

**Empathy Modulation and Coalition Management:
An Evolutionary Theory of Moral Judgment and Prejudice**

Tim Marsh
BPsych(Hons).

Submitted in fulfilment of the requirements of a degree of Doctor of Philosophy (Ph.D.),

Department of Psychology,

Macquarie University

Table of Contents

Copyright Notice	4
ABSTRACT	5
List of Papers Submitted and Published During the Course of this Thesis	6
Statement of Authentication and Ethical Accordance	7
Acknowledgements	8
CHAPTER 1 Introduction	9
Overview of the Thesis	14
A note on ‘Thesis by Publication’	18
CHAPTER 2 Early Perspectives	19
Declaration for Thesis Chapter 2	22
Applying Evolutionary Theory to Individual Differences: Insights from Moral Psychology	23
Discussion for Thesis Chapter 2	40
CHAPTER 3 Framing the Theory and Approach	43
Understanding Disunity in Psychology	44
Understanding Epistemology in Psychology	46
Understanding Evolutionary Explanations	49
Developing an Integrated Approach	54
CHAPTER 4 The Difficulties of Integration in Psychology	59
Declaration for Thesis Chapter 4	61
Unifying Psychology: Shared Ontology and the Continuum of Pragmatic Assumptions	62
Discussion for Thesis Chapter 4	120
CHAPTER 5 The Evolutionary Approach to Individual Variation	122
Declaration for Thesis Chapter 5	125
Evolutionary and Differential Psychology: Conceptual Conflicts and the Path to Integration	126
Discussion for Chapter 5	185
CHAPTER 6 The Implicit Measurement of Racial Prejudice	187
Declaration for Thesis Chapter 6	190
Evaluative Attitudes and Identification with Light- and Dark-Skinned Racial Groups	191
Discussion for Thesis Chapter 6	236

CHAPTER 7 The Development of the SATEST Measure	238
Declaration for Thesis Chapter 7.....	244
Sympathy vs. Social Rule Adherence: A New Measure of Interpersonal Empathy ..	245
Discussion for Thesis Chapter 7.....	323
CHAPTER 8 General Discussion and Conclusion.....	329
The Interplay of Theory and Measurement	330
Future Directions.....	334
Conclusion.....	338
References for the Unpublished Sections of this Thesis	340
APPENDIX A: The SATEST Scenarios	361
APPENDIX B: Final Ethics Approval	404

Copyright Notice

Under the Copyright Act of 1968, this thesis must only be used under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyrighted content to my work without the owner's permission.

ABSTRACT

In order to generate novel theory and predictions concerning the empathetic mechanisms underlying prejudice, this thesis aimed to integrate a range of findings and insights from the largely distinct fields of social, differential, and moral psychology. Despite the enduring negative consequences of prejudice and discrimination in the modern world, the psychological mechanisms underlying group-motivated social conflicts remain poorly understood. Many obstacles to exploring these phenomena can be traced to historical divisions of subject matter between conceptually distinct research traditions, most notably between the fields of intergroup social psychology and moral psychology. Beginning with an appreciation of the neglected overlaps between these approaches, the goal of this thesis was to synthesise the insights of multiple fields in order to generate new, more conceptually robust methods of exploring human prejudice. In addition to a published introductory book chapter, the main contributions of this thesis are made by four journal articles. The first article addresses the broad topic unification in psychology, proposing some means of overcoming a range of conceptual issues that have slowed integration efforts in the past. The second article focuses more closely on the theory-driven difficulties and recent practical successes in overcoming barriers to integration in the field of individual differences, through the application of methods and principles refined by evolutionary psychologists. The third article lays the empirical groundwork for the testing of implicit racial attitudes across more than two racial groups. The fourth article builds upon these earlier tools and principles, outlining the development of a novel approach to measuring prejudice, which focuses on the modulation of subjects' sympathetic behaviour and attributions by their perception of racial group-membership. The thesis concludes with reflection on the disparate insights from social, differential, moral, and evolutionary psychology that made the development of this measurement tool possible, and the prospects of similar integrative insights in the future.

List of Papers Submitted and Published
During the Course of this Thesis

- Marsh, T. & Boag, S. (2010). Applying Evolutionary Theory to Individual Differences: Insights from Moral Psychology. In R.E. Hicks (Eds.), *Personality and Individual Differences: Current Directions* (pp. 123-134). Bowen Hills: Australian Academic Press.
- Marsh, T., & Boag, S. (under review). Unifying Psychology: Shared Ontology and the Continuum of Pragmatic Assumptions. Submitted 9/2013 to *Review of General Psychology*.
- Marsh, T., & Boag, S. (2013). Evolutionary and differential psychology: conceptual conflicts and the path to integration. *Frontiers in Psychology*, 4(655), 1-15.
- Marsh, T., & Boag, S. (under review). Evaluative Attitudes and Identification with Light- and Dark-Skinned Racial Groups. Submitted 10/2013 to *Personality and Social Psychology Bulletin*.
- Marsh, T., & Boag, S. (under review). Sympathy vs. Social Rule Adherence: A New Measure of Interpersonal Empathy. Submitted 10/2013 to *Journal of Experimental Social Psychology*.

Statement of Authentication and Ethical Accordance

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, in full or in part, for a degree at this or any other institution.

All human research carried out in this thesis was approved by the Macquarie University Human Research Ethics Committee and in accordance with the American Psychological Association guidelines for research with human subjects.

Tim Marsh

Signature: _____

Date: _____

Acknowledgements

There are a handful of people who have been instrumental in the completion of this thesis.

First, and foremost, thank you to Simon Boag, my indispensable supervisor without whom I couldn't have possibly completed a project of this magnitude. Thank you for taking chance on an undisciplined kid with big ideas, thank you for coaching me and inspiring me through the challenges of theoretical psychology, and thank you for never giving up on me while I was still finding my academic voice.

Thank you to my always wonderful partner, Belinda, love of my life, who has suffered more impatient headaches throughout the course of this thesis than I have, but who has been a bottomless fountain of support and encouragement. We finally made it!

A special thank you to Robert Maynard whose programming skills, design guidance, and tenacious bug-chasing made most of the methodological achievements of this thesis possible. I couldn't have done this without you, or at least, not nearly so stylishly.

Thank you, also, to Francis Vivek, who has helped me sharpen my love of cognitive and behavioural science since my high school days, and who remains an inexhaustible source of probing conversation.

Lastly, thanks to my remaining family and friends, Mom and Dad, Dean and Neville, and Li my sister from a shared academic lineage. I doubt I could've kept my focus without such a wealth of sympathetic ears to turn to. I am lucky to have met each and every one of you.

CHAPTER 1

Introduction

Throughout recorded history, few social and cultural phenomena have had so pervasively negative an impact on human well-being as prejudice and discrimination (Bromley, 1987). From the perspective of victims, prejudice represents often insurmountable obstacles to the crucial social resources of support, cooperation, and bonding beyond the family, in addition to an either stated or implied threat of deprivation or violence. Conversely, from the perspective of perpetrators, discrimination and prejudicial judgments are experienced as appropriate reactions to undesirable others, typically with little-to-no consideration of the subjective pain or displeasure of victims. When occurring across group boundaries, as in the cases of racism and nationalism, prejudice has provided the ideological contexts for raids, wars, and programs of forced segregation. When occurring between dissimilar members of common groups, as in the cases of sexism and classism, prejudice serves to justify inequalities in power and representation that undermine the interests of the oppressed. Even between analogous members within a group, prejudicial standards are sometimes internalised, and culturally legitimise the policing of normative behaviours and beliefs with threats of ostracism or loss of social status, as in the application of heteronormative standards to sexual minorities. The uniting characteristic of prejudice, in its broadest conception, is possessing (typically negative or reductive) category-delimited attitudes, judgments, beliefs and dispositions that are applied to others in advanced of, or in priority relative to, considerations of their individual character or circumstances (Klineberg, 1968). Though the concept of prejudice and the harm it perpetuates has served as a motivating theme in the writing of this thesis, so too have the past and present difficulties in approaching the topic of prejudice from the perspective of psychological science (Duckitt, 1994). While the meaning and implications of prejudice are readily understood on the level of the layperson, the social, cultural, and

psychological complexity of these phenomena demands a great deal of conceptual clarity and elaboration if any aspect is to be addressed with scientific rigour.

The earliest scientific explorations of prejudice in psychology emerged during the growth of psychometrics at the beginning of the 20th century, though many of these efforts were, themselves, attempted confirmations of prejudicial standards such as white supremacy (reviewed in Garth, 1930). As social sympathy for racial and lifestyle minorities grew in the Western world, the partially clinical study of prejudiced attitudes as rigid (or perhaps pathological) personality characteristics emerged between the 1930s and 50s (Duckitt, 1994), to eventually find grounding in Gordon Allport's theories, which connected intergroup prejudice to categorical thinking in general (Allport & Kramer, 1946; Allport, 1954). Most of the psychological study of prejudice, particularly by social psychologists, builds upon the foundations of (or the reactions to) Allport's early investigations (Duckitt, 1992), and to this day an observation made by Allport in his book, *The Nature of Prejudice* (1954), remains descriptive of social prejudice research: "as a rule [theories of prejudice are] advanced by their authors to call attention to ... one important causal factor, without implying that no other factors are operating" (p. 207). As a complex, and culturally-embedded social phenomenon, one's understanding of prejudice depends largely on the level of analysis chosen to begin one's investigations. For example, during the 1950s (particularly in the USA), the prevailing approach to the study of prejudice was the search for individual personality constructs that fostered intolerance and authoritarianism (reviewed in Condor & Brown, 1988; Milner, 1983). This gave way in the 60s and 70s to an emphasis on cultural elements driving social perception of outgroup members (reviewed in Fairchild & Gurin, 1978), which in turn yielded to the dominance of cognitive-process accounts in the late 20th century (based on automatic categorisation theories, as in Tajfel, 1982). Each level of analysis contributes a small, but presumably causally significant, part of the explanation as to why some individuals

are more prejudiced than others, though typically with little consideration of how these levels of analyses may interact.

Some theorists, most notably John Duckitt (1992; 1994), have made concerted efforts to organise existing approaches to the social psychology of prejudice into four mutually influential explanatory levels; Fundamental cognitive processes that foster categorical thinking, individual differences at the level of personality and beliefs, social pressures concerning the prevailing norms of the ingroup and exposure to outgroups, and the wider social and cultural context of how groups and classes of people currently interact and compete. Though these integrative efforts were exhaustive and instructive, the majority of investigations directly focusing on prejudice remain limited to single forms of prejudice (e.g. racism, classism, etc.), and a single explanatory level (Dovidio, Glick & Rudman, 2005). Motivated by the observation that prejudiced individuals appear disposed towards several kinds of prejudice simultaneously, a subset of prejudice research attempts to specifically explore individual dispositions towards prejudice in general (see Schaller, Boyd & Yohannes, 1995; Reynolds et al., 2001; Duckitt & Sibley, 2010; Sibley et al., 2010; Akrami, Ekehammar & Bergh, 2011), but these approaches rely primarily on theories of personality, and lack the intergroup explanatory value of stereotype content employed extensively in other approaches (Haslam & Wilson, 2000).

As such, the notion that there are general psychological processes that underpin all forms of prejudice has merit, but the nuances of such processes remain under-explored by prejudice researchers within social psychology, for two distinct reasons. The most obvious reason is that social psychology is typically concerned with reactions to and interactions between human subjects, with a strong methodological preference for measuring genuine, ecologically valid social responses. As such, prejudice must be expressed for subjects to socially engage with, which requires researchers to choose a particular form of prejudice to focus on.

Whatever psychological processes are common to all forms of prejudice must be intrapersonal in nature, and can only be expressed in some specified form to be detectable in a social context. The second reason for the neglect of general mechanisms of prejudice does not concern the methods of social psychology as a field, but rather, concerns the field's perceived subject matter. As mentioned above, many efforts to explore the intrapersonal underpinnings of expressions of prejudice are perceived as straying from the purview of the social psychology field, straying into the disciplinary 'territory' of personality and individual differences, a field with conspicuously less contemporary interest in prejudice as a topic. In a similar vein, the field of moral psychology also draws upon prejudicial social interactions as a means of exploring general moral character and judgment (for example, Monin & Miller, 2001), conceiving of such behaviours as moral failings in accordance with theories wholly divorced from those employed in both social and personality psychology. In this sense, the causal complexity of prejudice as a phenomenon places it at an uncomfortable cross-roads between several fields of psychology which only partially overlap, resulting in multiple approaches to the topic that are too conceptually and theoretically different to be integrated in any meaningful way. To some extent, conflicts like this are to be expected in sciences like psychology, where, in the words of Paul Meehl, "what is one psychologist's subject matter is another psychologist's error term" (1978, p. 808).

The central focus of this thesis is on the general psychological mechanisms underlying prejudice in humans, specifically those relating to the experience of empathy, and how empathetic feelings and their subsequent evaluations are modulated by one's subjective categorisation of others. In approaching this topic, however, it has been necessary to contend with how this common set of psychological phenomena has been addressed by three theoretically and methodologically disparate fields: social psychology, personality and individual differences, and moral psychology. The history of each field contributes some

insights typically lost to the others (as explored in Chapter 7), but express these insights as embedded in a theoretical framework largely incompatible with the dominant theories of the other fields. Fortunately, despite their vastly disparate conceptual histories, all three fields have, in recent years, undergone considerable challenges and innovations owing to the expansion of evolutionary approaches to psychology. Though the adaptationist approach of evolutionary psychology was introduced to these fields at different times (decades ago in social psychology, but only in recent years for individual differences), and with different degrees of impact (the evolutionary and biologically based approach to moral judgment is now the dominant approach in the field, while evolutionary personality approaches remain uncommon), the reinterpretations of the native theories of each field by evolutionary researchers has served as a common ground through which all three may be connected.

Using the adaptationist approach as a grounding meta-theory, this thesis sought to integrate the disparate understandings of the modulation of empathy provided by each of the three fields, into a single coherent (though preliminary) evolutionary model of one of the key mechanisms underpinning human prejudice. In integrating this wide range of conceptually diverse findings, a novel approach to the measurement of prejudice was theorised, and the development and testing of this new methodology comprises the empirical research of this thesis. This new measure, referred to here as the Sympathetic Attribution Towards Emotive Social Transgressors (SATEST) approach, measures carefully controlled social expressions of sympathy, justification, and attribution, to provide information on the baseline moral character of subjects, and detect expressions of prejudice as apparent lapses in empathetic processes.

Overview of the Thesis

Following the introductory chapters, intended to give adequate framing to the problems addressed in the later sections, this thesis can be broken into two connected halves. The first half consists of theoretical contributions, two extensive journal articles in which conceptual challenges of integrating several disparate fields with a common evolutionary approach are addressed, and lay the groundwork for what is to follow. The second half consists of empirical contributions, two journal articles which collectively account for the four studies (total $N = 1016$, including preliminary testing), the final three of which concern the development and testing of the SATEST approach. Given the diversity of topics covered, no single section of this thesis is dedicated to the reviewing of background literature. Rather, each of the four major papers is written to provide an extensive literature review relevant to the topics immediately discussed, with the most general of background details fleshed out in Chapter 3, to account for any specific lack of familiarity on the part of the reader.

Chapter 2 is based upon a short published book chapter written by the attendees of an individual differences conference near the very beginning of this thesis. Though the quality of writing and analysis is predictably lower than the following four primary papers, which were completed some years later, the chapter serves two key purposes in the overall narrative of the thesis: first, to provide a more typical perspective on the overlap between individual differences and moral psychology, that foreshadows the more in-depth analyses required later in the thesis, and second, to give a simple preliminary account of the role evolutionary approaches have to play in both fields.

Chapter 3 is written specifically to support the narrative of this thesis, filling in critical details that bridge between the perspective in which Chapter 2 was written, and the perspective in which the subsequent 4 chapters were written. Each of the subsequent chapters, being

composed as contributions to pressing issues in various fields, requires some additional conceptual background for the reader to fully appreciate their contribution to the problems addressed by this thesis.

Chapter 4 is based upon a journal article submitted to *Review of General Psychology*, which addresses the general topic of unification and integration between the various fields of psychology. In addition to providing instructive background on the types of incongruities commonly found between different fields, this paper introduces concepts and terminology that will prove essential to the integrative efforts discussed in later chapter.

Chapter 5 is based on a journal article published in the journal *Frontiers in Psychology*, specifically the *Evolutionary Psychology and Neuroscience* specialty field. It explores the recent interface between evolutionary psychology and the study of individual differences, reviewing a wide range of literature to outline the recently emerging insights into systematic psychological variation in humans, and how such variation interacts with selection pressures. Furthermore, the paper draws on literature from the wider philosophy of science to outline the integrative strengths of evolutionary approaches in the developing explanatory theories, elements crucial to the empathy synthesis discussed in later chapters.

Chapter 6 focuses on a journal article submitted to *Personality and Social Psychology Bulletin*, concerning the measurement of implicit racial identity and both implicit and explicit racial attitudes. Beyond the novel racial identity methodology explored in this paper, it contributes to the topic of the thesis in two ways: first, by establishing the background literature on a range of implicit and explicit racial attitude measures that will appear in the third testing phase of the final paper, and second, by empirically verifying a modification of the Implicit Association Test's categorical comparisons that would later prove essential in demonstrating some aspects of convergent validity in the SATEST measure.

Chapter 7 concerns the final journal article in the thesis, submitted to the *Journal of Experimental Social Psychology*, which provides the supporting literature review, development, and empirical validation of the SATEST measure. Variants of the SATEST approach are tested across three studies, separately demonstrating its reliable psychometric properties, its veracity as a moral judgment measure, and finally as a measure of racial prejudice. The specifications of how and why the SATEST performs as it does are discussed, as are the preliminary assessments of the methodology's wider potential applications, and the support its success offers to the underlying evolutionary theory from which its predictions were drawn.

Chapter 8 concludes the thesis with a summary of the theoretical synthesis and empirical accomplishments of this research project, both with regards to the individual papers and the collective discoveries of the thesis as a whole. Future applications of the SATEST methodology are discussed, as are a range of alternative means of expanding upon and testing the evolutionary-moral synthesis that the SATEST approach relies on. The final appendices include full descriptions of the SATEST vignettes and branching interaction trees.

The central research contribution of this thesis is the evolutionary theory of moral judgment and prejudice outlined in full and empirically investigated in Chapter 7, with implications explored in Chapter 8. In its simplest form, this theory approaches the breaches in socio-moral behaviour collectively understood as 'prejudice', as the historically mismatched legacy of our ancestral past. Since the prevailing, successful social organisation for most of identifiable human history consisted of small, kin and non-kin cooperating groups, in both direct and indirect competition with rival groups of comparable size, the mechanisms of the human mind require features that both increase the benefits and likelihood of ingroup trust and cooperation, while simultaneously allowing for the reliable identification of, mistrust of, and if necessary hostility towards, members of outgroups. The relative fluidity, complexity

and conditionality of human coalitional groups necessitates a set of management mechanisms sensitive to a wide range of fitness-relevant factors, and if prosocial tendencies are to express adaptively, one's ingroup-favouring empathetic responses must be largely modulated by the real-time evaluations of this coalition-management apparatus. This approach affords a wide range of specific and useful predictions concerning social conflicts with categorical 'others' (key examples in Chapter 8), and is amenable to complex and multifaceted psychometric exploration, which the SATEST tool developed in this thesis addresses.

However, because this theory and measure are the products of a broad range of insights, drawn from the fields of moral, social, differential and evolutionary psychology, their development crosses several traditional disciplinary boundaries, and in doing so encounters many conceptual complications that threaten to confuse the premises upon which the theory and its predictions are grounded. As such, the early portions of this thesis are dedicated to painstakingly disambiguating and defining the conceptual foundation upon which this theory stands. After the general background is introduced in Chapters 2 and 3, Chapter 4 clarifies the problems inherent to interdisciplinary integration efforts in general, and introduces the philosophical framework that the adaptationist approach of the main theory is based in. Chapter 5 addresses a relatively contentious element in the literature concerning how theories in evolutionary psychology can meaningfully engage with the study of individual differences, which must be established before the SATEST can be introduced as an evolution-guided measure of moral individual differences. Lastly, Chapter 6 introduces many of the elements of prejudice research that do not fit comfortably into the narrative of the paper in Chapter 7, and offers empirical support for one of the main testing techniques (the light-dark manipulation of the racial attitudes Implicit Association Test) employed to verify the SATEST's predicted psychometric properties. Though this structure is unorthodox, it is necessary to minimise the confusion of readers when approaching a complex,

interdisciplinary theory, and has given rise to a collection of journal articles that make meaningful contributions in and of themselves.

A note on ‘Thesis by Publication’

The Higher Degree Research programs of Macquarie University strongly encourage PhD candidates to complete their theses via ‘thesis by publication’, which includes submitting chapters written with intent to be published as independent journal articles. Four refereed journal articles, and an additional peer-reviewed published book chapter, comprise 5 of the chapters of this thesis, and as such some degree of overlap and repetition is to be expected between publications discussing common topics. The components of this thesis appearing immediately before and after each publication are intended purely to contextualise the contributions of each paper, and help maintain the logical and narrative flow of the thesis as a whole.

CHAPTER 2

Early Perspectives

Before introducing (in Chapter 3) the key concepts and background literature that underpin the more sophisticated analyses of the later chapters, it will prove instructive for some readers to first review a more broad account of how the fields of moral, differential, and evolutionary psychology interrelate, and how the insights of one field may clarify the difficulties of another. To this end, this chapter includes a short paper published years before the four main journal articles that follow, which touches upon the relationships between the three fields on a more introductory level.

From the earliest months of this thesis, it became clear that the study of general mechanisms underlying prejudice had already been undertaken in three distinct styles, by three different fields of psychology: social psychology, personality and individual differences, and moral psychology. Although the concept of prejudice is most typically associated with social psychology, providing the social literature with the widest range of approaches to specific manifestations of prejudice and discrimination (e.g. racism, classism, homophobia, etc.), much of this research (particularly the fruitful literature concerning stereotyping, see Judd & Park, 1993; Greenwald et al., 2002; Cox et al., 2012, for reviews) focuses on the unique characteristics of specific prejudicial judgments and ideologies. Exceptions are most readily found in the social cognition literature, which has made some significant discoveries concerning the fundamental cognitive tendencies thought to make prejudiced attitudes and beliefs possible, most notably the automatic categorisation of others as ingroup or outgroup members (reviewed in Tajfel, 1982) and the subsequent biases that predictably follow (see McGregor, Haji & Kang, 2008, and Tarrant, Dazeley & Cottom, 2009 for key examples and Riek, Mania & Gaertner, 2006 for a meta-analytic overview). The study of stereotypes and their ideological propagation has also generated a range of theories concerning how more

obviously harmful positions are socially justified, including realist conflict theory (Sherif et al., 1988), integrated threat theory (Stephan et al., 2000) and the justification-suppression model (Crandall & Eshleman, 2003), each of which have been applied successfully to the social contexts of specific forms of discrimination. Though these intrapersonal and interpersonal processes are no doubt crucial in understanding the real world occurrence of prejudice, these social psychology theories share a common neglect of a key element of prejudicial phenomena: how do prejudiced individuals vary from less-prejudiced individuals, and what processes make these differences possible? By Duckitt's (1992; 1994) taxonomy of levels of analysis in prejudice discussed above, the social psychology literature has focused extensively on fundamental mechanisms of cognitive categorisation, social pressures and the context of intergroup relations, in order to describe the behaviour of those engaging in prejudice, while largely neglecting the individual-level differences that separate those prone to discriminatory thoughts and behaviours from those who are not.

This differential aspect of prejudice falls more under the disciplinary purviews of moral psychology and the study of personality and individual differences. As was described in Chapter 1, attempts to account for differences in prejudicial tendencies with reference to personality trait theories have been somewhat successful (examples include Schaller, Boyd & Yohannes, 1995; Reynolds et al., 2001; Duckitt & Sibley, 2010; Sibley et al., 2010; Akrami, Ekehammar & Bergh, 2011), though most theories of this sort function on a purely descriptive level, with little explanatory value beyond the inference of trait heritability between generations (notably Sibley et al., 2010). This stands in sharp contrast with many approaches in moral psychology, which owing to the once dominant cognitive-developmental theories pervading the field (see Rest et al., 1999; 2000; Narvaez & Bock, 2002), focused more extensively on how differences in moral character and moral judgment ability are acquired as an interaction between social experiences and personal dispositions. Acts of

prejudice, in this moral conception, are understood as failures to act in a moral manner, owing to undertrained deficiencies in perspective, self-control or appreciation of social nuances.

This book chapter was written following a conference presentation, concerning these differences of approach between moral psychology and the study of individual differences, at the *Australian Conference for Personality and Individual Differences*. Contributors to the conference were invited to submit chapters to be peer-reviewed for inclusion in a book intended to provide a cross-section of the current directions individual differences research was taking in the Australasian region. At the time of writing the field of moral psychology was beginning to reflect significant theoretical changes owing to the growing popularity of evolutionarily-grounded biopsychological theories, most notably the Social-Intuitionist approach introduced by Jonathan Haidt (2001) and colleagues, but the field of personality and individual differences was still perceived as being largely at odds with the methods and theories of evolutionary psychology. In addition to exploring several of the conceptual and explanatory differences between moral psychology and differential psychology, from a perspective more closely resembling that of typical members of these fields than the later works in this thesis, this book chapter offers a simple framing of some of the integrative difficulties that the theoretical half of this thesis endeavours to address.

The following book chapter was published as Chapter 12 in *Personality and Individual Differences: Current Directions*, edited by R.E. Hicks in 2010, with the title ‘Applying Evolutionary Theory to Individual Differences: Insights from Moral Psychology’.

Declaration for Thesis Chapter 2

In the case of the book chapter featured in Chapter 2, the nature and extent of my contribution to the work, and the contributions of the other listed co-authors is as follows:

<i>Name</i>	<i>Nature of Contribution</i>	<i>Contribution</i>
Tim Marsh	Decision concerning the topic of the paper	90%
	Search and review of the literature	
	Principle writing and editing of the manuscript	
Simon Boag	Advice on topic and approach	10%
	Assistance with editing and cutting	
	Suggestions for the refinement of the manuscript	

**Applying Evolutionary Theory to Individual Differences: Insights from Moral
Psychology**

Tim Marsh
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: timothy.marsh@mq.edu.au

Simon Boag
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: simon.boag@mq.edu.au

ABSTRACT

Evolutionary theory and adaptation-based explanations in psychology have become increasingly common within the last 20-30 years, and researchers are beginning to utilise the evolutionary paradigm in the study of personality and individual differences. This paper examines several evolutionarily-oriented personality trait studies to highlight apparent tendencies towards simple genetic determinism, and contrasts this method with the more developmentally sensitive approach pursued in some recent works in moral psychology. Salient theoretical concerns addressed include the direct inheritance of behavioural predispositions, and the environmental and developmental calibration of universally inherited potentials. The implications of these issues are discussed.

Keywords: Evolution; Genetics; Individual Differences; Morality; Personality; Selection

Psychology can be conceptualised as the behavioural and cognitive arm of biological science (Burghardt, 2009). In the neurosciences and psychology explanations of the configurations and functions of tissue structures require an understanding of the conditions under which these contemporary structures developed (cf., Symons, 1990). For this reason, the theory of evolution by natural selection is regarded as an explanatory cornerstone of biological science. In taking the ‘adaptationist’ approach to a biological system, one supplements the particularly challenging and detailed historical question of “How did this develop?” with the more functional concern of “*Why* did this develop?”. It is an approach wherein one views mechanisms as adaptations (cf., Gigerenzer, 2002; Tooby & Cosmides, 2005).

While this approach has proven instrumental to the development of informative explanatory theories, the effectively speculative nature of the method raises a number of theoretical and empirical concerns. One relates to thoughts about a remote prehistory where *the environmental and contextual details of interest are at best educated guesses*. Subsequently there are problems in formulating testable and falsifiable hypotheses based on a well-supported explanatory theory rather than ‘storytelling’ (Symons, 1990).

The goal of this chapter is to examine evolutionary psychology in relation to the domain of personality and individual differences. This paper therefore discusses the broad application of evolutionary theory in psychology, and then gives examples of how the evolutionary paradigm has been recently applied to the study of personality traits. These *methods* will be compared to recent evolutionary work in the field of moral psychology.

Evolution in Psychology

Evolution by natural selection –the fitness-based non-random selection of individual differences– is a theoretical framework that attempts to organise and explain the morphological and functional features observed in all biological systems. The theory

postulates that organisms naturally diversify developing phenotypic properties that permit heightened success in the given environment, becoming ‘better adapted’ to the specific environmental challenges faced (cf., Gigerenzer, 2002).

The growth of evolution-dependant theories in psychology has been evident over the past two decades in particular. The evolutionary paradigm was proposed to consist primarily of the identification of modular, adaptive psychological mechanisms in humans, evolving from our distant hominid and pre-hominid ancestors. This modularity theory meshed well with empirical findings in human cognition, and provided fresh explanatory insights into many of areas in psychology (cf., Buss, 1999; Figueredo et al., 2005). Examples concern (1.) seeing emotions as circumstance-specific motivational states (Turner, 1996) (e.g., anger and jealousy are theorised to play an important role in social and reproductive success—Buss, Larsen, Westen, & Semmelroth, 1992; Cosmides, 1989); (2.) mating preferences and intuitive mating strategies (cf., Buss & Dedden, 1990); and (3.) the perception of attraction and aggression in facial expressions and structures (cf., Langois & Roggman, 1990).

Overall, humans display an innate talent for processing concerns of reciprocal social exchange, social hierarchies, and cheat-detection in social circumstances (Kyl-Heku, 1990). They also seem to possess in-born language acquisition talents, evidence for which appears extensive, so much so that competing non-evolutionary theories, such as domain-general social-learning models, only hold a fraction of the explanatory power (cf., Pinker, 2002). Today evolutionary psychology plays pivotal theoretical roles in cognitive neuroscience (Krill, Platek, Goetz, & Shackelford, 2007), learning (Jiménez-Díaz, Sancho-Bielsa, Gruart, López-García, & Delgado-García, 2006), and notably developmental psychology generally (e.g., Carroll, 2008; Moore, 2008; Whittle, Allen, Lubman, & Yucel, 2006).

Variation in adaptive mechanisms is best understood as differential levels of efficacy or totality, though developing research methods to examine the complexity of the adaptive

variations in the environmental contexts is not easy and is a central challenge of researchers seeking to apply the power of the evolutionary paradigm to personality and individual differences.

Personality and Adaptation

Critics such as de Jong and van der Steen (1998) have suggested that there is a fundamental incongruity between explanations relying on evolutionary theory, and the study of personality and individual differences in psychology, primarily because evolved mechanisms are generally regarded as species-wide solutions to age-old problems, and interpersonal variation appears to be neither ubiquitous nor genetically-contingent enough to be anything more than residual ‘noise’ to an evolutionary study. Such concerns mirror similar objections originating within the social sciences, where the concept of an inherited ‘human nature’ conflicts with popular theories that rely exclusively on social-learning (Pinker, 2002).

Despite this, in recent decades, research into personality and individual differences within an evolutionary paradigm has been expanding (Nettle, 2008). The majority of research in this area has focussed on traits, partly since trait-theory is one of the most extensively studied and applied domains in personality psychology, and also because traits are argued to be among the most stable of personality factors (Digman, 1990). Furthermore, trait levels are correlated with a number of key life outcomes, such as life-expectancy and marital stability, and thus are expected to have a reasonable impact on selective fitness (cf., Friedman et al., 1995).

In the early 1990s, Buss (1991) articulated a set of guiding principles for how one may conceptualise the evolutionary origins of common personality variation. He outlined a number of theoretically plausible Darwinistic origins for the kind of variance observed in personality research including: 1. *competing strategies based on inherited genetic*

predispositions; 2. environmental or developmental calibration of a standardised set of inherited potential strategies; and 3. non-adapted variance in personality constituting selectively neutral 'noise'. Buss and Grieling (1999) later refined this further, offering methodological advice focussing primarily on the identification of genetic influences in the absence of environmental factors during development.

Recent evolutionary personality psychology studies have attempted to discover the direct heritability of genetic predispositions in personality. Bouchard (1994), and MacDonald (1995) indicated strong biological determination implying adaptive strategies characterised by varying levels on the Five Factor dimensions (cf., Canli, 2004; Plomin & Nesselroade, 1990). Bouchard and Loehlin (2001) also suggested a high heritability of personality traits, though they also note that determining non-shared environmental influences is an ongoing challenge in genetic personality research. Building off MacDonald's work, Nettle (2008) has formulated a detailed theoretic framework to explain the apparent equilibrium of presumably highly heritable personality traits.

While offering many valuable insights, particularly into transmission of personality traits within families, an issue with these studies is their high reliance on what appear to be simplistic assumptions of high genetic determinism. Since there is ample evidence to suggest that there are strong continuities in personality traits through one's lifetime, it is possible this has been interpreted as evidence that personality traits are a biological 'fixture', and thus are likely to have directly genetic causes. These elements need to be examined in more detail in current research.

The Impact of Evolution in Moral Psychology

Another field of individual differences that has recently received extensive insights from evolution theory is moral psychology, particularly its most well researched domain, moral

judgement. The work that arguably founded the empirical field of moral judgement was that of Kohlberg (1969), which was influenced heavily by the earlier developmental work of Piaget and the moral philosophy of John Rawls.

In his works Kohlberg developed a model comprising of six ordinal steps, intended to hierarchically represent both the progression most people go through as they morally develop from childhood, and also account for the differences in sophistication in adult moral reasoning. Kohlberg's work founded a tradition of cognitive-developmental focus in the moral judgement literature that is now referred to as the Neo-Kohlbergian paradigm (See Rest, Narvaez, Thoma, & Bebeau, 2000). This paradigm retains the core components of focussing on cognitive deliberation, personal 'upward' development, and the vital distinctions between rule and norm-based 'conventional' reasoning, and principle and consideration-based 'post-conventional' reasoning.

There are, however, numerous weaknesses to this approach to moral judgement, with its focus on conscious deliberation being continually challenged by empirical findings, such as Reber's (1993) work on the centrality of implicit, rapid decision making. During this same time there was a rise in biologically-based research concerning how emotional reactions guide moral reactions and judgements. The works of Haidt (e.g., Haidt, 2001, 2004; Haidt, Koller, & Dias, 1993) provided evidence for the primary role of affective and intuitive responses in determining the majority of moral judgements, including a number of studies that specifically demonstrated the failure of *post hoc* cognitive justifications to explain initial reactions to taboo yet harmless scenario stimuli (cf., work by Narvaez & Vaydich, 2008; O'Neill & Petrinovich, 1998; Pizarro & Bloom, 2003). This affective-intuitive paradigm based on evolutionary accounts of the formation of pro-social, anti-taboo intuitions challenges the cognitive-developmental paradigm (see Krebs & Denton, 2006; Krebs & Hemingway, 2008).

This movement in moral psychology towards biologically-integrated theories guided heavily by evolution has yielded strong insights into the function of the underlying mechanisms of moral judgement (cf., Hauser, 2006). For instance, using neuro-imaging methodology Anderson, Bechara, Damasio, Tranel, and Damasio (1999) confirmed the theorised functional links between moral judgement and moral action. Further neuro-imaging work has extensively mapped the regions of the brain associated with as diverse components of moral functioning as judgement, interpreted context, altruism and punishment cues, and the perception of transgressions (Borg, Lieberman, & Kiehl, 2008; Harbaugh, Mayr, & Burghardt, 2007; Moll, Eslinger, & de Oliveir-Souza, 2001; Moll, de Oliveir-Souza, Krueger, & Graftman, 2002; Moll, Zahn, de Oliveir-Souza, Krueger, & Graftman, 2005). Evolutionary theories of modular specialisation in moral functioning have also been supported by research into disgust and moral judgement (Danovitch & Bloom, 2009; Gutierrez & Giner-Sorolla, 2007; Jones & Fitness, 2008; Schnall, Haidt, Clore, & Alexander, 2008).

While initially adversarial, the cognitive-developmental and affective-intuitive paradigms of moral psychology are now moving towards a synthesis; for example, see Narvaez's (2008) *Triune Ethics Theory*. The result of this detailed synthesis is a model that maximises the explanatory power of inputs from the cognitive-developmental and the affective-intuitive approaches, to produce the individual differences observed in moral functioning. More research is needed using appropriate investigative techniques (cf., Caldwell & Millen, 2008), but the successful theoretic synthesis of prior social-learning and childhood development-based findings into a biologically robust evolutionary account of underlying mechanisms serves as an example to other fields of psychology, such as personality.

Beyond ‘Selecting For’

Several conceptual approaches can be adopted when investigating evolutionary accounts of personality and individual differences. The trait-based evolutionary personality psychology investigations (e.g., MacDonald, 1995; Nettle, 2008) provide bases for a more comprehensive understanding of the evolution of personality. The assessment of heritability of personality traits is complicated, however, since one’s theoretic interpretation of the causal path between genetic inheritance and manifested trait behaviour makes it possible to prematurely commit to a model of high direct heritability.

Both Gottlieb (e.g., 2004) and Finlay (2007) note that among many researchers employing the evolutionary psychology paradigm, there is a common functional assumption of the high genetic determinism. Under such a conception, genetic inheritance is seen as the primary cause of the manifest behaviours described by personality traits, and, based on the situational benefits and liabilities of the given traits, the related genes are differentially ‘*selected for*’.

Finlay (2007) argued that if psychologists are going to integrate the theories and frameworks of Darwinism into their analyses, especially in the assessment of individual differences, then great care in theory building will be needed. This is because there is a lack of species-wide standardisation of observable phenotypes, sharply increasing the number of theoretic ‘stories’ that can be speculated to explain them. Making the key explanatory distinction between developmental-configuration and directly inherited predispositions is confounded further still when it is likely that both mechanisms are true to different degrees.

First, research to clarify the contributions of genetic and developmental influences is hampered because minor genetic changes can influence the phenotypic, behavioural manifestations (Nettle, 2008; Rowe & Houle, 1996). But shared environmental (e.g., family) influences could also give the illusion of higher genetic influence than is truly present and

careful research is needed to unravel the respective (nature-nurture) contributions (cf., Gerhard & Kirschner, 1997).

Second, when considering the calibration of psychological mechanisms by ontogenic, developmental and culturally-inherited factors, it is important to keep in mind that one of the most potent selective pressures applied to organisms in highly dynamic environments is *selection in favour of intra-generational adaptation* (Narvaez, 2008; Tooby & Cosmides, 1992, 2005). This poses further challenges to tracing the reliability of genetic determinants of behaviour, since even casually shared kin environments may elicit the development of very similar calibrated strategies (Lickliter, 2008).

Third and last, while the proposition that the differential value of (say) the inherited Five Factor personality traits, can generate testable predictions (cf., Canli, 2004), the statistical analyses can potentially be confounded by other sources of equilibrium-maintenance in a gene-pool. Livnat, Papadimitriou, Dushoff, and Feldman (2008) argued that pairing many genes and their alleles in the selection process into complex combinations *over many generations*, leads to an overall selective pressure to produce genes that work well in multiple genomic contexts. The subsequent selective pressure towards genes and gene complexes that work well in multiple contexts is likely to obscure the effects of a simple system of moderate-selection due to benefits and trade-offs, and is made more confounding under circumstances where genetically ubiquitous sets of strategies are being ontogenically calibrated by developmental and environmental factors (Schwartz & Begley, 2003).

Consequently, as seen in these three aspects, the theoretical challenges and sheer number of practical considerations facing an evolutionary psychologist are magnified when studying personality and individual differences.

Final words

This chapter has reviewed several recent endeavours to conceptualise and study major areas in personality and individual differences through the lens of evolutionary science. In comparing and contrasting examples from both personality and moral psychology, the goal has been to demonstrate the theoretic differences between developmental calibration-based and genetic determinism-based accounts of individual differences, and in doing so to clarify how a highly sophisticated understanding of the nuances of evolutionary biology is necessary to experimentally and conceptually distinguish between the two. Far from seeking to discourage research into directly heritable predispositions in personality variance, the present chapter seeks to suggest essential considerations that can elevate the quality of generated hypotheses in all domains of evolutionarily-focussed individual differences research. It is also essential to the success of further research in this field that more evolutionarily-guided structural and neuro-imaging research is done into the potential, functionally-distinct psychological mechanisms of personality differentiation. In this way it may be possible to achieve the aspirations of researchers such as Buss, Toobey, and Cosmides, in integrating a well-supported developmental and evolutionary framework, as the functional core of personality psychology.

References

- Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, 2(11), 1032-1037.
- Borg, J. S., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience*, 20(9), 1529-1546.
- Bouchard, T. (1994). Genes, environment, and personality. *Science*, 264, 1700-1701.
- Bouchard, T. J. & Loehlin, J. C. (2001). Genes, evolution and personality. *Behavior Genetics*, 31, 243-273.
- Burghardt, G. M. (2009). Darwin's legacy to comparative psychology and ethology. *American Psychologist*, 64(2), 102-110.
- Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypothesis testing in 37 cultures. *Behavioral and Brain Sciences*, 12, 1-49.
- Buss, D. M. (1991). Evolutionary Personality Psychology. *Annual Review of Psychology*, 42, 459-491.
- Buss, D. M. (1999). *Evolutionary Psychology: The New Science of the Mind*. Boston, MA: Allyn & Bacon.
- Buss, D. M. & Dedden, L. (1990). Derogation of competitors. *Journal of Cross Cultural Psychology*, 21, 5-47.
- Buss, D. M. & Greiling, H. (1999). Adaptive individual differences. *Journal of Personality*, 67, 209-243.
- Buss, D. M., Larsen, R., Westen, D., & Semmelroth, J. (1992). Sex differences in jealousy: Evolution, physiology and psychology. *Psychological Science*, 3, 251-255.

