanted, or were in a position to evaluate those individuals. Please describe this situation in which you had power—what happened, how you felt, etc.”

Low-power: “Please recall a particular incident in which someone else had power over you. By power, we mean a situation in which someone had control over your ability to get something you wanted, or was in a position to evaluate you. Please describe this situation in which you did not have power—what happened, how you felt, etc.”

## Appendix N: Information and Consent form shown to participants on MTurk at the start of Study 4.



## Appendix O: Confession probe presented to participants in Study 5.



## Appendix P: Forgiving God prime (Studies 6 & 7).

There was a rich man in the town that dressed in fine clothes and lived in luxury every day.  At his gate the poor of the village came to beg for food, desperate to feed their starving children. He would turn them away scornfully and lived his life apart from God. The time came when the rich man died. His coffin was carried through the town and the people lined the streets to see. A young boy turned to his father and asked what would become of the rich man’s soul. His father replied “Our Lord is merciful and forgiving, my son”.

## Appendix Q: Punishing God prime (Studies 6 & 7).

There was a rich man in the town that dressed in fine clothes and lived in luxury every day.  At his gate the poor of the village came to beg for food, desperate to feed their starving children. He would turn them away scornfully and lived his life apart from God. The time came when the rich man died. His coffin was carried through the town and the people lined the streets to see. A young boy turned to his father and asked what would become of the rich man’s soul. His father replied, “God will punish the world for its evil, and the wicked for their sins, my son”.

## Appendix R: Original punishment scenario shown to participants in Study 6 (taken from Carlsmith et al, 2002).

John was a corporate executive who worked in a section of his international corporation that dealt with currency conversion. The section was responsible for converting local currencies to U.S. dollars to be returned as profits to the company headquarters. The system worked by tracking what are called “bands” of conversions ratios on a daily basis for each conversion, and making the conversions at a rate somewhere within those bands. John needed money, and he arranged to cheat on the system to get it. He made the conversions at a slightly below-average level in the daily band. He did this to create small surpluses of money that he then kept. This occurred before the profits were entered on the corporate books. The reason he needed money was for debts he had run up with loan sharks. He had a habit of betting on football games and had started to run losses. To keep betting he borrowed from these loan sharks, while continuing to maintain his extravagant style of living.

## Appendix S: Victim-directed punishment scenario (Study 7).

John is a handyman who lives and works in a medium-sized town in the U.S. He is well known in the local area and has for many years been employed by residents, especially the more elderly ones. They trust him because they know him well and as a result he is regularly employed to carry out various repairs on their homes that they can't do themselves. John needs money and steals from his elderly clients to get it. He has figured out that they are the most likely to keep large sums of cash in their homes rather than in bank accounts. Once he's employed by them, he can then easily access this cash while he's working in their homes and steal it.

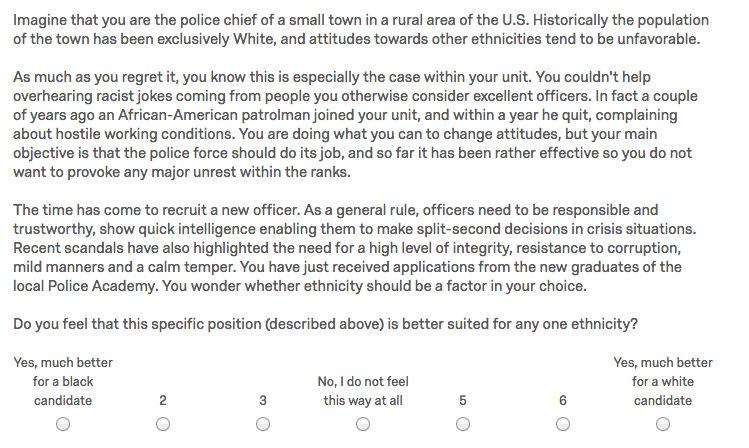
## Appendix T: Forgiving God scale presented to participants (Studies 8 & 9).

All items are measured on the following scale:

(1) Strongly disagree, (2) Somewhat disagree, (3) Neither agree nor disagree, (4) Somewhat agree, (5) Strongly agree.

1. I see God as merciful
2. I believe people will be forgiven for what they have done
3. I see God as a harsh judge
4. I believe God is all forgiving
5. Despite their shortcomings, I believe people will be forgiven
6. Some people will go to hell when this life is over
7. I think God will punish people for what they have done
8. God will ultimately condemn those who do wrong

## Appendix U: Job hiring task shown to participants in Studies 10 & 11 (taken from Kouchaki, 2011).

**

## Appendix V: Verbal fusion measure (taken from Gómez et al., 2011).

To what extent do you agree with the following statements? (from 0 = totally disagree, to 6 = totally agree)

1. The USA is me
2. I am one with the USA
3. I feel immersed in the USA
4. I have a deep emotional bond with the USA
5. I am strong because of the USA
6. I’ll do more for the USA than other citizens would do
7. I make the USA strong

## Appendix W: Moral/immoral/neutral in-group information presented to participants (Study 11).

Please read the following:  
A team of international researchers from 15 universities recently conducted a large cross-cultural investigation of people’s moral behaviour. Data were collected from participants in over 40 different countries over a period of five years.

The main aim of the research was to learn more about why people engage in certain moral behaviours (e.g., charitable giving, volunteering etc.) and how they form moral judgments.

The researchers had anticipated some variation across different countries, but this was not found. For example, relative to citizens of other countries, US citizens scored *higher/lower/nor higher nor lower* on measures of morality (such as moral attitudes and engagement in moral behavior).

## Appendix X: Moral/immoral/neutral salient out-group information presented to participants (Study 11).

Please read the following:A team of international researchers from 15 universities recently conducted a large cross-cultural investigation of people’s moral behavior. Data was collected from participants in over 40 different countries over a period of five years.

The main aim of the research was to learn more about why people engage in certain moral behaviors (e.g., charitable giving, volunteering etc.) and how they form moral judgments.

The researchers had anticipated some variation across different countries but noted that, relative to citizens of other countries, citizens from Iran exhibited consistently higher/lower/nor higher nor lower scores on measures of morality (such as moral attitudes and engagement in moral behaviour