- Caldwell, C. A. & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, 29, 165-171.
- Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: Learning from individual differences in emotion processing. *Journal of Personality*, 72, 1105-1131.
- Carroll, S. B. (2008). Evo-Devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell*, 134(1), 25-36.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition*, 31, 187-276.
- Danovitch, J. & Bloom, P. (2009). Children's extension of disgust to physical and moral events. *Emotion*, 9(1), 107-112.
- de Jong, H. L. & van der Steen, W. J. (1998) Biological thinking in evolutionary psychology: Rockbottom or quicksand? *Philosophical Psychology*, 11(2), 183 – 205.
- Digman, J. (1990). Five factor model. *Annual Review of Psychology*, 41, 417-440.
- Friedman, H. S., Tucker, J. S., Schwartz, J. E., Martin, L. R., Tomlinson-Keasey, C., Wingard, D. L. & Criqui, M. H. (1995). Psychosocial and behavioural predictors of longevity: The aging and death of the "Termites." *American Psychologist*, 50, 69–78.
- Figueredo, A. J., Sefcek, J. A., Vasquez, G., Brumbach, B. H., King, J. E., & Jacobs, W. J. (2005). Evolutionary personality psychology. In D. M. Buss (Ed.), *Handbook of Evolutionary Psychology* (pp. 851-877). Hoboken, NJ: Wiley.
- Finlay, B. L. (2007). Endless minds most beautiful. *Developmental Science*, 10(1), 30-34.
- Gerhart, J. & Kirschner, M. (1997). *Cells, Embryos and Evolution*. Malden, MA: Blackwell Science.
- Gigerenzer, G. (2002). *Adaptive Thinking: Rationality in the Real World*. Oxford, England: Oxford University Press.

- Gottlieb, G. (2004). Normally occurring environmental and behavioral influences on gene activity: From central dogma to probabilistic epigenesis. In C. G. Coll, E. L. Bearer, & R. M. Lerner (Eds.), *Nature and Nurture: The Complex Interplay of Genetic and Environmental Influences on Human Behavior and Development* (pp. 85-106). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Gutierrez, R. & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion*, 7(4), 853-868.
- Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613-628.
- 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. (2004). The emotional dog gets mistaken for a possum. *Review of General Psychology*, 8(4), 283-290.
- Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, 316(5831), 1622-1625.
- Hauser, M. (2006). *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*. NY: Harper Collins.
- Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. NY: Cambridge University Press.
- Jiménez-Díaz, L., Sancho-Bielsa, F. J., Gruart, A., López-García, C., & Delgado-García, J. (2006). Evolution of cerebral cortex involvement in the acquisition of associative learning. *Behavioral Neuroscience*, 120, 1043-1056.
- Jones, A. & Fitness, J. (2008). Moral hypervigilance: The influence of disgust sensitivity in the moral domain. *Emotion*, 8(5), 613-627.

- Kohlberg, L. (1969). Stage and sequence: The cognitive developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory* (pp. 347-480). Chicago: Rand.
- Krebs, D. L. & Denton, K. (2006). Explanatory limitations of cognitive-developmental approaches to morality. *Psychological Review*, 113(3), 672-675.
- Krebs, D. L. & Hemingway, A. (2008). The explanatory power of evolutionary approaches to human behavior: The case of morality. *Psychological Inquiry*, 19(1), 35-38.
- Krill, A. L., Platek, S. M., Goetz, A. T., & Shackelford, T. K. (2007). Where Evolutionary Psychology meets Cognitive Neuroscience: A précis to Evolutionary Cognitive Neuroscience. *Evolutionary Psychology*, 5, 232-256.
- Kyl-Heku, L. (1990). *Effects of Context and Sex on Hierarchy Negotiation*. Unpublished doctoral dissertation, University of Michigan.
- Langois, J. H. & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, 1, 115-121.
- Lickliter, R. (2008). The growth of developmental thought: Implications for a new evolutionary psychology. *New Ideas in Psychology*, 26, 353-369.
- Livnat, A., Papadimitriou, C., Dushoff, J., & Feldman, M. W. (2008). A mixability theory for the role of sex in evolution. *Proceedings of the National Academy of Sciences*, 105, 19803-19808.
- MacDonald, K. (1995). Evolution, the 5-factor model, and levels of personality. *Journal of Personality*, 63, 525-567.
- Moll, J., Eslinger, P. J., & de Oliveir-Souza, R. (2001) Frontopolar and anterior temporal cortex activation in a moral judgment task. *Arquivos de Neuro-Psiquiatria*, 59(3-B), 657-664.

- Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, 16(3), 696-703.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). Opinion: The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), 799-809.
- Moore, D. S. (2008). Integrating development and evolution in psychology: Looking back, moving forward. *New Ideas in Psychology*, 26(3), Dec 2008, 327-331.
- Narvaez, D. (2008). Triune ethics: The neurobiological roots of our multiple moralities. *New Ideas in Psychology*, 26, 95-119.
- Narvaez, D. & Vaydich, J. L. (2008). Moral development and behaviour under the spotlight of the neurobiological sciences. *Journal of Moral Education*, 37(3), 289-312.
- Nettle, D. (2008). Putting ethology (back) into human personality psychology. *European Journal of Personality*, 22(5), 464-465.
- O'Neill, P. & Petrinovich, L. (1998). A preliminary cross-cultural study of moral intuitions. *Evolution and Human Behavior*, 19, 349-367.
- Pinker, S. (2002). *The Blank Slate*. NY: Viking Press.
- Pizarro, D. A. & Bloom, P. (2003). The intelligence of the moral intuitions: A reply to Haidt (2001). *Psychological Review*, 110, 193-198.
- Plomin, R. & Nesselroade, J. R. (1990). Behavioral genetics and personality change. *Journal of Personality*, 58, 191-220.
- Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge*. NY: Oxford University Press.
- Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (2000). A neo-Kohlbergian approach to morality research. *Journal of Moral Education*, 29(4), 381-395.
- Rowe, L. & Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 263, 1415-1421.

- Schnall, S., Haidt, J., Clore, G. L., & Alexander, J. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096-1109.
- Schwartz, J. & Begley, S. (2003). *The Mind and the Brain: Neuroplasticity and Power of Mental Force*. NY: Harper Perennials.
- Symons, D. (1990). On the use and misuse of Darwinism in the study of human behaviour. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. NY: Oxford University Press.
- Tooby, J. & Cosmides, L. (1992). The Psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. NY: Oxford University Press.
- Tooby, J. & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 5-67). Hoboken, NJ: John Wiley & Sons Inc.
- Turner, J. H. (1996). The evolution of emotions in humans: A Darwinian--Durkheimian analysis. *Journal for the Theory of Social Behaviour*, 26(1), 1-33.
- Whittle, S., Allen, N. B., Lubman, D. I., & Yucel, M. (2006). The neurobiological basis of temperament: Towards a better understanding of psychopathology. *Neuroscience and Biobehavioral Reviews*, 30, 511-525.

Discussion for Thesis Chapter 2

This book chapter contributes two introductory points to the overall arc of this thesis. First, it emphasises the significance of the rise of the Social-Intuitionist approach in moral psychology, and the relative decline of the Neo-Kohlbergian approach, acknowledging the new avenues of theoretical and experimental innovation that this shift made possible. Second, the chapter offers reasons to be cautious about the more simplistic evolutionary approaches to personality and individual differences that received attention in the years preceding, insisting upon more sophisticated accounts of systematic variation that mesh more meaningfully with the explanatory methods of evolutionary psychology. Each of these two points directly inspired one of the major journal articles featured later in this thesis (the theoretical approach of Chapter 7 and the problems addressed in Chapter 5, respectively), while the interdisciplinary comparisons of the chapter begin to scratch the surface of the integration issues addressed in Chapter 4. However, as is to be expected of a piece written so early in this thesis, both key points of this paper are presented briefly and simplistically, and an appreciation of the nuances overlooked will prove instructive to understanding the directions that subsequent thesis chapters have taken.

With regards to the issue of moral psychology gradually favouring affective and intuitive approaches over a well-established tradition of cognitive-developmental approaches, the book chapter rightly praises the efforts of moral psychologists pursuing the social-intuitionist approach in attempting to salvage and integrate many of the deliberative elements of the Neo-Kohlbergian tradition. The Neo-Kohlbergian tradition, in itself, represents a perfect example of a theory that was handicapped by its initial ideological commitments. Having committed to the assumption that human moral choices must be strictly rational, the Neo-Kohlbergian tradition's central weakness was its *a priori* exclusion of any emotional, intuitive or affective influences on moral judgments, conceiving of these as sources of error that merely obscure

the conscious, deliberate reasoning thought to primarily determine moral positions (see Rest et al., 1999). The powerful demonstrations (most notably Haidt's 'moral dumbfounding' work reported in 2001) of the Social Intuitionist approach exposed this core assumption of the Neo-Kohlbergian approach to be fundamentally mistaken, but subsequent research took pains to acknowledge the empirical advances made under this false assumption, endeavouring to reinterpret successful findings in light of the newer theories (more on this concept in Chapter 4). While later Social-Intuitionist experiments, most notably those undertaken by Joshua Greene and associates (see Greene & Haidt, 2002, and Greene, 2013, for overviews), extensively accounted for the supplementary role of Kohlbergian-like deliberate conscious reasoning in particular moral decision making contexts, the underlying cognitivist structure of the Neo-Kohlbergian approach has been firmly refuted by evidence of the centrality of intuitive responses in the vast majority of moral evaluations. The preceding book chapter draws upon moral psychology to provide insights on how evolutionary psychology may more constructively approach the personality and individual differences field, extolling the merits of the calibration-based approaches often used to account for variation in moral variables. In doing so, the book chapter fails to acknowledge that the undermining of the Neo-Kohlbergian approach and its strictly cognitive-developmental assumptions also undermines the models of moral differentiation traditionally employed by moral psychologists to explain these differences. The Social-Intuitionist perspective preserves some of these explanatory concepts only with regards to the deliberate reasoning subcomponents preserved from the earlier tradition, and indeed faces precisely the same difficulties in accounting for variation in (presumably evolved) affective and intuitive mechanisms as the examples of evolutionary individual differences outlined in the book chapter.

As such, the titular 'insights from moral psychology' are largely undermined by the conceptual difficulties addressed by the paper's discussion of then-popular evolutionary

explanations of individual differences. Thus, the central limitation of this book chapter is its failure to acknowledge that the problems inherent to evolutionary explorations of personality and individual differences intrinsically carry over to any evolutionary account of systematic variation, including the variability in intuitive responses and affective reactions observed in Social-Intuitionist approaches to moral psychology. At the time this book chapter was written, many innovations in the evolutionary study of individual differences were being published amongst various psychological fields (the most illustrative of which appear in behavioural genetics studies), but these conceptual framings were not yet synthesised closely enough to permit appreciation by wider readerships. The precise character of the issues facing evolutionary approaches to individual differences, and the subsequent implications for the explanatory role of evolutionary theories in psychology is the topic addressed in Chapter 5 of this thesis.

CHAPTER 3

Framing the Theory and Approach

The following Chapters 4 through 7 are comprised of four journal articles which represent the major theoretical and empirical accomplishments of this thesis. Chapters 4 and 5 outline, review and constructively address the conceptual and theoretical problems raised in Chapter 2 and its discussion. By reframing and analysing the related issues of disunity in psychology and the specific conceptual conflicts between evolutionary and differential psychology, these theoretical contributions outline an adaptation-guided approach to the underlying empathetic mechanisms of prejudice, which permits the novel integration and synthesis of the diverse insights of social, moral, and differential psychology that are explored in the subsequent chapters. Chapters 6 and 7 recount the four major studies undertaken in the course of this thesis, the first of which established a new race-based manipulation of the Implicit Association Test required for later testing phrases, while the final three studies jointly explored the psychometric properties, moral measurement capabilities and prejudice measurement capabilities of the new *Sympathetic Attributions Towards Emotive Social Transgressors* (SATEST) task. These chapters (Chapter 7 in particular) draw key predictions from a synthesis of the relevant social, moral, and differential psychology literatures, made possible by the conceptual tools introduced in the theoretical chapters of this thesis. As such, this thesis, as a whole, is structured to illustrate the value of sophisticated theoretical and conceptual analyses, both in understanding the apparent conflicts of diverse research literatures and in generating innovating new approaches to empirical research.

By necessity, the following four chapters of this thesis cover a wide variety of topics within psychology, with each stand-alone journal article written to address researchers and theorists hailing from different fields, each representing distinct cohorts of presumed background knowledge. In the interest of making this thesis as coherent and accessible as possible to

readers of any scientific background, this chapter was written to provide the general contextual details necessary to engage with Chapters 4 through 7. Although each of these four journal articles, in the forms presented here, contain extensive literature reviews in themselves, each chapter (with the possible exclusion of Chapter 6) presupposes a familiarity with a general topic or issue that could be safely assumed of the readership of the corresponding journal, but not necessarily assumed for psychology professionals in general. What follows are concise general introductions to the key issues of psychological disunity and epistemology, explanation in evolutionary psychology, and the integration of insights from diverse fields, necessary for interpreting the contributions of Chapters 4 through 7. Elements of these subsections were once tentatively submitted as part of a review article to the journal *Review of General Psychology* early in the thesis, but were subsequently repurposed when the theoretical contributions of Chapters 4 and 5 were developed.

Understanding Disunity in Psychology

In order to appreciate the goals and contributions of the journal article featured in Chapter 4, it is necessary to have a sense of what is often labelled the disunity ‘crisis’ literature that both Chapter 4 and the special issue of *Review of General Psychology* that it was written in response to address. For a period of approximately 50 years (going back at least to Gladin, 1961) recurring criticisms have echoed through the psychology literature raising issue with the lack of integration between different schools of psychology. The central and reoccurring theme of this ‘crisis’ literature frames this disunity as a result of the conspicuous absence of a prescriptive, unifying framework to tie the behavioural and cognitive sciences together. Critics have written extensively (Kantor, 1979; Staats, 1983; 1999; Yanchar & Slife, 1997; Goertzen, 2008), with what almost appears to be a growing sense of incredulity, about the

problems faced by the fragmented science of psychology (de Groot, 1990). The central theme of these issues can be summarised as a 'wastefulness' that is not reflected in other natural sciences (Goertzen, 2008). In the words of Paul Meehl (1978), "it is simply a sad fact that in ... psychology theories rise and decline, come and go, more as a function of baffled boredom than anything else; and the enterprise shows a disturbing absence of that cumulative character that is so impressive in disciplines like astronomy, molecular biology, and genetics" (p. 807). As Goertzen's (2008) shrewd analysis of the 'crisis' literature observes, this issue of disunity is easy to construe as a threat to the 'legitimacy' of psychology as a whole. As Staats (1999) phrases it, "in fact, several have argued that psychology is a "would be" science because, unlike the "true" sciences of physics and biology, it has been unable to generate a consensually agreed upon conceptual framework that guides its scientific endeavours" (p. 4).

In response to the perceived 'crisis' of psychology, there are diverse camps in the literature, who pursue a variety of strategies in dealing with the perceived problems of disunity. There are those who acknowledge the fractured nature of psychology, but consider it an acceptable or desirable expression of mere specialisation (Dixon, 1983; Bower, 1993; Neisser, 1995; Kelly, 1998), in sharp contrast to those who suggest psychology is sufficiently unified as is, be it by 'clusters' of similar theories, or simply by methodological conventions (Baars, 1984, 1985; Matarazzo, 1987, 1992; Kassino, 2002; Stam, 2004). There is also a collection of opponents who attempt to rebuke the relative standard used, claiming that higher profile natural sciences (physics, biology, etc.) are merely concealing disunity problems akin to those in psychology (Overmeier, 1989; Viney, 1996).

In addition to these positions, there have also been many attempts over the years to meet the problems of disunity head-on, with a diverse series of unifying theories, or at least unifying grounding principles (Gilgen, 1987; Newell, 1990; Staats, 1996; Kimble, 1996; Anderson, 1996; Magnusson, 2000; Sternberg & Grigorenko, 2001; Henriques 2003, 2004, 2008; Gintis,

2007; see also the 2013 special issue of *Review of General Psychology*). While there has been great variety among these approaches with regard to what principles they advance as being central to psychology, several of the most recent attempts have been organised around core principles drawn directly from evolutionary theory. As briefly reviewed by Fitzgerald and Whitaker (2010), despite sources of consistent criticism from several sub-disciplines, the paradigm of evolutionary psychology appears to be growing in acceptance across psychology as whole, laying the seeming foundations of a unified psychology that will 'crystallise' around what is currently viewed as a discrete approach. While this is no doubt due to in no small part to the work of evolutionary psychologists who seek to reduce the paradigm's bad press and public misinformation (see Confer et al, 2010, which addresses many of these common misunderstandings), many evolutionary psychologists (e.g., Tooby & Cosmides, 2007; Daly & Wilson, 2008) call this spread predictable, due to the sheer utility of the paradigm's adaptationist approach. Since the theoretical and conceptual apparatus of evolutionary psychology serves as the uniting meta-theory that permits the integration of insights from social, differential and moral psychology in this thesis, the following subsection explicitly details how the adaptation approach to psychology is applied, and the explanatory power entailed in its use.

Understanding Epistemology in Psychology

To appreciate the conceptual and explanatory value of the adaptationist approach of evolutionary psychology, it is first necessary to frame the problem that explanatory theories in psychology seek to address. At the heart of this problem are the epistemological difficulties facing psychology as a science, which as Chapter 4 explores, are key to understanding the aforementioned disunity crisis and the value of theoretical common ground

between fields. The journal articles featured in both Chapter 4 and Chapter 5 describe the unique difficulties of psychological inquiry with reference to the engineering science concept of a 'black box'. As such the subsections of those chapters which frame the black box concept entail a degree of redundancy with this section of the thesis and with each other, though each paper applies the black box conceptual tool to slightly different ends. The account offered here is intended to provide general background on how the epistemological limitations of any scientific investigation present as uniquely problematic when investigating psychological phenomena.

Although views on what precisely constitutes 'science' vary considerably between sources, science generally consists of the systematic observation/description of, and the theory-guided explanation of, variations in a particular set of natural phenomena. The terms 'observation' and 'description' are defined here as having any empirical access to the phenomena in question, so as to permit an understanding of their occurrence, frequency and properties (observation), and being able to detail this process systematically (description). Explanation, on the other hand, is defined as establishing relations between elements or events, in such a way as to account for their causation over time (see Boag, 2011). To avoid some confusions, 'natural' in this context is used in the broad sense, to mean any commonly caused phenomenon that has not been contrived specifically for measurement. In the words of the unification theorist Gregg Henriques (2003), speaking of the distinctions made in E.O. Wilson's book *Consilience* (1998), "the goal of science, at least in theory, is to factor out human values and to develop representations of reality that are as accurate as possible" (p. 172), though researchers may disagree on whether said accuracy is defined in pragmatic or literal terms. With this variation-based definition of science in mind, it is instructive to frame the issue of epistemological access in science with regards to the black box conceptual tool.

In the engineering sciences, 'black box' is the catch-all term for any system that has traceable outputs, and generally traceable inputs, but of which one can gain little to no direct insight into the internal processes that bridge between inputs and outputs (Nairne, 1997; Sober, 1998). A black box system poses few difficulties for the task of observation/description, as these are generally concerned with the system's inputs and outputs, the elements to which we have direct access. Black boxes do, however, pose substantial challenges to the other central task of science, viz. explanation. Since explanation involves specifying elements and events, in order to then give an account of the causal relations between them (a phenomenon can only be 'explained' via reference to its antecedents which, in the past, caused its current state), black box systems present the undesirable situation of having observable phenomena (the outputs and inputs of the black box) which have causal relations to elements and/or objects that cannot be observed. Any explanatory account of the inputs or outputs of a black box must by necessity contain a space of incompleteness or speculation.

Acknowledgement of this philosophical stance provides the explicit grounding for one of the defining characteristics of the scientific method, hypothesis-testing. Hypothesis-testing is an algorithmic process comprised of both the generative and selective phases that most diagnostic procedures rely on (Fisher, 1925). It is common in the investigation of natural phenomena, particularly when seeking to describe and then explain the variations and patterns of variation thereof, to be faced with a situation where only a subset of the phenomena is available to you. While closed systems that cannot be viably penetrated by any available means are the prototypical examples of black boxes, black boxes can also be understood as relational to an investigator. How 'closed' a system is to investigation varies circumstantially, and as such black box limitations need not be defined only by what is contemporarily possible, but also by what is possible for particular researchers. Such limitations can range from momentary pragmatics (such as an ornithologist who cannot see

the tops of trees while on foot), to temporal limitations (the causal antecedents of some non-recurring event are now lost in the past), to the limitations of the physically unobservable (the original trajectories of uncertain quantum particles, which are deflected when photons are 'bounced' off them), but from the perspective of the investigator they all represent black boxes, in that they are amenable only to the hypothesis-testing of peripheral phenomena.

The central limit of hypothesis-testing is that a theory can only be supported definitively via the exhaustive disproving of all possible alternative hypotheses. For most kinds of black box situations, there are a functionally infinite number of alternative hypotheses for what may be the case in the hidden sections, and heuristics that guide investigators toward testing the most 'likely' or 'plausible' hypotheses are the saving grace that render hypothesis-testing even remotely practical. Such heuristics are generally drawn from theory, however, and black boxes become increasingly difficult to understand the more extensive or multi-layered the black box space is, and are also increasingly difficult the more diverse or baseline variable the measurable inputs and outputs are. It is with regards to the construction of explanatory theories, in such a manner as to provide instructive constraints on hypothesis testing, that the adaptationist approach of evolutionary psychology distinguishes itself as uniquely useful. Particular examples of how the adaptationist approach can integrate diverse findings into a coherent explanatory theory are outlined at the end of this chapter, and in Chapter 7. What follows below is a broad, background account of how explanations in evolutionary psychology can be understood in general.

Understanding Evolutionary Explanations

The defining aspect of any work of evolutionary psychology, whether named so or not, is the application of the adaptationist approach (Buss, 2005). While 'adaptationism' is sometimes

used to describe an ideological or philosophical stance privileging adaptations over other evolutionary forces and phenomena, the adaptationist approach in evolutionary psychology generally refers to heuristic methodologies oriented around the identification or disconfirmation of adaptations (Sober, 2000). In this context, an adaptation (when used as a noun) is understood to be a feature or set of features of an organism, the apparent design or concerted complexity of which suggest that it is a product of natural selection, and as such represents a relational calibration of said organism to its recurring environmental challenges (Tooby & Cosmides, 2005). The heart of the paradigm of evolutionary psychology is the suggestion that the species-typical behavioural and cognitive regularities of animals (usually humans), likely consist of or are shaped by adaptations.

In contrast to some arguments for a wider philosophical adaptationism (Sober, 2000), evolutionary psychology focuses on adaptations primarily for pragmatic reasons. Firstly, it must be acknowledged that while all organisms are the products of natural selection (with the addition of artificial selection in domesticated species), not all features of organisms are adaptations. In the words of John Tooby and Leda Cosmides (2005):

"The cross-generationally recurrent design of an organism can be partitioned into (1) adaptations, which are present because they were selected for, (2) by-products of adaptations, which were not themselves targets of selection but are present because they are causally coupled to or produced by traits that were, and (3) noise, which was injected by the stochastic components of evolution" (pp. 25-26).

For reasons of logical necessity, it is nearly impossible to use any positive criteria to confirm that some biological or psychological characteristic is either a by-product or phylogenetic noise. However, a feature can only be identified as an adaptation when it shows evidence of

'good design' with relation to the adaptive problem or problems it is presumed to address (Buss, 2005). Adaptations are, by their very nature, relations between organism characteristics and the fitness demands which statistically favoured those characteristics in the gene-pool (Sober, 2000), and no trait can be accurately described as an adaptation in the absence of this feature-problem matching. For this reason, the adaptationist approach begins with the postulation of adaptations, moving on to the possibilities of by-products and noise when the evidence for adaptation is inadequate (Tooby & Cosmides, 2005). The tell-tale signs of biological design are the clues used by evolutionary psychologists to generate and refine theories about the probable structure and development of a psychological adaptation, utilising the intrinsic relationships between the form and function of a well-designed system. Investigations of this sort are appropriately referred to as 'reverse-engineering' (Buss, 2005).

As was discussed above, when formulating explanatory theories regarding the hidden processes within a black box (in this case, the hidden psychological processes of the minds of humans and other animals), we rely on hypothesis-testing to disprove and discard those theories whose predictions are incompatible with the input and output phenomena we observe. In psychology, however, the observable outputs are highly interpretable (De Los Reyes & Kazdin, 2008), and the number of possible competing hypotheses for any given causal sequence are potentially infinite (Jaszczolt, 1996), so we must rely extensively on methods that refine our theories, such that only the most probable theories earn the investment of empirical testing.

In principle, there are three means of informing an explanatory theory prior to (or in conjunction with) prediction testing of the inputs and outputs. The first and usually most difficult option is to attempt to directly measure the contents of the black box. In psychology, the various methods of neuroimaging (and controlled lesioning, in the case of animal models) serve as our only direct indicators of the internal workings of living brains. While in the past

40 years neuroimaging and related direct measurement methods have been responsible for essential insights in cognitive science (Stevenson & Goldworth, 2002, and Tashiro, 2004 for overviews), their usefulness is ultimately limited. Beyond the contemporary pragmatic limitations of immense cost and technical difficulty, neuroimaging technologies only provide us with activity patterns, which while potentially closely correlated to the information-transformations of the mind, do not constitute measurement of the actual phenomena in question (Caplan, 2009). Even if neuroimaging were so refined as to accurately discern specific action potentials and the dynamic dendrite configurations of individual neurons, the interpretation of these patterns into meaningful psychological content could likely only be achieved following detailed correlation with some other source of insight into the processes in question. Thus while highly useful, neuroimaging can only be taken alongside psychological observations as means of testing and refining existing hypotheses (Bennett & Hacker, 2003).

The second option for refining theories independent of testing involves using logical inference to determine what must be the necessary minimum requirements of the systems in question, assuming that the systems are internally consistent. This method is extensively employed in computational cognitive psychology (Fodor, 1975) and is the central guiding heuristic of all computational models. While insufficiently discriminative in their own right, such logical inferences become vastly more powerful when supplied with alternative insights into the limitations of the psychological processes in question (for example, basic neurological insights into the properties of neurons, and regional clusters of the brain). Thus we are left to rely on the final option of refining explanatory theories, the independent discovery of design details. In mechanical and electrical engineering, such insights may take the form of acquiring early blueprints, learning what materials and tools were available to manufacturers, or learning what objectives the systems were designed to implement. In a

manner wholly analogous to the design of modern machines by human engineers, abundant evidence (Dawkins, 2009) suggests that all organisms were designed, through deep geological time, by passive biological forces of selection. In the design of human-made machines, imagination and memory are drawn upon to generate diverse forms and possibilities, which are selected among on the basis of production possibility and pragmatism, and in doing so, matching a design to the demand characteristics of a project. Similarly, in Darwinian biological evolution, diverse forms are generated by random mutation and recombination, which are in turn acted upon by the many situational forces of natural selection, in effect designing (by refining) the characteristics of organisms to match the survival and reproductive demands of their environment. While embracing the adaptationist approach is not strictly necessary to gain some of the crucial benefits of this third option (any biological, medical, and developmental insights into the properties of nervous-systems provide powerful tools for use with this second option), the adaptationist approach is designed to draw as much theory-guiding information as possible from the reciprocal relationships of form versus function (of what minds *'are'* versus what they *'do'*), and of *how* they operate versus *why* they operate (Hodgson & Knudsen, 2008). As Henriques (2003) summarises, the inference of adaptations plays directly into the development of constrained explanatory theories:

“If the presence of functional design is reasonably inferred, one then posits an adaptive problem that might account for the selection pressure that resulted in the present design. As with a detective who must establish motive, means, and opportunity for a suspect, a reverse engineer must effectively argue that the selection pressure was significant and that the design feature could have evolved given the phylogenic history. The explanation should be fundamentally consistent with available evidence, serve as a useful heuristic,

offer a parsimonious account of the evidence available, and ultimately make falsifiable predictions” (p. 168).

Developing an Integrated Approach

Much of the literature review of the journal article featured in Chapter 7 concerns the evolutionary synthesis of findings from social, differential, and moral psychology, to generate novel predictions and methods of measurement concerning the empathetic mechanisms underlying prejudice. By necessity, however, the literature reviewed in Chapter 7 endeavours to construct as straightforward a narrative of supporting research as possible so as not to distract readers from the insights that informed the development of the SATEST measure. What follows is a general account of how an evolutionary approach to the psychological phenomena shared by prejudice research in social, differential, and moral psychology, enables a insights from all three fields to be integrated within a common conceptual space. In continuation of the distinctions drawn in the discussion of Chapter 2, differential and moral psychology will often be grouped together for the sake of comparison with the large prejudice research body in intergroup social psychology.

Although sometimes studied side-by-side (e.g., Monin & Miller, 2001), the area of intergroup prejudice in social psychology and related constructs in differential and moral psychology descend from widely separated research traditions. Moral psychology has its roots in classical moral philosophy (Kohlberg, 1969; Rest et al, 2000; Morrow, 2009; Kristjánsson, 2010) and up to the present retains its primary focus on individuals and their moral-domain thoughts and behaviours. Though some moral psychology literature focuses on an individual’s holistic moral character (Dweck & Leggett, 1988; Dweck, Chiu & Hong, 1995; Feather & Atchison, 1998; Miller, Burgoon & Hall, 2007), the majority of the moral psychology literature

concerns ability-like recognition, reaction, and reasoning processes working in a generalised moral domain. Focusing primarily on cognitions that precede morally-loaded behaviour, both excluding (Rest et al, 2000; Narvaez & Bock, 2002; Krebs & Denton, 2005) and more recently including (Haidt, 2001; 2004; 2007; Nichols, 2002; Tsang, 2002; Ellis, 2005; Huebner, Dwyer & Hauser, 2009) emotional and intuitive components, moral psychology emphasises an individual's particular moral capacities, and the lifetime development thereof (Kohlberg, 1969; Rest et al, 1999; 2000; Narvaez, 2001; Harenski, 2010). These moral theories, particularly those pertaining to moral character and judgment, closely resemble a set of theories in the field of personality and individual differences, which conceive of interpersonal variations in one's general treatment of others as stable dispositions, or 'traits' in the personality sense (Schaller, Boyd & Yohannes, 1995; Reynolds et al., 2001; Duckitt & Sibley, 2010; Sibley et al., 2010; Akrami, Ekehammar & Bergh, 2011).

Intergroup prejudice, on the other hand, is an area of social psychology research concerned with the characteristic conflicts known to occur between both actual and perceived groups across many levels of demographic dissimilarity (Bernstein et al, 2010; Akrami, Ekehammar & Bergh, 2011). The literature largely consists of either generalised, group-level social effects (Bizman & Yinon, 2001; McCoy & Major, 2003; Shapiro & Neuberg, 2008; McGregor, Haji & Kang, 2008; Binder et al, 2009; Barlow, Louis & Hewstone, 2009; Tarrant, Dazeley & Cottom, 2009; Page-Gould et al, 2010; Christ et al, 2010), including aspects of ingroup dominance (Pratto & Shih, 2000; Troop & Pettigrew, 2005; Chow, Lowery & Knowles, 2008) and outgroup stereotyping (Haslam & Wilson, 2000; Gabarrot et al, 2009), or the personal and social impact of targeted negative ideologies towards particular groups, such as racial minorities (McGrane & White, 2007; Cohrs & Asbrock, 2009; Plant, Devine & Peruche, 2010; Butz & Yogeeswaran, 2011; Kteily, Sidanius, & Levin, 2011), homosexuals (Dasgupta & Rivera, 2006; Terrizzi, Shook & Ventis, 2010; Callahan & Vescio,

2011), and both binary genders (Eagly & Mladnic, 1989; Carr & Steele, 2009; Navarrete et al, 2010). There is a degree of social psychology research into the development of prejudicial behaviours (Aboud, 2003; Nesdale et al, 2005; 2007; 2010), but for the most part, intergroup prejudice research conceptualises a range of group-level social effects, which interact with contextual information both cultural and anecdotal to produce specific suites of antagonistic beliefs, expectations and behaviours between groups (Bernstein et al, 2010).

Taking a step back from these distinctions, it becomes clear that the social psychology study of intergroup prejudice, and the study of interpersonal dispositions in differential and moral psychology, represent two disparate research traditions which essentially deal with the same subject matter. Both, in a literal sense, are concerned with one's attitudes towards and treatment of others in a range of social situations. The two approaches merely focus on different aspects of these interactions, with the former emphasising group-facilitated factors and specific cultural contexts, while the latter focuses on the broad spectrum processes and dispositions that one employs in general contexts when negotiating harmful or beneficial behaviour towards others. Prejudice research in social psychology is oriented to overlook individual-level differences in how one generally treats and regards others in order to distil the effects of both group-membership and ideology on social behaviour. Conversely, differential and moral psychology generally overlook the influences of both group-comparisons and framing ideological associations, in order to form generalised, individually descriptive accounts of interpersonal judgments and tendencies. Implicitly, both approaches produce conceptual blind-spots, guaranteeing that neither methodology can encompass all relevant aspects of the social situations of interest. Though it is tempting to view this simply as specialisation, these core differences in epistemological assumptions render both the methods and findings of both fields difficult to reconcile in any meaningful way.

The epistemological methods of the adaptationist approach, however, approach both of these domains from a singular perspective: searching for evidence of psychological adaptations that have evolved in response to reoccurring fitness demands. In this case specifically, adaptations that pose general solutions to the ubiquitous challenges of efficiently managing social negotiations, reputations, and resource conflicts with other humans, some with which it is possible to cooperate and form allegiances or coalitions, and others with whom it is not. The distinctions between the two approaches may appear to be intuitive from a western, academic cultural perspective when endeavouring to abstractly classify phenomena. This is likely because we tend to conceptualise the former as diffusely occurring on the ‘level’ of groups, while the latter is synonymous with judgements and descriptions of individual character. However, from an evolutionary epistemological stance the distinction is wholly arbitrary, as the stimuli presented, necessary evaluations, and probable outcomes are near-identical on the level of the organism. Even if there were reliable enough circumstantial differences between the two approaches to suggest their evolution under distinct and separate selective-pressures, in order to be present as phenotypically adaptive both varieties of adaptations must be capable of activating simultaneously and in concert. Sufficiently autonomous adaptations would be likely to cause maladaptive conflicts under conditions as multifaceted as real-world social conflicts (see Cervone, 1999, for an account of modular conflict).

The contemporary literature of these three fields are beginning to show evidence that not only can an adaptationist approach successfully encompass these fields, but that evolutionary analyses appear to intrinsically favour the dissolving of such arbitrary divisions. For example, there is a growing trend in moral psychology theories, particularly those focusing on the prevalence of largely innate moral intuitions, to employ function-driven conceptions of moral processes, which specify environmental sensitivities to group-level concerns (Broom, 2006; Haidt, 2007; Jayawickreme & Chemero, 2008; Huebner et al, 2009; Rai & Fiske, 2011).

Furthermore, increasingly many social psychology theories of intergroup prejudice are, by necessity, coming to focus on individually-grounded and individually varying processes, particularly when drawing from affect- and intuition-based theories (McCoy & Major, 2003; Kreindler, 2005; Gutsell & Inzlicht, 2010). Considering the rise of evolutionary insights in widely used prejudice constructs, such as social dominance (Hawley, 1999; Zuroff, Fournier, Patail & Leybman, 2010), this trend appears likely to continue. Future adaptation-based theories in these areas are likely to eventually engulf other immediately functionally related phenomena such as the regulation of emotional displays (Denton, McKinley, Farrell & Egan, 2009; Griskevicius, Shiota & Neufeld, 2010). An evolutionary synthesis of the possible mechanisms that modulate subjective empathy in social interactions (and the subsequent justifications and attributions) is discussed at length in the early sections of Chapter 7.

CHAPTER 4

The Difficulties of Integration in Psychology

In order to generate novel theory and predictions concerning the empathetic mechanisms underlying prejudice, this thesis aimed to integrate a range of findings and insights from the largely distinct fields of social, differential, and moral psychology. The basic conception of this synthesis under an adaptationist account of shared psychological phenomena was described briefly in the final section of the preceding chapter, and is expanded upon further in the introductory sections of the journal article featured in Chapter 7. The preceding chapter also introduced the literature concerning the ‘crisis’ of disunity in psychology as a science, offering a brief account of how disunity of this sort is fostered by the epistemological difficulties inherent to the subject matter of psychology. As much of the crisis literature laments, the sometimes vast conceptual and theoretical differences between the various fields of psychology often render the empirical discoveries of one field difficult or impossible to meaningfully interpret within the conceptual space of another. The widely varying founding assumptions and methodological commitments of social, differential, and moral psychology are no exception to these difficulties, and a coherent conceptual and theoretical common ground must be established between the three fields if their insights are to be incorporated into a single predictive theory. The separate application of the adaptationist approach to each of these fields provided valuable reference points that the evolutionary synthesis of this thesis ultimately built upon, but in order for the majority of relevant findings in each field to be meaningfully reinterpreted within a new organising theory, it was first necessary to devise a means of detailing the paradigmatic differences between fields. Only when the similarities and departures between the theories and conceptions of these fields are clearly understood and compared, can empirical findings generated within the context of one field be safely translated into conceptual framework of another.

The journal article featured in this chapter was written to address precisely this issue. Toward the end of 2012, the journal *Review of General Psychology*, which has published pieces contributing to and responding to the crisis literature in years past, began development of their 2013 annual special issue specifically addressing the topic of unification in psychology. The editors requested short articles from over twenty theorists, each representing a different theoretical or methodological tradition in psychology, concerning how their respective traditions view the issue of psychological disunity, and what their traditions recommend as constructive integration efforts. In the hope of preparing a prompt and relevant response to the special issue, three of the proposed contributors (representing vastly different theoretical traditions) were contacted in early 2013, and were kind enough to provide early versions of their manuscripts. Drawing on these three articles as primary examples, and expanding the scope as necessary when the completed special issue was published in July 2013, the present journal article was written as a reflection upon, and response to, the special issue. By proposing a means of organising the diverse positions and recommendations advanced in the issue, this paper offers insights on how multiple fields of psychology can be integrated through the explicit conceptual analysis of foundational theoretical assumptions. The two conceptual tools introduced, *the ontological common ground* and *the continuum of pragmatic assumptions*, allow the diverse research traditions of psychology to be viewed as a single, tentatively branching project of inquiry, and as such the terminology introduced in this chapter is explored extensively throughout the remainder of this thesis when discussing the integration of interdisciplinary findings.

The following article has been submitted for publication to the journal *Review of General Psychology*, under the title ‘Unifying Psychology: Shared Ontology and the Continuum of Pragmatic Assumptions’.

Declaration for Thesis Chapter 4

In the case of journal article featured in Chapter 4, the nature and extent of my contribution to the work, and the contributions of the other listed co-authors is as follows:

<i>Name</i>	<i>Nature of Contribution</i>	<i>Contribution</i>
Tim Marsh	Decision concerning the topic of the paper	90%
	Search and review of the literature	
	Principle writing and editing of the manuscript	
Simon Boag	Advice on topic and approach	10%
	Assistance with editing and cutting	
	Suggestions for the refinement of the manuscript	

Unifying Psychology: Shared Ontology and the Continuum of Pragmatic Assumptions

Running Title: *Unifying Psychology*

Tim Marsh
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: timothy.marsh@mq.edu.au

Simon Boag
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: simon.boag@mq.edu.au

Acknowledgements: We would like to thank Gregg Henriques, Robert Lickliter, Agnes Petocz and Nigel Mackay, for kindly providing us with advanced copies of their special issue manuscripts, allowing preparation for this paper to begin prior to the special issues' publication.

ABSTRACT

Critics have described psychology as a science impaired by disunity. The most recent special issue of *Review of General Psychology* sought to specifically address this concern, seeking perspectives from a wide range of theorists, each of whom offered their tradition's approach to how psychology as a whole may be integrated into a more unified whole. To continue this discussion, this paper draws upon examples from the special issue, the disunity crisis literature, and wider writings in the philosophy of science, to explore the theoretical and conceptual divisions that foster ambiguity, confusion, and apparent irreconcilable differences between the disparate fields of psychology. The authors conclude that the majority of contemporary, scientific psychology is oriented towards a shared physical ontology, which can serve as a common grounding point from which the conceptual and theoretical differences of disparate fields may be meaningfully framed and evaluated. To this end, this paper proposes that the various research traditions of psychology can be understood through their positions along a *continuum of pragmatic assumptions*, which embodies the inherent conflict between two scientific priorities: metaphysical certainty (the *safe* end of the continuum) and practical experimental predictions (the *risky* end of the continuum). Three theoretical perspectives offered in the unification special issue are examined under this framework: Situational Realism (a distinctly *safe* approach), Developmental Evolutionary Psychology (an intermediate approach), and the Tree-of-Knowledge Unified Theory (a relatively *risky* approach). The authors explore how the recommendations of each approach can be seen as a function of its position on the continuum of pragmatic assumptions, and the implications of this understanding for future integrative efforts is discussed.

Keywords: unified psychology, integration, ontology, realism, evolutionary psychology

Unifying Psychology: Shared Ontology and the Continuum of Pragmatic Assumptions

For a period exceeding 50 years (going back at least to Gladin, 1961) recurring criticisms have echoed through the psychology literature raising issue with the lack of integration between different schools of psychology. The central theme of this ‘crisis’ literature frames this disunity as a result of the conspicuous absence of a prescriptive, unifying framework to tie psychology (and the wider behavioural sciences) together (noteworthy examples include Kantor, 1979; Staats, 1983; de Groot, 1990; Yanchar & Slife, 1997; Goertzen, 2008, 2011; Sturm & Mülberger, 2012). As was observed by Mandler (2011), psychology and its historical antecedents have faced several such crises of disciplinary disunity, with the present crisis representing only the most recent step in the difficult transition between speculative philosophy and natural science. Most recently, this year’s special issue of the *Review of General Psychology* (July, 2013) was specifically dedicated to reviving and expanding interest in unification, bringing together submissions from a wide diversity of theorists and inviting them each to argue the case for integration from the perspective of, and on the terms of, their respective research paradigms. These short articles, and the problems they each propose to solve, provide an opportune platform from which to compare and contrast contemporary efforts at unification.

One may argue that presenting nineteen distinct approaches (which themselves do not constitute an exhaustive list) serves primarily to demonstrate the multitude of disparate approaches that sympathetic theorists must struggle to integrate. However, close examinations of each perspective reveals encouraging and recurring claims to some conceptual common ground. As has been explored in the most recent works of Goertzen (2008, 2010, 2011), those fields within psychology most explicitly dedicated to scientific and experimental inquiries have begun to converge around a small number of highly influential explanatory approaches (notably information-processing, developmental systems and

evolutionary theory), while more peripheral traditions are clarifying their foundational differences so as to distinguish their efforts from the empirical mainstream (see Goertzen, 2011 for further detail). Despite this progress, it appears that now more so than ever, the goal of integrating psychology seems beyond the plausible reach of individual theorists seeking to court others to their frameworks with promises of comprehensive singular unified theories. Rather, in the contemporary landscape of multiple, differentially-viable meta-theories, each an emissary of an established school of thought with their own foundational assumptions and preferred empirical approaches, the goal of integration seems now to rely most on the slow dissolving of barriers between sub-disciplines (Mandler, 2011; Trafimow, 2012). While concerted attempts to cannibalise entire fields into their stronger contemporaries are not likely to be abandoned (nor necessarily should they be), the literature is primed for the emergence of innovative hybrid perspectives that rely upon an acknowledgement of conceptual compatibility and common definitional assumptions.

The aim of this paper is to propose and explore a new means of understanding the conceptual and theoretical disunities of psychology, by reframing the popular Kuhnian (Kuhn, 1962; 1970; 1996) perspective of scientific revolutions to reflect the nuances and interrelations of specific assumptions. In undertaking this analysis, the authors are specifically addressing the mainstream empirical commitments of modern scientific psychology, and the ‘realist’ ontological and epistemological positions that this entails. By focusing on the suite of conceptual assumptions that comprise a scientific paradigm (as in Wertz, 1999; Ribes-Iñesta, 2003; Goertzen, 2008), rather than on a paradigm as a whole, the authors suggest that the disunities of psychology can be understood as a predictable consequence of unique ambiguities of subject matter associated with the study of behaviour and cognition. Phrased simply, each sub-discipline is a partial-paradigm, sharing many assumptions with the rest of

psychology, while adding additional assumptions that have proved fruitful within their specialised domain of inquiry. In noting that these assumptions do not cluster arbitrarily, but build upon one another in hierarchical arrangements, the authors suggest that all grounding theories in psychology can be arranged along a conceptual continuum of assumptions. This continuum, if made explicit, can serve as a guide to resolving conceptual and theoretical conflicts between sub-disciplines, in a manner made impossible under the classical Kuhnian framework of incommensurability. To illustrate, this paper draws attention to three of the proposed unifying approaches in the recent special issue of *Review of General Psychology*: Situational Realism (Petocz & Mackay, 2013), Developmental Evolutionary Psychology (Lickliter & Honeycutt, 2013), and the Tree-of-Knowledge Unified Theory (Henriques, 2013), which are explored as occupying increasingly ‘risky’ positions along the continuum of pragmatic assumptions. In detailing key threads of compatibility between these examples that may foster enhanced interdisciplinary collaboration and theory building, the authors seek to highlight an underutilised path to integration that may assist in the gradual emergence of a unified psychology.

Disunity, Revolution, and ‘Normal Science’

When scrutinising the historical development of any scientific discipline, Kuhn’s (1962; 1970; 1996) model of scientific revolution has endured as one of the most conceptually influential approaches in the philosophy of science (see Boyd, Gasper & Trout, 1991; Sturm & Mülberger, 2011). For the sake of clarity, it must be noted that Kuhn employed the word ‘paradigm’ in a variety of senses (see Masterman, 1970), and this paper relies primarily on the ‘disciplinary matrix’ sense of the word (outlined in the postscript of Kuhn, 1970). While Kuhnian accounts of paradigm-shifts have been offered for all major branches of natural

science, critics (most notably Driver-Linn, 2003) have observed that the language and underlying concepts of Kuhn's model are particularly well-accepted in psychology. The specific relevance of the Kuhnian model to psychology is unsurprising, given Kuhn's reliance on psychology for key examples of pre-paradigmatic scientific practice, and his early stipulation that the paradigm approach is only appropriate to the goals of 'natural' science, but not 'social' science. As Driver-Linn (2003) elaborates, "psychologists, uncomfortably straddling natural and social science traditions, reference [Kuhn's approach] ... because it presents an intermediate, naturalistic position in the war between relativist and rationalist views of scientific truth" (p. 269). In order to appreciate the disunities of psychology, these philosophical and practical tensions between natural and social scientific practice must be addressed, as they are both an unending source of conceptual confusion, and are inherent to the subject matter of psychology.

The Revolutionary and the Normal

Central to the Kuhnian approach to scientific development are punctuated periods of revolution, contrasted with stable periods of knowledge accumulation within the constraints of the current paradigm, called simply 'normal science' (Kuhn, 1962; 1970; 1996). While the greatest controversies generated by Kuhn's *The Structure of Scientific Revolutions* (SSR; 1962) emerged in response to his claims concerning incommensurability and the (potentially) non-cumulative nature of science (explored in Boyd, Gasper & Trout, 1991; Michell, 2000), Kuhn was similarly pessimistic in his account of normal science and its intrinsic handicaps in generating truly novel discoveries. In his most provocative phrasing, Kuhn (1962) states that, "[n]ormal science does not aim at novelties of fact or theory and, when successful, finds none" (p. 52). In essence, Kuhn observed that most researchers working within any scientific

discipline were recipients, typically via their initial schooling, of the background knowledge, beliefs, and expectations implicit to the dominant paradigm. As such, contrary to the constant critical appraisal emphasised and idealised by Popper (1959) and the like, the majority of scientists (according to Kuhn) never re-examine the core theories of their field. Instead, knowledge is accumulated in a manner that may be considered ‘mop-up work’, in that it merely fleshes out and increases the precision of details whose existence is already presumed by paradigm.

Kuhn did not intend for this account of normal science to wholly demean the scientific enterprise, but considered it a both inevitable and essential process, since one cannot test hypotheses beyond the scope of one’s theory, and one cannot discover unexpected results without exhaustively exploring a theory’s predictions (see chapters 6 and 7, 1962; 1970; 1996). As Kuhn (1977) summarised, “[u]nder normal conditions the research scientist is not an innovator but a solver of puzzles, and the puzzles upon which he concentrates are just those which he believes can be both stated and solved within the existing scientific tradition” (p. 234). In this vein, scientific revolution must be preceded by a period of increasing tension and disciplinary disunity, as unexpected discoveries which cannot be accounted for under the current paradigm accumulate and proliferate within the research community. Though the findings yielded under one scientific paradigm are in some senses incommensurable with the framework of a new paradigm (in their original forms, at least, as comparing results requires common conceptual distinctions), a post-revolution paradigm will typically “resolve some outstanding and generally recognized problem that can be met in no other way” and “preserve a relatively large part of the concrete problem solving activity” of its predecessor (Kuhn, 1962, p. 168).

As far as the science of psychology is concerned, the persistent disunity of its various fields and the absence of any easily identified unified past ensures that the ‘crisis’ described in the literature carries comparatively little tension. Critics including Staats (1999; 2004) and Goertzen (2008) have observed that deep incompatibilities between branches of psychology are noted yet seen as unproblematic by many researchers, who accept such disunity as part of the ‘background noise’ of the discipline. This perception lends credence to the Kuhnian notion that the science of psychology is ultimately still pre-paradigmatic, which Driver-Linn (2003) notes is taken to be both pejorative and delegitimizing of modern psychology. Driver-Linn’s (2003) analysis suggests that researchers are thus motivated to regard psychology as possessing sufficient paradigm-like unity by default, an appearance supported by high rate of knowledge accumulation within (and sometimes between) the various fields, which by Kuhnian standards ought to only be possible within the guiding constraints of normal science. As such, theorists including Staats (1999) and Driver-Linn (2003) suggest that modern psychology can best be understood as a collection of related partial-paradigms, each of which supplies the necessary background assumptions and grounding theories necessary for their particular field, but is not sufficiently distinct in concepts or subject matter to be regarded as a wholly unique area of scientific inquiry. Since each area embraces a partial sphere of independence, taking the incongruities between fields as normal, Staats summarises psychology as “a plethora of diverse and unrelated scientific products but with little investment in unifying those products” (2004, p. 273). It will therefore prove instructive to focus more closely on the ways in which the partial-paradigms of psychology differ and overlap on the level of specific conceptual and theoretical assumptions.

Assumptions within Paradigms

The Kuhnian perspective has strongly emphasised the incommensurability of paradigms, particularly in the first and second editions of SSR (1962; 1970), but Kuhn took pains in the third edition to clarify that he did not support the suggestion of a truly arbitrary relativism between any two paradigms (1996). These ambiguities concerning how paradigms may interrelate or transform gradually, led Driver-Linn (2003) to investigate Kuhn's responses to criticisms that his structure of revolutions prescribed a needlessly universal and 'stage-like' pattern of development, many of which were raised by his colleague and then Harvard president James Conant. Over letters reviewing Kuhn's manuscript, Conant had suggested to Kuhn that discipline-spanning singular paradigms that undergo total revolutions oversimplified much of scientific development. Kuhn ultimately acknowledged these criticisms, but nevertheless no substantial revisions ever emerged in the editions of SSR published prior to his death (see Driver-Linn, 2003, for further details). If one permits the constraints of total paradigms and revolutions a degree of definitional flexibility, psychology becomes comprehensible as a multi-paradigmatic science, whose crises of disunity are a product of disparate fields growing increasingly irreconcilable as they pursue their specific domains and methods. In this sense, the great generation of findings within each field only serves to compound the problems facing psychology as whole. As Staats (2004) similarly notes, "because of its modern productivity, psychology's task of unification is much more difficult than that faced by the physical or biological sciences in their early development" (p. 273). Consequently, the dividing assumptions that initially separate any two fields can be expected to grow more entrenched in each field's practitioners as their respective research trajectories gain institutional momentum.

To understand how the partial-paradigms of psychology can relate to each other, yet exhibit many of the characteristic incongruities described in Kuhn's accounts of incommensurability (1962; 1970; 1996), one must clarify what precisely a paradigm consists of. Though Kuhn goes into some detail on the key elements of a 'disciplinary matrix' (a more technical term intimately related to 'paradigm'), particularly in the postscript of the second edition of *SSR* (1970), employing the four umbrella terms "symbolic generalisations" (p. 182), "metaphysical presumptions" (p. 184), "values" (p. 185) and "exemplars" (p. 187), the concept that shines the greatest light on paradigmatic conflict in psychology is described by Kuhn as "tacit knowledge" (p. 191). While the former four refer to firm conceptual and theoretical commitments that a researcher must (to some degree) be explicitly taught as part of their education in their particular field, tacit knowledge describes the unstated standards, distinctions and presumptions that one internalises through training and application of their research skills. For the sake of simplicity, this paper refers to all of these specific components of a paradigm with the general term 'assumptions', but the key point to draw from Kuhn's description of tacit knowledge and clashing norms is that an assumption need not be explicitly stated, nor even easily consciously identified by those who possess it, to engender rifts in understanding between paradigms.

In the case of psychology, with many partial-paradigms operating independently but linked by a shared name and history, large clusters of assumptions are shared across related fields (such as the assumptions related to neural information-processing in all areas of cognitive psychology; see Sternberg & Sternberg, 2012), while other assumptions offer unique theoretical grounding for specialised areas of inquiry (such as the assumptions of intra-psychic partitions in many areas of psychodynamic theory, including the famous Id, Ego and Superego—see Freud, 1923). While many such assumptions are explicitly described by

theorists as points of contention or divergence between fields (see Staats, 1999), assumptions that remain typically unstated, and which are therefore embraced and enacted uncritically, can in principle produce confusions and incongruities between fields that no party is capable of addressing directly. To this end, critics such as Kashdan and Steger (2004) have criticised the willingness of modern psychologists to tolerate apparent conceptual incongruities between fields, despite their shared subject matter, suggesting that “paradigmatic rigidity is retained without methodological rigor and creativity” (p. 272). In seeking a solution to this issue, it is necessary to audit the subject matter of psychology in general, in order to take account of why the assumptions that divide fields emerge, and how a defensible common ground may be identified and expanded.

The Subject Matter of Psychology

Describing psychology as a multi-paradigmatic science, with each partial-paradigm representing a suite of both explicit and unstated assumptions (some shared between fields and others diverging wildly), allows one to appreciate the unique disunity of modern psychology as compared to the histories of the strictly physical and biological sciences (Staats, 1999; Henriques, 2003, 2011). That said, theorists including Goertzen (2008) and Mandler (2011) have argued that this high degree of persistent disunity cannot be attributed solely to accidents of historical circumstance, but rather, must be understood as a predictable consequence of the difficulties inherent to the subject matter of psychology. Psychology sits at the border between the natural and social sciences (in the senses employed by Kuhn, 1970, and followers), with much of psychology seeking to understand social phenomena as a special instance of material natural phenomena (see Henriques, 2003, for a brief overview), and in that sense, must simultaneously negotiate the challenges of both perspectives, with

appropriately inclusive conceptual apparatus. The anthropologist Clifford Geertz (2000) summarised the problem as follows: “[psychology] has not just been troubled with a proliferation of theories, methods, arguments, and techniques. That was only to be expected. It has also been driven in wildly different directions by wildly different notions as to what it is, as we say, “about”—what sort of knowledge, of what sort of reality, to what sort of end it is supposed to produce” (p. 187).

Of course, some variation is to be expected, since the various fields focus on different aspects of the sum behavioural, social and cognitive phenomena broadly associated with psychology. However, the goal of this discussion is to probe beyond this expected variance in methods and approaches to address the subject matter of psychology on the level of ontology, which is to say, in what form do researchers understand the phenomena in question to exist. Though there are peripheral areas of psychology for which this may be a decided non-issue (some of which are discussed in Goertzen, 2010), the entirety of experimental and scientific psychology is, by purpose, committed to the investigation of real phenomena (Staats, 1999; Henriques, 2003, 2004, 2011). When comparing across contemporary fields this commitment cannot easily be met, since terms such as ‘constructs’, ‘traits’ and ‘mental representations’ are used in a variety of ontologically distinct manners across the research literature (see Michell, 2000, 2003a, 2006, 2013; Boag, 2011; Marsh & Boag, 2013, for several critiques). Many researchers use such terms as mere shorthand, to describe configurations or activity patterns in physical neurological structures (as argued in Anderson, 2010), while many others use such terms in a literal sense, to describe a ‘mental’ or abstractly ‘psychological’ character that merely relates to, or superimposes upon, physical phenomena (see Charles, 2011). What is troubling is the strikingly widespread occurrence of researchers employing both senses of such terms (or partial forms somewhere in-between), signifying perhaps confusion over the implied ontological status, or more commonly, a simple lack of concern over what real

phenomena are ultimately being described. To account for the persistence of these issues within psychological science, it is instructive to briefly consider the relationship between the scientific process itself and the domain of natural phenomena it is employed to explore. To this end, it is essential to frame the practical limitations human investigators must struggle with in order to measure elusive phenomena, a problem most succinctly summarised with the metaphor of a 'black box'.

Lost in a Black Box

In the engineering sciences, a 'black box' is the catch-all term for any system that has traceable outputs, and generally traceable inputs, but of which one can gain little to no direct insight into the internal processes that bridge between these inputs and outputs (Nairne, 1997; Sober, 1998). While such black box systems pose few difficulties for the scientific tasks of observation and description, as these are generally concerned with the system's inputs and outputs (the elements to which we have direct access), black boxes do, however, pose substantial challenges to the other central task of science, viz. explanation. To explain a set of phenomena, one seeks to establish relations between elements or events, in such a way as to account for causality over time (see Boag, 2011). Systems of phenomena that must be regarded as black boxes typically present the undesirable situation of having observable phenomena (the outputs and inputs of the black box) which have causal relations to elements and/or objects that cannot be observed (that which is 'inside' the figurative box). Thus, any explanatory account of the inputs or outputs of a black box must by necessity contain a space of incompleteness or speculation.

To address this difficulty, the scientific method places hypothesis-testing at the heart of its theoretical enterprise. Any explanatory theory is only as strong as the testable hypotheses it generates, and only failures to disprove these hypotheses provide evidence that the theory should be tentatively accepted. The corollary problem is that for most kinds of black box situations, there are a functionally inexhaustible number of alternative hypotheses for what may be the case in the hidden sections, and heuristics that guide investigators toward testing the most 'likely' or 'plausible' hypotheses are the saving grace that render hypothesis-testing even remotely practical. Such heuristics are generally drawn from theory, however, and black boxes become increasingly difficult to understand the more extensive or multi-layered the black box space is, and are also increasingly difficult to understand the more diverse or variable the measurable inputs and outputs are (see Sober, 1998, for further discussion). As such, for any explanatory theory, or indeed for multiple theoretical frameworks seeking to share data, a clear understanding of what classes of phenomena must constitute the intermediary stages is indispensable, for these distinctions will fundamentally shape what testing approaches are likely to bear fruit.

It is here that the characteristic difficulties of psychology (and behavioural science in general) become clear. The subject matter of psychology, while easy to state in pre-scientific terms (i.e., terms in general use such as 'behaviour', 'thought', etc.), has defied straightforward clarification for much of history (Kaitaro, 2004). To focus on the 'behaviours' of humans and other animals (itself a problematically vague concept), one confronts the sheer interpretability and framing problems of complex, subtle, and continuous actions (see Jaszczolt, 1996; De Los Reyes & Kazdin, 2008, for details), a task that is obscured and complicated rather than simplified by pre-scientific and folk intuitions (see O'Donohue, Callaghan & Ruckstuhl, 1998). Conversely, to focus on the intermediate factors between perceptions and behaviours

requires the reverse-engineering of some of the most delicate, inscrutable, and bafflingly complex structures in the natural world: the brains and nervous systems of sophisticated animals (Walsh, 2000). While the explicit Cartesian superstitions of early inquiries are in continuous decline due to the growth of our physical and biological understandings (in scientific practice, at least; see Kerr, 2008), much of the original uncertainty concerning the possible forms of causal intermediaries has endured.

In attempting to guess at the contents of the black box of psychological phenomena, one must make educated assumptions about what one expects to encounter, and evaluate and revise these assumptions on the character of subsequent findings (see Kuhn, 1977, for general discussion). However, given the vastness of the black box and the interpretability of the measurable outputs implicit to the subject matter of psychology, any assumptions that a research tradition adopts to begin their inquiries will invariably shape the criteria against which these and subsequent assumptions are assessed and revised (Jaszczolt, 1996; De Los Reyes & Kazdin, 2008). Thus, it would appear that much of the disunity in psychological science emerged due to the disparate starting assumptions embraced by different domains of inquiry (perception, memory, personality, etc.), and that the entrenchment of these assumptions, both explicit and unstated, now maintains the apparent incompatibilities between psychology's various approaches and fields.

If the early and foundational assumptions that have shaped each field of psychology could be thought of as simply extended hypotheses to be tested, the disunity in question would be purely theoretical in nature, and one might expect disparate theories to gradually fuse and combine as each field continues to accrue knowledge and facts on related topics. Indeed,

theories have converged and fused in precisely this manner, though typically only within single fields (see Sternberg, 2005, for discussion, and Mandler, 2011, for examples). However, in an a manner alluded to by Kuhn's (1996) account of incommensurability, the assumptions embraced by each research area provide more than theories or models to be tested, but also provide the more basic conceptual distinctions and categories that such theories are built upon. For example, much of modern cognitive psychology employs concepts of signals and representations (as discussed in Bechtel & Abrahamsen, 1991) derived from insights in information theory and computation, as tools in understanding perception, thought, and behaviour. The use of these concepts in a theory necessitates an often unstated commitment to a particular ontological account of *how* and *where* psychological phenomena exist. For any theorist who has not embraced (or is at least aware of) this suite of cognitive psychology assumptions, these conceptual categories and any theories built upon them are at best confusing and inaccessible. What is worse, given the parallels of terminology that can be found across disparate fields (with general terms such as 'beliefs', 'reactions', etc.), there is an omnipresent risk that incompatible conceptions hailing from distinct traditions may become easily confused by researchers drawing from, or writing to, several fields at once. Such confusions, though seemingly innocuous and easily committed from the perspective of any one researcher, are cumulatively disastrous to our collective understanding of how disparate fields do, and do not, overlap.

With this risk in mind, one can appreciate the value of mutual vigilance between researchers and between research fields. As Machado, Lourenço, and Silva (2000) have argued, the subject matter of psychology necessitates a balanced epistemological approach, attending to facts, theories and concepts with comparable degrees of care. As a highly productive but deeply disunified modern science (at least, according to Staats, 2004), which still suffers from

threats to its institutional legitimacy (Michell, 2003a), psychology has placed an excessive emphasis on accruing ‘measurable’ facts across its fields, while neglecting much of the theoretical and conceptual analyses necessary for these findings to meaningfully converge on a shared understanding of psychological phenomena. In the interest of making such convergence possible, it is crucial that researchers establish a well-defined set of shared conceptual distinctions and theoretical commitments, for without such a common ground, our ever-growing wealth of empirical findings are doomed to isolation on increasingly disparate trajectories of inquiry.

Common Ground

Although the very nature of psychological inquiry has cast researchers adrift in a sea of interpretive possibilities, Valsiner (2009) notes that not all starting assumptions are equally arbitrary and several viable points of convergence have crystallised throughout the literature over its history. As noted earlier, three general and compatible explanatory approaches stand at the focal points of the most successful integration efforts in experimental scientific psychology (Mandler, 2011). Two of these, perhaps understandably, are an inheritance from the successes of the biological sciences, namely, *evolutionary adaptationism* (Buss, 1984, 1995; Cosmides & Tooby, 1989; Tooby & Cosmides, 1989) and *lifespan development* (Richardson, 1998; Lickliter & Honeycutt, 2003; Michel & Tyler, 2007). The third has emerged with the aid of technological insights into physical computation, namely, the *information-processing approach* (Fodor, 1975, 1983; Hilbert, 2009). With regards to explanation, each of these perspectives offers researchers a grounding insight into how and why key elements of psychological phenomena exist (i.e., matching organism-environment characteristics, emergence of abilities through maturation, generation of complex responses,

etc.). Through these insights, and their pervasive connections to the other natural sciences, modern psychologists are at last in a position to offer a definition of the subject matter of psychology that is immediately grounded in a concrete, material ontology. Under this suite of assumptions, the nervous systems of animals, including humans, are thought to be comprised of neuronal tissues whose cells connect in dynamic patterns to process information. The basic organisation of these structures emerges from an evolved genetic inheritance, which interacts with the environment over the course of ontogeny to produce individual configurations capable of ongoing calibration and learning. The overt reactive behaviours of such organisms are the result of both real-time sensory stimulation, and acquired biases and variations in neural structures owing to past experience. From this increasingly influential perspective, as Gazzaniga (2010) notes, it is these functional patterns and organisations that are the definitive domain of ‘psychological’ phenomena, over and above what may be considered merely neurobiology.

Nevertheless, despite the apparent successes, disputes remain. While the broad facts of this ontological common ground appear uncontroversial in most of scientific psychology, it is not at this most basic level where disagreements tend to emerge. Rather, disagreements between the disparate schools of psychology tend to focus on the perceived differential relevance of this basic ontology to their respective phenomena of interest. For example, as Vul (2011) argues, the hard details of neurophysiology and cognitive computation are understood to form the basis of the interactions studied by social psychologists, but a social psychologist would consider only certain emergent activities of such cognitive systems (particularly those expressed between persons) as their relevant subject matter. From a strictly pragmatic perspective, there is merit to objections of this sort, but with regards to theory, to adopt the position that the subject matter of other fields should not encroach on your field’s subject

matter (and thus is beyond your field's concern), is to handicap the prospect of meaningful integration *a priori*. For any area of psychological inquiry that pragmatically eschews the content of other fields (which, depending on how fields are defined, could be the majority of areas), we can expect to see the fundamental black box limitations of psychological phenomena re-emerge. The gulf of details between the base matter of the brain and its associated behavioural manifestations is so wide, that every niche of psychological inquiry must make some assumptions concerning the conceptual intermediaries that comprise the subject matter of other fields. This is to say, while social psychologists may seek to eschew the details of neurophysiology, and neurophysiologists in turn may seek to eschew the details of social contexts and interactions, each field invariably makes general theoretical commitments concerning the form that these eschewed phenomena are likely to take. Even fields of psychology as conceptually distant as these two examples cannot remain truly 'agnostic' with regard to the defining questions of other fields, because their position as part of a larger whole is the key overlap of their founding assumptions (Vul, 2011). Substantial innovations, or perhaps even revolution, within any partial-paradigm of psychology will not only affect the field in question, but will change the character of the assumed intermediaries that grounded the division of subject matter between fields in the first place. For example, if the information-processing perspective of cognitive psychology were overturned tomorrow, not only would the field of social psychology change with it, but the grounding assumptions concerning how and why it is possible to study social processes at all would need to be revised as well. Thus, when differences in founding assumptions produce multiple viable candidates for the intermediary phenomena or processes that may occupy some section of the psychological black box (such as the historical competition between the 'reactive' and 'hydraulic' models of aggression), the impact of such conflict extends beyond the fields in question, to influence any other field that adopts one stance or another as a background

assumption. As such, it is vital for researchers to remain explicitly aware of the assumptions that tie their field to the empirical status of others, for these shared assumptions offer guidance as to what other areas of psychology do and do not share a conceptual common ground.

For the sake of compatibility with the aforementioned suite of founding assumptions, ideally all intermediary phenomena would represent hypothetical organisations of neurological structures, defined by either their relevant functions or their literal anatomy. However, the diverse research goals and histories of the various traditions of psychology have given rise to innumerable postulated *psychological* entities that were not conceived to fit this ontological framework (such as ‘constructs’, ‘traits’, and ‘mental representations’, mentioned above). Indeed, there are many such proposed concepts that may be ill-suited for any ontological specification at all (such as those thought to exist exclusively between-persons, which exist as relations, but have no independent substance). Since different research traditions demonstrate differential degrees of explicit commitment to the aforementioned ontological common ground (or in some cases, to any ontology at all), the current literature is saturated with convenient common terms (such as ‘traits’ and ‘representations’) which are employed in distinct, often incompatible senses.

Unseen, Confused, or Ignored Distinctions

To illustrate the problems that can emerge from a lack of ontological grounding, the present authors reviewed a contemporaneous cross section of published psychology research, to gauge the degree to which each paper demonstrated referential vagueness, confusion, or

evident contradiction, concerning the ontological status of its subject matter (single issues selected randomly from the year 2012). Terminology was judged as being problematically vague when the ontological status of the phenomena described (that is, some account of whether it is to be understood as a literal object, a functional abstraction, or a descriptive metaphor) remained unaddressed throughout the length of the paper. Similarly, papers which reference or imply multiple accounts of the ontological status of a single term were taken to be confused, or as contradictory when at least two of these accounts were mutually exclusive (as in a tension between literal and metaphorical meanings). In the interest of fairness, the three journals selected were all highest-tier APA or APS publications, each with a strong focus on experimental empirical science: *Psychological Bulletin* (Volume 138, Issue 2), *Psychological Review* (Volume 119, Issue 1), and *Psychological Science* (Volume 23, Issue 7). Focussing only on those articles depicting full research results or reviews, a total of 31 papers were assessed. Of those 31 articles, 13 (or 42%) employed key terms or concepts that were used in an ontologically confused or contradictory manner, inconsistently regarding common terms as both literal and metaphorical in separate instances. For example, Freund and Kasten (2012), in their study of self-estimates of cognitive ability, take great care in much of their terminology, but employ the general term ‘cognitive ability level’ as sometimes representing an abstract aggregate of tested behaviours and outcomes (e.g., p. 297), and other times representing an actual level of some causative phenomenon within an individual, particularly when generalising the practical implications of their findings (e.g., p. 314). Beyond this, 23 (or 74%) of the articles cited and built upon at least some previous research papers drawn from both literally and metaphorically defined usages of common terms. A clear example can be found in model proposed by Kruglanski et al. (2012), which employs a conception of ‘mental resource’ that is interchangeably informed by highly non-literal approaches, such as Lewin’s (1951) and Deutsch’s (1968), as well as more process-oriented

and materialist approaches such as those in Schmeichel, Vohs and Baumeister (2003). Finally, perhaps most troubling were the 17 (or 55%) of articles which, in their own descriptions and explanations, identified either no ontological leaning concerning their subject matter, or made ontological references so vague as to permit interpretation in any combination of literal or figurative definitions of psychological terms. These broad trends signify both a lack of attention and a lack of concern among many psychology researchers regarding what the subject matter of their studies is presumed to be, and what underlying assumptions would inform these judgements. Readers should not, however, take these figures as a condemnation of the authors in the journals described, but rather as a conservative indication of the magnitude of this problem. By a considerable margin all three of the journals examined here demonstrate far greater scrutiny and higher scientific standards concerning these and related conceptual issues than can commonly be found in the literature as a whole, making the problem all the more striking.

As was explored above, these conceptual and definitional problems represent an ongoing source of disunity for psychology, as they not only impair integration efforts between potentially complimentary fields, but can also create the illusion of integration between perspectives that are, in fact, ontologically incompatible. This is a major cause of concern since conflicts in scientific perspective can only be fruitfully addressed when directed towards a common, objective subject matter. As Henriques (2003) writes, “agreement about the phenomena under examination is needed prior to healthy scientific disagreement about particular issues. Without such prior agreements, opponents cannot agree on the questions to ask, which greatly limits the value of answers offered by the empirical process” (p. 152).

The Continuum of Pragmatic Assumptions

As the recent special issue of *Review of General Psychology* (July, 2013) demonstrates, many theorists from competing traditions seek to establish the particular suite of assumptions inherent to their approach as the most fitting arbiter for all of psychology. These disparate approaches are pitted against one another as representing alternative conceptions of the subject matter of psychology, complete with theoretical and methodological recommendations specially tailored to the subject matter as they see it. While all such proposals are certainly not equivocal (with some appearing to offer a more comprehensive framework than others), critics such as Goertzen (2008; 2011) note that these top-down attempts to convince researchers to abandon their existing assumptions and methods are unlikely to succeed. This is because, as Driver-Linn (2004) observes, adopting the background and practices of a particular field requires that researchers “pick a side (against their colleagues)” (p. 271), and maintain their commitment to the traditions of their fields by perceiving the problems that their framing may yet solve. Thus, while institutional changes to the training and education of psychology researchers could potentially eventually elevate a single perspective to a position of dominance over all other perspectives, there are far-reaching benefits to the gradual dissolving of disciplinary incommensurability. The most obvious benefit, of course, is the preservation and greater utilisation of the talented minds and successful research projects that are currently immersed in the assumptions of their various fields. Beyond this, dissolving the boundaries between incommensurable disciplines allows the competitive strengths of different approaches to be explored in active discourse, and under ideal circumstances, may take considerably less time than the generations required to silence dissenting voices through old age (as Kuhn sometimes alluded to, e.g. p. 151 and 152, 1970).

How can disciplinary incommensurability be addressed? A first step is to clarify the subject matter of psychology. It is noteworthy that for many the perceived incompatibilities between various fields of psychology are often described and understood as differences of subject matter (Neisser, 1995; Kelly, 1998; Stam, 2004), and indeed in some cases this is undeniable (such as the differences between mechanistic and deeply constructivist approaches; see Stam, 1990; Botella & Gallifa, 1995). However, as was explored above, in the largest areas of modern scientific psychology disagreements do not typically concern what kinds of phenomena are thought to exist. Rather, incompatibilities of theory and concept can be understood as differences in the assumptions embraced by disparate fields and traditions, many of which have become implicit and remain unstated to their adherents, and as such cannot be easily called upon to explain and resolve points of confusion. As critics such as Machado, et al. (2000) have argued, a greater degree of theoretical and conceptual analysis could allow such clashes to contribute meaningfully to scientific development and the interpretation of findings. However, the first step in such a process would require that every tradition in psychological science closely examine its ontological and epistemological commitments, in order to make its entire suite of assumptions clear and available to explicit scrutiny. To do so would not merely clarify the true parameters of divergence between any two theories one may wish to compare, but would make the research findings of competing research fields interpretable as the tentative results of an elaborately explored set of hypotheses.

The present authors propose then that many of the unifying frameworks that have been recently offered, and indeed many unrepresented theories in the wider literature, may be brought into a mutually acknowledged common conceptual space via their acceptance of, and commitment to, a shared ontology concerning the subject matter of psychology (as outlined

above). To this end, we suggest that the defining distinctions of each theoretical approach be regarded not as dogmatic necessities, but rather as extended tentative hypotheses along a *continuum of pragmatic assumptions*. This notion of a continuum is grounded in the observation that the patterns of assumptions embraced by different traditions in psychology are not arbitrary, but instead can be thought of as hierarchically arranged, with the more complex and tenuous assumptions built upon the more basic and certain ones. For example, branches of cognitive psychology, including the majority of evolutionary psychology, rely on the concept of functionally-delimited cognitive ‘modules’ in generating hypotheses about psychological processes (see Barrett & Kurzban, 2006, for an overview of the concept). In doing so, these researchers are relying upon an assumption concerning how neuronal systems are likely to be organised, particularly as a result of natural selective pressures. This assumption does not stand alone, however, as it is inextricably grounded in a range of computational assumptions that are more widely embraced throughout cognitive psychology (Fodor, 1975; 1983), which in turn are based upon a set of assumptions concerning the physiology of the human nervous system that are more widely embraced still (Dewsbury, 1991). These hierarchical connections can be thought to extend in branching paths, from those fundamental assumptions, generally well-supported so as to be regarded as ontologically certain and ubiquitous (such as the facts concerning the physical composition of human beings), through to the most tenuous and niche-specific assumptions embraced only within particular fields.

The conservative nature of scientific practice ensures that any novel assumption advanced by a research tradition is likely to be only an incremental extension beyond what that tradition has taken to be reasonably certain. Furthermore, as Kuhn (1970, chapter 9) reflected upon in his account of framing new paradigms, new assumptions are typically introduced as a

possible means of addressing problems that previous framings struggle with. In the aforementioned example, researchers who embrace the assumptions of cognitive modularity gain a powerful new means of structuring their theories and generating testable hypotheses. Furthermore, as is often the case when employing hypothesis-testing to chart a vast black box (Sober, 1998), the most productive means of exploring the truth or viability of a logically coherent possibility (such as that of a specific cognitive module) is to tentatively assume its existence, and examine the results derived from this assumption for contradictions and inconsistencies. As such, while adhering only to the more basic and well-verified of assumptions entertained in psychology is a viable strategy to avoid wasting one's time fleshing out possibilities that may ultimately prove false, to do so would be to embrace a relative handicap in the generation of new theories and hypotheses, as compared to traditions that have accrued a more adventurous suite of assumptions within their niche. That said, from an interdisciplinary perspective, is it crucial that these less certain assumptions be embraced as tentative and conditional upon competitive verification, in acknowledgement of the wide range of possible assumptions that could conceivably provide a superior alternative in explaining psychological phenomena. In this sense, diverse traditions that rely on collections of assumptions not shared by their disciplinary alternatives can indeed be regarded as in extended hypotheses competition, vying to construct reliable new insights upon a common ontological base.

There is obviously insufficient space within a single journal article such as this to offer an extended treatment of each of the three examples explored hereafter. As such, each example shall be addressed primarily with regard to the unique assumptions defining their approach, and both the integrative prospects and implied incompatibilities that commitment to these assumptions suggests. In comparing these examples, we shall draw attention to the range of

conflicts that emerge when the pragmatic assumptions underpinning a theory are rejected (or simply questioned) by others. In this sense, the two opposing extremes on the continuum of pragmatic assumptions can be regarded as the metaphysically *safe* end, characterised by theories which make few uncertain assumptions but incur empirical disadvantages, and the metaphysically *risky* end, characterised by theories built upon many potentially false assumptions but which gain empirical advantages within a theoretical niche. Suffice it to say, a theory's position along the continuum of pragmatic assumptions will prove instructive in understanding both the theory's recommendations for integrative change, and in predicting which other approaches the advocating theorist will likely disapprove of. In aligning these examples along the continuum, and exploring whether their unification strategies are calls toward philosophical safety, or towards pragmatic risks, we aim to demonstrate the underlying compatibilities of many fields of psychology, whilst illustrating the indispensable role of clear assumptions and conceptual analysis in unifying psychology.

A Pull towards Safety – Situational Realism

As was outlined in Petocz and Mackay (2013), Situational Realism is a psychological research tradition that has emerged from the intellectual legacy of the philosopher John Anderson (see also, Mackay & Petocz, 2011, for a detailed cross-section of the current state of Situational Realism). While there is some degree of conceptual overlap between Situational Realism and other contemporary philosophically realist traditions in psychology (compare, for instance, Charles, 2013, Heft, 2013, and Tonneau, 2013), the Andersonian approach can be distinguished by its particularly staunch commitment to: (i) a strictly monistic material ontology; (ii) the conceptual emphasis placed on the infinite complexity of real situations, and; (iii) the centrality of the distinction between objects and relations. In this

view, all acts of cognition and knowing in humans (and other animals) are construed to be relations (or complex combinations of relations) between an organism (or relevant systems comprising the organism, e.g., drives and the perceptual apparatus) and a real situation (or specific aspects comprising a situation). Although potentially compatible with organism-environment interaction accounts offered by the other aforementioned realist and ecological approaches (particularly those of the Gibsonian and Holt traditions), this emphasis on relation allows one to conceive of conventionally ‘mental’ events without a need to postulate ontologically questionable or untenable entities (Maze, 1991). Rather, ontologically real spatio-temporal things (or particular aspects thereof) are understood to be the objects of cognition, constrained and subject to error on the part of the knowing subject by the physical and causal structures that make the relation possible (such as the fallible apparatus of an animal’s eyes and ears).

As Petocz and Mackay (2013) note, the approach of Situational Realism is not well-known in international circles, and has thus far contributed primarily theoretical contributions and conceptual clarifications, rather than empirical findings (though contributions focusing on the issue of measurement are particularly noteworthy, see Michell, 2006). This situation reflects perhaps the most distinctive characteristic of the Situational Realist approach, an unwavering commitment to strict logical and conceptual forethought, and a subsequent reluctance to embrace theoretical and methodological assumptions that stand upon uncertain metaphysical foundations (e.g., Maze, 1991). This commitment is not made purely on principle, but is suggested as a solution to the insidious conceptual problems that abound in psychological research (Michell, 2000), due to the misleading character of popular terms (for example, ‘ultimate’ causes in evolutionary theory can be construed teleologically; references to ‘mental resources’ can be taken as subscribing to Cartesian dualism, etc.). According to Situational

Realists, allowing the use of such metaphysically uncertain terms cultivates needless confusion, and offers potentially false findings built upon logically unsupported assumptions (Hibberd, 2009). As such, with regards to the black box nature of psychological subject matter, Situational Realism seeks to avoid many of the aforementioned metaphysical risks of postulating hypothetical causal intermediaries, by focusing instead on the logically necessary components of any process that is conceived as a relation (typically, as subject and object terms, see Maze, 1991).

On the continuum of pragmatic assumptions, it is clear that few research traditions stand as close to the *safe* extreme as Situational Realism. As such, the primary benefits of adopting this conservative stance are described by Petocz and Mackay (2013) as “clarification and redirection” of the mainstream efforts of psychologists and cognitive scientists away from contemporary research trajectories in “cognitive neuroscience and information processing” (p. 217). Instead, what is advocated is a wholesale shedding of the many of the assumptions made in experimental psychology, and in their place assuming the more defensible and basic propositional and relational terminology of realism, in the hopes of readdressing all psychological phenomena in a manner less encumbered by unsupported and confused theoretical and conceptual baggage. The existing wealth of empirical research findings in psychology would not be discarded, on this view, but rather carefully re-examined and reinterpreted paying close attention to the set of assumptions under which the original hypotheses were proposed. This approach is considered viable, since regardless of the initial intentions or interpretations of researchers, all empirical findings are ultimately accounts of real spatio-temporal situations (Petocz & Mackay, 2013).

The present authors agree that the founding assumptions of many psychological fields are both wrongfully regarded as certainties, and possess potentially confusing conceptual terminology. As such, the metaphysically cautious approaches embodied in Situational Realism offer theorists a range of valuable conceptual tools (most notably, binary and ternary cognitive relations), with which to scrutinise questionable conceptions in psychology (such as the many senses of the term ‘mental state’) without committing to their distinctions prematurely. However, while it is generally wise to err on the side of caution, this conceptual conservatism is itself vulnerable to some of the terminological misunderstandings discussed earlier in this paper. When the pragmatic assumptions at the core of a research tradition become implicitly and unreflectingly accepted by its adherents, confusions and ambiguities may come to be tolerated in that field’s literature, cultivating a false impression among outside readers as to how well-formed these founding assumptions truly are. For example, the field of Freudian psychodynamics employs the process of ‘repression’ as an explanatory concept in understanding human psychology. Within the psychodynamic field, the concept of repression was refined considerably from its origins in Freud’s early writings, yet the popular conception of ‘repression’ to researchers outside the field typically represents a distorted caricature of repression’s simpler origins, fostering persistent misinterpretations of the contemporary psychodynamic literature. Indeed, this conceptual disjunct is sometimes so pronounced as to draw into question whether or not some critics of contemporary psychodynamics have even read the very literature they criticise (see Boag, 2006, for details). As this example suggests, the conventions of theory and terminology within a field can be expected to foster more charitable interpretations of ambiguous or confused publications among its adherents, as compared to outside readers.