1. Variants of this ritual existed across the globe. Dutch settlers to America in the 17th century, for example, brought with them the practise of eating *doed-koecks* at funerals and until the early 20th century “burial-cakes” were made in parts of Northern England and Bavaria, all surviving vestiges of the sin-eating ritual. [↑](#footnote-ref-1)
2. So economised and complex was this system that during the Middle Ages, the Church administered absolution literally like a bank (Swanson, 2007). [↑](#footnote-ref-2)
3. Pope John Paul II authorised bishops to offer plenary indulgences in 2000 to mark the third millennium of the Church (Vitello, 2009). [↑](#footnote-ref-3)
4. Western, Educated, Industrialised, Rich and Democratic (see Henrich, Heine & Norenzayan, 2010). [↑](#footnote-ref-4)
5. e.g., The cognitive reflection task (Frederick, 2005), which was designed to measure the ability to engage in reflective though. Participants must disregard intuitive but incorrect answers. For example, *A bat and a ball cost $110 in total. The bat costs $100 more than the ball. How much does the ball cost?* (Answer: 5 cents). [↑](#footnote-ref-5)
6. i.e., *Journal of Personality and Social Psychology* (JPSP); *Journal of Experimental Psychology: Learning, Memory, and Cognition* (JEP: LMC); *Psychological Science*. [↑](#footnote-ref-6)
7. This is due in part to empirical papers with statistically significant results (i.e., *p <* 0.05 in psychology) being more likely to be published than those with non-significant results or negative findings (Ioannidis, 2005). Non-significant findings are relegated to a “file drawer” never to see the light of day, leading to the literature being flooded with inflated results (i.e., studies reporting only significant findings). [↑](#footnote-ref-7)
8. Experimenter bias takes place when the researcher unconsciously affects the behaviour of participants in an experiment with their expectations for a certain outcome. [↑](#footnote-ref-8)
9. I adopted this practise in the later studies in this thesis by using the online service AsPredicted (https://aspredicted.org/) to pre-register studies (Studies 6-11). [↑](#footnote-ref-9)
10. Here the journal agrees to publish based on the rationale and methods not the outcome of the results. Pre-registration per se is no guarantee of publication. [↑](#footnote-ref-10)
11. There are now several templates and platforms available online to help researchers with pre-registration (e.g., van’t Veer & Giner-Sorolla, 2016; see The Center for Open Science, https://cos.io/blog/tag/preregistration/) [↑](#footnote-ref-11)
12. Several high impact journals (e.g., *Journal of Experimental Social Psychology*, *Journal of Personality and Social Psychology*, *Psychological Science*) now publish both successful and unsuccessful replications. [↑](#footnote-ref-12)
13. In 2016 The Netherlands recently became the first country to set up a national fund (estimated $3.3 million) dedicated to replication studies. The 3-year project is attempting to reproduce important findings from the medical and social sciences (‘NWO makes 3 million available’, 2016). [↑](#footnote-ref-13)
14. In this economic game (Forsythe, Horowitz, Savin & Sefton, 1994) the first player, or “dictator”, decides how to split an allocation of resources, normally a sum of money, between himself and a recipient. The recipient does not play an active role in the game and simply receives the allocation that has been determined by the dictator. [↑](#footnote-ref-14)
15. The 91 studies included in the data set consisted of 83 published studies (that were reported in 22 published or forthcoming journal articles) and 8 unpublished manuscripts. [↑](#footnote-ref-15)
16. Of the 106 studies included in this analysis; 34 were single study outcomes reported by Blanken et al. (2015); 61 single study outcomes from 23 published journal articles and 11 unpublished single study outcomes. [↑](#footnote-ref-16)
17. For example, a survey conducted in April 2018 by the Pew Research Center of 4,700 US adults found that 80% responded they believed in God (though not necessarily the one depicted in the bible) (Pew Research Center, 2018a). In contrast, levels of religious non-affiliation reach up to 48 % in areas of Western Europe (e.g., The Netherlands) (see Pew Research Center, 2018b). Moreover, religiously unaffiliated Americans exhibit nearly equal levels of spirituality as self-identified Christians in Western Europe (Pew Research Centre, 2018) [↑](#footnote-ref-17)
18. Using G\*Power, Blanken et al. (2015) found that 165 participants would be needed in each condition to achieve 80% power to find an effect of *d* = .31. [↑](#footnote-ref-18)
19. Mazar and Zhong (2010) found that purchasing green products acts as a moral license. [↑](#footnote-ref-19)
20. Or, the “Lady Macbeth effect”, as it should be. [↑](#footnote-ref-20)
21. Of these replications attempts only Earp et al. (2014) was a direct replication (of Zhong and Liljenquist, 2006; Study 2); it was not pre-registered. [↑](#footnote-ref-21)
22. As a point of comparison, 91 studies were included in Blanken et al’s meta-analysis on moral licensing (2015) yet only five were included in their earlier meta-analysis on moral cleansing effects (Blanken et al., 2014). [↑](#footnote-ref-22)
23. In response to the faces *vs.* eyes effects, it is true that both contain eyes so one may expect no difference between the two. However, Baillon, Selim & Van Dolder (2013) reported that faces were less effective than eye images alone. Although, as Saunders et al. (2016) point out, Baillon et al. (2013) varied several features of the images such as the orientation and number of faces *vs.* eyes (see Saunders et al., 2016). [↑](#footnote-ref-23)
24. However, erring on the side of caution is indeed wise, but only to a degree. It is certainly better to make fewer costly errors than errors that are less costly – but ensuring that one errs completely on the side of caution is not the best way to minimize overall costs (McKay, Ross, O’Lone & Efferson, 2017). [↑](#footnote-ref-24)
25. However, the characterisation of the HADD as a single, domain-specific cognitive module (McCauley, 2011) is contested (see Van Leeuwen & van Elk, 2018). For example,more than onebrain area is implicated in processing a) facial stimuli (e.g., Bernstein & Yovel, 2015; Collins & Olson, 2014; Parvizi et al., 2012) and b) other types of agentive stimuli (Allen, Larøi, McGuire, & Aleman, 2008; Kanwisher & Yovel, 2006; Scholl & Tremoulet, 2000). [↑](#footnote-ref-25)
26. An indexical is a linguistic expression, which denotes a particular referent (e.g., pronouns such as *I, you, my;* adverbs such as *tomorrow, now, today;* possessive adjectives such as *his, theirs, hers*). The referent of an indexical shifts from context to context (Perry, 1998). How the referent of the indexical is assigned is a matter of some debate in linguistics and philosophy. One prominent model of communication, Relevance Theory, proposes that reference assignment is governed via a process of inference to the best explanation (Sperber & Wilson, 1995). [↑](#footnote-ref-26)
27. Here I rely on van Elk and Aleman’s (2017) distinction between religious and spiritual beliefs as separate phenomena characterised by different fundamental features:

    “Religious beliefs typically refer to the more institutionalized aspects of belief in supernatural beings, such as within traditional religious communities. In contrast, spiritual beliefs refer to the individual and personalized beliefs regarding the transcendent or sacred, which are often based on personal experience rather than tradition.” (van Elk & Aleman, 2017, p. 361). [↑](#footnote-ref-27)
28. Equating morality and prosociality is problematic (see McKay & Whitehouse, 2015). In this thesis I take the view that prosociality is a subcategory of morality (see Preston, Salomon & Ritter, 2014). [↑](#footnote-ref-28)
29. More recently Martin Luther King said: “Our world hinges on moral foundations. God has made it so. God has made the universe to be based on a moral law. So long as man disobeys it he is revolting against God” (Martin Luther King, Jr, 1954). [↑](#footnote-ref-29)
30. The literal meaning of Islam for instance is “submission to God” and believers are required to submit entirely to His will. The notion of *Qadar* exemplifies this and refers to divine destiny where all and everything is pre-determined by the will of God. [↑](#footnote-ref-30)
31. A real-world example of this occurred recently with the case of the profoundly disabled infant Isaiah Haastrup. A judge ruled that his life support machine should be turned off. His deeply religious mother told him "For me, I don't think it is right to say who should live or who should die. If God wants to take the person, He will." (Isaiah Haastrup: ‘God should decide’, 2018) [↑](#footnote-ref-31)
32. Some examples: *“There is only one lawgiver and judge, he who is able to save and destroy. But who are you to judge?”* (James, 4:12); *“But it is God who executes judgement, putting down one and lifting up another”* (Psalm, 75:7); *“Why do you pass judgment on your brother? Or you, why do you despise your brother? For we will all stand before the judgment seat of God”* (Romans, 14:10). [↑](#footnote-ref-32)
33. Miller argues that intelligence, as a heritable dimension of individual variation, has “quasi-moral” status when assessed in social (and sexual) interaction (Miller, 2007). In a culture where intelligence is highly valued this attribute might undergo a process of moralisation (Rozin, 1999) becoming a (quasi) moral trait rather than just a general positive attribute. Therefore, the assumption that licensing effects are not restricted to the moral domain is misleading if information about one’s own or others’ intelligence has moral value. [↑](#footnote-ref-33)
34. Over six decades the German government’s compensation programme has paid $89 billion to Jewish victims of the Nazis. When the issue of whether, after 60 years, the repayments had been sufficient one leader of the programme stated “We will have done enough when no more survivors remain. As long as they live we will uphold our responsibility” (Eddy, November 12th, 2012). [↑](#footnote-ref-34)
35. For example, during the Nuremberg trials (1946) one of the most noted pleas by the defendants was *“Befehl ist befehl”*, (“I was only following orders”), a statement they believed would release them from lawful responsibility of the crimes they were accused of. This statement in the context of a morality using the foundation of Loyalty/Authority would have signalled an essentially moral quality. In the context of the 1946 trials at Nuremberg however, the “superior orders” defence plea was at fundamental odds with the moral foundations of Harm and Fairness that guided the evaluations of the atrocities laid before the court. [↑](#footnote-ref-35)
36. At the time of the original investigation, one man made four separate visits to the LAPD and confessed on each occasion, earning himself the nickname “confessin’ Tom” from detectives (Corwin, 1996). [↑](#footnote-ref-36)
37. The apparent “common sense” of this false belief is so alluring that juries tend to weigh a confession (even if recanted after legal counsel is provided) as the single most compelling piece of evidence. It is the most powerful factor influencing a jury’s decision to vote guilty, even in the absence of other evidence (Kassin & Neumann, 1997).  [↑](#footnote-ref-37)