Through this lack of explicit policing of ambiguities and subtle errors within a field (at least in any manner published, and thus visible to outsiders), large research traditions may come to cultivate grave misimpressions among researchers in other fields, with regards to what founding assumptions characterise their theories, and how their central concepts are defined. Misunderstandings of this sort are likely often tolerated by researchers in other fields out of simple uninterested courtesy (Driver-Linn, 2003), but approaches that strive for metaphysical certainties on the continuum of pragmatic assumptions, such as Situational Realism, are more likely to target the appearance of conceptual inadequacies as grounds to dismiss the field in question. As Petcoz and Mackay (2013) outline in their article, Situational Realism recommends a degree of disciplinary withdrawal from the current prominence of “cognitive neuroscience and information processing” (p. 217) approaches in psychology. This stance reflects a wider rejection of theories of ‘information processing’ and ‘mental representation’ in the realist literature (see McMullen, 2011). However, the present authors argue that these rejections are largely premature on the part of Situational Realists, stemming from ambiguities and confusions in the cognitive psychology literature, which foster a conception of information and mental representation that few cognitive scientists would willingly endorse. This issue can be illustrated by considering that nature of information and representation.

Relation, Information, and Representation

The relational view of Situational Realism insists upon never reducing a cognitive relation to anything less than a combined consideration of both knowing subject and known object (Michell, 1988). In doing so, Petcoz and Mackay (2013) suggest that psychologists may “extricate the legitimate concerns of representation in the information sciences from

incoherent epistemological representationism” (p. 218). At the heart of this view is the philosophical insistence that cognitive access to the real objects of the world must be ‘direct’, in contrast to models proposing ontologically literal intermediary representations, which degrade into a homunculus-style infinite regress by relying on a kind of ‘Cartesian theatre’ (see Maze, 1991). Critics of such ‘direct’ realist accounts (reviewed in Maclachlan, 1989) often take issue with this framing, drawing attention to the innumerable intermediate physical steps that must be chained into a causal sequence for any act of perception or cognition to occur. Such criticisms, however, mistake the sense in which cognitive relations are described as being ‘direct’. Realists do not deny that any cognitive relation is comprised of a great multitude of causally linked physical events (thus making the subject-object interaction ‘indirect’ in a strictly physical sense), but instead seek to call attention to the lack of meaningful semiotic intermediary states between the knowing subject and the known object (thus making the subject-object interaction ‘direct’ in a psychological sense). This distinction between physical and psychological senses in which many relational terms are used is elusive, both among schools of psychological realism, and more troublingly, among many cognitive psychologists.

The concept of ‘information’, which lies at the core of all information sciences (including all cognitive approaches to psychology), is typically employed in one of two general senses (Floridi, 2010). In most social and educational contexts, the word ‘information’ is used in the psychological sense, referring broadly to ‘that which is gained’ when something is learned, or when knowledge is acquired. Although information is routinely used as a noun in this sense (in English, at least), as if reifying an object that is moved, altered or duplicated, this psychological definition of information is only coherent when describing a relation between a knowing subject and something that is known (typically a variant of the propositions ‘ x knows y ’, or ‘ y is a state of affairs that x may come to know’). In this psychological sense, a

state of affairs is regarded as ‘information’ only insofar as it is something about which some subject is, or in principle could be, ‘informed’. For example, a set of coherent instructions written on a page is considered information, in this sense, as some subject could, in principle, become informed of their meaning. Conversely, a page filled with a truly random jumble of nonsensical characters does not, in this sense, ‘contain’ any information, for the patterns do not represent or embody any meaningful state of affairs that a subject may be informed of.

This contrasts sharply with the sense of information employed extensively in some branches of physics and in most of computer science, which refers to patterns of physical configurations in absolute terms. Information, in this physical sense, is any event or state of affairs that can affect a semiotic system. This usage of the term ‘information’ can thus apply contextually to any aspect or pattern in reality that, through interaction, can change the state of something else (be it photons affecting valance electrons, voltage changes affecting semiconductor-states, or a pencil inscribing a mark on a sheet of paper). Due to its tremendously broad range of potential applications, the information is typically employed in the physical sense in the specific context of a system that is sensitive to only certain changes, such as the sensitivity of neuronal dendrites to neurotransmitters, or the sensitivity of plant growth systems to the orientation of the sun (Arbib, 2002). Discussions in psychology, and in cognitive psychology in particular, routinely walk a tenuous line between the psychological and the physical senses of the word information, because many explanatory theories and process models seek to understand subjective cognition (transformations of information in the relational, psychological sense), as a product or emergent property of objective patterns of neurological activity (transformations of information in the systemic, physical sense).

Confusions and ambiguities between these two senses of information are understandably common across the psychology literature following the 1960s' 'cognitive revolution' (Miller, 2003), but are most problematic when employed in discussions of intermediate semiotic stages in psychological processes, often referred to by the umbrella term 'mental representations'. The "incoherent epistemological representationism" (p. 218) that Petcoz and Mackay describe in their piece (2013), refers to the seemingly widespread acceptance amongst cognitive psychologists of theories of cognition which place 'mental representations' as metaphysically impenetrable barriers and arbiters between the systems of the brain which 'know', and the worldly states of affairs that 'are known' (McMullen, 1996). This surely is a damning accusation, assuming that the cognitive theorists in question are indeed describing mental representations that are information constructs in the psychological sense. While it is likely that there are many cognitive psychologists who do mean exactly this, reflections on the evolution of the term 'mental representation' over the theoretical history of computational approaches to human psychology (see Smith, 1996; 1998) suggest that the foundational assumptions inherited by modern cognitive psychologists typically define mental representations, and indeed information in general, more so in the physical sense (see also Chomsky, 1980; Fodor, 1983, for early explicit disambiguations). With this in mind, it appears likely then that if the terminology and conceptual foundations of the (now expansive) field of cognitive psychology were more explicitly stated and strictly observed, the majority of the logical criticisms that drive Situational Realists to reject information processing approaches may be rendered moot.

With this in mind, the present authors suggest that many of the noteworthy conceptual conflicts between contemporary realist and representationalist approaches are grounded in confusions and ambiguities, rather than firm commitments to distinct ontologies. Although

realist traditions in psychology are sometime criticised for being insensitive to details and empirically restrictive (as outlined in Maclachlan, 1989), the prioritisation of metaphysical certainty among Situational Realists does little to shake their theoretical commitment to the common ground scientific ontology discussed earlier in this paper. The Situational Realist approach does not deny that the neural structures of an individual interact with external objects (and indeed, among themselves) via a vast contextualized network of causal influences (McMullen, 2011). Nor do they deny that the character of any cognitive relation is fundamentally changed, often to the point of error, by disruptions or inequities in its physical causal sequence (Rantzen, 1993). Their sole logical objection to accounts of mental representations concern the psychological sense of the term, wherein a distinct ‘mental’ intermediary interrupts the relation of access between the knower and what is known. Such suggestions are logically untenable, for ‘mental representations’ of this psychological sort are defined solely in what they do (i.e., represent) and there is no independent ontological account of what they actually are (McMullen, 1996).

With these false obstacles dissolved, the value of Situational Realism to psychology as a whole can be better appreciated. As an approach situated toward the far *safe* end of the continuum of pragmatic assumptions, Situational Realism (as with any compatible realist approach) provides a well-articulated philosophical grounding point, against which the relative certainty and logical viability of less certain assumptions can be assessed. Rather than petitioning other fields to cast off the conceptions and theoretical assumptions that facilitate proactive empirical testing (thus discarding valuable explorations of the psychological black box), Situational Realists would do well to encourage conceptual analysis of these assumptions, to clarify their tentative nature and explore the degree to which they have been supported by evidence. The present authors sympathise with Petcoz and Mackay’s (2013)

suggestion that “realism can integrate the traditional areas of psychology ... while also sustaining a number of different ‘alternative’ unifying approaches (albeit some suitably modified)” (p. 221). Though, precisely how many pragmatic assumptions should be discarded as a ‘suitable modification’ is no simple matter to judge, when many productive fields of psychology inadvertently cultivate poor reputations, due to their poor internal policing of confusions and ambiguities.

A Pull towards the Centre –Developmental Evolutionary Psychology

In sharp contrast to the aforementioned case of Situational Realism, Lickliter and Honeycutt’s (2013) proposal concerning Developmental Evolutionary Psychology does not outline the details of their theoretical framework exhaustively. Rather, their proposal builds upon the presumed existing familiarity of the reader with the adaptationist paradigm of evolutionary psychology, an oft-cited but controversial contender for an indispensable meta-theory in unifying psychology (Tooby & Cosmides, 2007; Webster, 2007; Daly & Wilson, 2008; Buss, 2009). In referencing this paradigm, Lickliter and Honeycutt (2013) have taken several bold steps towards *risky* on the continuum of pragmatic assumptions, when contrasted with cases like Situational Realism. Evolutionary psychology relies upon a network of conceptions and assumptions which, while presently quite well supported (both empirically and institutionally; see Fitzgerald & Whitaker, 2010), are on far less certain ground than the observable states of affairs discussed above (see, however, Richardson, 2007, for a dissenting position). Beyond reliance upon the computational information processing theories questioned by more conservative approaches, evolutionary psychology employs a specific adaptationist methodology which makes a range of probabilistic assumptions about the necessary role of natural selection in any set of complex biological designs (see Tooby &

Cosmides, 2005, for a detailed account). Building on their earlier work in the same vein (2003), Lickliter and Honeycutt (2013) draw special attention to a set of assumptions that evolutionary psychology has inherited from the ‘Modern Synthesis’ of evolutionary biology, concerning a heavy emphasis on the influence of genetics, to the detriment of the role of development (Mayr, 1982). Stated briefly, the Developmental Evolutionary Psychology approach endorses the entirety of the contemporary evolutionary psychology paradigm, with one key exception. They contend that the standard assumptions of evolutionary psychology separate genetic and developmental influences as distinct sources in organism formation and variation. Subsequently, evolutionary psychology privileges the role of genes as the ‘primary’ influence, with the role of development as supplemental, and assumption which Lickliter and Honeycutt describe as untenably preformationist and genetically deterministic. Citing a wide literature concerning recent discoveries in developmental systems and epigenetics, Lickliter and Honeycutt (2013) assert that these preformationist assumptions have become antiquated, and are now biologically indefensible. They instead propose a fundamental reframing of this component of the adaptationist approach, wherein evolutionary influences and developmental factors must always be considered as a complex whole. Stated directly: “it is not biologically meaningful to discuss gene activity and its influences without also referring to the broader context within which genes are activated and expressed ... genetic and environmental factors cannot be meaningfully partitioned” (Lickliter & Honeycutt, 2013, p. 185).

Dissolving Dichotomies, in Practice or Principle?

As a unification proposal, the Developmental Evolutionary Psychology submission follows a similar strategy to the Situational Realism submission, but to a far more moderate degree.

Rather than seeking to pull back the pragmatic assumptions of all other researchers to the far

safe end of the continuum, Lickliter and Honeycutt (2013) endorse the bulk of assumptions employed by evolutionary psychology, seeking only to pull researchers back from those assumptions concerning the distinctness and prioritisation of genetic and developmental influences. Just as with Situational Realism, the suggestion is that the assumptions targeted for redaction are logically and empirically untenable, and that researchers would do well to completely avoid these assumptions, eschewing the related distinctions in all future inquiry. Or, stated differently, this articulation of the Evolutionary Developmental approach seeks to strategically withdraw from several risky, pragmatic assumptions embraced by wider Evolutionary Psychology, drawing closer to the ‘safe’ end of the figurative continuum.

However, as was argued above, the pragmatic assumptions employed in various theories represent a delicate cost-benefit analysis between metaphysical certainty and empirical utility. As Buss and Reeve (2003) explore in their rebuttal to prior claims by Lickliter and Honeycutt (2003), the paradigm of evolutionary psychology is committed, at least in principle, to a ‘deeply interactionist’ conception of genetic and developmental influences. In their later writings on the topic, Lickliter and Honeycutt (2009) reflect on this professed acknowledgement, but insist that gene-privileging dichotomies remain practically entrenched in the concepts and hypotheses of most evolutionary psychology research, despite any theoretical claims to the contrary. The general endurance of at least partial favouring of genetic influences in analysis was attributed by Tooby, Cosmides, and Barrett (2003) as a matter of practicality. Since assuming a simplified, directional interaction between genes and development typically captures the majority of important design details in most situations, making this assumption is the most practicable option, only warranting reconsideration when contradictory evidence dictates.

It is here where one must weigh the costs and benefits of Lickliter and Honeycutt's position. While it seems undeniable that their critique is both conceptually insightful and (at least partially) empirically supported, does it warrant the dissolution of the conceptual genes-development distinction in practice, or merely in principle? Given the great difficulties inherent to analysing complex organism-environment interactions without at least starting with simplifying assumptions (which may be subsequently revised as necessary; see Buss & Reeve, 2003, for a review), must researchers sacrifice utility in strict observance to the complexity of situations? When viewed in the context of the continuum of pragmatic assumptions, a clear answer may be presented. If researchers within evolutionary psychology were to explicitly acknowledge their preformationist assumptions as strictly tentative, adopted for the sake of practicality, the fear that phenomena with radically diverse developmental trajectories may go unnoticed can be addressed, without needlessly impairing their research methodologies. Indeed, the intractable problems that Lickliter and Honeycutt (2003; 2013) caution against can only come to pass in a research environment where pragmatic assumptions are either embraced dogmatically, or are accepted as implicit and unstated, thus obscuring their position as both tenuous and tentative. The present authors argue that researchers need not give up the empirical promise of such pragmatic assumptions, so long as one's assumptions remain clearly stated and amenable to conceptual analysis.

An Adventure in Risky Pragmatics—Tree of Knowledge Unified Theory

In referencing the continuum of pragmatic assumptions, we are able to appreciate the unification strategies central to the prior two examples as revolving around a common theme. Both target particular theoretical traditions in contemporary psychology (though Situational Realism targets a considerably wider range than does Developmental Evolutionary

Psychology), and challenge what are viewed as either the logically or empirically untenable assumptions underwriting them. They also suggest a common solution: the wholesale abandonment of these assumptions to improve the technical and metaphysical accuracy of subsequent investigations. This metaphysical *safety* is brokered at the cost of pragmatic methodologies and the findings thus generated, but it is suggested that any lost knowledge can be reacquired all-the-better without the need for *risky* assumptions. Both approaches, in this sense, are deconstructive and anti-pragmatist, focussing on the perceived and potential errors of a vast existing literature, and concerned more with reversing the missteps of their peers than offering avenues to new discoveries.

In sharp contrast, the Tree of Knowledge (ToK) Unified Theory advanced by Henriques (2003, 2004, 2005, 2008, 2011, 2013) focuses less on dissolving problematic distinctions, but instead seeks to harvest and combine the functional cores of many such conceptions (from diverse sources in the literature) into a more focused and inclusive set of pragmatic assumptions. Building upon its characteristic Tree of Knowledge model, which centres on the emergence of complexity in the natural world, Henriques' theory assumes a top-down, but fundamentally pragmatic, approach to scientific discovery within psychology. While the complete ToK Unified Theory makes many targeted suggestions and integrative comparisons (covered exhaustively in Henriques, 2011), in employing it as an example the present authors wish to draw attention to two components that can be best construed as bold sets of pragmatic assumptions: *Behavioural Investment Theory* and the *Justification Hypothesis*. In reviewing both the pragmatic gains and the metaphysical risks of such propositions, the necessity of tentative assumptions can be better appreciated.

Behavioural Investment Theory: The Hopeful Chimera

The ToK Unified Theory, in a manner consistent with the robust ontology described above, regards the phenomena of psychology as a level of organisation that ‘emerges’ (in the philosophical sense) from biological phenomena, as the actions of the functionally-coordinated nervous systems of animals. Rather than relying upon the wide range of topic-specific assumptions scattered among other research traditions, Henriques condenses and distils the defining insights of many approaches into the 6 principles of Behaviour Investment Theory (BIT). As Henriques (2013) describes, “BIT starts with the proposition that the nervous system is an action control system that computes the investment of work effort on a cost-benefit ratio that evolves inter-generationally via evolutionary processes and is further moulded via experience during the life of the animal ... integrating evolutionary, neuroscience, behavioural science, and cognitive science perspectives” (p. 170). In a manner reminiscent of the two approaches discussed above, BIT emphasises the complex and interactive nature of the organism-environment system. However, rather than insisting on a single, simplified conception of how to describe and analyse such a system, BIT seeks to cultivate the rich diversity of pragmatic assumptions that have underpinned various research approaches (including biological, developmental, and cognitive approaches) and render them as compatible and reciprocal within a common conceptual framework. This approach regards the motivated nature of all cognition as central in framing: (i) the computational and neural constraints; (ii) evolved biological drives, and; (iii) learned and developmental calibrations, of any organism’s nervous system (see Henriques, 2011, chapter 3, for details).

While primarily a collection and integration of prevailing assumptions within experimental scientific psychology, the central value of the BIT lies in its organisation of these

assumptions into a coherent whole, which can serve as a ‘check-list’ of influences that researchers with diverse backgrounds must consider (Geary, 2005; Quackenbush, 2008). However, some critics, most notably Katzko (2008), have observed that this itemised combination of diverse assumptions blurs some ontological and epistemological distinctions that are more readily appreciated in the fields from which these assumptions emerge. For example, as the Lickliter and Honeycutt (2003; 2013) approach reviewed above emphasises, genetic and developmental influences are often deeply intertwined, and may interact in non-obvious ways when considering the evolution of a psychological process (see also Viney, 2004, for further discussion). Despite this, for practical purposes, BIT regards its range of assumptions as mostly independent and equally metaphysically certain, but in doing so obscures the interrelations between some postulates, such as the reliance of evolutionary modularity upon a particular conception of neural computation (also discussed in Pinker, 1997). This apparent equivalence is potentially misleading, and misses an opportunity to frame the network of assumptions in a manner that acknowledges their hierarchical interrelations and tentative position the pragmatic continuum. A more explicit account of the relations and dependencies between the assumptions of BIT would enhance the unifying appeal of the ToK Unified Theory, by allowing researchers to qualify any particular postulates they deem inappropriate, without having to discard the framework as a whole. For example, since the founding principles of evolutionary psychology draw upon insights from cognitive psychology, but not vice versa, researchers seeking to address cognitive phenomena while addressing or limiting evolutionary concerns would presently have no choice but to reject BIT in its entirety. If BIT were expressed in a manner that qualifies which assumptions are being relied upon, such a research project may be expressed in the language of BIT despite the exclusion of evolutionary concerns, and could thus in principle be meaningfully

compared to other BIT proposals along whatever dimensions are shared (for related perspectives, see Yanchar, 2004; Kirschner, 2006).

Justification Hypothesis: A Key to Culture

The most bold, and arguably most innovative, contribution of the ToK Unified Theory concerns the pragmatic assumptions underlying the Justification Hypothesis (discussed in Shaffer, 2005; Anchin, 2008; Quackenbush, 2008). The Justification Hypothesis is proposed as the central heuristic for understanding the social-symbolic characteristics unique to the psychology of humans (in particular, the manner in which humans describe, understand, and communicate beliefs and decisions), by proposing that much of our social cognitive apparatus are evolved adaptations which address the demands of predicting, coordinating and describing one's actions in a manner that can be justified to others. That is to say, in-line with perspectives in social psychology such as Haidt's Social Intuitionist model (Haidt, 2001; 2012), the Justification Hypothesis regards the primary adaptive function of most forms of deliberate human reasoning, as providing socially defensible justifications and rationalisations for our beliefs and actions, so as to guard the many benefits of cooperation and social status (see Henriques, 2011, chapter 5 for a full account). Thus, in the ToK framework, hypotheses concerning the function and organisation of many social psychological processes (particularly those involving self-awareness and intention) can be generated by considering the adaptive demands of social justification, particularly in ancestral environments. Henriques (2011) argues that these guiding constraints provide both a potentially instructive means of understanding reflective and meta-cognitive psychological systems, and also a unique means of analysing the emergence and adaptive function of many human cultural phenomena (such as norms, traditions, and historical narratives), which may

be regarded as socially distributed ‘justification systems’. Any proposition concerning so wide a range of phenomena as the Justification Hypothesis must be subjected to careful scrutiny, because while the approach offers an enticing means of organising a tangle of complex problems, its scientific merit can only be evaluated via the presence or absence of empirical support of its predictions.

Despite receiving some critical support for the wide sphere of potential insights it affords (Stanovich, 2004; Gilbert, 2004; Katzko, 2004; Haaga, 2004; Shealy, 2005), the Justification Hypothesis serves as an illustrative example of a theory built upon pragmatic assumptions. While it is perhaps possible that, indeed, all of human culture may be best understood as justification systems, constrained and operated by particular psychological mechanisms, counter-arguments against such a complete account are already emerging. Both Katzko (2008) and Shaffer (2008), for instance, argue that several social-phenomena require mind-culture bridgings that exceed the projected theoretical role of the Justification Hypothesis. Others, notably Vazire and Robins (2004), argue that the obvious utility of the Justification Hypothesis may be better understood as the result of several distinct adaptations, each with an alternative set of evolutionary origins to those proposed by Henriques (see also Katzko, 2004; Stanovich, 2004; Shealy, 2005). Regardless of which perspective ultimately triumphs, exploring the Justification Hypothesis as an example exposes the value of tentative pragmatic assumptions that the present authors wish to draw attention to. That is, the disagreements concerning the utility and validity of the Justification Hypothesis could only be addressed by tentatively pursuing research based upon its assumptions, and then comparing the value of its findings to those of alternative perspectives (Calhoun, 2004). Within the groundless black box of psychological phenomena, promising assumptions must be tentatively adopted and

empirically tested to evaluate their worth, for *a priori* speculation will always preclude those findings which run contrary to our intuitions, and thus teach us the most (Kuhn, 1970; 1996).

Conclusion

Theorists such as Goertzen (2008; 2011) and Trafimow (2012) have openly lamented the lack of attention given to the underlying conceptual and philosophical assumptions that proliferate within psychological research. The black box limitations of the subject matter of psychology ensure that traditions built upon ill-acknowledged assumptions invariably lose direction, and gradually become increasingly incompatible with alternative traditions based on different foundations. By embracing the increasingly accepted physical ontology underlying the organism-environment interactions of psychological phenomena, researchers are in a position to organise their theories and empirical explorations along a continuum of pragmatic assumptions. With a shared definitional basis, the metaphysical certainty of any scientific theory of psychology can be regarded as a tentative postulate in a network of related assumptions, ranging from those with the greatest certainty (but with vague applicability), to those assumed for practical purposes, which must be evaluated by the strength of their results.

The three examples explored above, Situational Realism, Developmental Evolutionary Psychology, and the Tree of Knowledge Unified Theory, can be understood as increasingly *risky* increments along the continuum of pragmatic assumptions. The recommendations of their advocates regarding wider unification can be best understood as a function of their position along the continuum, but all three approaches share an ultimate commitment to the realist ontology at the heart of contemporary scientific psychology (Mandler, 2011). We have

outlined how the explicit acknowledgement of foundational assumptions, and the appropriate designation of these assumptions as tentative (pending empirical exploration), can permit approaches presently at-odds to integrate and overlap wherever conceptual compatibilities permit. However, the prospect of this form of unification is contingent upon the expansion of both theoretical development and conceptual analysis in the practice of psychological research (Machado, Lourenco & Silva, 2000), two practices which grow increasingly neglected in modern academic institutions (Michell, 2003a; 2003b).

References

- Anchin, J.C. (2008). The critical role of the dialectic in viable metatheory: A commentary on henriques' tree of Knowledge System for Integrating Human Knowledge. *Theory and Psychology, 18*(6), 801-816.
- Anderson, J.R. (2010). *Cognitive psychology and its implications*. New York, NY: Worth Publishers.
- Arbib, M. A. (2002). *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press.
- Barrett, H.C., & Kurzban, R. (2006). Modularity in Cognition: Framing the Debate. *Psychological Review, 113* (3), 628-347.
- Bechtel, W., & Abrahamsen, A. (1991) *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Cambridge: Blackwell.
- Boag, S. (2006). Freudian repression, the common view, and pathological science. *Review of General Psychology, 10*(1), 74-86.
- Boag, S. (2011). Explanation in personality research: 'verbal magic' and the Five-Factor Model. *Philosophical Psychology, 24*, 223-243.
- Botella, L. & Gallifa, J. (1995). A constructivist approach to the development of personal epistemic assumptions and worldviews. *Journal of Constructivist Psychology, 8*, 1-18.
- Boyd, R., Gasper, P., & Trout, J.D. (1991). *The Philosophy of Science*. Blackwell Publishers, Cambridge, MA.
- Buss, D. M. (1984). Evolutionary biology and personality psychology: Toward a conception of human nature and individual differences. *American Psychologist, 39*, 1135-1147.

- Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, 6, 1-30.
- Buss, D.M. (2009). The Great Struggles of Life: Darwin and the Emergence of Evolutionary Psychology. *American Psychologist*, 64, 140-148.
- Buss, D.M. & Reeve, H.K. (2003). Evolutionary psychology and developmental dynamics. *Psychological Bulletin*, 129, 848-853.
- Calhoun, L.G. (2004). The unification of psychology: A noble quest. *Journal of Clinical Psychology*, 60(12), 1283-1289.
- Charles, E. P. (2011). Ecological psychology and social psychology: Continuing Discussion. *Integrative Psychological and Behavioral Sciences*, 46(2), 249-258.
- Charles, E.P. (2013). Psychology: The empirical study of epistemology and phenomenology. *Review of General Psychology*, 17(2), 140-144.
- Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.
- Cosmides, L. & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II. Case study: A computational theory of social exchange. *Ethology & Sociobiology*, 10, 51-97.
- Daly, M. & Wilson, M. (2008). Is the "Cinderella effect" controversial?: A case study of evolution-minded research and critiques thereof. In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology*. (383-400). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.
- de Groot, A.D. (1990). Unifying psychology: A European view. *New Ideas in Psychology*, 8, 309-320.

- De Los Reyes, A. & Kazdin, A.E. (2008). When the evidence says, "Yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science*, 17, 47-51.
- Deutsch, M. (1968). Field theory in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 412– 487). Reading: Addison Wesley.
- Dewsbury, D. A. (1991). Psychobiology. *American Psychologist*, 46, 198-205.
- Driver-Linn, E. (2003). Where is psychology going? Structural fault lines revealed by psychologists' use of Kuhn. *American Psychologist*, 58(4), 269-278.
- Fitzgerald, C. J. & Whitaker, M. B. (2010). Examining the acceptance of and resistance to evolutionary psychology. *Evolutionary Psychology*, 8, 284-296.
- Floridi, L. (2010). *Information: A Very Short Introduction*. Oxford University Press, Oxford.
- Fodor, J. (1975). *The Language of Thought*. Harvester Press.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Freud, S. (1923). *The Ego and the Id*, Joan Riviere (trans.), Hogarth Press and Institute of Psycho-analysis, London. Revised (1961) for *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, James Strachey (ed.), W.W. Norton and Company, New York.
- Freund, P.A. & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2), 296-321.
- Gazzaniga, M. (2010). *Psychological Science*. New York: W.W. Norton & Company.

- Geary, D.C. (2005). The motivation to control and the origin of mind: Exploring the life-mind joint point in the tree of knowledge system. *Journal of Clinical Psychology*, 61(1), 21-46.
- Geertz, C. (2000). *Available light: Anthropological reflections on philosophical topics*. Princeton, NJ: Princeton University Press.
- Gilbert, P. (2004). A much needed macro level view: A commentary on Henriques' "psychology defined". *Journal of Clinical Psychology*, 60(12), 1223-1226.
- Gladin, L.L. (1961). Toward a unified psychology. *Psychological Record*, 11, 405-421.
- Goertzen, J.R. (2008). On the Possibility of Unification: The Reality and Nature of the Crisis in Psychology. *Theory & Psychology*, 18, 829-852.
- Goertzen, J. R. (2010). Dialectical pluralism: A theoretical conceptualization of pluralism in psychology. *New Ideas in Psychology*, 28(2), 201-209.
- Goertzen, J.R. (2011). Further problematizing the potential for a more unified experimental, scientific psychology: A comment on Mandler. *Journal of Theoretical and Philosophical Psychology*, 31(4), 247-249.
- Haaga, D.A.F. (2004). Defining psychology: What can it do for us? *Journal of Clinical Psychology*, 60(12), 1227-1229.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York, NY: Pantheon.

- Heft, H. (2012). An ecological approach to psychology. *Review of General Psychology*, 17(2), 162-167.
- Henriques, G. (2003). The tree of knowledge system and the theoretical unification of psychology. *Review of General Psychology*, 7, 150-182.
- Henriques, G.R. (2004). Psychology defined. *Journal of Clinical Psychology*, 60, 1207-1221.
- Henriques, G.R. (2005). Toward a useful mass movement. *Journal of Clinical Psychology*, 61(1), 121-139.
- Henriques, G.R. (2008). The problem of psychology and the integration of human knowledge: Contrasting Wilson's consilience with the Tree of Knowledge System. *Theory & Psychology*, 18, 731-755.
- Henriques, G. (2011). *A New Unified Theory of Psychology*. Springer Publishing, New York.
- Henriques, G. (2013). Evolving from methodological to conceptual unification. *Review of General Psychology*, 17(2), 168-173.
- Hibberd, F. J. (2009). John Anderson's development of (situational) realism and its bearing on psychology today. *History of the Human Sciences*, 22(4), 63-92.
- Jaszczolt, K. (1996). *Relevance and infinity: Implications for discourse interpretation*. *Journal of Pragmatics*, 25, 703-722.
- Kaitaro, T. (2004). Brain-mind identities in dualism and materialism: A historical perspective. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 35(4), 627-645.
- Kantor, J.R. (1979). Psychology: Science or nonscience? *The Psychological Record*, 29, 155-163.

- Kashdan, T.B., & Steger, M.F. (2004). Approaching psychological science with Kuhn's eyes. *American Psychologist*, 59, 272-273.
- Katzko, M.W. (2004). Psychology's dilemma: An institutional neurosis? *Journal of Clinical Psychology*, 60(12), 1237-1241.
- Katzko, M.W. (2008). Pruning the Tree of Knowledge. *Theory and Psychology*, 18(6), 817-828.
- Kelly, R.J. (1998). The crisis in psychology: Trouble in the temple. *Journal of Social Distress and the Homeless*, 7, 211-223.
- Kerr, C.E. (2008). Dualism redux in recent neuroscience: "Theory of mind" and "embodied simulation" hypotheses in light of historical debates about perception, cognition, and mind. *Review of General Psychology*, 12, 205-214.
- Kirschner, S.R. (2006). Psychology and pluralism: Toward the psychological studies. *Journal of Theoretical and Philosophical Psychology*, 26(1), 1-17.
- Kruglanski, A.W., Bélanger, J.J., Chen X., Köpetz, C., Pierro, A. & Mannetti L. (2012). *Psychological Review*, 119(1), 1-20.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd Ed.). Chicago: University of Chicago Press.
- Kuhn, T.S. (1977). Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension* (pp. 320–339), Chicago: University of Chicago Press.
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd Ed.). Chicago: University of Chicago Press.

- Lewin, K. (1951). *Field theory in social science*. Harper & Brothers, New York.
- Lickliter, R., & Honeycutt, H. (2003). Developmental dynamics: toward a biologically plausible evolutionary psychology. *Psychological Bulletin*, 129(6), 819-835.
- Lickliter, R. & Honeycutt, H. (2009). Rethinking epigenesis and evolution in light of developmental science. In: M. Blumberg, J. Freeman, S. Robinson (Eds.), *Oxford Handbook of Developmental Behavioral Neuroscience* (pp. 30-50). New York: Oxford University Press.
- Lickliter, R., & Honeycutt, H. (2013). A developmental evolutionary framework for psychology. *Review of General Psychology*, 17(2), 184-189.
- Machado, A., Lourenço, O., & Silva, F.J. (2000). Facts, concepts, and theories: The shape of psychology's Epistemic Triangle. *Behavior and Philosophy*, 28, 1-40.
- Mackay, N. & Petocz, A. (2011). (Eds.), *Realism and psychology: Collected essays*. Leiden: Brill.
- Maclachlan, D.L.C. (1989). *Philosophy of Perception*. Prentice Hall, New Jersey.
- Mandler, G. (2011) From association to organization. *Current Directions in Psychological Science*, 20(4), 232-235.
- Marsh, T., & Boag, S. (2013). *Evolutionary and differential psychology: conceptual conflicts and the path to integration*. *Frontiers in Psychology*, 4(655), 1-15.
- Masterman, M. (1970). The nature of paradigms. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 61-65). New York: Cambridge University Press.
- Mayr, E. (1982). *The growth of biological thought*. Cambridge, MA: Harvard University Press.

- Maze, J. R. (1991). Representation, realism and the redundancy of "Mentalese". *Theory & Psychology, 1*, 163-185.
- McMullen, T. (1996). Psychology and realism. In C.R. Latimer & J. Michell (Eds.), *At once scientific and philosophic: A festschrift for John Philip Sutcliffe* (pp. 59–66). Brisbane, Australia: Boombana.
- McMullen, T. (2011). "Out there", not "in here": A realist account of concepts. In N. Mackay & A. Petocz (Eds.), *Realism and psychology: Collected essays* (pp. 325-356). Leiden: Brill.
- Michel, G. F., & Tyler, A. N. (2007). Developing human nature: "Development to" vs. "Development from". *Developmental Psychobiology, 49*, 788–799.
- Michell, J. (1988). Maze's direct realism and the character of cognition. *Australian Journal of Psychology, 40*, 227-49.
- Michell, J. (2000). Normal science, pathological science, and psychometrics. *Theory & Psychology, 10*, 639-667.
- Michell, J. (2003a). The quantitative imperative: Positivism, naïve realism and the place of qualitative methods in psychology. *Theory and Psychology, 13*(1), 5-31.
- Michell, J. (2003b). Pragmatism, positivism and the quantitative imperative. *Theory and Psychology, 13*(1), 45-52.
- Michell, J. (2006). Psychophysics, intensive magnitudes, and the psychometricians' fallacy. *Studies in the History and Philosophy of Biological and Biomedical Sciences, 17*, 414–432.

- Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology*, 31(1), 13-21.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7, 141–144.
- Nairne, J. S. (1997). *Psychology: The adaptive mind*. Brookes/Cole, Pacific Grove.
- Neisser, U. (1995). Criteria for an ecological self. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 222-231). Elsevier, Amsterdam.
- O'Donohue, W.T., Callaghan, G.M., & Ruckstuhl, L.E. (1998). Epistemological barriers to radical behaviorism. *Behavior Analyst*, 21, 307-320.
- Petocz, A., & Mackay, N. (2013). *Unifying psychology through situational realism. Review of General Psychology*, 17(2), 216-223.
- Pinker, S. (1997) *How the Mind Works*. New York: W.W. Norton.
- Popper, K. (1959). *The Logic of Scientific Discovery*. New York, NY: Routledge Classics.
- Quackenbush, S.W. (2008). Theoretical unification as a practical project: Kant and the tree of knowledge system. *Theory and Psychology*, 18(6), 757-777.
- Rantzen, A. J. (1993). Constructivism, direct realism and the nature of error. *Theory & Psychology*, 3, 147-171.
- Ribes-Iñesta, E. (2003). What is defined in operational definitions? The case of operant psychology. *Behavior and Philosophy*, 31, 111-126.
- Richardson, K. (1998). *The origins of human potential: Evolution, development, and psychology*. London: Routledge.

- Richardson, R. (2007). *Evolutionary Psychology as Maladapted Psychology*. Cambridge, MA: The MIT Press.
- Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, 85, 33–46.
- Shaffer, L.S. (2008). Religion as a large-scale justification system: Does the Justification Hypothesis explain animistic attribution? *Theory & Psychology*, 18, 779–799.
- Shealy, C.N. (2005). Justifying the Justification Hypothesis: Scientific-humanism, Equilintegration (EI) Theory, and the Beliefs, Events, and Values Inventory (BEVI). *Journal of Clinical Psychology*, 61(1), 81–106.
- Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, 70(5), 893–912.
- Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, & S. T. Fiske (Eds.), *The handbook of social psychology*, vol. 2 (4th ed), (pp. 391–445). New York, NY, US: McGraw-Hill.
- Sober, E. (1998). Black box inference: When should intervening variables be postulated? *British Journal for the Philosophy of Science*, 49, 469–498.
- Staats, A.W. (1983). *Psychology's crisis of disunity: Philosophy and method for a unified science*. New York: Praeger.
- Staats, A. W. (1999). Unifying psychology requires new infrastructure: Theory, method, and a research agenda. *Review of General Psychology*, 3, 3–13.
- Stam, H.J. (1990). Rebuilding the ship at sea: The historical and theoretical problems of constructionist epistemologies in psychology. *Canadian Psychology*, 31, 239–253.

Stam, H.J. (2004). Unifying psychology: Epistemological act or disciplinary maneuver?

Journal of Clinical Psychology, 60, 1259-1262.

Stanovich, K.E. (2004). Metarepresentation and the great cognitive divide: A commentary on

Henriques' "Psychology Defined". *Journal of Clinical Psychology*, 60, 1263-1266.

Sternberg, R.J. (2005). Unifying the field of psychology. In R. J. Sternberg (Ed.), *Unity in*

psychology: Possibility or pipedream? (pp. 3–14). Washington, DC: American

Psychological Association.

Sternberg, R. J., & Sternberg, K. (2012). *Cognitive psychology*. (6th ed). Belmont, California:

Wadsworth.

Sturm, T., & Mülberger, A. (2012). Crisis discussions in psychology: New historical and

philosophical perspectives. *Studies in History and Philosophy of Biological and*

Biomedical Sciences, 43, 425-521.

Tonneau, F. (2013). Neorealism: Unifying cognition and environment. *Review of General*

Psychology, 17(2), 237-242.

Tooby, J. & Cosmides, L. (1989). Evolutionary psychology and the generation of culture,

Part I. Theoretical considerations. *Ethology & Sociobiology*, 10, 29-49.

Tooby, J. & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D.

M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (5-67). Hoboken, NJ: Wiley.

Tooby, J. & Cosmides, L. (2007). Evolutionary psychology, ecological rationality, and the

unification of the behavioral sciences. Comment on, A framework for the unification

of the behavioral sciences, by Gintis. H. *Behavioral and Brain Sciences*, 30, 42-43.

Tooby, J., Cosmides, L., & Barrett, H. (2003). The second law of thermodynamics is the first

law of psychology: Evolutionary developmental psychology and the theory of

- tandem, coordinated inheritances: Comment on Lickliter and Honeycutt (2003). *Psychological Bulletin*, 129, 858-865.
- Trafimow, D. (2012). The role of mechanisms, integration and unification in science and psychology. *Theory & Psychology*, 22, 696-703.
- Valsiner, J. (2009). Integrating psychology within the globalizing world: a requiem to the post-modernist experiment with Wissenschaft. *Integrative Psychological & Behavioral Science*, 43, 1-21.
- Viney, W. (2004). Pluralism in the sciences is not easily dismissed. *Journal of Clinical Psychology*, 60(12), 1275-1278.
- Vazire, S. & Robins, R.W. (2004). Beyond the justification hypothesis: A broader theory of the evolution of self-consciousness. *Journal of Clinical Psychology*, 60(12), 1271-1273.
- Vul, E. (2011). Reductionism and Practicality. *Psychological Inquiry*, 22(2), 137–138.
- Walsh, V. (2000). Reverse engineering the human brain. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 358, 497-511.
- Webster, G.D. (2007). Evolutionary theory's increasing role in personality and social psychology. *Evolutionary Psychology*, 5(1), 84-91.
- Wertz, F. (1999). Multiple methods in psychology: Epistemological grounding and the possibility of unity. *Journal of Theoretical and Philosophical Psychology*, 19, 131-166.
- Yanchar, S.C. (2004). Some discontents with theoretical unification: A response to Henriques' "Psychology Defined". *Journal of Clinical Psychology*, 60(12), 1279-1281.
- Yanchar, S.C. & Slife, B.D. (1997). Pursuing unity in a fragmented psychology: Problems and prospects. *Review of General Psychology*, 1, 235-255.

Discussion for Thesis Chapter 4

Although the approach to integration advanced in this article was written as a general contribution to the unification of psychology literature, the two conceptual tools that the article introduces are essential to the evolutionary synthesis of prejudice that is the main goal of this thesis. The clarification of the ontological common ground at the heart of most of contemporary scientific psychology, while the more minor of the two key contributions, is necessary as both a means of orienting the continuum of pragmatic assumptions, and as an affirmation of the shared values and goals that make scientific psychology possible. As the paper explores, the use of metaphor in psychological theory-building is often essential when facing unintuitive phenomena, but is also implicitly risky when the constraints of one's literal expectations are unclear. By historical necessity, useful theorists' fictions, such as quantitative 'constructs', 'behavioural tendencies', 'mental-states', and the simplest conceptions of psychological 'processes', abound in many fields of psychology, a practice which can foster great ambiguity and confusion when the metaphorical use of such terms is obscured by ubiquitous use. For any metaphor, or similar illustrative approximation, to function as intended, its users must remain clear with regard to the senses in which the term is apt, and those senses in which it is not (for example, a personality 'trait' functions as a descriptive simplification of how someone is typically expected to act, but cannot be meaningfully specified as a literal object 'within' a person). The ontological common ground outlined in this paper specifies a tentative account of what literal, physical objects and entities researchers can realistically expect any psychological phenomenon to rely upon or consist of. Any functionally or metaphorically defined concept that cannot be accounted for in this ontology (for example, the crudest conceptions of 'psychical energy') is unlikely to hold any scientific explanatory value, and those concepts most easily translated into literal terms can be regarded as more plausible than their alternatives. As Chapter 5 explores, one of the key

benefits of adaptationist approaches to psychology is the intrinsic explanatory utility afforded by their consistent adherence to this realist ontology.

The continuum of pragmatic assumptions, and its opposing ends of *risky* and *safe*, are applied to a range of key comparisons later in this thesis, both between the three fields of social, differential, and moral psychology, and between the evolutionary and non-evolutionary approaches within each (most notably in framing the new theory of Chapter 7). Although, due to the order in which the publications were submitted, this terminology could not be introduced and employed in the published journal article featured in Chapter 5, the underlying concepts of metaphysical certainty working in opposition to the pragmatically testable are crucial to understanding the implications of Chapter 5's explanatory analyses to the thesis as a whole. The discussion of the empirical findings of this thesis also employs the terminology of the continuum of pragmatic assumptions in framing the differential certainty of several related interpretations of the data, demonstrating the value of these distinctions in both the generative and interpretative phases of theoretical and conceptual analysis in psychological science.

CHAPTER 5

The Evolutionary Approach to Individual Variation

As the book chapter featured in Chapter 2 (and its subsequent discussion) identified, the intrapersonal psychological mechanisms (such as the empathetic processes focused on in this thesis) underlying individual variation in prejudiced behaviour and beliefs are largely neglected in social psychology approaches to prejudice, but are of specific concern in the prejudice-related research done in moral and differential psychology. This thesis employs the adaptationist approach of evolutionary psychology to integrate the findings of these three fields, taking as a starting point the promising evolutionary work that has emerged in each of these fields within the last 12 years (introduced in Chapters 2 and 3, and reviewed extensively in Chapter 7). This task is complicated considerably, however, by the conceptual difficulties that evolutionary psychology has faced since its inception in meaningfully addressing systematic individual differences in psychological characteristics within the human species.

As several theorists and critics noted in the year when the book chapter featured in Chapter 2 was published (most notably Confer et al., 2010, and Fitzgerald & Whitaker, 2010), evolutionary and differential psychology had come to embody two opposing emphases in behavioural and cognitive research. The field of differential psychology, routinely described as the study of ‘personality and individual differences’, is primarily concerned with the ways in which variation between individuals can occur, understood sometimes at the level of the individual, and more often understood as systematic patterns at the level of the population. Evolutionary psychology, in contrast, had until recently encountered the vast majority of its success in the direct application of the adaptationist approach to understanding human psychological features that were reliably species-typical. This emphasis on species-typical characteristics was grounded in the strong and easily interpreted evidence for selection that is offered by traits whose survival and reproductive advantages are so unambiguous that they

became ubiquitous and genetically fixated prior to the major geographical divergences of modern human populations (typically presumed to be the early Pleistocene epoch). The extensive study of these unambiguously adaptive features in the early two decades of evolutionary psychology gave rise to the conception (shared even by some evolutionary researchers at the time, as detailed in Confer et al., 2010) that any feature of human psychology that systematically varied in all human populations cannot be the distinct product of natural selective forces, for any heritable trait appreciably enhancing reproductive success is expected to eventually outcompete its conspecifics and reliably dominate the gene pool. Several evolutionary approaches to personality and individual differences that follow this simple logic of selection are explored and critiqued in Chapter 2, which lean heavily on uncertain assumptions of selective neutrality to explain the persistence of many population variants.

The journal article featured in this chapter was inspired by the breakthroughs that evolutionary psychologists have made in recent years with regards to the study of personality and individual differences (early examples of which were collected in Buss & Hawley, 2011), and was written to further elaborate upon the key conceptual innovations to the adaptationist approach that an appreciation of these breakthroughs permits. In particular, the evolutionary mechanisms through which individual variation within a population can be produced and maintained, allow for the explanatory value of the study of individual differences to be cast in a new and more constructive light. The journal article featured in this chapter extensively reviews the explanatory and descriptive strengths of traditional and evolutionary approaches to differential psychology, and demonstrates the superiority of the latter for a wide range of integrative efforts in cognitive and behaviour science, including the synthesis at the heart of this thesis.

The following article was published in journal *Frontiers in Psychology*, assigned to the specialty section *Evolutionary Psychology and Neuroscience*, under the title ‘Evolutionary and Differential Psychology: Conceptual Conflicts and the Path to Integration’.

Declaration for Thesis Chapter 5

In the case of journal article featured in Chapter 5, the nature and extent of my contribution to the work, and the contributions of the other listed co-authors is as follows:

<i>Name</i>	<i>Nature of Contribution</i>	<i>Contribution</i>
Tim Marsh	Decision concerning the topic of the paper	90%
	Search and review of the literature	
	Principle writing and editing of the manuscript	
Simon Boag	Advice on topic and approach	10%
	Assistance with editing and cutting	
	Suggestions for the refinement of the manuscript	

**Evolutionary and Differential Psychology: Conceptual Conflicts and the Path to
Integration**

Running title: *Evolutionary and Differential Psychology*

Tim Marsh
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: timothy.marsh@mq.edu.au

Simon Boag
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: simon.boag@mq.edu.au

ABSTRACT

Evolutionary psychology has seen the majority of its success exploring adaptive features of the mind believed to be ubiquitous across our species. This has given rise to the belief that the adaptationist approach has little to offer the field of differential psychology, which concerns itself exclusively with the ways in which individuals systematically differ. By framing the historical origins of both disciplines, and exploring the means through which they each address the unique challenges of psychological description and explanation, the present article identifies the conceptual and theoretical problems that have kept differential psychology isolated not only from evolutionary psychology, but from explanatory approaches in general. Paying special attention to these conceptual problems, the authors review how these difficulties are being overcome by contemporary evolutionary research, and offer instructive suggestions concerning how differential researchers (and others) can best build upon these innovations.

Keywords: bottom-up explanation, differential psychology, evolutionary psychology, individual differences, integration, top-down explanation, unification of psychology

Evolutionary and Differential Psychology: Conceptual Conflicts and the Path to Integration

Psychology has been described as a science impaired by disunity (Gladin, 1961; Meehl, 1978; Kantor, 1979; Staats, 1983; 1999; de Groot, 1990; Yanchar & Slife, 1997; Henriques, 2003, 2004, 2011; Goertzen, 2008, 2010; Mandler, 2011). There is, however, disagreement over precisely how large a problem theoretical and institutional disunity is for psychologists and behavioural scientists in general (Dixon, 1983; Baars, 1984, 1985; Matarazzo, 1987, 1992; Bower, 1993; Neisser, 1995; Kelly, 1998; Kassinove, 2002; Stam, 2004). Nevertheless, integration is widely considered a desirable course of action, if only for the potential benefits of combining disparate theories and findings within a common conceptual space (Staats, 1999; Henriques, 2003, 2012; Goertzen, 2008). In recent years, the *adaptationist* approach of evolutionary psychology has emerged as strong candidate for central inclusion in a unifying meta-theory of psychology (Penke, Denissen & Miller, 2007a; 2007b; Tooby & Cosmides, 2007; Webster, 2007; Daly & Wilson, 2008; Buss, 2009). Stated briefly, adaptationism is a paradigm for analysing the physical and behavioural characteristics of organisms by focussing on functionally complex features which can only arise through natural selective pressures (see Daly & Wilson, 1999 for a brief review of the origins of adaptationism in behavioural science). Despite some enduring camps of resistance (Rose & Rose, 2000; Buller, 2005; Richardson, 2007), the literature shows a trend of increasing acceptance of adaptationism in diverse fields of psychology (Confer et al, 2010; Fitzgerald & Whitaker, 2010). Many recent unification efforts orient around evolutionary theories and approaches (Sternberg & Grigorenko, 2001; Henriques, 2003, 2004, 2008, 2011; Gintis, 2007). Nevertheless, although the adaptationist approach can and has been readily applied to an extensive range of psychological phenomena, as highlighted in Confer et al. (2010) some areas of psychology pose unique theoretical and methodological difficulties, which

evolutionists must (befittingly) adapt to. Perhaps the largest category of phenomena that demands a revision of traditional adaptationist analyses is the systematic occurrence of variation in normative psychological characteristics, the domain of personality and individual differences (Buss, 2009; Buss & Hawley, 2011). As Confer et al (2010) summarise: “Evolutionary psychology has been far more successful in predicting and explaining species-typical and sex-differentiated psychological adaptations than explaining variation within species or within the sexes” (p. 123). The recent innovations on this front discussed later in this article are best appreciated relative to the history and present state of traditional differential psychology.

Over the past century, the study of normative individual differences in thought, behaviour and ability (hereafter referred to by the umbrella term ‘differential psychology’) has become one of the largest and most popular arms of psychological science (Lubinski, 2000; Borghans, Golsteyn, Heckmann & Humphries, 2011). Differential psychology has intimate ties to multiple fields of applied psychology, including psychometric assessment, developmental and educational psychology, lifestyle and vocational adjustment, and our shifting conceptions of psychopathology (Lubinski, 2000). Despite this, differential psychology has a long history of remaining largely theoretically autonomous from related sub-disciplines of psychology. To this day, there is little cross-pollination between even the largest areas of differential psychology and their immediately adjacent research fields (Mischel, 1963, 1973; Cervone, 1991; Borsboom, 2005; Cramer, Waldorp, van der Maas, & Borsboom, 2010). To illustrate the point, the differential psychology domain of cognitive ability/intelligence testing has developed largely independent of the insights of functional cognitive psychology (Cronbach, 1957; Neisser et al 1996; Garlick, 2002, 2003; Anderson, 2004). Additionally, differential trait theories have become a prevailing approach in the study of personality, whilst remaining predominantly separate from other leading conceptions and

models within personality psychology (see Block, 1989, contrasted with 2010, for examples both before and after the rise of the Five Factor Approach). Evolutionary psychology represents only the most recent theoretical approach that must now struggle to integrate with the relatively independent niche carved out by the traditions of differential psychology.

While both evolutionary psychology and differential psychology are immensely diverse and heterogeneous fields, the arguments of this paper seek to cast as wide and as relevant a net as possible. As such, primary focus shall be given to fundamental conceptual and methodological elements that are near-ubiquitous characteristics of the respective fields, with more specific examples drawn from the most relevant and representative research areas available. By utilising some often overlooked distinctions from the wider philosophy of science, examining the fundamental scientific tasks of *description* and *explanation* (and beyond this, *forms* of explanation), the authors seek to explore the apparent theoretical isolation of differential psychology, and argue that integration is possible only when descriptive efforts are designed to inform causal explanations. In approaching this contentious topic from a neglected theoretical perspective, this paper contributes a new argument to the collective evolutionary-differential integration efforts started by David Buss almost 30 years ago (1984), an argument designed to address the fundamental conceptual concerns echoed by some critics of evolutionary psychology (Buller, 2005; Richardson, 2007). The current state of integration efforts and possible future avenues for individual differences research will also be discussed.

A Common Ancestor

During the formative period of the late 1800s, the precursors of both evolutionary and differential psychology were initially proposed as means to a common end. Methodologies emphasising species-typical features and those emphasising between-subjects variation share a number of common ancestors, perhaps the most illustrative of which is the career of Sir Francis Galton (Galton, 1889; Allen, 2002). Whereas vaguely evolutionarily-guided biological insights have shaped such influential theories as those of Sigmund Freud (Young, 2006) and B.F. Skinner (Skinner, 1966, 1984), Galton (a half-cousin to Charles Darwin) focused very specifically on the application of several Darwinian principles to studying the human species (Forest, 1995).

Galton is of central relevance to the history of personality and individual differences (Bynum, 2002), having pioneered the psychometric assessment of both abilities and dispositions, first articulating the paradigm of ‘nature vs. nurture’, and developing statistical methods oriented around correlation and the use of regression towards the mean with standard deviations (Simonton, 2003). Though now remembered poorly for his advocacy of eugenics, Galton’s endeavours in measuring variability and supplementing selective pressures in human populations were two necessary components of a single ambition: to preserve and aid the evolution of the human species, with particular regard towards human intelligence and character (Jensen, 2002; Seligman, 2002).

Galton understood that the most crucial aspects of Darwin’s theory of evolution by natural selection can be broken down into two discrete concepts. Firstly, that all populations of organisms contain some meaningfully heritable variation, and second, that the differential efficacy of these variants with regard to the demands of survival and reproduction produce a form of selection (Darwin, 1859; 1871). The representative properties of any species will

reliably change over time, in such a manner as to increase their contextual reproductive success, so long as sufficient variation and selection can occur. In the introductory pages of their recent edited book, Buss and Hawley (2011) state flatly: “Individual differences are indispensable for natural selection. Without heritable variants, natural selection – the only known process capable of creating and maintaining functional adaptations – could not occur.” (p. ix).

From this perspective, we can appreciate, in much the same manner as Galton and his contemporaries, that the study of population variation and the study of selective pressures are two sides to the same coin. Both aspects are necessary to understand the history and present-state of biological and psychological functioning, and our richest insights must be born of complex interactions between the two. Thus, to understand the apparent rift that has since formed between these two philosophically congruent fields, one must turn an eye to their separate trajectories of historical implementation.

Contrasting Focuses and Conflicting Methods

The technological progression of the past 150 years has precluded the study of human variation and the study of human evolutionary design from developing hand-in-hand. Darwin’s original articulation of evolutionary theory was inhibited from its inception by a lack of insight into the molecular mechanisms of heredity. While basic inheritance of biological traits had been well-observed, it was not until more than 50 years later, when Mendel’s theories of genetics and Morgan’s chromosome theory were integrated, that biologists were in a position to undertake meaningful investigations into the propagation of

varying traits throughout a population (Huxley, 1942; Dennett, 1995a; Bowler, 2003; Olsson, Hobfeld, & Breidbach, 2006). From the beginning of the 20th century, the study of selective pressures was impaired for many decades, awaiting both the gradual stockpiling of heritability data, and the development of molecular-genetic and computer-modelling techniques.

During this time, several disciplines focussing on measuring and predicting population variation thrived (Stern, 1911), most notably the burgeoning field of differential psychology (Lamiell, 2003; Bergman & Trost, 2006; Uher, 2008). These early endeavours did not suffer at all in the absence of a study of selection, for the findings themselves were considered simply a cross-section of a presumably changing population. Since selection can only occur between generations, only measures of variation spanning over two or more generations would require insights into selection to be understood. It is during this period, while selection-focused sciences were still handicapped by technology, that differential psychology flourished.

The early differential techniques fed strongly into many of the experimental psychology approaches of the era (Tucker, Sinclair, & Thomas, 2005), enduring the dominance of behaviourism to then be reinvigorated by the cognitive revolution that followed (Block, 1989; Baum, 1994; Mandler, 2002; Miller, 2003). During this time, differential psychologists distanced themselves from the rapidly shifting theories in related fields, and came to rely heavily on their robust statistical constructs and improving ability to predict outcomes (Lubinski, 2000; Maltby, Day & Macaskill, 2007). Growing beyond initial interests in improving the process of military recruitment, differential psychology forged close relationships with many areas of applied psychology (Tyler, 1965). The domains of personality- and intelligence-testing in particular, grew ever-more prominent in predicting and informing outcomes including educational development, vocational selection, risk-

management, and mental and physical health outcomes, to name only a few (Karasek, 1979; Lubinski, 2000; Marks et al, 2005; Reisner, 2005; Maltby, Day & Macaskill, 2007).

From the late 1980s to the present day, differential psychology has fortified its position as a central pillar of psychological science, with influential constructs such as the ‘g’ factor of intelligence and trait models of personality standing at the centre of decades of empirical support (Chamorro-Premuzic & Furnham, 2006; Reeve & Charles, 2008; Block, 2010). Contemporary personality and individual differences research is defined by constructs that rely little on grounding theories, but rather, are built on robust statistical data drawn from large populations (Borsboom, 2005). One might thus presume that researchers would regard differential psychology constructs as having limited or strictly instrumental use, relative to the explanatory theories they diverge from. To the contrary, however, trends in the literature suggest that differential constructs are thriving, while theory-based and qualitative research approaches are systematically disfavoured (Rogers, 2000). One explanation for this bias is the ‘quantitative imperative’ (Michell, 1990; 2003a; 2003b): “The quantitative imperative is the view that in science, when you cannot measure, you do not really know what you’re talking about, but when you can, you do” (Michell, 2003a p.5). According to Michell (2005) this quantitative imperative acts both as an explicit principle and as a subtle network of social and institutional biases. Through such influences, the individual differences field has come to embrace its historical overspecialisation in nomothetic statistics as evidence of true scientific validity (Borsboom, Mellenbergh & Van Heerden, 2004; Borsboom, 2005).

In contrast, early attempts to address human psychological phenomena with reference to selective pressures only began to emerge in the latter half of the 20th century, under the umbrella-term ‘sociobiology’ (Hamilton, 1954; Wilson, 1975). These attempts ultimately proved conceptually inadequate, as many were highly reminiscent of the genetic-determinist theories then-prevalent in ethology and zoology, or depended intimately on the then-

controversial prospect of group-selection (Gould & Lewontin, 1979; Gould, 1981; Vining, 1986). Only in the late 1980s did the adaptationist paradigm of evolutionary psychology fully emerge (Buss, 1984, 1995; Cosmides & Tooby, 1989; Tooby & Cosmides, 1989), requiring another decade of development before the approach became widely acknowledged (Confer et al, 2010; Fitzgerald & Whitaker, 2010). Evolutionary psychologists established a refined adaptationist approach, drawing from contemporary cognitive psychology to strongly emphasise the modularity and domain-specificity of hypothesised psychological adaptations (Cosmides & Tooby, 1987, 1997; Pinker & Bloom, 1992; Pinker, 1997; Nesse & Lloyd, 1992). Evolutionary psychology specifically targeted those features of psychological functioning which are species-typical mechanisms that evolved in response to the recurring survival and reproductive challenges of Pleistocene epoch human ancestors (Buss, 1999; 2005). Such species-typical features offer an important means of empirical hypothesis-testing, as only ubiquitous, biologically-based features are likely to exist in similar forms cross-culturally (Buss, Abbott & Angleitner, 1990; Barkow, Cosmides & Tooby, 1992; Tooby & Cosmides, 1992; Buss, 2005).

As of the beginning of the 21st century, an apt summary of the two fields was that evolutionary psychology focuses on the features which are shared across our species, while differential psychology focuses on the ways in which the members of our species systematically differ (Borghans, Golsteyn, Heckmann & Humphries, 2011). Given their shared origins, one might presume that findings of the two approaches must be intrinsically disposed to integration. However, despite some attempts dating back to the formative years of evolutionary psychology (Buss, 1984; 1991), integration efforts have faced theoretical and practical difficulties, to a degree that some view as an interdisciplinary hostility (Anderson, 2004; Muncer, 2011).

To understand this divide, it is necessary to explore some of the unique conceptual challenges that afflict psychology research more so than almost any other field of science. These conceptual difficulties lend disproportionate weight to variations in approach and methods, and are a driving force behind the characteristic rifts between the sub-disciplines of psychology (see Goertzen, 2008 for a diverse account). Moreover, an exploration of these issues can offer an insight into the asymmetrical unification attempts between evolutionary and differential psychology especially (Pinker, 2002; Tooby, Cosmides & Barrett, 2005; Rodeheffer, Daugherty & Brase, 2011), and provide specific means through which such conflicts may, and must, be overcome.

The Unique Challenges of Psychological Inquiry

In order to discuss the challenges that psychology faces as a science, it is necessary to first clarify precisely what is meant by ‘science’. While views on what constitutes ‘science’ vary (Salmon, 1989; Gaukroger, 2006), the scientific enterprise generally consists of two major elements: the systematic observation and description of a particular set of natural phenomena, and the theory-guided explanation of the causes of said phenomena (Wilson, 1998; Cervone, 1999; Boag, 2011). In employing this definition, the authors seek to approximate the position advocated by Wilson (1998), and emphasise that the role of science is to ‘factor out human values’ through procedural error-checking, with the goal of developing ‘representations of reality that are as accurate as possible’.

A Science within a Black Box

To understand the conceptual difficulties of psychological inquiry, it is illustrative to regard all aspects that cannot be immediately observed as existing within a figurative ‘black box’. In the engineering sciences, a ‘black box’ is the catch-all term for any system that has traceable outputs, and at least somewhat traceable inputs, but of which one can gain little to no direct insight into the internal processes that bridge between them (House, 1991; Nairne, 1997; Astbury & Leeuw, 2010). The black box nature of psychological phenomena poses few difficulties for the tasks of observation and description, as these are generally concerned with the system's inputs and outputs (behaviours, levels of activity, etc.). Black boxes do, however, pose substantial challenges to the task of explanation.

Since a phenomenon can only be explained via reference to those related antecedents which, in the past, caused its current state, black box systems concern observable phenomena (the outputs and inputs of the black box) which have causal relations to elements and/or objects that cannot be observed (Kitcher, 1985; Salmon, 1989; Cervone, 1999; Ketelaar & Ellis, 2000; Hüttemann & Love, 2011). Any explanatory account of the inputs or outputs of a black box must by necessity contain some incomplete space, permitting nothing more concrete than speculation. As an example, consider an alarm clock, with the standard inputs (an electric power cord) and outputs (patterns of light and sound). While one might reasonably presume that the device contains electrical circuits that keep time, we must acknowledge that without opening the box, one can only speculate as to precisely what form these internal components take. Relying only on the inputs and outputs, we have no means with which to distinguish between multiple options that achieve the same overt patterns, for any mechanism capable of keeping time, regardless of method, would be functionally identical.

This limitation underpins one of the defining characteristics of the scientific method: hypothesis-testing, which acts as an algorithmic process comprised of both the generative and selective phases that most diagnostic procedures rely on (Fisher, 1925; Kaplan, 1964). It is common when investigating natural phenomena to only have a subset amenable to direct measurement. As such, hypothesis-testing is employed to interpret predictive patterns in that which is observed, to infer the possible characteristics of the variables that cannot be observed (Bunge, 1963; Beizer, 1995). The black box metaphor need not only refer to physical limitations, but rather, a situation can present as a ‘black box’ relative to the means of the investigator. Any situation is figuratively a black box, if vital explanatory details are amenable only to the hypothesis-testing of peripheral phenomena.

From a methodological perspective, the fundamental limit to the utility of hypothesis-testing is that a theory could only be definitively ‘proven’ via the exhaustive disproving of all possible alternative hypotheses. For most kinds of black box situations, there are an effectively infinite number of alternative hypotheses concerning the character of the hidden sections. Thus, heuristics that guide investigators toward testing the most likely or plausible hypotheses are the saving grace that renders actual hypothesis-testing possible. Such heuristics are generally drawn from theory, however, and the more extensive or multi-layered the black box is, the more potentially inscrutable the input-output contingencies become (Fisher, 1925; Bunge, 1963; Kaplan, 1964; Beizer, 1995; Kaplan & Craver, 2011; also see Cervone, 1999 for psychology-specific discussion).

This fundamental challenge of constructing explanatory theories for complex black box phenomena is the central philosophical and conceptual difficulty that defines psychology as a science. To a degree largely unshared by any other natural science, the black box phenomena comprising the information-processing systems of humans and other animals are near-insurmountably complex. The subject matter of psychology concerns highly interpretable

stimuli, passing through immensely long, largely immeasurable, variable and internally-referential causal sequences (Jaszczolt, 1996), to emerge as behavioural outputs that are themselves highly interpretable (De Los Reyes & Kazdin, 2008).

As an immediate consequence of this, sub-disciplines of psychology are particularly vulnerable to sectarianism and disunity. Most fields of psychology have, understandably, built their theories and explanatory models using those insights most conducive to answering their specific research questions (Matarazzo, 1992; Kelly, 1998). As a result many fields of psychology make dissonant or contradictory pragmatic assumptions about those aspects of the mental black box that they are not presently addressing. The mutually incompatible assumptions that characterise different research approaches appear to be responsible for the majority of institutional disunity in psychology, including the rift between evolutionary and differential psychology.

Refining Explanatory Theories

Although description is fundamentally necessary for explanation to occur, explanation is arguably the highest goal of science (Wilson, 1998; Cervone, 1999; 2005). As such, a research approach in psychology is perhaps the best judged in terms of its ability to constrain theories and predictions, so as to reliably draw maximum utility out of practical hypothesis-testing (see Kaplan, 1964, chapter 2 for general elaboration; for discussion specific to evolutionary psychology, see Resnick, 1996; Sober, 2000; Lewens, 2002).

There are, in general, three means of informing an explanatory theory prior to (or in conjunction with) prediction-testing of input-output contingencies (Bunge, 1963). The first,

and often most difficult, option is to attempt to directly measure the contents of the black box.

In psychology, this may be achieved in two ways: directly, through the use of various neuroimaging technologies, and analogously, through the invasive (generally surgical) manipulations of non-human animals. While there is not nearly sufficient space here to discuss the valuable psychological insights that have been gathered through these respective methods (for some key topics, see: Stevenson & Goldworth, 2002; Bennett & Hacker, 2003; Tashiro, 2004; Filler, 2009; Dietrich & Kanso, 2010), for the specific purpose of theory-building their utility is none-the-less akin to that of standard observation-based methods.

While many intuitively assume that the real-time outputs of fMRI scans provide privileged access to the content of the mind, neuroimaging technologies only provide us with activity patterns, which while potentially closely correlated with the information-transformations of the mind, do not constitute direct measurement of the phenomena in question (Caplan, 2009). Even if neuroimaging techniques were so technologically refined as to accurately discern specific action potentials and the dynamic dendrite configurations of individual neurons, the interpretation of these patterns into meaningful psychological content could still only be achieved via detailed correlation with some other source of insight into said processes (Bennett & Hacker, 2003). Though immensely instructive, these methods cannot side-step the fundamental difficulties of hypothesis-testing, but rather can only be taken alongside psychological testing as means of refining existing hypotheses (Caplan, 2009; Filler, 2009).

The second option for refining theories independent of testing involves the use of logical inference to determine what must be the necessary minimum requirements of the systems in question, assuming that said systems are physically internally consistent. This method is extensively employed in computational cognitive psychology (Fodor, 1975; 1983), and is the guiding heuristic of all computational models (Neisser, 1967; Boden & Mellor, 1984; see chapter 4 of Boden, 2006 for a wider historical context). While insufficiently discriminative

in their own right, such logical inferences become vastly more powerful when supplied with alternative insights into the limitations of the psychological processes in question (for example, basic neurological insights into the properties of neurons and regional clusters of the brain).

The third, and perhaps final option for refining explanatory theories, is the independent discovery of design details (Lewens, 2002). In mechanical and electrical engineering, such insights may take the form of early blueprints, listing all available materials and tools, or learning what objectives a system was designed to implement (Dorst & Cross, 2001). In a manner wholly analogous to engineered design by humans, abundant evidence suggests that all organisms were designed (Dawkins, 2009), over a geological timescale, by a range of algorithmic evolutionary forces (see chapter 8 of Dennett, 1995a, for further details).

Reliance upon details indicative of the design process is the central principle of the adaptationist approach, and is thus the heuristic core of evolutionary psychology (Buss, 2005). While embracing the adaptationist approach is not strictly necessary to gain some of the crucial benefits of the third aforementioned option (indeed any biological, medical, and developmental insights into the properties of the nervous-system provide powerful tools for use with the second and third), the adaptationist approach is designed to draw as much theory-guiding information as possible from the reciprocal relationships of form versus function. Stated simply, adaptationist heuristics regard what a mind *is* (structurally) as being intimately related to what a mind *does* (functionally), by in turn acknowledging that *how* a mind functions has been shaped by *why* it functions, in a Darwinian sense (Hodgson & Knudsen, 2008).

Reverse-Engineering and Adaptationism

The paradigm of evolutionary psychology is primarily concerned with explanation, and this orientation has formed the basis of its conceptual incompatibility with the most prominent domains of differential psychology. By examining the explanatory methods employed in evolutionary research, the authors will demonstrate, by contrast, the explanatory short-cuts that have become entrenched in differential psychology, which keep differential researchers at odds not merely with adaptationists, but with theoretically robust psychology in general.

The Guidance of Design

The adaptationist approach is the defining aspect of any work of evolutionary psychology (Sober, 2000; Buss, 2005). An adaptation (when used as a noun) is understood to be a feature or set of features of an organism, the apparent design or concerted complexity of which suggest that it is a product of natural selection, and thus represents a relational calibration of said organism to its ancestor's recurring environmental challenges (Tooby & Cosmides, 2005). At the heart of this paradigm is the suggestion that the species-typical behavioural and cognitive regularities of animals (of humans in particular), likely consist of, or are actively shaped by, adaptations.

Evolutionary psychologists focus on adaptations primarily for pragmatic, explanatory reasons. While all organisms are the products of natural selective forces (and artificial selection in domesticated species), not all features of organisms are adaptations. In the words of Tooby and Cosmides (2005, p. 25-26):

The cross-generationally recurrent design of an organism can be partitioned into (1) adaptations, which are present because they were selected for, (2) by-products of

adaptations, which were not themselves targets of selection but are present because they are causally coupled to or produced by traits that were, and (3) noise, which was injected by the stochastic components of evolution.

For reasons of logical necessity, it is nearly impossible to use any positive criteria to confirm that some biological or psychological characteristic is either a by-product or phylogenetic noise. However, a feature can generally be identified as an adaptation when it shows contextual evidence of 'good design' with relation to the adaptive problem or problems it is hypothesised to address (Dennett, 1995a; Buss, 2005).

Functionally speaking, adaptations are relations between organism characteristics and the fitness demands which statistically favoured those characteristics in their gene-pool (Dennett, 1995a; Sober, 2000; Dawkins, 2009). Thus, no trait can be accurately described as an adaptation in the absence of this feature-problem matching. For this reason, adaptationists approach complex features with the postulation of possible adaptations, moving on to the possibilities of by-products and noise when the evidence for adaptation is lacking or inadequate (Tooby & Cosmides, 2005). The tell-tale signs of biological design are the clues used by evolutionary psychologists to generate and refine theories about the probable structure and development of a psychological adaptation, utilising the intrinsic relationships between the form and function of a well-designed system (Dennett, 1995a; Pinker, 1997; 2002; Tooby & Cosmides, 2005). Investigations of this sort are appropriately referred to as 'reverse-engineering' (Buss, 2005), though it is worth noting that in seeking to gain insight into black box structures through inference from observable input-output contingencies, one could easily argue any psychologist employing explanatory theories is, by necessity, a reverse-engineer (Dennett, 1995b).

Literal Structures Defined by Function

Evolutionary theory regards the ‘mind’ as a coordinated system of fitness-enhancing problem-solving apparatus. These hypothesised adaptations are specified to strictly consist of computational neurophysiological structures (Crawford, 2000; Cosmides & Tooby, 2001; Tooby & Cosmides, 2005). The existence, performance, and related properties of these adaptations are predicated upon the function they were selected for (Sappington, 1990; Keri, 2003). This focus on information-processes clearly lends itself to many process models in psychology, while many other targeted phenomena in psychology, such as intrinsic ‘traits’ (Church et al, 2006), internal representations (Fuhrman & Boroditsky, 2010), and qualitative mental states (Markus, 1998), can be understood as calibrated components, products, and observation-level descriptions of psychological processes (see Buss, 2005 for further detail). In contemporary evolutionary psychology, such structures are defined as psychological mechanisms, commonly further designated into processing ‘modules’ (see Buss, 1995; Cosmides & Tooby, 2005, concerning the Massive Modularity Hypothesis).

This account of causally-integrated psychological mechanisms is vital to the conceptual lexicon of evolutionary psychology, and sets a clear yet inclusive standard for the compatible expression of any scientifically viable explanatory psychological construct (including those not thought to be adaptations). The viability of proposing such structures depends largely on evidence found in concerted phenotypic function. As such, the adaptationist approach also provides a unique means of bridging the gap between literal and non-literal construct-based theories, because any construct that is defined by its function is conceivable and testable as a literal, neurophysical psychological mechanism (Dennett, 1995b). Despite these evident benefits, it is precisely this concept of psychological mechanisms, and the detailed explanatory approach that such a conception demands, that is responsible for the much of

incompatibility between the theories and approaches of evolutionary and differential psychology.

Top-Down Explanations and Descriptive Constructs

There is perhaps no more fitting a characterisation of differential psychology than as a field that endeavours to be descriptive. The methodologies and conceptual-tools of differential psychology are supremely well-adapted to the tasks of summarising and extracting the statistical cores of behaviourally-recurrent trends in populations. With such immense statistical credentials, differential psychology is considered perhaps the greatest beneficiary of the above-mentioned quantitative imperative in behavioural science (Michell, 2003a). Indeed, researchers routinely seek to establish the real-world relevance of theory-based explanatory models (particularly concerning cognitive abilities and personality traits) through the use of differential descriptive constructs. It is telling that the opposite is only very scarcely the case.

The most prominent constructs in differential psychology, the general factor of intelligence ‘g’ and the largely orthogonal personality trait dimensions of the Five Factor Model, were founded with few-or-no explanatory tasks in mind (Meehl, 1998; Lubinski, 2000), and have built their reputations instead on robust statistical properties and impressive correlations with life-outcomes. The ‘g’ construct is an illustrative example, for contrary to common opinion, g is not an explicitly (linguistically) defined construct that is supported by a nexus of covarying statistical trends between many measures. Instead, ‘g’ is simply a name given to a robust statistical nexus of covariation (Lubinski, 2000). Similarly, the orthogonal factor structures of the Five Factor Model of personality take precedence over any worded definition of the factors in question, in a sense making the definition of the factors

intrinsically and permanently subject to interpretation (Cattell, 1996; McCrae & Costa, 1999; Grucza & Goldberg, 2007).

The esteem and popular use of such descriptive constructs has, however, led to their insertion into domains that do not match their original intentions or conceptual strengths. While differential descriptive constructs have proven their value through predictive correlations with achievement and outcome measures (Lubinski, 2000), in recent decades the literature has seen the rise and growing acceptance of individual differences papers which employ said descriptive constructs as proposed causative agents in simple explanatory theories (see Boag, 2011 for a detailed account). This form of explanation-description substitution produces a range of far-reaching conceptual problems, particularly with regard to circular reasoning and reification. As the following examples demonstrate, there are limited circumstances in natural science where empirical inquiry into antecedent causes cannot continue, and detailed description is embraced as a surrogate form of explanation. This explanatory approach is viable for only a minority of natural phenomena, and is intrinsically ill-suited for psychology and cognitive science.

Limiting Cases

When utilising descriptive constructs in the role of causative agents, one is relying upon the assumption that reliable trends in observable behaviours are indicative of specific causal forces, be they agents or merely ‘laws’ of expression (Boring, 1950). While this assumption is far from unheard of in some natural sciences, the subject matter of many scientific fields are not nearly as obfuscated by the black box limitations inherent to psychology. For two examples, consider the well-regarded fields of classical molar chemistry and moderate-scale Newtonian physics (Kitcher, 1985). These two fields have enviably few ambiguities in their

subject matter, provided they are measured with sufficiently accurate instruments. Subsequently, both molar chemistry and Newtonian physics are founded upon reliable explanatory 'laws', such as Gay-Lussac's law or the Law of Universal Gravitation, all of which were discovered essentially atheoretically through the logical induction of observable trends. While these inquiries yielded theories, they did not require any assumed theoretical framework to undertake. In the terminology advocated by Cervone (1999; 2004; 2005), the explanatory method employed in these two examples, and subsequently misemployed when employing descriptive psychological constructs in explanatory roles, is referred to as *top-down* explanation (see also Kitcher, 1985; Salmon, 1989; Glennan, 2002).

Top-down explanation relies upon the induction of reliable, structural trends and distinctions, based purely on observational regularities. Of particular interest to psychologists, research programs that employ a top-down explanatory approach are directly compatible with population-level data, as inductions are best made statistically from a wide pool of nomothetic observations. In some sciences, such as chemistry and physics, sufficiently robust observational trends can be reliably assumed to correlate with fundamental causal forces, but such accounts are minority cases not to be confused with the wider sense of explanation, which relies upon giving accounts of causal antecedence (explored in-depth in *Kitcher, 1985*).

In the example of the Law of Universal Gravitation, Newton described in great detail the patterns of relative moment between bodies with mass, and ascribed the name gravity to the consistencies observed (Keesing, 1998). Thus in Newton's model, it is true that positing the force of gravity successfully explains the movement of objects with mass (within particular limits), but the phenomena of gravity itself remains merely described, and not explained at all. To this day, physicists struggle with competing theories in an effort to give a substantial antecedent-based explanation of gravitation and mass, but in Newton's era the viable limit of

inquiry had been reached, and it was enough to say that the explanatory effort could end at a detailed description of the most fundamental accessible cause. Though such reasoning is inescapably circular, this description-explanation substitution was accepted due to the immense regularity of the patterns being observed, and because the phenomena in question are so fundamental and causally inscrutable, that the act of reification would not result in the premature dismissal of accounts of true causal antecedents. In psychology, however, this is far from the case.

Misapplication

The most prominent contemporary example of descriptive constructs being invoked as top-down explanations of behaviours, are those centred on the Five Factor Model of personality (McCrae & Costa, 1994; 1997). The problems with attempting to use super-ordinate traits in this manner are two-fold: Firstly, psychological phenomena do not meet the conditions of simplicity and observational clarity required to employ an empirically coherent top-down analysis, as most relevant behaviours demand some interpretation or contextual inference to be studied (De Los Reyes & Kazdin, 2008). Human (and animal) behaviours are the result of many cumulative causal forces, whose patterns and configurations cannot in any way be directly induced from observable behavioural trends (Cervone, 2004; 2005). Second, these super-ordinate personality traits are proposed as explanations of the very behaviours that they are aggregated from. This represents internally-inconsistent circular reasoning, as a discrete phenomenon cannot be coherently understood to cause itself (Skinner, 1953; Hanson, 1958; Nozick, 1981; Bandura, 1999; Cervone, 2005; for a more complete treatment of the logical inconsistencies and reification errors in personality trait models, see Boag, 2011).

While the aforementioned conceptual problems are readily identified by those familiar with cases of circular reasoning, attention must also be drawn to the practical and methodological barrier between said constructs and explanatory theories in psychology. Although differential psychologists can and do utilise repeated-measures and other within-persons approaches, the majority of popular descriptive constructs are derived nomothetically, based upon between-persons patterns within sampled populations, and are thus befittingly labelled ‘difference variables’ (Lubinski, 2000). Generally, these population-level variables are presumed to serve as indicators of some intrapersonal factor that determines an individual’s contribution to the variation within a group, but as is pointed out by Borsboom & Dolan (2006), such assumptions cannot be embraced without empirical support. To simply presume equivalence between hypothetically related variables, when one exists on the individual-level and the other on the population-level, is conceptually unsound. These conceptual problems compound even further the more aggregated or abstracted a construct is from direct behavioural measurements. A clear example of this conceptual error can be found in the works of Kanazawa (2010a), which investigate ‘intelligence’ as an adaptation for negotiating evolutionarily novel stimuli, while relying methodologically upon the general factor *g* (Kanazawa, 2006a; 2006b; 2006c; 2007; Lynn & Kanazawa, 2008; Kanazawa & Perina, 2009; Kanazawa & Reyneirs, 2009; Kanazawa, 2010a; 2010b). Kanazawa’s theories presuppose the existence of a mechanism of general problem-solving, which is further assumed to correlate with population-level intelligence-differentials so closely that the *g* construct can be taken as its direct measure. As Borsboom and Dolan (2006) demonstrate, neither the probable existence of this mechanism, nor its presumed correlation to *g* have any substantive empirical or theoretical support. Conversely there are also a number of compelling reasons to believe that domain-general problem-solving mechanisms of the sort described cannot exist coherently in a computational framework (see

Penke et al, 2011 for details). Kanazawa's use of *g* illustrates precisely the kinds of conceptual errors that arise when the untenable 'top-down' explanatory approach native to differential psychology attempts direct integration with more robust theories, which rely upon a 'bottom-up' approach to explanation.

Bottom-Up Explanations and Process Models

In contrast to top-down explanatory methods, Cervone (1999; 2005) also speaks of their conceptual opposite, called simply 'bottom-up' explanation. This is the form of explanation predominantly referenced throughout this paper, and is the approach required by adaptationism. Bottom-up explanations are comprised of either literally specified causal antecedents, or functionally-defined approximations of possible literal causal antecedents, hypothesised to underlie the phenomena of interest (Cervone, 2005). To varying degrees, all process models in psychology (specified at the level of an individual) are designed to employ a bottom-up explanatory approach, as they rely upon establishing the counterfactual causes of the phenomena in question (Edwards & Jaros, 1995). There are, however, two key conceptual limitations to the use of classical process models in seeking bottom-up explanations. The first issue concerns the relative completeness of a process account, while the second concerns the difficulty in addressing the first issue via the integration of multiple models.