38. A real-life example of this took place in the infamous case of the “Iceland Six” in 1974. Six people falsely confessed to a double murder due to suffering varying levels of memory distrust syndrome. As it turned out, the two victims had not been murdered, both were, in fact, accidental deaths, and they were unrelated (Abraham, 2017). [↑](#footnote-ref-38)
39. There is a primary-secondary distinction in the psychopathy literature. Unlike primary psychopaths, secondary psychopaths demonstrate (generalised) feelings of guilt and remorse (Hare, 1975). [↑](#footnote-ref-39)
40. *Integral* emotions on the other hand, are those triggered by the current situation (Gino & Schweitzer, 2008). [↑](#footnote-ref-40)
41. For example, patients with a religious subtype of OCD called *Scrupulosity* exploit the cleansing function of confession to relieve pathological guilt. One of the features of this condition is the over-perception of sin and an accompanying compulsion to confess (Abramowitz, Huppert, Cohen, Tolin & Cahill, 2002). [↑](#footnote-ref-41)
42. Mean implicit guilt scores are as follows: Religiously affiliated (*N* = 41), *M* = -.27, *SD* = .35; non-affiliated (*N* = 59), *M* = -.15, *SD* = .22. [↑](#footnote-ref-42)
43. I use the term ‘ethical dissonance’ in the sense defined by Barkan et al. (2012, p. 6): “the inconsistency between one’s unethical behavior and the need to maintain a moral self-image”. [↑](#footnote-ref-43)
44. This is a very similar emotion to guilt, and the two are often confused. Despite their similarities they are distinct emotions, which have both phenomenological differences and divergent behavioural consequences (e.g, a tendency towards reparative behaviours vs. avoidance) (see Tangney & Fischer, 1995). [↑](#footnote-ref-44)
45. The presence of a camera is an established and well-validated manipulation used to induce feelings of being watched that has consistently yielded reliable results (e.g., Alden, Teschuk & Tee, 1992; Davies, 2005; Ellett & Chadwick, 2007; Yao & Flanagin, 2006). [↑](#footnote-ref-45)
46. During the debriefing all participants in the surveillance condition indicated that they believed the camera had been on and recording. [↑](#footnote-ref-46)
47. Van Elk (2013) found that traditional religious beliefs (and belief in witchcraft) were unrelated to illusory agency detection. However *paranormal* beliefs were strongly associated with illusory agency detection (see also Krummencher, Mohr, Haker & Bruger, 2010; Riekki et al., 2013). The finding that beliefs in witchcraft were unrelated to illusory agency detection but paranormal beliefs were is surprising given that beliefs in witchcraft belong to a set of beliefs that one would label as paranormal; “Paranormal beliefs refer to an eclectic range of New Age beliefs and practices, involving belief in Psi, precognition, witchcraft, superstition, as well as telekinesis and channeling” (van Elk, 2017, p. 313). However, the distinction between paranormal and supernatural (and religious) beliefs is, at best, fuzzy and Lindeman and Svedholm conclude that there is no reason to separate the concepts *paranormal, supernatural, superstitious and magical* (Lindeman & Sveldholm, 2012). [↑](#footnote-ref-47)
48. This equates to an attrition rate of 27%. According to Zhou and Fishbach (2016) as a general rule anything higher than 20% is potentially problematic for internal validity (see Zhou & Fishbach, 2016). This issue is addressed in the “Limitations” section of this chapter. [↑](#footnote-ref-48)
49. Transgression recall condition (*M* = 98.87, *SD* = 17.38); Neutral condition (*M* = 95.04, *SD* = 18.79) [↑](#footnote-ref-49)
50. This figure is large but in line with general rates of attrition for MTurk studies (31.9% -51%) (Zhou & Fishbach, 2016). [↑](#footnote-ref-50)
51. Ego depletion effects have come under scrutiny during the “replication crisis” (see Converse & DeShon, 2009; Tuk, Zwang & Sweldens, 2014; Xu et al., 2014). A recent meta-analysis of the ego depletion effect concluded that a probable value for the effect was zero and revealed evidence of a *small study* bias. This suggests that the published studies included in the meta-analysis were underpowered. This implies that the likelihood of finding so many large significant effects was unlikely and improbable (Carter, Kofler, Forster & McCullough, 2015). A later meta-analysis by Hagger et al. (2016) revealed that the ego-depletion effect was small with confidence intervals that included zero, *d* = 0.04, 95% CI [-0.07, 0.15]. [↑](#footnote-ref-51)