Incomplete and Incompatible

Process models have been proposed to describe innumerable specific domains of cognition: the expression of innate temperaments (Richards, 1986; Eysenck, 1994; Mauer & Borkenau, 2007; Aron et al, 2010), the formation of attitudes (Tybout & Scott, 1983; Park et al, 2007),

detail-extraction in perception (Marslen-Wilson & Warren, 1994; Vandenbroucke et al, 2009; Wascher & Beste, 2010), and in social learning processes in general (Bandura, 1986; 1989), to name only a few. Each of these examples demonstrates that strong theories of probable internal operations can (and must) be induced from a wide variety of formative and design-related clues. However, each theory is also fundamentally incomplete when considering the black box nature of the mind. In order to be reliably scrutinised via hypothesis-testing, a theory should account for at least some form of influence at all relevant stages of information-transformation between input stimuli and behavioural output. For example, a process account of reacting to a perceived stimuli should give some consideration to each stage of influence, from perception, to recognition, to motivation, to contemplation, and finally to expression, because variations at any of these levels would fundamentally change the observed input-output contingencies. While such a task may be impossible to achieve in exhaustive detail, and no theorist could be reasonably held to so lofty a standard, the more complete a theory's account of the causal sequence is, the lower the chances that some overlooked variable might skew or invalidate the results.

An intuitive solution to this issue would be to rely on existing process models of related psychological phenomena to supplement those points in a model where intervention would be meaningful. Unfortunately, the persistence of this problem can be largely attributed to issues of terminology, which present an obstacle to integration. Even those processes whose causal pathways of interest may appear mutually compatible are often kept separate by the incompatible referential terminologies of the fields from which they originate (Henriques, 2003). For example, Ho and Fung (2011) published a detailed process model of forgiveness, designed to account for some cultural influences on when and how forgiveness occurs and is displayed. By defining the process of forgiveness in terms of changes in affect and appraisal towards a transgressor, Ho and Fung adopted a functional approach well-suited to cross

cultural comparisons, allowing for the simultaneous consideration of emotion, motivation, and other cognitions (for background on this approach, see Enright & Fitzgibbons, 2000). While this model does well by considering a wide variety of potential points of influence in the forgiveness process, some stages (deliberation and expression, in particular) are construed in such a manner as to leave their relationship to other published models vague. Rather than indicating how related models overlap with the stages described, or alternatively, justifying why the existing distinctions prevalent in the literature are inappropriate in this context, both interpretations appear potentially viable. For example, the model (p.79) defines a process of ‘dialectical thinking’ as a major stage in forgiveness, but gives limited elaboration on what this consists of. From the descriptions, dialectical thinking appears to involve comprehension and attribution, cognitions that have also been addressed with cognitive process models in recent years (Rossett, 2008; Ali, Chater & Oaksford, 2011). Unfortunately, the authors neither acknowledge this potential overlap, nor explain why the terminology used is to be preferred. It seems that the possibility of integration was simply not considered, and that the distinctions employed in this model are idiomatic to the research task. Similarly, the forgiveness model accounts for cultural sources of variance in the emotion-negotiation and expression of forgiving sentiments, but not in a manner immediately compatible with prevailing process models of emotion-regulation (Ochsner & Gross, 2008; Thiruchselvam et al, 2011). It seems that with several basic changes to the defining terminology, this model of forgiveness could potentially be integrated with models of related phenomena, to yield testable predictions in far more substantive detail. Such conceptual clashes are par-for-the-course in psychology research, with only a minority of new theories showing explicit aspirations for wider integration (see Sheldon, 2011, as an example).

Integration through Adaptationism

The paradigm of evolutionary psychology offers a valuable potential solution: the standardisation of referential language into the terminology of modern computational cognitive psychology (Cosmides & Tooby, 2000). An adaptationist theory must be either functionally-oriented toward behavioural outcomes, or hypothesise directly about literal psychological mechanisms. As such, employing evolutionary terminology ensures that effectively any process theory can be expressed in a manner highly compatible with many (and potentially all) other psychological mechanisms (Buss, 2005). Unlike other more abstract procedural concepts, adapted psychological mechanisms are conceptually primed to integrate on the basis of function (see Tooby & Cosmides, 2005 for further discussion). Beyond this, adaptationists can qualify meaningful predictions purely on the level of manifest behaviour, because any well-designed adaptation must not interfere with the successful engagement of other mechanisms, except in explicit conditions of evolutionary mismatch (explained further in Tooby & Cosmides, 2001). In these two ways, the grounding theories of evolutionary psychology allow for potentially any process-based theory to be incorporated into more complete, conceptually sound, bottom-up theories. As such, adaptationist theories demonstrate a conceptual interplay between descriptive and explanatory tasks not commonly seen psychological science.

The Evolution of Individual Differences

As was explored in the preceding sections, the prevailing methods in differential psychology cater specifically to the scientific task of description, and are thus not only theoretically-impoverished with regard to explanation, but appear irreconcilable with more theoretically-robust approaches (Anderson, 2004; Muncer, 2011). These arguments are not to be taken as a

general indictment of differential psychology, which remains a highly successful and instructive descriptive enterprise, but merely as a warning and reminder that top-down explanations are scientifically ill-suited to psychological phenomena.

The descriptive nomothetic data provided by prominent differential psychology constructs are commonly designed for highly generalised predictions of outcomes, rather than to provide details that disambiguate the mysteries of particular explanatory models (Lubinski, 2000). This explanatory neutrality represents the primary obstacle to researchers hoping to harness statistically powerful descriptions in aid of explanatory hypothesis-testing. Such researchers must struggle to interpret the meaning of quantitative differences that, as explored above, often do not easily map onto linguistic definitions (Cervone, 1999; 2004; 2005). If theorists hope to modify descriptive constructs to better inform causal explanations, population-level behavioural variations must be measured in a manner more indicative of the intrapersonal variables suspected to cause them (Borsboom & Dolan, 2006). That is to say, individual-differences measures must be adjusted so as to preserve (rather than control or mask) individual-level details that map onto the relevant features of explanatory theories. Without such considerations, any research paradigm seeking to bridge the gap between its specific hypotheses and the wider observations of differential psychology, must struggle in vain to match those elements in their explanatory theories thought to produce systemic variations, to the form said variation is expected to take on a generalised behavioural level.

Though some integration efforts have endured for decades (Buss, 1984; 2009), only in recent years have leading evolutionary psychologists embraced the task of modifying and expanding traditional adaptationist theories, in order to account not only for sources of random variation, but also variations preserved or arising from selective forces (Tooby & Cosmides, 1990; Confer et al, 2010; Buss & Hawley, 2011). The following section briefly

details some recent expansions of evolutionary theory into areas once thought to be the exclusive purview of classical differential psychology.

When Selection Maintains Variation

Since the infancy of evolutionary psychology, David Buss (1984; 1991; 1995; 2009) has explored the concept that a species may evolve a species-typical suite of adaptive interaction strategies (rather than a single ‘one size fits all’ strategy), which are activated or deactivated developmentally as a means of calibrating an individual to the particular adaptive challenges of their lifetime (see also, Marsh & Boag, 2010). Despite the promise of this conception with regard to understanding personality psychology, this model presupposed a complex adapted system whose existence must be second-order to, and in principle shaped by, the more basic selective influences thought to also produce systemic variation (Buss & Greiling, 1999). As such, the greatest advances over the past 10 years of variation-focussed evolutionary psychology have comprised a range of sophisticated conceptual and empirical syntheses, aimed at exploring nuanced and often-overlooked Darwinian effects on the cognitive and dispositional properties of human individuals (Michalski & Shackelford, 2010). Speaking broadly, three largely distinct selective phenomena have been refined as viable sources of systemic individual differences in evolved psychology: First, that some dispositions and tendencies represent selectively-neutral or frequency-dependent fitness tradeoffs (as in the case of some personality traits). Second, that some abilities vary due their configural sensitivity to mutation-selection balance (as in the variables of human intelligence). Lastly, in accordance with Buss’s founding insights, some psychological phenomena may vary as a function of niche-selecting mechanisms, be they cognitive or epigenetic. This final conception of variation remains largely in its infancy, and will not be

discussed at length here (for a wide overview of the potential impact of this perspective on both addressing and redefining psychopathology, see Kennair, 2011).

With regards to fitness tradeoffs, early research (see Buss, 1995) investigated the influence that highly flexible, rapidly-changing environments, would likely have on the slow inter-generational process of trait-favouring selection in a population. Analysis suggests that some human ancestral environments may appear selectively-neutral by virtue of selective pressures either frequently changing, or being too contingent on intra-generational factors (see Belsky, 1999 for summary). This analysis was enriched by increasingly sophisticated tradeoff theories, hypothesising that the fitness optima of highly variant traits are in fact their ‘moderate’ as opposed to ‘high’ levels, since extremes along many trait continuums are likely to confer maladaptive side-effects (see Keller & Miller, 2006; Nettle, 2006; Penke, Denissen, & Miller, 2007a, 2007b; Ellis, Figueredo, Brumbach & Schlomer, 2009, for details). Building on these insights, theorists were able to account for the selective value of some seemingly disadvantageous, yet common, behavioural tendencies (such as those related to both competitive and altruistic social compulsions) via the inclusion of costly signalling theory (see Miller, 2007 for review) and life-history considerations (see Kaplan & Gangestad, 2005 for relevant discussion). These investigations gave rise to the study of frequency-dependant selection, wherein some variations are understood to be differentially effective based on the distribution of the same and other strategies employed by other members of the population (see Penke et al, 2007 for an introduction). With this wealth of insights, evolutionary psychologists now possess a sufficiently nuanced understanding of the selective pressures that likely underlie much of the systematic variation in personality and preference (Keller & Miller, 2006; Penke, 2011; Nettle, 2011).

In contrast, the traditional conception of adaptive optimisation still appears to be relevant in studying the variations found in cognitive abilities and intelligence. Unlike variations in

personality or preference, there appear to be very few tradeoffs or contingent circumstances that render higher levels of ability anything but an unambiguous enhancement of global fitness (Penke, Denissen, & Miller, 2007a). Fortunately, technological and analytical advances in population genetics have allowed the once-elusive concept mutation-selection balance to be applied to the study of cognitive ability (Keller & Miller, 2006; Penke, Denissen, & Miller, 2007a; Penke, Denissen, & Miller, 2007b). It has long been understood that the vast majority of natural mutations between the generations of a species tend to impair the collective functioning of their evolved adaptations. It is the ongoing role of natural selection to counteract this accretion of deleterious mutations by selecting against individuals with the greatest accumulation of impairments (individuals with a high effective ‘mutation-load’). The specific relevance of this phenomenon to cognitive abilities is due to the vulnerability of complex neurological adaptations to relatively small genetic impairments (Michalski & Shackelford, 2010). Since the configuration and optimisation of complex psychological adaptations rely upon many structural and developmental provisions, the collective influence of many coordinated genes and expression-factors contribute to the formation of the delicate final product. Small changes to structural characteristics or enzyme efficiencies can thus result in measurable reductions in the calibrated efficiency of the overall mechanism (Keller & Miller, 2006). Thus, mutation-selection balance suggests that the majority of ordinal variation observed in the heritable characteristics of ‘intelligence’, are due largely to negative influences of mutation-loads not yet ‘filtered-out’ by the omnipresent pressures of selection (Penke, Denissen & Miller, 2007b), which in-turn partially explains some once-mysterious correlates of intelligence, including general health, vascular development, and body symmetry (Penke, 2011).

Finding Variation within Mechanisms

The conceptual tools are now available to other researchers, including career differential psychologists, to begin bridging the divide between evolutionary intra-personal models and traditional individual differences methods. By engaging with explanatory process models, and building upon the elements of those models which permit of individual variations (both as heritable genetic biases or ontogenically calibrated strategies), new causally-relevant hypotheses can be tested with only minor modifications to existing psychometric techniques. Although the above-discussed modes of variation will likely be alien to those without an evolutionary background, it is now well within the reach of differential psychologists to apply their methodological expertise, on both individual and population levels, to enriching even relatively simple process-based evolutionary theories.

The key to such efforts, however, is to embrace the lack of relevance most popular differential psychology constructs have to explanatory hypothesis-testing, and working to produce intermediary measurement tools and approaches that can bridge between the predicted variations within a process model, and what form said variation can be expected to take on an overt behavioural level. A strong example of this kind of research can be found in the social rank/dominance and social-exchange measures developed by Leybman and associates (Zuroff, Fournier, Patall & Leybman, 2010; Leybman, Zuroff, Fournier, Kelly & Martin, 2011; Leybman, Zuroff, & Fournier, 2011). Although the various incarnations of these measures resemble, both in presentation and in statistical verification, traditional differential psychometrics, fundamental design distinctions were taken directly from existing evolutionary process models of how humans negotiate reputation-sensitive social exchanges. Rather than creating and factor-analysing a pool of items, with the goal of retroactively assigning descriptive titles to the factors that emerge, each element of the measures was intended to capture particular sources of intra-personal variation in the theorised

psychological mechanisms, and their statistical validity was judged by how well response-patterns reflected this. Not only are these measures of dominance and social-exchange primed for testing hypotheses pertinent to the explanatory theory they are inspired by, their correlations with other descriptive constructs designed purely on the population-level can further inform an understanding of how intra-personal variations shape (and in the case of frequency-dependent selection, interact with) the overall diversity of the population (Leybman, Zuroff, & Fournier, 2011).

In addition to providing more causally-relevant theoretical structures for the examination of variations already explored at the individual and population level, evolutionary-differential integration may also, on occasion, permit insightful conceptual revisions of some individual differences phenomena that have otherwise eluded explanation. For example, by expanding beyond the initial efforts of J.P. Rushton (1985; 2000; 2004), A.J. Figueredo and colleagues have developed a new approach to studying the General Factor of Personality (GFP), which utilises life history strategy as the ultimate factor organising the seemingly diffuse traits and behaviours observed (see Figueredo et al, 2005; Figueredo & Rushton, 2009). Beyond offering an account of the general organisation of personality traits relating to social functioning, this approach has yielded a range of novel predictions concerning how ontogenic calibrations of life history strategy, such as degree of parental support, shape variation in GFP (van der Linden et al, 2012). Similar life history approaches have recently been applied to other domains of normative variation that have eluded simple explanation, including the clustering of several cognitive aptitudes and personality traits (as explored in Woodley, Figueredo, Brown & Ross, 2013), and the human stress response system (see Del Giudice et al, 2011, for theoretical framing of the model, and Del Giudice et al, 2012, for promising empirical support). Each of these examples demonstrates a collection of psychological phenomena that had been successfully identified top-down as a reliable pattern of variation

by differential psychologists, but which eluded explanation and a source of novel predictions in the absence of a functional account of evolved psychological mechanisms.

Conclusion

In closing, this article has explored both the historical origins and contemporary impact of a perceived incompatibility between differential and evolutionary psychology, within the wider context of the unique challenges psychology faces as a science. The core of this incompatibility can be traced to confusions over, and a lack appreciation for, the distinct scientific tasks of description and explanation. Exclusive specialisation in quantitative descriptive statistics has left differential psychology institutionally powerful, but theoretically impoverished and conceptually isolated, with only limited means of applying its descriptive prowess to causal explanatory models. Evolutionary psychology has demonstrated a range of empirical and conceptual strengths that support its suitability as an integrating platform for functional cognitive and behavioural science. This strength has most recently manifested as a series of sophisticated and highly successful attempts to expand into the territories of differential psychology, thus establishing a range of innovative new means of describing and explaining the underlying causes of individual differences.

Researchers now have the foundations laid for them to develop new, theoretically-rich descriptive tools which can contribute directly to the hypothesis-testing of explanatory process models. Particularly when utilising the heuristic tools of evolutionary psychology, even researchers inexperienced with adaptationism can work to bridge the conceptual gaps between our theories of functional, psychological mechanisms, and our accounts of tendencies and abilities in individuals and the population at large.

References

- Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: Mental models or causal models. *Cognition*, 119 (3), 403-418.
- Allen, G. (2002). The measure of a Victorian polymath: Pulling together the strands of Francis Galton's legacy to modern biology. *Nature*, 145(3), 19-20.
- Anderson, M. (2004). Marrying intelligence and cognition: A developmental view. In R.J. Sternberg (Eds.), *Cognition and Intelligence: Identifying the Mechanisms of the Mind* (pp. 268-287). Cambridge University Press.
- Aron, A., Ketay, S., Hedden, T., Aron, E.N., Markus, H.R., & Gabrieli, J.D.E. (2010). Temperament trait of sensory processing sensitivity moderates cultural differences in neural response. *Social Cognitive and Affective Neuroscience*, 5(3), 219-226.
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), 363-381.
- Baars, B. (1984). View from a road not taken. *Contemporary Psychology*, 29, 804-805.
- Baars, B. (1985). The logic of unification. *Contemporary Psychology*, 30, 340.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), *Annals of Child Development*, 6: *Six theories of child development* (pp. 1–60). Greenwich, CT: JAI Press.
- Bandura, A. (1999). Social cognitive theory of personality. In D. Cervone & Yuichi Shoda (Eds.), *The coherence of personality: Social-cognitive bases of consistency, variability, and organization* (pp. 185–241). New York: Guilford Press.

- Barkow, J.H., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York, NY: Oxford University Press.
- Baum, W.M. (1994). *Understanding behaviorism: Science, behavior, and culture*. New York: Harper Collins.
- Beizer, B. (1995). *Black-box testing: Techniques for functional testing of software and systems*. New York: John Wiley & Sons. Inc.
- Belsky, J. (1999). Modern evolutionary theory and patterns of attachment. In J. Cassidy and P. R. Shaver (Eds) *Handbook of Attachment: Theory, Research, and Clinical Applications* (pp. 141-161). Guilford, New York.
- Bennett, M. R., & Hacker, M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.
- Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly*, 52, 377-389.
- Block, J. (1989). Critique of the Act Frequency Approach to Personality. *Journal of Personality and Social Psychology*, 56(2), 234-245.
- Block, J. (2010). The Five-Factor Framing of Personality and Beyond: Some Ruminations. *Psychological Inquiry*, 21, 2-25.
- Boag, S. (2011). Explanation in personality psychology: 'verbal magic' and the Five-Factor Model. *Philosophical Psychology*, 24(2), 223-243.
- Boden, M.A. (2006). *Mind as Machine: A History of Cognitive Science Vol. 1*. Oxford: Oxford University Press.

- Boden, M.A., & Mellor, D.H. (1984). What is computational psychology? *Proceedings of the Aristotelian Society, Supplementary Volumes*, 58, 17-35+37-53.
- Borghans, L., Golsteyn, B. H. H., Heckman, J., & Humphries, J. E. (2011). Identification problems in personality psychology. *Personality and Individual Differences*, 51(3), 315-320.
- Boring, E. G. (1950). *A history of experimental psychology*. New York: Appleton-Century-Crofts.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.
- Borsboom, D., & Dolan, C. V. (2006). Why g is not an adaptation: A comment on Kanazawa. *Psychological Review*, 113, 433–437.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Bower, G.H. (1993). The fragmentation of psychology? *American Psychologist*, 48, 905-907.
- Bowler, P.J., (2003). *Evolution: The History of an Idea (3rd Ed)*. University of California Press, Berkeley and California.
- Buller, D.J. (2005). (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge, MA, US: MIT Press; US.
- Bunge, M. (1963). A general black box theory. *Philosophy of Science*, 30 (4), 346-358.
- Buss, D.M. (1984). Evolutionary biology and personality psychology: Toward a conception of human nature and individual differences. *American Psychologist*, 39(10), 1135-1147.

- Buss, D. M. (1991). Evolutionary personality psychology. *Annual Review of Psychology*, 42, 459-491.
- Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, 6, 1-30.
- Buss, D.M. (1999). *Evolutionary psychology: The new science of the mind*. Boston: Allyn & Bacon.
- Buss, D.M. (Eds.), (2005). *The handbook of evolutionary psychology*. Hoboken, NJ, US: John Wiley & Sons Inc; US.
- Buss, D.M. (2009). How can evolutionary psychology successfully explain personality and individual differences? *Perspectives on Psychological Science*, 4 (4), 359-366.
- Buss, D. M., Abbott, M., & Angleitner, A. (1990). International preferences in selecting mates: A study of 37 cultures. *Journal of Cross-Cultural Psychology* 21(1), 5-47.
- Buss, D. M., & Greiling, H. (1999). Adaptive individual differences. *Journal of Personality*, 67, 209 -243.
- Buss, D. M., & Hawley, P. H. (2011). *The evolution of personality and individual differences*. New York: Oxford University Press.
- Bynum, W. F. (2002). The childless father of eugenics. *Science*, 296, 472.
- Caplan, D. (2009). Experimental design and interpretation of functional neuroimaging studies of cognitive processes. *Human Brain Mapping*. 30 (1), 59-77.
- Cattell, H.E.P. (1996). The original big-five: A historical perspective. *European Review of Psychology*, 46(1), 5-14.
- Cervone, D. (1991). The two disciplines of personality psychology. *Psychological Science*, 6, 371 - 77.

- Cervone, D. (1999). Bottom-up explanation in personality psychology: The case of cross-situational coherence. In D. Cervone & Y. Shoda (Eds.), *The coherence of personality: Social-cognitive bases of personality consistency, variability, and organization* (pp. 303-341). New York: Guilford Press.
- Cervone, D. (2004). The architecture of personality. *Psychological Review*, *111*, 183- 204.
- Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, *56*, 423-452.
- Chamorro-Premuzic, T., & Furnham, A. (2006). Intellectual competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology*, *10*, 251-267.
- Church, A.T., Katigbak, M.S., Del Prado, A.M., Ortiz, F.A., Mastor, K.A., Harumi, Y., Tanaka-Matsumi, J., De Jesus Vargas-Flores, J., Ibanez-reyes, J., White, F.A., Miramontes, L.G., Reyes, J.A.S, then can you get me a bike, Cabrera, H.F. (2006). Implicit theories and self-perceptions of traitedness across cultures: Toward Integration of Cultural and Trait Psychology Perspectives. *Journal of Cross-Cultural Psychology*, *37*(6), 694-716.
- Confer, J. C., Easton, J. A., Fleischman, D. S., Goetz, C. D., Lewis, D. M., Perilloux, C., & Buss, D. M. (2010). Evolutionary Psychology: Controversies, Questions, Prospects, and Limitations. *American Psychologist*, *65*, 110-126.
- Cosmides, L., & Tooby, J., (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (Ed.) *The Latest on the Best: Essays on Evolution and Optimality* (pp. 276-306). Cambridge, MA: MIT Press.

- Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II. Case study: A computational theory of social exchange. *Ethology and Sociobiology*, 10(1-3), 51-97.
- Cosmides, L. and Tooby, J. (1997) The modular nature of human intelligence. In A.B. Scheibel & J.W. Schopf (Eds.) *The Origin and Evolution of Intelligence* (pp. 71-101). Sudbury, MA: Jones and Bartlett.
- Cosmides, L. & Tooby, J. (2000). Evolutionary psychology and the emotions In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions, 2nd Edition*. (pp. 91-115.) NY: Guilford.
- Cosmides, L. & Tooby, J. (2001). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In R. J. Sternberg & J. C. Kaufman (Eds.), *The evolution of intelligence*. (pp. 145-198). Hillsdale, NJ: Erlbaum.
- Cosmides, L. & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Eds.), *The Handbook of Evolutionary Psychology* (pp. 584-627). Hoboken, NJ: Wiley.
- Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, 33, 137-193.
- Crawford, C. (2000). Evolutionary psychology: Counting babies or studying information processing mechanisms. *Annals of the New York Academy of Sciences*, 907, 21-38.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Daly, M. & Wilson, M. (1999). Human evolutionary psychology and animal behaviour. *Animal Behavior*, 57, 509-519.

- Daly, M. & Wilson, M. (2008). Is the "Cinderella effect" controversial?: A case study of evolution-minded research and critiques thereof. In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology*. (pp. 383-400). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.
- Darwin, C. (1859). *The origin of species*. Oxford: Oxford University Press, 1996.
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: Gibson Square Books, 2003.
- Dawkins, R. (2009). *The Greatest Show on Earth: The Evidence for Evolution*. Free Press.
- de Groot, A.D. (1990). Unifying psychology: A European view. *New Ideas in Psychology*, 8(3), 309-320.
- Del Giudice, M., Ellis, B.J., & Shirtcliff, E.A. (2011). The adaptive calibration model of stress responsivity. *Neuroscience and Biobehavioral Reviews*, 35, 1562-1592.
- Del Giudice, M., Hinnant, J.B., Ellis, B.J., & El-Sheikh, M. (2012). Adaptive patterns of stress responsivity: A preliminary investigation. *Developmental Psychology*, 48 (3), 775-790.
- De Los Reyes, A. & Kazdin, A.E. (2008). When the evidence says, "Yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science*. 17 (1), 47-51.
- Dennett, D.C. (1995a). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster, New York.
- Dennett, D.C. (1995b). Cognitive science as reverse engineering several meanings of "Top-down" and "Bottom-up". *Studies in Logic and the Foundations of Mathematics*, 134, 679-689.

- Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin*, 136(5), 822-848.
- Dixon, R.A. (1983). Theoretical proliferation in psychology: A plea for sustained disunity. *The Psychological Record*, 33, 337-340.
- Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem-solution. *Design Studies*, 22, 425-437.
- Edwards, L. & Jaros, G.G. (1995). Psychology, a discipline with a structure-based history and a process-based future. *Journal of Social and Evolutionary Systems*, 18 (1), 67-85.
- Ellis, B.J., Figueredo, A.J., Brumbach, B.H., & Schlomer, G.L. (2009). Fundamental dimensions of environmental risk: the impact of harsh versus unpredictable environments on the evolution and development of life history strategies. *Human Nature*, 20, 204-298.
- Enright, R.D., & Fitzgibbons, R.P. (2000). *Helping clients forgive: An empirical guide for resolving anger and restoring hope*. American Psychological Association, Washington.
- Eysenck, H. J. (1994). The Big Five or giant three: Criteria for a paradigm. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 37-51). Hillsdale, NJ: Erlbaum.
- Figueredo, A.J., & Rushton, J.P. (2009). Evidence for shared genetic dominance between the general factor of personality, mental and physical health, and life history traits. *Twin Research and Human Genetics*, 12, 555-563.

- Figueredo, A.J., Vásquez, G., Brumbach, B.H., Sefcek, J.A., Kirsner, B.R., & Jacobs, W.J. (2005). The K-factor: Individual differences in life history strategy. *Personality and Individual Differences*, 39, 1349–1360.
- Filler, A. G. (2009). The history, development, and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, DTI. *Nature Precedings*. doi: 10.1038/npre.2009.3267.5.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Fitzgerald, C. J., & Whitaker, M. B. (2010). Examining the acceptance of and resistance to evolutionary psychology. *Evolutionary Psychology*, 8(2), 284-296.
- Fodor, J. (1975). *The Language of Thought*, Harvester Press.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Forest, D. (1995). Francis Galton (1822-1911). In R. Fuller (Ed.), *Seven pioneers of psychology: Behavior and mind* (pp.1-19). Routledge: London and New York.
- Fuhrman, O. & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science: A Multidisciplinary Journal*. 34(8), 1430-1451.
- Galton, F. (1889). *Natural Inheritance*. Macmillan: London.
- Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, 109(1), 116-136.
- Garlick, D. (2003). Integrating brain science research with intelligence research. *Current Directions in Psychological Science*, 12(5), 185-189.
- Gaukroger, S. (2006). *The Emergence of a Scientific Culture*. Oxford: Clarendon Press.

- Gintis, H. (2007). A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences*, 30, 1-61.
- Gladin, L.L. (1961). Toward a unified psychology. *Psychological Record*, 11, 405-421.
- Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, 69, 342-353.
- Goertzen, J.R. (2008). On the Possibility of Unification: The Reality and Nature of the Crisis in Psychology. *Theory & Psychology*, 18(6), 829-852.
- Goertzen, J.R. (2010). Dialectical pluralism: a theoretical conceptualization of pluralism in psychology. *New Ideas in Psychology*, 28(2), 201-209.
- Gould, S. J. (1981). *The Mismeasure of Man*. New York: W.W. Norton & Co.
- Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London, B*, 205, 581-598.
- Grucza, R.A., & Goldberg, L.R. (2007). The comparative validity of 11 modern personality inventories: Predictions of behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment*, 89 (2), 167-187.
- Hamilton, W. D. (1954). The genetical evolution of social behaviour. I & II. *Journal of Theoretical Biology*, 7, 1-52.
- Hanson, N.R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge University Press, Cambridge.
- Henriques, G. (2003). The Tree of Knowledge System and the theoretical unification of psychology. *Review of General Psychology*, 7, 150-182.
- Henriques, G.R. (2004). Psychology defined. *Journal of Clinical Psychology*, 60, 1207-1221.

- Henriques, G.R. (2008). The problem of psychology and the integration of human knowledge: Contrasting Wilson's consilience with the Tree of Knowledge System. *Theory & Psychology*, 18, 731-755.
- Henriques, G.R. (2011). *A New Unified Theory of Psychology*. New York: Springer.
- Ho, M.Y., & Fung, H.H. (2011). A Dynamic Process Model of Forgiveness: A Cross-Cultural Perspective. *Review of General Psychology*, 15(1), 77-84.
- Hodgson, G.M. & Knudsen, T. (2008). In search of general evolutionary principles: Why Darwinism is too important to be left to the biologists. *Journal of Bioeconomics*, 10(1), 51-69.
- House, E. R. (1991). Realism in research. *Educational Researcher*, 20, 2-9.
- Hüttemann, A. & Love, A. C. (2011). Aspects of reductive explanation in biological science: Intrinsicity, fundamentality, and temporality. *British Journal for the Philosophy of Science*, 62(3), 519-549.
- Huxley, J. (1942). *Evolution: the Modern Synthesis*. Harper & Brothers.
- Jaszczolt, K. (1996). Relevance and infinity: Implications for discourse interpretation. *Journal of Pragmatics*, 25(5), 703-722.
- Jensen, A. (2002). Galton's legacy to research on intelligence. *Journal of Biosocial Science*, 34, 145 – 172.
- Kanazawa, S. (2006a). Why the Less Intelligent May Enjoy Television More than the More Intelligent. *Journal of Cultural and Evolutionary Psychology*, 4, 27-36
- Kanazawa, S. (2006b). Mind the Gap... in Intelligence: Reexamining the Relationship between Inequality and Health. *British Journal of Health Psychology*, 11, 623-642.
- Kanazawa, S. (2006c). IQ and the Wealth of States. *Intelligence*, 34, 593-600.

- Kanazawa, S. (2007). The Evolutionary Psychological Imagination: Why You Can't Get a Date on a Saturday Night and Why Most Suicide Bombers Are Muslim. *Journal of Social, Evolutionary, and Cultural Psychology, 1*, 7-17.
- Kanazawa, S. (2010a). Why Liberals and Atheists Are More Intelligent. *Social Psychology Quarterly, 73*, 33-57.
- Kanazawa, S. (2010b). Evolutionary Psychology and Intelligence Research. *American Psychologist, 65*, 279-289.
- Kanazawa, S. & Perina, K. (2009). Why Night Owls Are More Intelligent. *Personality and Individual Differences, 47*, 685-690.
- Kanazawa, S., & Reyniers, D.J. (2009). The role of height in the sex difference in intelligence. *American Journal of Psychology, 122*(4), 527-536.
- Kantor, J.R. (1979). Psychology: Science or nonscience? *The Psychological Record, 29*, 155-163.
- Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. Scranton, PA: Chandler Publishing Co.
- Kaplan, D.M., & Craver, C.F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78* (4), 601-627.
- Kaplan, H. S., & Gangestad, S. W. (2005). Life history theory and evolutionary psychology. In D. M. Buss (Ed.), *Handbook of evolutionary psychology*, pp. 68-95. New York: Wiley
- Karasek, R.A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24*, 285-307.

- Kassinove, J.I. (2002). As defined, unification is inevitable. *American Psychologist*, 57, 1127.
- Keller, M. C., & Miller, G. F. (2006). Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behavioral and Brain Sciences*, 29, 385-452.
- Kelly, R.J. (1998). The crisis in psychology: Trouble in the temple. *Journal of Social Distress and the Homeless*, 7, 211-223.
- Kennair, L.E.O. (2011). The problem of defining psychopathology and challenges to evolutionary psychology theory. In D. M. Buss and P. H. Hawley (Eds.), *The evolution of personality and individual differences* (pp. 451-479). New York: Oxford University Press.
- Keri, S. (2003). Genetics, psychology, and determinism. *American Psychologist*, 58(4), 319.
- Keesing, R.G. (1998). The history of Newton's apple tree. *Contemporary Physics*, 39(5), 377-391.
- Ketelaar, T., & Ellis, B. (2000). Are evolutionary explanations unfalsifiable?: Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry*, 11, 1-21.
- Kitcher, P. (1985). Two approaches to explanation. *Journal of Philosophy*, 82, 632-639.
- Lamiell, J. T. (2003). *Beyond individual and group differences: Human individuality, scientific psychology, and William Stern's critical personalism*. Thousand Oaks, CA: Sage Publications.
- Lewens, T. (2002). Adaptationism and engineering. *Biology and Philosophy*, 17(1), 1-31.

- Leybman, M.J., Zuroff, D.C., & Fournier, M.A. (2011). A five-dimensional model of individual differences in social exchange styles. *Personality and Individual Differences, 51* (8), 940-945.
- Leybman, M.J., Zuroff, D.C., Fournier, M.A., Kelly, A.C., & Martin, A. (2011). Social exchange styles: Measurement, validation, and application. *European Journal of Personality, 25* (3), 198-210.
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "sinking shafts at a few critical points". *Annual Review of Psychology, 51*, 405-444.
- Lynn, R. & Kanazawa S. (2008). How to Explain High Jewish Achievement: The Role of Intelligence and Values. *Personality and Individual Differences, 44*, 801-808.
- Maltby, J., Day, L., & Macaskill, A. (2007). *Personality, Individual Differences and Intelligence*. London: Pearson Education.
- Mandler, G. (2002). Origins of the cognitive revolution. *The Journal of the History of the Behavioral Sciences, 38*, 339-353.
- Mandler, G. (2011). Crises and Problems Seen From Experimental Psychology. *Journal of Theoretical and Philosophical Psychology, 31*(4), 240–246.
- Marsh, T. & Boag, S. (2010). Applying Evolutionary Theory to Individual Differences: Insights from Moral Psychology. In R.E. Hicks (Eds.), *Personality and Individual Differences: Current Directions* (pp. 123-134). Bowen Hills: Australian Academic Press.
- Markus, K.A. (1998). Psychological processes and mental stability. *American Psychologist, 53*(9), 1077-1078.

Marks, D. F., Murray, M. P., Evans, B., Willig, C., Sykes, C. M., & Woodall, C. (2005).

Health Psychology: Theory, Research & Practice. London: Sage Publications.

Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101(4), 653-675.

Matarazzo, J.D. (1987). There is only one psychology, no specialties, but many applications. *American Psychologist*, 42, 893-903.

Matarazzo, J.D. (1992). The unity or diversity of psychology: Concluding remarks. *International Journal of Psychology*, 27, 327-330.

Mauer, N. & Borkenau, P. (2007). Temperament and early information processing: Temperament-related attentional bias in emotional Stroop tasks. *Personality and Individual Differences*, 43(5), 1063-1073.

McCrae, R. R., & Costa, P. T. (1994). The stability of personality: Observation and evaluations. *Current Directions in Psychological Science*, 3, 173-175.

McCrae, R., & Costa, P. (1997). Personality trait structures as a human universal. *American Psychologist*, 52, 509-516.

McCrae, R. R. & Costa, P. T. Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 139-153). New York: Guilford Press.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.

- Meehl P.E. (1998). *The Power of Quantitative Thinking*. Washington, DC: American Psychological Society Cattell Award Address.
- Michalski, R. L., & Shackelford, T. K. (2010). Evolutionary personality psychology: Reconciling human nature and individual differences. *Personality and Individual Differences*, 48, 509-516.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.
- Michell, J. (2003a). The quantitative imperative: Positivism, naïve realism and the place of qualitative methods in psychology. *Theory and Psychology*, 13(1), 5-31.
- Michell, J. (2003b). Pragmatism, positivism and the quantitative imperative. *Theory and Psychology*, 13(1), 45-52.
- Michell, J. (2005). The Meaning of the Quantitative Imperative. *Theory and Psychology*, 15 (2), 257-263.
- Miller, G.A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7, 141–144.
- Miller, G.F. (2007). Sexual selection for moral virtues. *Quarterly Review of Biology*, 82(2), 97-125.
- Mischel, W. (1968). *Personality and assessment*. Wiley, New York.
- Mischel, W. (1973a). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80, 252 – 83.
- Muncer, S. J. (2011). The general factor of personality: Evaluating the evidence from meta-analysis, confirmatory factor analysis and evolutionary theory. *Personality and Individual Differences*, 51, 775-778.

- Nairne, J. S. (1997). *Psychology: The adaptive mind*. Pacific Grove, CA: Brooks/Cole.
- Neisser, U. (1967). *Cognitive psychology*. New York, NY: Meredith.
- Neisser, U. (1995). The unity of psychology and of persons. *International Newsletter of Uninomic Psychology*, 15, 6-12.
- Neisser, U., Boodoo, G., Bouchard Jr., T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J., & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*, 51(2), 77-101.
- Nesse, R. M., & Lloyd, A. T. (1992). *The evolution of psychodynamic mechanisms*. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.) *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 601-624). New York: Oxford University Press.
- Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, 61, 622-31.
- Nettle, D. (2011). Evolutionary perspectives on the five-factor model of personality. In: Buss, D.M., Hawley, P.H. (Eds.) *The Evolution of Personality and Individual Differences*. Oxford University Press, New York, pp. 5–28.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge Belknap Press of Harvard University Press, Cambridge.
- Ochsner, K.N., & Gross, J.J. (2008). Cognitive emotion regulation: insights from social cognitive and affective neuroscience. *Current Directions in Psychological Science*, 17, 153–158.
- Olsson, L., Hobfeld, U., & Breidbach, O. (2006). Preface: From evolutionary morphology to the modern synthesis and "evo-devo": Historical and contemporary perspectives. *Theory in Biosciences*, 124(3-4), 259-263.

- Park, H.S., Levine, T.R., Westerman, C.Y.K., Orfgen, T., & Foregger, S. (2007). The Effects of Argument Quality and Involvement Type on Attitude Formation and Attitude Change: A Test of Dual-Process and Social Judgment Predictions. *Human Communication Research*, 33(1), 81-102.
- Penke, L. (2011). Bridging the gap between modern evolutionary psychology and the study of individual differences. In D. M. Buss and P. H. Hawley (Eds) *The Evolution of Personality and Individual Differences* (pp. 243-279). New York: Oxford University Press.
- Penke, L., Borsboom, D., Johnson, W., Kievit, R.A., Ploeger, A., & Wicherts, J.M. (2011). Evolutionary Psychology and Intelligence Research Cannot Be Integrated the Way Kanazawa (2010) Suggested. *American Psychologist*, 66 (9), 916-917.
- Penke, L., Denissen, J.J.A., & Miller, G.F. (2007a). Evolution, genes, and inter-disciplinary personality research. *European Journal of Personality*, 21(5), 639-665.
- Penke, L., Denissen, J.J.A., & Miller, G.F. (2007b). The evolutionary genetics of personality. *European Journal of Personality*, 21(5), 549-587.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Pinker, S., & Bloom, P. (1992). Natural language and natural selection. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.) *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 451-494). New York: Oxford University Press.
- Reeve, C. L., & Charles, J. E. (2008). Survey of opinions on the primacy of g and social consequences of ability testing: A comparison of expert and nonexpert views. *Intelligence*, 36, 681-68.

- Reisner, A. (2005). The common factors, empirically validated treatments, and recovery models of therapeutic change. *The Psychological Record*, 55(3), 377–400.
- Resnik, D. (1996). Adaptationism: Hypothesis or Heuristic? *Biology and Philosophy*, 12 (1), 39-50.
- Richards, M. (1986). Relationships between the Eysenck Personality Questionnaire, Strelau Temperament Inventory and Freiburger Beschwerdenliste Gesamtform. *Personality and Individual Differences*, 7 (4), 587-589.
- Richards, M. (1986). Relationships between the Eysenck Personality Questionnaire, Strelau Temperament Inventory and Freiburger Beschwerdenliste Gesamtform. *Personality and Individual Differences*, 7, 587-589.
- Richardson, R. (2007). *Evolutionary Psychology as Maladapted Psychology*. Cambridge, MA: MIT Press.
- Rodeheffer, C. D., Daugherty, J. R., & Brase, G. L. (2011). Resistance to evolutionary psychology as a continuation of conflicts over scientific integration. *Futures*, 43(8), 777-786.
- Rogers, A.G. (2000). When methods matter: Qualitative research issues in psychology. *Harvard Educational review*, 70(1), 75–85.
- Rose, H., & Rose, S. (2000). *Alas Poor Darwin: Arguments Against Evolutionary Psychology*. New York: Harmony Books.
- Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition*, 108 (3), 771-780.
- Rushton, J.P. (1985). Differential K theory: The sociobiology of individual and group differences. *Personality and Individual Differences*, 6, 441–452.

- Rushton, J.P. (2000). *Race, evolution and behavior: A life history perspective* (3rd ed). Port Huron, MI: Charles Darwin Research Institute.
- Rushton, J.P. (2004). Placing intelligence into an evolutionary framework, or how g fits into the r-K matrix of life history traits including longevity. *Intelligence*, 32, 321–328.
- Salmon, W.C. (1989). Four decades of scientific explanation. In P. Kitcher & W.C. Salmon (Eds.), *Minnesota studies in the philosophy of science: Vol. XIII Scientific explanation*. Minneapolis: University of Minnesota Press.
- Sappington, A.A. (1990). Recent psychological approaches to the free will versus determinism issue. *Psychological Bulletin*, 108(1), 19-29.
- Seligman, D. (2002). Good breeding. *National Review*, 54(1), 53-54.
- Sheldon, K.M. (2011). Integrating Behavioral-Motive and Experiential-Requirement Perspectives on Psychological Needs: A Two Process Model. *Psychological Review*, 118 (4), 552-569.
- Simonton, D. K. (2003). Francis Galton's Hereditary Genius: Its place in the history and psychology of Science. In R. J. Sternberg (Ed.), *The anatomy of impact: What makes the great works of psychology great* (pp. 3-18). American Psychological Association: Washington, D.C.
- Skinner, B. F. (1953). *Science and behavior*. New York: The Free Press.
- Skinner, B. F. (1966). The phylogeny and ontogeny of behavior. *Science*, 153, 1203-1213.
- Skinner, B. F. (1984). The evolution of behaviour. *Journal of the Experimental Analysis of Behavior*, 41(2), 217-221.
- Sober, E. (2000). *Philosophy of biology*. Boulder, CO: Westview Press.

- Staats, A.W. (1983). *Psychology's crisis of disunity: Philosophy and method for a unified science*. New York: Praeger.
- Staats, A. W. (1999). Unifying psychology requires new infrastructure: Theory, method, and a research agenda. *Review of General Psychology*, 3, 3–13.
- Stam, H.J. (2004). Unifying psychology: Epistemological act or disciplinary maneuver? *Journal of Clinical Psychology*, 60, 1259-1262.
- Stern, W. (1911). *Differential psychology in its methodological foundations*, 2nd Edition. Leipzig, Germany: Barth.
- Sternberg, R. J., & Grigorenko, E. L. (2001). Unified psychology. *American Psychologist*, 56, 1069–1079.
- Stevenson, D. & Goldworth, A. (2002). Ethical considerations in neuroimaging and its impact on decision-making for neonates. *Brain and Cognition*, 50(3), 449–454.
- Tashiro, M. (2004). Impacts of Neuroimaging on Psycho-Oncology. *Psycho-Oncology*. 13(7), 486-489.
- Thiruchselvam, R., Blechert, J., Sheppes, G., Rydstrom, A., & Gross, J.J. (2011). The temporal dynamics of emotion regulation: An EEG study of distraction and reappraisal. *Biological Psychology*, 87 (1), 84-92.
- Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, part I. Theoretical considerations. *Ethology and Sociobiology*, 10(1-3), 29-49.
- Tooby, J. & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, 58, 17-67.

- Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.) *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19-136). New York: Oxford University Press.
- Tooby, J. & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 5-67). Hoboken, NJ: Wiley.
- Tooby, J. & Cosmides, L. (2007). Evolutionary psychology, ecological rationality, and the unification of the behavioral sciences. Comment on A framework for the unification of the behavioral sciences, by Gintis. *Behavioral and Brain Sciences* 30(1), 42-43.
- Tooby, J., Cosmides, L., & Barrett, H.C. (2005). Resolving the debate on innate ideas: learnability constraints and the evolved interpenetration of motivational and conceptual functions. In: P. Carruthers, S. Laurence, S. Stich (Eds.), *The Innate Mind: Structure and Contents* (pp. 305-337). Oxford University Press, New York.
- Tucker, J.S., Sinclair, R.R., & Thomas, J.L. (2005). The multilevel effects of occupational stressors on soldiers' well-being: Organizational attachment, and readiness. *Journal of Occupational Health Psychology*, 10, 276–299.
- Tybout, A.M. & Scott, C.A. (1983). Availability of well-defined internal knowledge and the attitude formation process: Information aggregation versus self-perception. *Journal of Personality and Social Psychology*, 44(3), 474-491.
- Tyler, L.E. (1965). *The psychology of human differences*. New York: Appleton Century Crofts.
- Uher, J. (2008). Comparative personality research: Methodological approaches. *European Journal of Personality*, 22, 427-455.

- Vandenbroucke, M.W.G., Scholte, H.S., van Engeland, H., Lamme, V.A.F., & Kemner, C. (2009). A new approach to the study of detail perception in Autism Spectrum Disorder (ASD): Investigating visual feed forward, horizontal and feedback processing. *Vision Research*, 49(9), 1006-1016.
- van der Linden, D., Figueredo, A.J., de Leeuw, R.N.H., Scholte, R.H.J., & Engels, R.C.M.E. (2012). The general factor of personality (GFP) and parental support: testing a prediction from Life History Theory. *Evolution and Human Behavior*, 33, 537-546.
- Vining, D. R. (1986). Social versus reproductive success: The central theoretical problem of human sociobiology. *Behavioral and Brain Sciences*, 9, 167–215.
- Wascher, E. & Beste, C. (2010). Spatial representations as an emergent feature of perceptual processing: Evidence from human electrophysiology. *Journal of Psychophysiology*, 24(3), 161-172.
- Webster, G.D. (2007). Evolutionary Theory's Increasing Role in Personality and Social Psychology. *Evolutionary Psychology*, 5 (1), 84-91.
- Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge*. Alfred A. Knopf, New York.
- Woodley, M.A., Figueredo, A.J., Brown, S.D., & Ross, K.C. (2013). Four successful tests of Cognitive Differentiation-Integration Effort hypothesis. *Intelligence*, <http://dx.doi.org/10.1016/j.intell.2013.02.002>.
- Yanchar, S.C., & Slife, B.D. (1997). Pursuing unity in a fragmented psychology: Problems and prospects. *Review of General Psychology*, 1, 235-255.
- Young, A. (2006). Remembering the evolutionary Freud. *Science in Context*, 19(1), 175-189.

Zuroff, D.C., Fournier, M.A., Patall, E.A., & Leybman, M.J. (2010). Steps toward an evolutionary personality psychology: Individual differences in the social rank domain. *Canadian Psychology*, 51, 58-66.

Discussion for Chapter 5

As the historical context explored in this article suggests, the study of evolved physical and behavioural characteristics demands an appreciation of both how phenotypic variations arise, and how selective pressures act upon complex phenotypes to adjust population gene-frequencies over generations. In acknowledgement of the related considerations of variation and selection, this article sought to challenge the once popular conception that the methods of differential and evolutionary psychology are fundamentally incompatible, and encourage collaborative efforts from researchers in both areas to build upon the breakthroughs that have recently emerged in the evolutionary literature. This journal article also builds on the appreciation for the separate scientific tasks of description and explanation introduced in Chapters 3 and 4, and applies this understanding to account for the atheoretical traditions that have come to dominate much of differential psychology, and to explain why the explanatory approaches adopted by these traditions struggle with issues of reification. The explanatory language introduced in this chapter (specifically the top-down versus bottom-up framing), adapted primarily from the work of Daniel Cervone (1999; 2004), is not only useful in describing differences of scientific practice between fields, but allows for an instructive degree of nuance in interpreting empirical results, particularly those relying upon theories that employ largely untested pragmatic assumptions. This utility is employed directly in service of the goals of this thesis, when addressing the more ‘risky’ possible interpretations of the findings discussed in Chapters 7 and 8.

Furthermore, beyond providing a more detailed account of the conceptual and theoretical strengths of well-conducted evolutionary syntheses in psychology, this journal article catalogued the range of Darwinian selective phenomena that can be brought to bear on a sophisticated adaptationist account of psychological individual differences. In addition to the simple adaptive trade-offs and selective neutralities discussed in Chapter 2, *frequency-*

dependent selection can account for otherwise unpredictable patterns of variance in socially influential aspects of human personality. The generational process of *mutation-selection balance* allows for an evolutionarily nuanced account of individual differences in fitness-relevant cognitive abilities, and their volatile connection to general biological signals of developmental resources and *mutation-load*. Perhaps most intriguingly (particularly with regards to individual differences in moral characteristics), the burgeoning new understandings concerning possible *niche-selecting mechanisms* in the neuronal, developmental, and epigenetic systems of animals, afford evolutionary theorists a novel means of conceptualising the complex strategic calibrations of an organism (or lineage of organisms) to the demands of its environment. These evolutionary insights into the possible ultimate and proximal causes of systematic human variation inform the much of the specific theory-building discussed in Chapters 7 and 8, particularly with regard to the strategic role empathy modulation is likely to play in the negotiation of fluid, but highly consequential, coalitional groups and alliances.

CHAPTER 6

The Implicit Measurement of Racial Prejudice

This chapter marks the conclusion of the theoretical ‘half’ of this thesis, and the beginning of the empirical ‘half’, which details the 4 major research studies undertaken during this thesis. The study described in the journal article featured in this chapter represents the chronologically earliest research arc in this thesis, excluding the pilot studies and program bug-testing phases mentioned briefly in the introduction to Chapter 7. Unlike the three subsequent studies focused on in Chapter 7, which collectively represent the development, refining, and preliminary testing of the new SATEST measurement tool, this study fits more comfortably in the standard tradition of social cognitive psychology research into explicit and implicit attitudes.

Although the psychological mechanisms of interest in this thesis (those concerning the modulation of social empathy and subsequent justifications and attributions) are conceived as underpinning the general belief and behaviour patterns common to all typical expressions of prejudice, the measurement of actual prejudicial responses (as explained in Chapters 1 and 2) demands that a particular form of prejudice be selected for study. Since the issue of racial prejudice has received by far the most extensive and diverse treatment in the social psychology intergroup prejudice literature, and because it is a form of prejudice that is expected to show at least moderate variance in most undergraduate samples (Greenwald et al., 2009), racial prejudice was selected as the focus of this study, and subsequently the third phase of investigation described in Chapter 7. Given that the sample used in this study consisted of university undergraduate students in a research participation context, expressions of aggressive or overt racially-motivated animosity were regarded as unlikely, and as such racism was operationalised in only two forms in this study: as implicit attitude preferences between contrasted race categories, as measured by response latencies on an Implicit

Association Test (IAT; Greenwald, Nosek & Banaji, 2003), and as explicit survey responses to variations of the *Modern Racism Scale* (MRS; McConahay, 1986).

The primary goal of the present study, with regard to the overarching goals of the thesis, was to test the psychometric veracity of a manipulation of the standard (face-image stimuli) racial attitudes Implicit Association Test (RA-IAT), that grouped together multiple racial/ethnic categories together into contrast groups of ‘light’ and ‘dark’ skinned individuals, as opposed to the typical established approach of contrasting precisely two racial groups as a dichotomy. Beyond the possible independent scientific merit of this investigation (specifically, with regards to parochial ‘light vs. dark’ Eurocentric evaluations in many parts of the world, see Blair, Judd, Sadler, & Jenkins, 2002, and Maddox, 2004 for reviews), the racially ambiguous target stimuli employed in the most recent study of the SATEST measurement tool could only be meaningfully compared to a measure of implicit racial attitudes that utilised comparably general race evaluations.

The secondary research goal of this study was to explore the psychometric and predictive properties of a new IAT methodology, designed to measure implicit identification with light- and dark-skinned racial groups, which utilised carefully selected celebrity ‘others’ to produce more extensive and multifaceted ‘self and other’ response stimuli than previous identity-focused implicit measures (see Knowles & Peng, 2005 for an example). As the article describes, this methodology appears successful in its designed purpose, but its demonstrated relationships to the other variables studied suggested it would not prove of sufficient additional value to the study of racial prejudice to warrant its inclusion in subsequent studies in this thesis. As is discussed below, it is likely that the ‘celebrity other’ design of the racial identity Implicit Association Test (RI-IAT) may demonstrate greater predictive power if specified to individual racial ingroup-outgroup comparisons, rather than the general light- and dark-skinned categories explored in this study.

The following article has been submitted for publication to the journal *Personality and Social Psychology Bulletin*, under the title ‘Evaluative Attitudes and Identification with Light- and Dark-Skinned Racial Groups’. The presentation of the article in this thesis varies slightly from the version of the manuscript submitted for publication, as several sections were added to the paper’s introduction (totalling approximately 2000 words), offering detailed background concerning implicit measurement in social psychology research. These sections are included here to provide better context for readers of this thesis, but were removed from the publication manuscript both due to concerns for the overall word-length, and due to the presumed familiarity of the journal’s typical readership with the social cognitive literature in general.

Declaration for Thesis Chapter 6

In the case of journal article featured in Chapter 6, the nature and extent of my contribution to the work, and the contributions of the other listed co-authors is as follows:

<i>Name</i>	<i>Nature of Contribution</i>	<i>Contribution</i>
Tim Marsh	Decision concerning the topic of the paper	90%
	Search and review of the literature	
	Design and programming of measurement tools	
	Administration of study and data collection	
	Analysis and interpretation of data	
	Principle writing and editing of the manuscript	
Simon Boag	Advice on topic and approach	10%
	Assistance with editing and cutting	
	Suggestions for the refinement of the manuscript	

Evaluative Attitudes and Identification with Light- and Dark-Skinned Racial Groups

Running Title: *Racial Attitudes and Identification*

Tim Marsh
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: timothy.marsh@mq.edu.au

Simon Boag
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: simon.boag@mq.edu.au

ABSTRACT

The study of racial prejudice and discrimination has focussed heavily on the evaluative attitudes, beliefs and stereotypes directed toward members of other racial groups, particularly minorities. Methodological advances in social cognition have led researchers to draw clear divisions between consciously deliberated explicit attitudes, and implicit attitudes that can be detected through indirect cognitive associations. Similar explicit-implicit divisions have recently been established in identification with racial groups, but little is known about how implicit racial identity may interact with racial attitudes. The present study explores the relationships between explicit racial prejudice, and implicit racial attitudes and identity, operationalised as preferences between light- and dark-skinned racial groups. The sample consisted of 261 university undergraduate students, participating through an online interface. Results indicate that implicit racial identity serves as a statistically significant predictor of implicit, but not explicit, racial attitudes. Conceptual and methodological issues with light versus dark skin tone comparisons are discussed.

Keywords: racial attitudes, racial identity, modern racism, implicit association test, prejudice

Evaluative Attitudes and Identification with

Light- and Dark-Skinned Racial Groups

Though many forms of intergroup conflict are endemic to the lives of social animals (Maynard-Smith & Parker, 1976), the perceived boundaries of ‘kin’ and ‘kind’ typically support the most consequential of distinctions between ingroup and outgroup (see Markham, Alberts & Altmann, 2012 for examples). It is of little surprise, then, that beyond the immediate concerns of breeding partners (Miller, 2000) and familial relations (Dixson, 1998), much of the enduring social conflict throughout recorded human history has revolved around group distinctions of ethnicity and race (Hewstone, Rubin & Willis, 2002). While the scientific merit of such distinctions has been called into question (e.g., Lieberman et al., 2004), empirical evidence attests to the distinction’s relative ubiquity (Neuberg, 1989) and the apparent perceptual primacy (Messick & Mackie, 1989) of the categorisation of others along prototypical racial distinctions, which are in turn often presumed to signal particular ethnic backgrounds (Maddox, 2004) and stereotypical dispositions (Shapiro & Neuberg, 2007).

Decades of research in psychology and the social sciences have further explored the contemporary consequences of racially-delineated thought, with regards to both interpersonal and institutional discrimination and prejudice (see Crandall & Schaller, 2005, for a historical review). While studies of discrimination have assumed a wide range of social perspectives and theoretical stances (see Bobo & Fox, 2003 for a brief overview), it was only towards the end of the 20th century that a long-neglected element of racial conflict began to integrate into prejudice literature: the psychology of social group membership and its implications for racial identity (Sellers et al., 1998). The present study was undertaken to explore the racial attitudes (both explicit and implicit) of participants with regards to light or dark racial skin tones,

while accounting for both implicit and explicit expressions of self-identification with either group.

Racial Attitudes and Prejudice

The study of racial prejudice in western psychology initially emerged as a facet of authoritarianism (Allport, 1954; Altemeyer, 1981) and the socialisation of conservative values (Carlson & Iovini, 1985; see Katz, 2003 and Duriez & Soenens, 2009, for recent overviews focusing on race). Although problematically embedded in the cultural norms of the time and place of its inception (Rubinstein, 1996), the theoretical legacy of authoritarianism provided the conceptual bases from which the two most successful traditions of prejudice research have arisen: the study of evaluative attitudes towards groups (Dovidio et al, 2002), and the study of group stereotyping (see Duckitt, 1992; Judd & Park, 1993; Operario & Fiske, 2001 for overviews). Concerning these approaches, while the current study has focused primarily on attitudes and the relative evaluation of racial differences, the relations of mutual influence between racial stereotypes and racial attitudes are well-established (see Cox et al., 2012, for a review), and will be afforded some consideration in interpreting the findings.

With some exceptions concerning the topic of identity, which shall be addressed below, contemporary social psychologists conceptualise racial attitudes as the affect- and belief-driven evaluations of racially-defined outgroups, relative to at least one racially-defined ingroup (Stanley, Phelps & Banaji, 2008). As an example of typical ingroup-outgroup distinctions, individuals are expected to demonstrate a range of biases favouring their own race, within the wider context of the stereotyped expectations associated with each group (Operario & Fiske, 2001; see Bobo, 2001, for a detailed review centred on the USA). Much racial prejudice and discrimination is understood to emerge from negative attitudes towards

other races (Crandall & Schaller, 2005; Duriez & Soenens, 2009), a relation that is enhanced by widely divergent relative evaluations between racial ingroups and outgroups (Williams & Eberhardt, 2008), and greater acceptance of racial stereotypes (Cox et al., 2012) and outgroup homogeneity (Yzerbyt, Judd & Corneillo, 2003). That said, research findings concerning both the development (Nosek, Greenwald & Banaji, 2007), variability (Nosek, Greenwald & Banaji, 2005) and expression (McConnell & Leibold, 2001) of racial attitudes have all but eliminated the possibility that individuals possess singular, conscious evaluations of each racial group that they are exposed to (see Greenwald et al, 2009, for an overview). Rather, the current literature, and the present study, conceptualises racial biases as occurring in two distinct (but interrelated) ways: as explicit attitudes, and as implicit attitudes.

Measuring Racial Bias

While ‘explicit’ and ‘implicit’ have become the preferred terminology employed in the social psychology literature to refer to the two modes of measurement discussed below, it is instructive to first clarify the conceptual distinction being made. The opposing terms explicit and implicit were primarily popularised by the study of memory effects by cognitive psychologists (see Schacter, 1987). In this context, the tested effects of a memory condition were established by the experimenters via detectable changes in performance on subsequent tasks, which corresponded to details in some earlier stimulus. For example, participants who demonstrated improved performance in a word-recognition task following some prior exposure to these words (e.g., Richardson-Klavehn & Bjork, 1988; Roediger, 1990), were designated by the researchers as displaying memory of the prior exposure. The term ‘explicit’ memory was subsequently applied to instances where the participant would report conscious recollection of their prior exposure to the information, whereas the term ‘implicit’ memory

was used to describe those instances where the participant performed as if they recalled the information in question, but reported no corresponding conscious memory. As such, the words explicit and implicit carry for many a connotation that they distinguish between phenomena that are conscious and non-conscious respectively (within this literature, at least). This is not, though, the distinction denoted by the use of ‘explicit’ and ‘implicit’ in the attitudes literature, although it is likely that many participants happen to be unaware of their implicit attitudes at the time of testing. In the discussion below, as in the social psychology literature in general, explicit and implicit are terms intended to denote a style of measurement that in the former utilises deliberate reflection and linguistic expression, and in the latter specifically avoids this (Fazio & Olson, 2003). A perhaps more appropriate verbal distinction would be between ‘direct’ and ‘indirect’ methods of measurement, terms which, when employed in this article, are to be taken as synonyms of the widely accepted ‘explicit’ and ‘implicit’, respectively.

Explicit Measurement

As is common throughout the history of social psychology, the measurement of racial attitudes began with the administration of self-report questionnaires and exploratory interviews (Carlson & Iovini, 1985; Bobo, 2001). In the decades closest to the authoritarian roots of the topic (Allport, 1954; see also Rubinstein, 1996, for a retrospective account), overt and ideologically motivated racial discriminations were confessed readily by participants, and were subsequently treated as presumably truthful admissions of consciously held biases. However, widespread shifts in the public egalitarian values of western nations (perhaps most notably in the United States Civil Rights movements of the 1960s) gradually shifted normative social climates so as to discourage racial prejudice and discrimination (Bobo,

2001; Crandall, Eshleman, & O'Brien, 2002). Although evidence suggests that these and related social changes truly reduced mean levels of negative attitudes towards racial minorities by some degree (Saucier, Miller & Doucet, 2005), these revised social norms provided a greater incentive for individuals to misrepresent their subjectively held attitudes and evaluations, so as to avoid the potential stigma such sentiments may attract (Crocker & Major, 1989).

Subsequently, the concerns of impression management and social desirability that present difficulties for all self-report methodologies have become a defining challenge for the study of racial attitudes (Fazio et al., 1995). In most modern populations, particularly in the western world, researchers are unlikely to obtain a reliable representation of a participant's negative racial attitudes by prompting their agreement with overtly racist statements. Beyond the strong influences of social desirability, some evidence (such as Bobocel et al., 1998) suggests that even people with distinctly negative racial attitudes and stereotypical beliefs have internalised a rejection of classical racism due to social stigma, resulting in far more subtle and indirect expressions of negativity (see also, McConahay, 1986).

Examining the racial attitudes that participants are willing to consciously and deliberately express retains great conceptual value, despite the omnipresent desire to over-represent one's egalitarianism, as such sentiments can ideally reflect an individual's rational engagement with wider social biases and the content of stereotypes (Cox et al., 2012). To this end, the contemporary study of explicit racial attitudes has been conceptually refined, so as to specify approximately four ways in which negative racial attitudes and beliefs are expressed (see Pearson, Dovidio & Gaertner, 2009 for an overview). *Dominative racism*, first distinguished by Kovel (1970), describes the kind of overt racial bias that more recent social pressures target, wherein one directly acts upon stereotyped and bigoted beliefs. *Symbolic racism* (Sears, 1988; Sears, Henry & Kosterman, 2000), refers to a more subtle consequence of

pervasive stereotyping, wherein specific individuals of a particular racial group are not targeted as examples, but the group itself is none-the-less taken to symbolise various negative attributes as a whole. *Modern racism*, outlined by McConahay (1986) is seemingly the most commonly occurring and widely studied contemporary expression, acknowledging the past oppression of racial minorities, but nevertheless suggesting that minorities are either too aggressive in their present attempts to obtain societal resources and opportunities, or are otherwise undeserving of those they currently possess. The final expression of negative racial attitudes, *aversive racism*, was originally conceived by Gaertner and Dovidio (1986) as a behavioural aversion to members of a disfavoured race, despite the conscious endorsement, and perhaps truly held belief, of egalitarian principles. Unlike the aforementioned three, the concept of aversive racial bias has proven a poor candidate for explicit measurement, but now stands as the dominant perspective in the newer domain of implicit measurement of racial attitudes.

Implicit Measurement

As was outlined above, implicit measurement in the context social psychology refers to methods of measurement which avoid directly soliciting deliberation and explicit expression from the participant. Research employing this brand of implicit testing has its roots in the exploration of social priming effects (e.g., Gaertner & McLaughlin, 1983; Fazio et al., 1986; Greenwald et al., 1989; Perdue et al., 1990), the principles of which have come to inform the great expansion in indirect testing that has emerged over the past 20 years (see Fazio et al., 1995, for an early example specifically targeting racial attitudes). Many implicit methodologies are based upon reliable response tendencies that participants have little insight into, such as the linguistic tendencies to describe events consistent with their expectations in

a manner more abstract (e.g., von Hippel et al., 1995) and less detailed (von Hippel et al., 1997; Sekaquaptewa et al., 2003) than those that defy these expectations, or a person's subtle preference for letters that appear in their own names (Nuttin, 1985; Koole et al., 2001; Jones et al., 2002; Pelham et al., 2002). Some others grant participants insight into what is being studied, but measure physiological responses such as event-related brain potentials (Cacioppo et al., 1993; Crites et al., 1995; Ito & Cacioppo, 2000), cardiovascular excitation (Blascovich et al., 2001), or eye-blink startle response (Phelps et al., 2000; Amodio et al., 2003) to avoid reliance on explicit answers. By far, however, the most popular and robustly established implicit measures are those that rely upon time-pressured decision tasks to measure the strength of cognitive associations, such as cross-coding measures (De Houwer & Eelen, 1998; De Houwer et al., 2001), the Go/No-go Association Test (GNAT; Nosek & Banaji, 2001), and most influential of all, the Implicit Association Test (IAT; Greenwald, McGhee & Schwartz, 1998).

Although it has since been applied to a wide range of topics and stimuli, from its earliest inception the IAT methodology has been employed to investigate relative evaluative proportions of racial attitudes (Greenwald, McGhee & Schwartz, 1998; Greenwald et al., 2003; 2009). In accordance with the aversive racism approach mentioned above, the form of implicit racial attitudes targeted by associative measures like the IAT represent biases in evaluative regard (affective warmth, positive familiarity, etc.) when comparing one racial group to another. Typically, a racial attitudes or racial prejudice IAT will employ four classes of stimuli, representing opposing ends of two comparative dimensions. The first dimension consists of both valences of a vague judgment (positive vs. negative words, violent vs. calm images, etc.), upon which the evaluation of the second dimension is based. The second dimension consists of two conceptually opposed or contrasted stimuli that are to be evaluated. In the test phases, the respective components of the judgment dimension and the dimension to

be evaluated are paired, in both possible configurations over separate testing blocks, so as to contrast the participant's pairing accuracy and speed between the two conditions. Stimuli and evaluations that are congruent pairings in the general experience and cognition of the participant have been demonstrated to result in faster reaction times on average than relatively incongruous pairings, providing evidence for the pre-existing association of particular stimuli with particular valenced judgments (Greenwald et al., 2003). As a consequence of this approach, however, it must be noted that IATs strictly only produce a difference score of relative evaluative preference between the paired sets of options tested. As such, the interpretation of IAT results hinges on the conceptual justification of contrasting any two tested categories, making the methodology vulnerable to the testing of false dichotomies (Blanton, Jaccard, Gonzales, & Christie, 2006).

In contrast with the research traditions surrounding explicit expressions of racial attitudes, the study of implicit racial attitudes has advanced fairly atheoretically (Greenwald et al., 2002; Blanton et al., 2006), relying primarily on the ability of methodologies such as the IAT to circumvent the social desirability limitations that plague explicit measures (Nosek, Greenwald & Banaji, 2007; Greenwald et al., 2009). When grounding justifications are offered for designing implicit measures around the unawareness of the participant or the time-pressure of their responses, many researchers cite theories such as Fazio's (1990) MODE model, which predicts that deliberative and effortful tasks such as masking one's true attitudes require both sufficient time and a suitable motivation. Also, the form of biases demonstrated in tests of implicit attitudes closely fit the defining characteristics of aversive racism (Pearson, Dovidio & Gaertner, 2009), but the reliance of associative methodologies on contrasting two categories, rather than measuring the attitudes towards one racial group in isolation, limit meaningful comparisons to those examples where antipathy towards one racial

group can best be understood as oppositional to another group (as in comparisons between White- and African-Americans; see Nosek, Greenwald & Banaji, 2007, for a review).

The 'Real' Bias?

During the early development of implicit measures of racial attitudes, methodologies such as the IAT were entertained as likely replacements for explicit measures of racial bias, capable of solving the 'problems' of impression management and social desirability (Greenwald, McGhee & Schwartz, 1998). However, the subsequent 15 years have demonstrated that the simple replacement of explicit measures would be a poor idea. The aversive racial attitude patterns detected by implicit measures have not been shown to simply exceed those of explicit alternatives, but rather, appear to constitute a largely non-overlapping realm of predictive value (see Greenwald et al., 2009, for a meta-analysis of predictive findings). Though it was initially expected, given concerns over participant honesty, that explicit and implicit measures of racial attitudes would correlate poorly, growing evidence now suggests that implicit and explicit measures of racial bias both possess good predictive validity of distinct sets of valuable outcomes, rather than implicit measures serving as total replacements for their explicit counterparts. Typically, implicit measures are found to be superior predictors of unreflective behavioural outcomes, whereas explicit measures serve as stronger predictors of behaviours that one may deliberately shape to cultivate an impression. For example, Dovidio et al. (2002) found that while the implicit racial attitudes of white participants was a strong predictor of their quality of non-verbal interaction with black participants, the most effective predictor of verbal behaviour trends between the two races were their explicit responses. Similar patterns have been observed in studies of room-sharing behaviours (Towles-Schwen & Fazio, 2006), the effect of anti-Muslim stereotyping on hiring

patterns (Rooth, 2010), and many others (Dovidio et al., 1997; McConnell & Leibold, 2001; Dovidio, Kawakami, & Gaertner, 2002; Greenwald & Krieger, 2006; Nosek, Greenwald & Banaji, 2007; Greenwald et al., 2009). As such, while it is almost certainly the case that implicit and explicit racial attitudes exert mutual influences, particularly since affective biases in implicit attitudes are likely to inform the deliberative acts that give rise to explicit declarations, neither can be claimed to be the ‘true’ attitude that researchers may wish to exclusively target.

Racial Identity

Despite the extensive research interest that racial prejudice and discrimination has received in psychology and the social sciences, much of this research has habitually taken racial groups, and participants’ membership within them, as a given (Maddox, 2004). Other research traditions, more concerned with the topic of social identity in general, have long given credence to the role of race and ethnicity in defining the self (see Cross, 1991, for an African-American example), but it was not until the ranks of western academics began to accommodate greater proportions of researchers from minority racial backgrounds that the specific issue of racial identity began to integrate into the study of both the occurrence and experience of racial discrimination (Sellers et al., 1998).

Multiple Dimensions of Racial Identity

The late-inclusion of racial identity into the wider prejudice literature gave rise to two fairly distinct research conventions: those approaches which continued to focus on social identity in general, drawing upon racial identity as a typical example (Phinney, 1990), and those approaches focusing on the unique cultural and experiential aspects of being a member of a

racial minority group (e.g., Cross, 1991). Although there are researchers who favour distinctly the former or the latter to this day, the majority (Ashmore, Deaux & Mclaughlin-Volpe, 2004) of current racial identity research adopts a multidimensional theoretical approach that represents a conceptual fusion of these traditions. While recently, exhaustive general models of collective identity such as Ashmore, Deaux and Mclaughlin-Volpe's (2004) multidimensional theory have grown in popularity, a great deal of racial identity research still builds upon the foundations laid by an earlier approach, the Sellers et al. (1998) Multidimensional Model of Racial Identification (MMRI).

The MMRI consists of four broad dimensions (each of which is paralleled and expanded upon in the Ashmore, et al., 2004, approach), some concerning immediate circumstances and others concerning trait-like dispositions, which collectively encompass the degree to which an individual feels that their racial group membership shapes their sense of self in any given moment (Sellers et al., 1998). The dimension of *Salience* simply concerns how obvious the racially-relevant elements of a given situation are to an individual in that situation.

Conversely, the dimension of *Centrality* describes an individual's enduring, trait-like disposition to consider their racial group membership an instructive element of who they are.

The dimension of *Regard* is divided into one's conception of the general attitudes that the public hold towards one's racial group, as well as one's private sense of how relatively valuable said group is. Lastly, the *Ideology* dimension describes an individual's beliefs, opinions and attitudes concerning how a prototypical member of one's racial group should

behave. While the MMRI itself, and the original measure based on its distinctions (the Multidimensional Inventory of Black Identity; MIBI), does not highlight any specific dimensions as being of greater empirical relevance than the others, the dimensions of

Centrality and Regard have been paid extensive attention in subsequent research (see

Knowles & Peng, 2005, for a discussion). Centrality has been targeted for its potential role as

an enduring individual difference between individuals disposed to greater or lesser degrees of racial self-identification, whereas Regard has been explored as a viable (though likely not conceptually distinct) connection to the wider literature concerning racial attitudes.

Furthermore, the conceptual similarities between the Ideology dimension and the literature addressing racial stereotyping leaves the MMRI dimension of Centrality as the prime candidate for an enduring element of racial identity that can contribute unique explanatory power to a study of racial biases born of disparate attitudes and stereotypes (Ashmore, Deaux & McLaughlin-Volpe, 2004).

Implicit Racial Identity

As with the case of attitudes, explicit expressions of identification with social groups are likely to reflect a degree of deliberation and impression management, and thus are likely to diverge in predictive utility from identification that is measured implicitly (see Devos & Banaji, 2003, for a comparative analysis). Relatively little empirical research has been undertaken to explore racial identity via implicit measurement (Aron et al., 1991; Smith & Henry, 1996; Coats et al., 2000; Devos & Banaji, 2003; Knowles & Peng, 2005; Craemer, 2008, 2010; Craemer et al., 2013). Of those few studies, many have employed a reaction-timed task developed by Aron et al. (1991), whose measurements are based upon the predictable self-other confusions known to occur in snap decisions when these categories are strongly associated. Others, most notably Knowles and Peng (2005), rely upon a modified IAT methodology, where the word categories of the evaluative dimension are approximate synonyms of ‘self’ and ‘other’. The advantage of the Knowles and Peng (2005) approach employs standard racial category stimuli used in IATs designed to measure racial attitudes (in their initial studies, names commonly associated with Caucasian- versus African-Americans),

and is conceptually oriented towards measuring the Centrality dimension of the MMRI, specifically. Their preliminary findings suggest that such an approach is comparably reliable to attitude-based IAT measures, making Centrality-focused racial identity IATs an ideal (though underused) tool for exploring the interactions between implicit racial identification and both implicit and explicit racial attitudes.

The Present Study

In the interest of exploring the predictive relations between racial attitudes and racial identity, the current study employed the IAT methodology (relying on the scoring calculations outlined in Greenwald et al., 2003) to produce implicit measures of both attitudes towards, and identification with, targets with relatively light or relatively dark skin tones. In contrast with the majority of existing IATs intended to measure racial biases (Nosek, Greenwald & Banaji, 2007; Greenwald et al., 2009), the IATs in the present study did not evoke an evaluate contrast between two distinct racial groups (such as those of European versus African descent). Rather, the two compared categories each included multiple races, whose prototypical skin tones separated along a more general division between lighter and darker colours (for example, Caucasian and East Asians possess relatively lighter skin tones than Africans and South Asians). This alternate approach was designed to target the broad preferences for light skin tones observed within several racial groups (Blair, Judd, Sadler, & Jenkins, 2002), a pattern that appears to overlap with the well-established Eurocentric intergroup preferences typically studied in the racial bias literature (see Maddox, 2004, for a review). The racial identity IAT also utilised a more diverse set of self and other evaluative stimuli than those used in the original Knowles and Peng (2005) study, the details of which are outlined below. Lastly, while the implicit measures were designed to avoid overt reliance

on comparing discrete pairs of racial groups, a survey scale of modern racism (McConahay, 1986) was used to measure explicit racial attitudes towards the three non-Caucasian minorities presented in the IATs.

Based on the established trends in the racial attitudes literature (Nosek, Greenwald & Banaji, 2007; Son Hing et al., 2008; Greenwald et al., 2009) and the predictive properties of prior implicit racial identity measures (Knowles & Peng, 2005; Craemer, 2008, 2013), it was hypothesised that a weak but significant positive correlation would be found between implicit attitude preferences for light-skinned stimuli and explicit racial prejudice towards groups with darker skin tones (specifically, people of African and South Asian descent). It was also hypothesised that participants' self-described categorisation by racial background would positively predict (for light-skinned participants) explicit negative racial attitudes towards dark-skinned groups, while showing no significant predictive relationship with implicit attitude preferences for either light- or dark-skinned racial stimuli. Finally, implicit racial identification with light-skinned groups was hypothesised to significantly positively predict implicit attitude preferences for light-skinned racial groups, but not explicit racial prejudice towards either light- or dark-skinned races.

Method

Participants

An initial 261 Australian psychology undergraduates were recruited via research participation pool in exchange for course credit. Of this initial number, as per the scoring criteria outline by Greenwald et al. (2003), four participants were excluded from the analysis for unacceptably high error and latency rates in the racial identity IAT, with one additional

participant excluded for unacceptable error rates in the racial attitudes IAT, leaving a total n of 256.

Ages within the sample ranged from 17 to 49 years, with a mean age of 21.05 ($SD = 5.809$), and presented a skewed gender ratio typical of psychology undergraduates with 193 (75%) females and 63 (25%) males. Participants indicated self-described racial background, with 144 (56%) identifying as Caucasian of European descent, 31 (12%) participants of Mediterranean descent, 26 (10%) of Middle-Eastern descent, 34 (13%) of East Asian descent, and 21 (8%) of South Asian descent. Participants also indicated whether they considered themselves members of a racial group with characteristically light (79%) or dark (21%) skin tones relative to other races. All participants identified English as their first language, and had lived in Australia for at least 5 years.

Measures and Procedure

Participants completed the study measure via a secure, online browser-imbedded computer interface constructed in *Adobe Flash*, on a PC of their choosing. Participants were first presented with a range of demographic questions (including those later employed in the racial identity IAT) followed by the three measures of interest in counterbalanced order (order of presentation showed no significant effects on the three measures, and as such has not been included in the subsequent analysis).

Explicit Racial Attitudes

Participants' general explicit evaluative attitudes towards various racial minorities were measured using three variants of the Modern Racism Scale (MRS; McConahay, 1986). To correspond with the three non-Caucasian racial groups represented in the image stimuli of the IAT measures, the MRS items were modified to

address attitudes towards people of East Asian, South Asian, and African descent, framed in an Australian national context. Since the construct of modern racism is grounded in the presumed conflating of a nation's dominant cultural perspective with its majority racial group (Caucasians of European descent, in Australia), negative explicit attitudes towards white Australians, to whatever degree they may exist, could not be measured in this way. To ensure clarity from the perspective of participants, racial groups were described with added reference to the world nations typically associated with the group in question (for example, that South Asian individuals generally have ancestry in Sri Lanka, India and/or Pakistan). Each of the three MRS subsections consisted of 10 statements (30 in total), such as "There are too many foreign students of East Asian descent being allowed to attend university in Australia" and "Discrimination against Africans is no longer a problem in Australia", which participants responded to with a 7-point Likert-scale ranging from "Strongly Disagree" to "Strongly Agree". The items were presented together, in random order, with responses to each subscale proving highly internally consistent (each Cronbach's $\alpha > .95$), and good average consistency between the subscales (with a collective Cronbach's α of .79). The distributions of the MRS subscales were also very similar, with all means between 30 and 34, and all standard deviations between 7.8 and 9.

Implicit Racial Attitudes Using 'positive' and 'negative' word lists as the evaluative categories (the standard in attitude IATs, provided by Greenwald, McGhee & Schwartz, 1998), and specially selected and cropped images of racially diverse adult human faces as the race-category stimuli (as in Nosek, Greenwald & Banaji, 2005), a *D*-scored IAT was designed to measure implicit racial attitudes along the broad dimension of light versus dark skin colour (hereafter referred to as the RA-IAT). The default congruent condition was designated as associating positive attitudes with light-skinned stimuli and negative attitudes with dark-skinned stimuli. As such, with scores ranging from -2 to 2, positive *D*-scores

reflect a relative evaluative preference for individuals with light skin tones, negative D-scores reflect a preference for individuals with dark skin tones, and D-scores close to zero indicate no significant preference towards either (Greenwald et al., 2003). The order of congruent and incongruent blocks was counterbalanced, and showed no significant influence on mean RA-IAT D-scores.

In order to reflect broader preferences between light and dark skin tones, as opposed to contrasting only two distinct racial categories as in most racial attitude IATs, the present RA-IAT grouped two races into each category based on the relative skin colour of prototypical members of each group (see Blair, Judd, Sadler, & Jenkins, 2002, for details on perceived racial prototypicality). Thus, the light-skinned category consisted of images of Caucasians of European descent and individuals of East Asian descent, while the dark-skinned images were comprised of individuals of African and South Asian descent. The lightness or darkness of skin tone in the selected images varied only mildly within each category for the sake of clarity, despite the relatively broad range of skin colour variation that truly occurs within each racial group (Maddox, 2004). Furthermore, all images used contained only minor variations in age and weight, were counterbalanced to present even gender proportions, and ensured equal numbers of each constituent race in each category. 20 distinct images (10 male, 10 female) were used to represent each of the four races, so as to minimise repetition and redundancy over the various trials.

Implicit Racial Identity

A second IAT measure (hereafter referred to as the RI-IAT) was employed to measure participants' implicit self-association with light or dark racial skin tones. Building on the initial work of Knowles and Peng (2005), this self-association was operationalised as an implicit element of the Centrality dimension of racial identity. The Knowles and Peng (2005) measure, like some of the earliest racial attitudes IATs (Greenwald, McGhee & Schwartz, 1998), relied upon the linguistic recognition of names

which, in the national context of the USA, were taken to signal either white or black racial group membership. For the purposes of this study, such specific racial dichotomies were not appropriate, and as such the category-stimuli employed to represent light- and dark-skinned racial groups in the RA-IAT were also employed to make the same distinction in RI-IAT.

With regards to the evaluative categories, the Knowles and Peng (2005) study merely used synonyms of ‘self’ and ‘other’ to represent these two cognitive categories. While this approach yielded promising results, as Greenwald & Farnham (2000) have outlined in the context of implicitly assessing self-esteem, ‘self’ and ‘other’ are a pair of categories that are not as distinctly contrasted as other evaluative dimensions like ‘positive’ and ‘negative’, since ‘other’ is defined simply as anything that is outside or separate from the more distinct ‘self’. Furthermore, words such as ‘I’ and ‘myself’, and in particular words such as ‘they’ and ‘them’, label the distinction of interest, but may not engage the cognitive familiarity with the targeted concept that diverse words such as ‘peace’, ‘joy’ and ‘wonderful’ do when employed as positive-valence words in attitude IATs (see Nosek, Greenwald & Banaji, 2005, for a discussion). Thus, in order to provide a pair of ‘self’ and ‘other’ evaluative categories are both distinctly defined, while possessing comparable variability to the word-lists used in implicit attitude measures, the present RI-IAT used demographic and autobiographic information provided by participants, such that they may distinguish terms that are descriptive of themselves from those descriptive of a specific other, a highly dissimilar celebrity (see Bruce & Valentine, 1985, concerning the use of celebrity images in priming tasks).