52. Zhou and Fishbach (2016) found that a mix of these safeguards did reduce attrition rates but it still remained at 20% (Study 4, Zhou & Fishbach, 2016). [↑](#footnote-ref-52)
53. None of the six source studies that Zhou and Fishbach (2016) replicated reported any information about participant attrition. [↑](#footnote-ref-53)
54. This urge starts early in life. In a parental report study by Kochanska, DeVet, Goldman, Murray, & Putnam (1994) 67% of mothers reported that their child had confessed to a wrongdoing; among 3-year-olds the confession rate rose to 95% (Kochanska et al., 1994).  [↑](#footnote-ref-54)
55. This association is reflected in scripture too, Weiner et al. (1991) write, “It is clearly stated that confession is the sine qua non condition for divine pardon. In the Saint John affirmation of the divine, it is written: "*If we confess our sins. He is faithful and righteous to forgive us"* (I John, 1)” (p. 283). [↑](#footnote-ref-55)
56. In a sample of 631 US police officers 92% stated they present false evidence to suspects (including telling them there are witnesses) at least some of the time (Kassin et al., 2007). [↑](#footnote-ref-56)
57. This is not the case in the UK however as police are not allowed to lie to suspects (Gudjonsson, 2003). [↑](#footnote-ref-57)
58. However, Prolific Academic users have also demonstrated less dishonesty than MTurk workers (Pe’er et al., 2017) rendering it something of a double-edged sword; participants on this platform might not cheat *enough*. On balance it seemed preferable for participants to believe the coin-toss task was genuine. [↑](#footnote-ref-58)
59. Pe’er et al. (2014) found that 18.79% - 24.60% of cheaters confessed (Studies 1 & 3 respectively). [↑](#footnote-ref-59)
60. In both Study 5 and Pe’er et al. (2014) participants were explicitly told they would not be penalised for over-reporting, either by deducting money or by receiving bad feedback from the experimenter on the research platform. Therefore, there was no credible fear of sanction or reputational threat that is presumably present in real-life interrogation scenarios. [↑](#footnote-ref-60)
61. Albeit a large increase in percentage terms. [↑](#footnote-ref-61)
62. For example, Yilmaz and Bahçekapili (2016) only made a binary distinction between a punishing and a ‘non-punishing’ God. A non-punishing God prime conveys information about the *absence* of a feature as opposed to information about an alternative feature, e.g., benevolence/forgiveness. [↑](#footnote-ref-62)
63. The God *Vishnu* is only forgiving if wrongdoers repent for their transgressions. The goddess *Lakshmi*, however, is forgiving even if wrongdoers show no remorse. [↑](#footnote-ref-63)
64. As discussed in Chapter 2, section 4.2, I agree with McKay and Whitehouse’s (2015; 2016) disagreement with the common construal of ‘prosocial’ behaviours as essentially ‘nice’. Punishment is undoubtedly an essential mechanism for reinforcing cooperation and therefore benefits the group at large, yet it is by no means *nice*. To the extent that punishment benefits the group at large it as a type of prosocial behaviour (see McKay & Whitehouse, 2015, 2016). [↑](#footnote-ref-64)
65. Some examples: “There is only one lawgiver and judge, he who is able to save and destroy. But who are you to judge?” (James, 4:12); “But it is God who executes judgement, putting down one and lifting up another” (Psalm, 75:7); “Why do you pass judgment on your brother? Or you, why do you despise your brother? For we will all stand before the judgment seat of God” (Romans, 14:10). [↑](#footnote-ref-65)
66. This effect size was based on the results of Laurin et al. (2012, Study 3, p. 3276). They reported a significant condition (salient vs. not-salient) x God beliefs interaction, *β* = -0.55*, t*(49) = 2.08, *p* = 0.04. Using this information as a guideline for entering parameters into an effect size calculator in G\*Power (Faul et al., 2009) an effect size of *Cohen’s* *f2=* 0.27 was computed. We used this effect size as a guide for all the studies reported in this article. [↑](#footnote-ref-66)
67. Four items I considered to be redundant were removed *(‘I am afraid God will judge me harshly for what I have done’, ‘I believe God is all merciful’, ‘My beliefs help me believe God will forgive my shortcomings’* and *‘God will judge me harshly one day’*). I also changed several items from the first person to the third person e.g., *‘Despite my shortcomings, I feel I will be forgiven’ 🡪 Despite their shortcomings, I feel people will be forgiven.* Finally, several verbs were changed e.g., ‘feel’ to ‘believe’. [↑](#footnote-ref-67)
68. Although identity fusion is related to group identification, it is thought to be an independent construct (e.g., Bortolini, Newson, Natividade, Vázquez & Gómez, 2018; Swann, Gómez, Seyle, Morales & Huici, 2009). [↑](#footnote-ref-68)
69. Moral Foundations theory (Haidt & Graham, 2007) proposes at least five different psychological foundations upon which people construct moralities; *harm, fairness, loyalty, authority and purity.*  [↑](#footnote-ref-69)
70. And those ingroup members who dissent would be considered immoral. [↑](#footnote-ref-70)
71. One participant alluded to this in the comments section of the study: “I wonder if you are studying whether we would think the new policeman to be hired should probably be white, because if an African-American candidate were hired they would have hostile working conditions and probably not be very happy”. [↑](#footnote-ref-71)
72. With this in mind, I recoded the DV (job suitability ratings) in Study 10 in terms of its absolute deviation from the midpoint (4). I ran a Kruskall Wallis H test to check for any differences in this new variable across conditions (i.e., moral, immoral and neutral). The mean ranks of absolute deviation from the midpoint were not statistically significantly different between groups, *Χ2*(2) = 2.74, *p* = .254. [↑](#footnote-ref-72)
73. e.g., “In one part of New Zealand an expiation for sin was felt to be necessary; a service was performed over an individual, by which all the sins of the tribe were supposed to be transferred to him, a fern stalk was previously tied to his person, with which he jumped into the river, and there unbinding, allowed it to float away to the sea, bearing their sins with it” (Frazer, 1890/2009, p. 563.). [↑](#footnote-ref-73)
74. Interestingly, Matthew Gibson’s suspicions that someone must have known about the murder and that someone was watching him occurred after he had developed an interest in religion. [↑](#footnote-ref-74)
75. e.g., “Nothing in all creation is hidden from God’s sight. Everything is uncovered and laid bare before the eyes of him to whom we must give account.” (Hebrews 4: 13). [↑](#footnote-ref-75)
76. “You have heard it said, ‘you shall not commit adultery’: but I say unto you, that everyone who looks on a woman to lust for her has committed adultery with her already in his heart” (Matthew, 5:27-28). [↑](#footnote-ref-76)
77. Kouchaki (2011) was published in Journal of Personality and Social Psychology; Newman and Brucks (2018) appeared in Journal of Consumer Psychology and Meijer et al. (2018) were published in Environment and Behavior. [↑](#footnote-ref-77)
78. In 2015 PLOS ONE launched a new collection called ‘Missing pieces’ which highlighted inconclusive, null findings or demonstrated failed replications of previous work. http://collections.plos.org/missing-pieces#id\_Research%20Article [↑](#footnote-ref-78)
79. Kurdi et al (in press) found that of the studies included in their meta-analysis (*N* = 36, 071) the median sample size was 40. As they note, this *N* is too small to reliably detect any effect below *r* = .24 (Cohen, 1992). An *N* of 40 only provides .40 power to detect the mean effect size reported by Greenwald, Peohlman, Uhlmann & Banaji (2009). These results led Kurdi et al. (in press) to state that these small levels of statistical power were so worrisome that “We can go so far as to say that many of the studies included in this meta-analysis should never have been undertaken given the potential for incorrect inferences about the population effect size” (Kurdi et al., in press, p. 22.) [↑](#footnote-ref-79)
80. Study 1 for example, used a rather elaborate false confession paradigm that had been designed specifically for the study and Study 4 used a novel surveillance perception measure (i.e., the portrait task). [↑](#footnote-ref-80)
81. For example, the 2009 annual report for the European Research Area Board set out their vision “to deliver excellence … where risk-taking in research, regardless of its

    public or private origin, will be the guiding principle for ERA policy.” (ERAB, 2009, p. 7). [↑](#footnote-ref-81)