To achieve this, participants were asked to answer a set of autobiographical questions, (such as their first name, surname, preferred nickname, country of birth, name of the town they presently live in, etc.), which were used by the interface to form a list of self-associated terms that, along with the words ‘self’ and ‘me’, comprised the stimuli for the evaluative ‘self’

category. These responses, in addition to their age, race, and gender responses in the demographics questions, were cross referenced against a database of 56 high-profile celebrities (actors, musicians and athletes), in order to rank the celebrities on the basis of their dissimilarity to the participant. For example, a participant who was a male, below the age of 45, of a light-skinned racial background, with the first name 'John', hailing from 'Sydney', would be preferentially matched with celebrities who were female, over the age of 45, of a dark-skinned racial background, with a first name other than 'John', and hailing from somewhere other than 'Sydney', such as actress Halle Berry. In an attempt to control for any particular fondness or distaste for any given celebrity, participants were presented with a short list (with accompanying picture) of the two or three celebrities calculated to be the most dissimilar to the participant (celebrities that perfectly matched any term to be used in the participant's lists were disqualified), and asked to select the person they felt the most dissimilar to. Participants were then shown their own list of identifying words alongside the corresponding answers for the selected celebrity, and given the option to change their decision if they felt any answers were too similar.

The subsequent RI-IAT was then generated with the default congruent condition of pairing self-related terms with the light-skinned image stimuli, and celebrity-other-related terms paired with dark-skinned image stimuli, though the actual presentation order of the blocks was counterbalanced, and as with the RA-IAT, showed no significant order effects. As such, for the RI-IAT, positive D-scores indicate greater association between one's self and light-skinned races (relative to a dissimilar celebrity), negative D-scores indicate greater association between one's self and dark-skinned races, and D-scores nearing zero indicating no strong implicit racial identification in either direction.

Results

Correlational Analysis

Bivariate correlations were explored between participant scores on five measures, in addition to relevant demographic details such as age, gender, and self-identification as belonging to a light- or dark-skinned racial group. As Table 1 shows, the D-scores of the RA-IAT significantly and positively correlated with both RI-IAT D-scores ($r = .176, p < 0.005$) and MRS-African scores ($r = .181, p < 0.005$), while approaching but not quite achieving a significant correlation with MRS-South Asian scores ($r = .107, p = 0.09$). All three MRS scales showed highly significant positive correlations of great effect size amongst themselves ($r = .738, p < 0.001$ for MRS-East Asian and MRS-African, $r = .931, p < 0.001$ for MRS-East Asian and MRS-South Asian, and $r = .818, p < 0.001$ for MRS-African and MRS-South Asian), with MRS-East Asian also correlating significantly but negatively with RI-IAT D-scores ($r = -.147, p < 0.05$). Self-described light- versus dark-skinned categorisation (dummy coded such that 0 = dark skin, 1 = light skin) correlated significantly with all three MRS scales (African, $r = .200, p < 0.005$, East Asian, $r = .178, p < 0.005$, and South Asian, $r = .196, p < 0.005$), and approached but did not achieve significance with D-scores on both the RA-IAT and RI-IAT ($r = .110, p = 0.086$, for both), and participant age ($r = -.121, p = 0.065$). Lastly, RI-IAT D-scores also showed a significant negative correlation with participant gender (dummy coded such that 0 = female and 1 = male, $r = .199, p < 0.005$), as well as a significant positive correlation with participant age ($r = .207, p < 0.001$).

Table 1*Pearson Correlations between MRS-scales, RA-IAT, RI-IAT, and demographic variables*

	MRS-EA	MRS-Afr	MRS-SA	RA-IAT	RI-IAT	L/D Skin
MRS-EA	-	.738***	.931***	.078	-.147*	.178**
MRS-Afr.	.738***	-	.818***	.181**	.030	.200**
MRS-SA	.931***	.818***	-	.107	-.038	.196**
RA-IAT	.078	.181**	.107	-	.176**	.110
RI-IAT	-.147*	.030	-.038	.176**	-	.110
Gender	.025	.034	.005	-.096	-.199**	-.083
Age	-.075	.075	-.055	.091	.207***	-.121

Note: $N = 256$ in all samples; Bonferroni correction for these comparisons is $\alpha = .05/3 = .016$.

* $p < .05$; ** $p < 0.016$; *** $p < 0.001$

Hierarchical Multiple Regression Analyses

Two hierarchical multiple regressions were conducted, with RA-IAT D-scores and MRS-African scores as the respective dependent variables, to examine the predictive overlap and shared variance between the above-noted correlates of each. A third hierarchical multiple regression was then conducted, targeting RI-IAT D-scores as the dependent variable but focussing only on a single racial group (Caucasians of European descent), in order to shed light on some of the unexpected patterns of significant correlations observed above.

Table 2*Hierarchical Multiple Regressions of MRS-African, RA-IAT and RI-IAT*

Dependent Variable	Step	Predictors	β	ΔR	ΔR^2_{Adj}
MRS-African score	1	Age	.098	.080	.006
		Gender	.478		
	2	SR Skin-Colour	5.782**	.207**	.043
	3	RI-IAT D-score	.009	.207	.031
RA-IAT D-score	4	RA-IAT D-score	1.036**	.264**	.070
	1	Age	0.22	.150	.014
		Gender	-.372		
	2	SR Skin-Colour	-.522	.185	.022
	3	RI-IAT D-score	.137*	.224*	.050
	4	MRS-African	.026**	.275**	.079
RI-IAT D-score	5	MRS-South Asian	-.015	.281	.076
	1	Age	.050***	.252***	.064
		Gender	-.427	.293	.073
	3	RA-IAT D-score	.157*	.330*	.110
	4	MRS-East Asian	-.148***	.460***	.230
	5	MRS-South Asian	.151*	.507*	.261
	6	MRS-African	-.017	.511	.257

Note: $N = 256$ for MRS-African and RA-IAT regressions, and $N = 144$ in RI-IAT regression;

* $p < .05$; ** $p < 0.01$; *** $p < 0.001$

As Table 2 shows, participants' self-reported racial skin tone (as either light or dark) served as a statistically significant predictor of MRS-African scores ($t = 3.034$, $p < 0.01$), when controlling for the influence of age and gender. RI-IAT D-scores provided no incremental

predictive improvement of self-reported racial skin tone at all, but even when controlling for RI-IAT, RA-IAT remained a statistically significant predictor of MRS-African scores ($t = 2.581, p < 0.01$). Due to the absence of a significant bivariate correlation between MRS-South Asian scores and RA-IAT D-scores, the predictive relationships between the two variables were not tested.

When taking RA-IAT D-scores as the dependent variable, neither age, gender, nor self-reported racial skin tone emerged as significant predictors, but RI-IAT D-scores did significantly positively predict RA-IAT D-scores ($t = 2.022, p < .05$), even when controlling for these first three variables (two of which were significantly correlated with RI-IAT). Beyond this, as Table 2 shows, MRS-African scores emerged as a significant predictor of RA-IAT D-scores ($t = 2.581, p < 0.01$), even when controlling for the previous four variables (including self-reported skin tone, itself a significant predictor of MRS-African scores). The inclusion of MRS-South Asian appeared to add negligible predictive value over and above MRS-African and the preceding variables.

Lastly, when regarding RI-IAT D-scores as the dependent variable, participant age was found to be highly significant predictive variable ($t = 3.082, p < 0.001$), but participant gender, which was also significantly correlated with RI-IAT on the bivariate level, did not emerge as a significant predictor when controlling for age. RA-IAT was also found to significantly predict RI-IAT scores ($t = 1.909, p < .05$), even controlling for age and gender, with MRS-East Asian scores showing strong incremental predictive significance over and above the preceding three ($t = -5.204, p < 0.001$). Despite showing only near-significant correlations with RI-IAT D-scores on the bivariate level, MRS-South Asian scores emerged as a marginally significant predictor ($t = 1.889, p < .05$), even controlling for the four preceding variables, including MRS-East Asian scores, with which it was highly correlated.

Discussion

Explicit and Implicit Racial Attitudes

This study aimed to explore the relationships between a novel measure of implicit racial identity, and measures of both implicit and explicit racial attitudes, when implicit racial distinctions are framed by contrasting dark-skinned and light-skinned stimuli. In support of the first hypothesis, concerning the correlation between implicit light-skinned preference and explicit racial prejudice against dark-skinned groups, a statistically significant positive correlation was found at the bivariate level between scores on the MRS-African measure of explicit racial attitudes, and the RA-IAT measure of implicit racial attitudes, indicating that participants who extolled more negative explicit attitudes towards persons of African descent also tended to show a stronger implicit evaluative preference for light-skinned faces over dark-skinned faces. Consistent with many earlier findings concerning explicit and implicit racial attitude measures, the effect size of the correlation was low (see Nosek, Greenwald & Banaji, 2005, 2007; Greenwald et al., 2009).

However, support for the first hypothesis must be regarded as incomplete, due to the results obtained from the other MRS subscale corresponding to a prototypically dark-skinned racial group, the MRS-South Asian scores. MRS-South Asian scores, like MRS-African scores, also positively correlated with RI-IAT D-scores within the approximate range of expected low effect sizes, but did not emerge on the bivariate level as statistically significant, simply approaching significance with a p -value of 0.09. It was for this reason that subsequent regression analyses focused solely on MRS-African scores as a more robust measure of explicit racial attitudes comparable to those explored in the earlier literature. Potential

methodological and contextual reasons for the stronger racial attitude results seen in the MRS-African over MRS-South Asian scales are discussed below.

The second hypothesis, concerning the predictive role of participants' self-reported categorisation as either racially light- or dark-skinned, was supported by the results of the first two regression analyses. Self-reported skin-colour category emerged as a statistically significant predictor of explicit negative racial attitudes towards dark-skinned individuals, as measured by MRS-African scores. This predictive significance was assessed having already controlled for age and gender, which were of potential concern due to the near-significant correlation observed between skin-colour category and age. Furthermore, the direction of the β coefficient was positive, indicating that in general, self-categorisation as a member of a light-skinned racial group predicted more negative explicit attitudes directed towards dark-skinned groups (specifically, those of African descent). The significance of this predictive relationship stands in stark contrast to that of implicit racial identification with light-skinned groups (as measured by RI-IAT D-scores), which added close to nothing to the cumulative variance explained by the model when fitted after self-reported skin-colour category. In contrast, when self-reported skin-colour category was similarly employed in the second regression analysis (following age and gender) predicting implicit racial attitudes, it was found to contribute no significant incremental predictive value. Thus, the predictions of the second hypothesis were supported in both regards, indicating that the present light- versus dark-skinned framing of the IAT methodology functions similarly to the single race dichotomies typically employed.

This finding is consistent with results reviewed by Nosek, Greenwald and Banaji (2005), who reasoned that the observed correlations between explicit racial attitude measures and self-reported categorical statements of group membership likely draw upon common deliberative evaluations that show poor overlap with the perceptual and behavioural biases targeted by

implicit measures (see Nosek, Greenwald & Banaji, 2007, for further discussion). It is also of note, when regarding the first multiple regression, that implicit racial attitudes (measured as RA-IAT scores) remained a significant predictor of explicit racial attitudes, even when the preceding variables were controlled (including RI-IAT, with which it correlates).

Furthermore, in a similar hierarchical multiple regression where MRS-South Asian scores were included in the third step, RA-IAT D-scores none-the-less retained a significant predictive relationship with MRS-African scores ($p < 0.05$), despite the strong correlations observed between both MRS scales. This regression was omitted from the results, as the immense item similarities between the MRS-African and MRS-South Asian subscales ($r = .818$) deceptively inflated the adjusted R^2 values by an increment of approximately 0.6.

The third and final hypothesis predicted that implicit racial identification with light-skinned groups (as measured by RI-IAT D-scores), would significantly predict implicit racial attitude scores (as measured by the RA-IAT), but not significantly predict explicit racial attitude scores towards dark-skinned individuals (for reasons outlined above, the analysis focussed on MRS-African scores). The non-significance of implicit racial identity as a predictor of explicit racial attitudes was discussed in the preceding treatment of the second hypothesis, with results consistent with the third hypothesis. Furthermore, in the second multiple regression focusing on implicit racial attitudes as the dependent variable, RI-IAT scores emerged as a significant predictor even after controlling for self-reported skin-colour category (which approached significance as a bivariate correlate of RI-IAT), age and gender (both of which were significant correlates of RI-IAT). The predictive value of implicit racial identity can also be considered distinct from that of explicit racial attitudes, since MRS-African scores maintained significance as a predictor after controlling for RI-IAT. In contrast, MRS-South Asian scores, found in other regressions to significantly predict implicit racial attitudes when fitted first, emerged as non-significant when fitted after MRS-African scores,

with which it was highly correlated. This support for the third hypothesis coincides with previous findings in the implicit attitudes literature addressing the expected relations between implicit and explicit measures (Nosek, Greenwald & Banaji, 2007; Greenwald et al., 2009), and is also consistent with the theoretical extrapolations concerning Centrality-based implicit racial identity and its conceptual distinctness from simple categorical assessments of racial group membership (see study 3 of Knowles & Peng, 2005).

Implicit Racial Identity

The RI-IAT methodology employed in the present study differed from previous IATs concerning racial identity, most notably that of Knowles and Peng (2005), in two ways. Firstly, both the evaluative and evaluated dimensions were modified, with the racial categories employing images of adult faces instead of group-typical first names (an alternative whose empirical viability has been established in papers such as Nosek, Greenwald & Banaji, 2005), and the self/other evaluative words replaced with a novel approach based on contrasting oneself with a dissimilar celebrity (though conceptually similar methods have been used in priming studies, see Bruce & Valentine, 1985). Second, rather than relying on an evaluative dichotomy between two specific racial groups, the RI-IAT (and present RA-IAT) employed multiple racial groups in order to distinguish racial categories based on prototypically light and dark skin tones (building upon the work of Blair, Judd, Sadler, & Jenkins, 2002; Maddox, 2004). As either of these modifications may have led to measurement inequalities that would render the present RI-IAT incomparable to previous measures such as Knowles and Peng's (2005) WICIAT, a third hierarchical multiple regression was conducted, taking RI-IAT scores as the dependent variable, to explore some of the atypical correlations observed between RI-IAT and other variables on the bivariate

level. However, due to the uneven distribution of racial groups in the present sample, the influence of racial intergroup variation could likely obscure such an analysis. As such, the third multiple regression was performed on only a single racial group, Caucasians of European descent, as this group comprised more than half (56%) of the total sample, and was the group most directly comparable with the samples used in the Knowles and Peng (2005) studies.

Of note, RI-IAT scores were the only measures employed in the present study which demonstrated significant bivariate correlations with participant age and gender. While age effects similar to those observed in this sample have been recorded in earlier IAT studies of racial biases, specifically that older participants sometimes display stronger preferential attitudes towards Caucasians (see Stewart, von Hippel & Radvansky, 2009, for an analysis), the present results were suspicious in that implicit racial identity correlated with age, but implicit racial attitudes did not. Though age effects were not directly explored in the Knowles and Peng (2005) studies, due to the very close age ranges of the participant samples, none of their analyses showed any significant gender effects, raising concern over the correlation between gender and RI-IAT in the present study. Considering the demographics data gathered, it appears that this unexpected correlation may be partially attributed to the higher proportion of female participants, as compared to male participants, that self-identified as belonging to a light-skinned racial group, which is likely a sampling property of the psychology undergraduate pool used. This possibility is discussed at greater length below, with regards to age.

Furthermore, in the third hierarchical multiple regression of the Caucasian subset of the sample, taking RI-IAT D-scores as the dependent variable, participant age was confirmed to be a highly significant predictor. However, with the variance predicted by age controlled for, gender subsequently failed to emerge as a significant predictor in its own right. This can

likely be explained in part by the near-significant correlation between gender and age in the sample ($r = .109, p = 0.093$). If the gender correlation is due primarily to trends in the wide age-range of the study, the results of the RI-IAT can still be meaningfully compared to Knowles and Peng's (2005) WICIAT, which observed no gender effects, but only in age-controlled samples. In addition, the crucial relations between implicit racial identity, and both implicit and explicit racial attitudes, were preserved in the regression after controlling for the collective influences of gender and age. RA-IAT scores remained a significant independent predictor, as did the MRS scales oriented towards both East Asians (which was highly significant) and South Asians (which was marginally significant despite possessing no significant bivariate correlation with RI-IAT). Furthermore, MRS-African scores, which were hypothesised and found to not relate to implicit racial attitudes in the preceding regression analyses did not emerge as a significant predictor, corroborating these findings.

With the bivariate correlation between implicit racial identity and gender largely accounted for, the mysterious age effects (now the primary unexpected finding) become potentially explicable when also considering the near-significant correlation observed between age and self-reported racial categorisation ($p = 0.058$). Closer examination of the relative frequencies of participants' racial backgrounds revealed a general trend, wherein all participants over the age of 26 happened to all identify as Caucasian of European descent. This higher relative proportion of older Caucasians (who were also disproportionately female, in line with the results above), would likely inflate the apparent influence of age on relative evaluations of implicit bias, though given the unique significance of the relation this is unlikely to explain the effect in its entirety. Therefore, while the results of this study are more comparable to those of Knowles and Peng (2005) than the initial bivariate correlations would imply, they must still be interpreted with caution, as it is unclear to what degree the observed age-effects

are due to the properties of the present sample versus the aforementioned tweaks in methodology.

Overall, the results of these analyses suggest that both implicit racial attitudes and implicit racial identity can be meaningfully assessed as more generalised comparisons between broad racial features (specifically the lightness or darkness of one's skin), and not merely as the two-race comparisons that dominate the literature. Furthermore, despite the unexpected gender and age correlations that appear to be a product of the sample used, this study provides strong preliminary support for an IAT approach to self-identification that is more sophisticated than the purely linguistic methodologies currently available. The use of recognisable celebrities, carefully selected for prominent differences relative to the participant, allowed participants to respond coherently to range of stimuli meaningfully associated with themselves and their own lives, while ensuring that the 'other' category of stimuli clustered together in a manner that could be related to skin-colour. This celebrity-other IAT methodology is potentially applicable to a many other self-identification constructs beyond racial identity, provided suitable celebrity candidates, recognisable within the cultural cohort being tested, can be identified who vary along the attribute under investigation (e.g. gender, nationality, etc.).

Methodological Limitations and Future Directions

While the key predictions of this paper have been supported by the results, and lend further support to both existing findings concerning implicit and explicit measures of racial bias (Nosek, Greenwald & Banaji, 2007; Greenwald et al., 2009), and the burgeoning study of implicit racial identification (Devos & Banaji, 2003; Knowles & Peng, 2005; Craemer, 2008, 2010; Craemer et al., 2013), several methodological limitations ensure that these results must

be interpreted with caution. Of primary concern, the present authors' decision to explore racial attitudes and racial identities via the generalised evaluations of light- and dark-skinned groups seeks to bridge the gap between much of the existing racial bias literature, and work focussing on more subtle Eurocentric evaluative biases (Blair, Judd, Sadler, & Jenkins, 2002; Maddox, 2004). In doing so the authors have, by necessity, limited their analyses to only those measures and constructs that can presently be administered without assuming dichotomous categories between pairs of racial groups (or in the case of stereotypes, beliefs about singular groups in the context of a dominant other; see Shapiro & Neuberg, 2007). Even the inclusion of the MRS, which, like most race measures targets specific groups in a relative context, was only possible due to the methodologically interchangeable nature of the scale items to multiple racial minorities in a single national context.

This shortcoming is most evident when reviewing the results of the multiple regression analyses, as despite the promising patterns of predictive significance and non-significance, the study's relative poverty of applicable measures has yielded models with relatively low adjusted- R^2 values, with the highest percentage of total variance explained reaching only 26%. The clear majority of the variance in each of the targeted dependent variables is left to be explained, and as reviews of the racial bias literature summarise (Hewstone, Rubin & Willis, 2002; Yzerbyt, Judd & Corneillo, 2003; Williams & Eberhardt, 2008), much of this variance could likely be explained by constructs which at present are only operationalised through racial dichotomies, such as racial stereotyping (Cox et al., 2012), social behavioural measures of racial bias (Saucier, Miller & Doucet, 2005), and both personal and media exposures to members of other races (Dixson, 2007). Future investigations of generalised light- versus dark-skin racial preferences would do well to employ a methodological overlap with multiple measures of specific racial dichotomies (such as Blacks compared with Whites,

East Asians compared with South Asians, etc.), in order to gain a more rich perspective on the interplay of both intergroup and intragroup racial biases.

References

- Allport, G.W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.
- Altemeyer, R.A. (1981) *Right-wing Authoritarianism*. University of Manitoba Press, Winnipeg.
- Amodio, D.M., Harmon-Jones, E., & Devine, P.G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink responses and self-report. *Journal of Personality and Social Psychology*, 84, 738 –753.
- Aron, A., Aron, E.N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in self. *Journal of Personality and Social Psychology*, 60(2), 241-253.
- Ashmore, R., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin*, 130, 80-114.
- Blair, I.V., Judd, C.M., Sadler, M.S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality & Social Psychology*, 83(1), 5-25.
- Blanton, H., Jaccard, J., Gonzales, P.M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, 42, 192-212.
- Blascovich, J., Spencer, S.J., Quinn, D.M., & Steele, C.M. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12, 225–229.

- Bobo, L. (2001). Racial Attitudes and Relations at the Close of the Twentieth Century. In *America Becoming: Racial Trends and Their Implications*, (Ed.) N Smelser, WJ Wilson, F Mitchell. Washington, DC: National Academy Press.
- Bobo, L.D. & Fox, C. (2003). Race, Racism, and Discrimination: Bridging Problems, Methods and Theory in Social Psychological Research. *Social Psychology Quarterly*, 66, 319-332.
- Bobocel, D.R., Son Hing, L.S., Davey, L.M., Stanley, D.J., & Zanna, M.P. (1998). Justice-based opposition to social policies: Is it genuine? *Journal of Personality and Social Psychology*, 75, 653-669.
- Bruce, V., & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology*, 76(3), 373-383.
- Cacioppo, J.T., Crites, S.L., Jr., Berntson, G.G., & Coles, M.G.H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, 4, 108-112.
- Carlson, J.M., & Iovini, J. (1985). The transmission of racial attitudes from fathers to sons: A study of Blacks and Whites. *Adolescence*, 20, 233-237.
- Coats, S., Smith, E.R., Claypool, H.M., & Banner, M.J. (2000). Overlapping Mental Representations of Self and In-Group: Reaction Time Evidence and Its Relationship with Explicit Measures of Group Identification. *Journal of Experimental Social Psychology*, 36, 304-315.
- Cox, W.T.L., Abramson, L.Y., Devine, P.G., & Hollon, S.D. (2012). Stereotypes, Prejudice, and Depression: The Integrated Perspective. *Perspectives on Psychological Science*, 7(5), 427-449.

- Craemer, T. (2008). Nonconscious Feelings of Closeness toward African Americans and Support for Pro-Black Policies. *Political Psychology*, 29(3), 407-436.
- Craemer, T. (2010). Ancestral Ambivalence and Racial Self-Classification Change. *Social Science Journal*, 47, 307–325.
- Craemer, T., Shaw, T.C., Edwards, C., & Jefferson, H. (2013). "Race Still Matters, However...": White Identification with Blacks, Pro-Black Policy Support, and the Obama Candidacy. *Ethnic and Racial Studies*, 36(6), 1047-1069.
- Crandall, C.S., Eshleman, A., & O'Brien, L.T. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82, 359-378.
- Crandall, C.S., & Schaller, M. (2005). *Social psychology of prejudice: Historical and contemporary issues*. Lawrence KS: Lewinian Press.
- Crites, S.L., Jr., Cacioppo, J.T., Gardner, W.L., & Berntson, G.G. (1995). Bioelectrical echoes from evaluative categorization: II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology*, 68, 997–1013.
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, 96, 608–630.
- Cross, W.E., Jr. (1991). *Shades of Black: Diversity in African American identity*. Temple University Press, Philadelphia.
- De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. *Cognition and Emotion*, 15, 189–206.

- De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion*, 12, 45–61.
- Devos, T., & Banaji, M.R. (2003). Implicit self and identity. In M. Leary & J. Tangney (Eds.), *Handbook of Self and Identity* (pp.153-175). The Guilford Press, New York.
- Dixson, A. F. (1998). *Primate Sexuality*. Oxford University Press, Oxford.
- Dovidio, J.F., Gaertner, S.L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity & Ethnic Minority Psychology*, 8, 88-102.
- Dovidio, J.F., Gaertner, S.L., Validzic, A., Matoka, K., Johnson, B., & Frazier, S. (1997). Extending the benefits of re-categorization: Evaluations, self-disclosure and helping. *Journal of Experimental Social Psychology*, 33, 401-420.
- Dovidio, J.F., Kawakami, K., & Gaertner, S.L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62–68.
- Duckitt, J.H. (1992). Psychology and prejudice: A historical analysis and integrative framework. *American Psychologist*, 47, 1182-1193.
- Duriez, B., & Soenens, B. (2009). The intergenerational transmission of racism: The role of right-wing authoritarianism and social dominance orientation. *Journal of Research in Personality*, 43, 906-909.
- Fazio, R.H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 75-109). Academic Press, New York.

- Fazio, R.H., Jackson, J.R., Dunton, B.C., & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327.
- Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.
- Gaertner, S.L., & Dovidio, J.F. (1986). The aversive form of racism. In J.F. Dovidio and S.L. Gaertner (Eds.), *Prejudice, Discrimination and Racism: Theory and Research* (pp. 61–89). Academic Press, Orlando.
- Gaertner, S.L., & McLaughlin, J.P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 46, 23–30.
- Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., & Mellott, D.S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109, 3–25.
- Greenwald, A.G., Klinger, M.R., & Liu, T.J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition*, 17, 35–47.
- Greenwald, A.G., & Krieger, L.H. (2006). Implicit bias: Scientific foundations. *California Law Review*, 94, 945–967.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.

- Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- Greenwald, A.G., Poehlman, T.A., Uhlmann, E., & Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17-41.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575-604.
- Ito, T.A., & Cacioppo, J.T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, 36, 660-676.
- Jones, J.T., Pelham, B.W., Mirenberg, M.C., & Hetts, J.J. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology*, 38, 170-177.
- Judd, C.M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109-128.
- Katz, P.A. (2003). Racists or tolerant multiculturalists? How do they begin? *American Psychologist*, 58(11), 897-909.
- Knowles, E.D., & Peng, K. (2005). White Selves: Conceptualizing and Measuring a Dominant-Group Identity. *Journal of Personality and Social Psychology*, 89, 223-241.

- Koole, S.L., Dijksterhuis, A., & Van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, 80, 669-685.
- Kovel, J. (1970). *White racism: A psychological history*. Pantheon, New York.
- Lieberman, L., Kaszycka, K.A., Martinez Fuentes, A.J., Yablonsky, L., Kirk, R.C., Strkalj, G., Wang, Q., & Sun, L. (2004). The race concept in six regions: variation without consensus. *Collegium Antropologicum*, 28(2), 907-921.
- Maddox, K.B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, 8, 383-401.
- Markham, A.C., Alberts, S.C., & Altmann, J. (2012). Intergroup conflict: Ecological predictors of winning and consequences of defeat in a wild primate population. *Animal Behaviour*, 84, 399-403.
- Maynard Smith, J. & Parker, G. A. 1976. The logic of asymmetric contests. *Animal Behavior*, 24, 159-175.
- McConahay, J.B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). Orlando FL: Academic Press.
- McConnell, A.R., & Leibold, J.M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435-442.
- Messick, D.M., & Mackie, D.M. (1989). Intergroup relations. *Annual Review of Psychology*, 40, 45-81.

- Miller, G. (2000). *The mating mind: how sexual choice shaped the evolution of human nature*. Heineman, London.
- Neuberg, S.L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology*, 56, 374-386.
- Nosek, B.A., & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition*, 19(6), 161-176.
- Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166-180.
- Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review (pp. 265-292). In J.A. Bargh (Ed.), *Automatic processes in social thinking and behavior*. Psychology Press, New York.
- Nuttin, J.M. (1985). Narcissism beyond Gestalt and awareness: The name–letter effect. *European Journal of Social Psychology*, 15(3), 353–361.
- Operario, D., & Fiske, S. T. (2001). Stereotypes: Processes, structures, content, and context. In R. Brown & S. Gaertner (Eds.) *Blackwell handbook in social psychology* (Vol. 4: Intergroup Processes, pp. 22-44). Cambridge, MA: Blackwell.
- Pearson, A.R., Dovidio, J.F., & Gaertner, A.L., (2009). The Nature of Contemporary Prejudice: Insights from Aversive Racism. *Social and Personality Psychology Compass*, 3, 1-25.

- Pelham, B.W., Mirenberg, M.C., & Jones, J.T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology*, 82, 469-487.
- Perdue, C.W., Dovidio, J.F., Gurtman, M.B., & Tyler, R.B. (1990). Us and them: social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology*, 59, 475-486.
- Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, 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), 729-738.
- Phinney, J. (1990). Ethnic identity in adolescents and adults: A review of research. *Psychological Bulletin*, 108, 499-514.
- Richardson-Klavehn, A., & Bjork, R.A. (1988). Primary vs. Secondary rehearsal in an imagined voice: Differential effects on recognition memory and perceptual identification. *Bulletin of the Psychonomic Society*, 26, 187-190.
- Roediger, H.L. III. (1990). Implicit memory: Retention without remembering. *The American Psychologist*, 45, 1043-1056.
- Rooth, D.O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3), 523-534.
- Rubinstein, G. (1996). Two peoples in one land: A validation study of Altemeyer's right-wing authoritarianism scale in the Palestinian and Jewish societies in Israel. *Journal of Cross-Cultural Psychology*, 27, 216-230.

- Saucier, D.A., Miller, C.T., & Doucet, N. (2005). Differences in helping Whites and Blacks: A meta-analysis. *Personality and Social Psychology Review*, 9, 2–16.
- Schacter, D.L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 501-518.
- Sears, D.O. (1988). Symbolic racism. In P. Katz & D. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53–84). Plenum Press, New York.
- Sears, D.O., Henry, P.J. & Kosterman, R. (2000). Egalitarian values and contemporary racial politics. In D. O. Sears, J. Sidanius, & L. Bobo (Eds.), *Racialized politics: The debate about racism in America* (pp. 75–117). University of Chicago Press, Chicago.
- Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, 39, 75-82.
- Sellers, R.M., Smith, M.A., Shelton, J.N., Rowley, S.J., & Chavous, T.M. (1998). Multidimensional model of racial identity: A reconceptualization of African American racial identity. *Personality and Social Psychology Review*, 2, 18-36.
- Son Hing, L.S., Chung-Yan, G.A., Hamilton, L.K., & Zanna, M.P. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology*, 94(6), 971-987.
- Shapiro, J.S., & Neuberg, S.L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, 11, 107-130.

- Smith, E.R., & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality and Social Psychology Bulletin*, 22, 635-642.
- Stanley, D., Phelps, E.A., & Banaji, M.R. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science*, 17(2), 164-170.
- Stewart, B.D., von Hippel, W., & Radvansky, G.A. (2009). Age, race, and implicit prejudice: Using process dissociation to separate the underlying components. *Psychological Science*, 20, 164-168.
- Towles-Schwen, T., & Fazio, R.H. (2006). Automatically-activated racial attitudes as predictors of the success of interracial roommate relationships. *Journal of Experimental Social Psychology*, 42, 698-705.
- von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. In M.P. Zanna (Ed.), *Advances in Experimental Social Psychology* (pp. 177-254). Academic, Orlando.
- von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology*, 33, 490-509.
- Williams, M.J., & Eberhardt, J.L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, 94(6), 1033-1047.
- Yzerbyt, V., Judd, C.M., & Corneille, O. (2003). *The psychology of group perception: Perceived variability, entitativity, and essentialism*. Psychology Press, New York.

Discussion for Thesis Chapter 6

The journal article featured in this chapter was designed to answer two empirical questions: does the RA-IAT continue to function as expected when employing more general evaluative categories than direct racial dichotomies, and can comparisons between ones' self and distinct celebrity-others generate meaningful implicit racial identity scores? With regards to the goals of this thesis, a positive result to the former question was necessary to justify a set of later empirical evaluations of the SATEST measurement tool (reviewed in Chapter 7), while the latter question represents a possible avenue of further analyses that did not warrant an extensive follow-up within the period of this thesis.

This study demonstrated that the light- versus dark-skinned manipulation of the RA-IAT methodology shows similar psychometric properties to more widely used dichotomous RA-IAT measures. The baseline reaction latencies of this manipulation showed the same general tendency towards favouring the socially dominant group ('light-skinned' persons in this study, 'white' Caucasian persons in many previous studies, see Nosek, Greenwald & Banaji, 2005) that has been consistently observed in earlier RA-IAT studies employing specifically compared pairs of racial groups, performed on culturally Western samples. More importantly, the reliably low (in terms of absolute effect size) but often statistically significant correlations and predictive relationships observed in previous studies between implicit and explicit measures of racial attitudes (including the MRS specifically, see Son Hing et al., 2008) were also replicated in this study using the light- and dark-skinned categories, and multiple forms of the MRS. Also consistent with prior studies, the present RA-IAT manipulation did not demonstrate any correlations with demographic variables such as gender and age. Of particular significance, the present RA-IAT manipulation grouped racial minorities (those of East-Asian descent, in this Western cultural context) into the same evaluative category as the white Caucasian majority. Despite the high correlations observed between all three versions

of the MRS used in this study (as is typical, Son Hing et al., 2008), the RA-IAT measure showed significant (or near-significant) correlations with the two dark-skinned groups, but no significant correlation with the MRS variant addressing East-Asians. This suggests that the present RA-IAT manipulation detects implicit biases in racial preferences to a degree comparable with previous RA-IAT measures, while remaining internally consistent with regards to the alternative comparative distinction it employs.

This present study also demonstrated empirical support for the efficacy of the celebrity-other RI-IAT methodology introduced, but the observed effect sizes and unique predictive value of this measure were not sufficiently impressive to justify its inclusion in later analyses within this thesis. Had the present RI-IAT measure demonstrated even greater predictive value than what was observed in this study, one possibility for the follow-up analyses (including those involving the SATEST measure) would have been to model implicit racial identity as either a mediating or moderating variable, operating between general empathetic dispositions and specific indicators of racial attitude biases. While such an analysis may yet be possible in the future, the results of the current study do not offer sufficient evidence to determine whether or not the shortcomings of this RI-IAT measure are strictly a product of its methodological design, or whether the construct of implicit racial identity is simply not well-suited to the general light- versus dark-skinned evaluative categories consistently applied throughout this study. It is possible, given the intimate nature of racial identity, that the self-other distinctions of many participants (particularly those of single-race backgrounds) may apply more meaningfully and reliably to identification with a specific racial ingroup, contrasted with a specific racial outgroup, than the more general comparison of light- versus dark-skinned groups.

CHAPTER 7

The Development of the SATEST Measure

The journal article featured in this chapter details the remaining three research studies comprising (along with Chapter 6) the empirical ‘half’ of this thesis. The introductory portions of this paper outline much of the supporting theory and literature concerning the evolutionary synthesis of empathy-modulating and coalition-management mechanisms that is of central importance to the goals of this thesis, drawing on insights from a range of findings across the study of prejudice in social, differential and moral psychology. As an extension of the theoretical positions introduced, this article also outlines the development of the SATEST task, a measurement tool designed to utilise several insights into the intrapersonal mechanisms of empathy and coalition management predicted by this tentative, integrated theory. The three reported studies empirically investigate the basic psychometric properties of the SATEST methodology, its viability as a measure of several elements of moral character and judgement, and finally its utility at measuring biases in sympathetic decisions, justifications and attributions in participants expressing prejudice (focusing on racial prejudice, building upon the study in Chapter 6).

As was first introduced in Chapter 1, the integrative theory of empathy-modulation and coalition-management developed in this thesis employs an adaptationist approach (inspired by evolutionary work in the three key prejudice fields cited) to conceptualise a common suite of adapted psychological mechanisms responsible for the lapses in sympathy, self-enhancing justifications, and moralistic attributions, thought to underpin most (if not all) prejudiced behaviours and evaluations. The following journal article elaborates on this conception, exploring both the predicted process-level features of the coalition-management mechanisms in question, and the related theories concerning the evolutionary selective pressures that potentially shaped them. It is hypothesised that humans (as with other primate species with

similar fluidity of social organisation; see O'Connell, 1995; de Waal, 2008) possess a suite of cognitive adaptations that evolved as a means of addressing the recurring fitness challenges of negotiating moderate-to-large coalitional groups of non-kin allies (Kurzban & Leary, 2001; Kurzban, Tooby & Cosmides, 2001; Cosmides, Tooby & Kurzban, 2003), whose precise memberships vary ambiguously over time and are subject to defection and free-riders (Cohen, 2012), but whose cohesion and cooperative benefits strongly influence any individual's survival and reproductive prospects (Nowak & Sigmund, 2005). In addition to the general primate concerns relating to social dominance hierarchies (Sidanius & Pratto, 1999) and intragroup competition for status and mating opportunities (Buss, 1988), the cohesion between coalitional ingroup members is predicted to operate in large part by virtue of similar empathetic (and sympathetic) emotional attachments as those observed in kin and mating partner bonds (de Waal, 2008; Krebs, 2008), however, these expressions of empathetic attachment are expected to express conditionally, depending directly upon evidence of reciprocity (Decety & Grézes, 2006; de Vignemont & Singer, 2006) and group-commitment (Kurzban, Tooby & Cosmides, 2001) in others. As such, it is theorised that due to the degree of intergroup conflict that appears to have been common in the social environments of our common human ancestors (Miller, Zielaskowski, Maner & Plant, 2012; McDonald, Navarrette & Van Vugt, 2012), empathy towards others is predicted to modulate strategically in response to others, with regards to their perceived categorical ingroup or outgroup status (Brewer, 1979; Krebs & Van Hesteren, 1994; Hart et al., 2000), their perceived similarity to oneself and apparent observance of group norms, goals and practices (de Vignemont & Singer, 2006), and contextual cues and beliefs concerning the opportunities and threats posed by known outgroups (Miller et al., 2012).

Beyond these affective, intuition-driven evaluations of others, it is theorised that social communication in the context of interpersonal and coalitional conflict is intrinsically

moralistic in nature, in service of the adaptive challenges of recruiting allies (Krebs, 2008), defusing potential dangerous confrontations (DeScioli & Kurzban, 2013), and cultivating a reputation of virtue and consistency of judgments (Nowak & Sigmund, 2005; Kurzban, DeScioli & Fein, 2012). As such, the down-regulation of empathy and sympathy towards categorical outgroup members (a distinction that can be provoked by any sufficiently salient cues signalling dissimilar characteristics or loyalties to one's own; see Cosmides, Tooby & Kurzban, 2003) is predicted to also motivate justifications for one's own behaviour that emphasise adherence to group norms or the prospect of conflict (McDonald, Navarrete & Van Vugt, 2012), in addition to attributions for the behaviour of others that emphasise violation of group norms or demonstrably poor moral character (often as grounds for moralistic punishment; see Haidt & Graham, 2007).

This integrated adaptationist theory of human empathy and coalition management provided a range of new predictions concerning how moral judgements and expressions of interpersonal social conflict are likely to be expressed. Contrary to traditional Neo-Kohlbergian moral judgment theories and similar cognitive-developmental approaches to variation in moral character (see Rest et al., 1999), which account for expressions of disregard for the well-being of others (including forms of prejudice such as racism) as reflections of differentially sophisticated moral principles (ranging from selfish, to socially conventional, to reflective and universal), the present theory takes affect-driven intuitive evaluations of group-membership and norm-adherence as the primary determinants of whether or not a subjects' mechanisms of empathy and sympathy will activate in reaction to another person (building on the central insights of Haidt, 2001; 2007; 2012). Subsequently, while earlier moral psychology theories regard the reasons offered for behaving poorly towards some other person as earnest (or perhaps simply exaggerated) explanations of the conscious reasoning motivating their decisions, the present evolutionary approach regards such justifications as

typically *post hoc* rationalisations of visceral evaluations, produced (often unknowingly) with the adaptive goal of cultivating a defensible moral reputation and recruiting the support of witnesses via appeals to group norms and stability-serving general sentiments (DeScoli & Kurzban, 2013). In accordance with this social goal, attributions concerning the behaviour and character of any person mistreated by the subject are predicted to serve as justifications of their mistreatment in accordance with these moralistic sentiments (Moll et al., 2002; 2005).

This account also contradicts many of the theoretical positions advocated in the intergroup prejudice social psychology literature, which construe the mistreatment of perceived outgroup members as the rational enactment of (often secretly held) beliefs concerning their stereotypical characteristics (such as possessing inferior abilities or posing a general threat; see Fiske et al., 2002). Rather than a product of entrenched, mistaken beliefs, which are kept secret in social contexts thought not to support them (i.e. concealed prejudices, see Crandall, Eshleman & O'Brien, 2002), the present theory regards the majority of initial evaluations as strategic reactions to one's perceived coalitional status, with subsequent justifications and attributions serving as *post hoc* rationalisations that predictably appeal to whatever moralistic standards or group norms are considered likely to be well-received by bystanders at the time.

Based on the predictions generated by this evolutionary model, the SATEST methodology was conceived as a means of measuring individuals' expressions of empathy, justification and attribution, by presenting participants with scenarios oriented around a hypothetical character who has broken an ultimately harmless social rule. By presenting the participant with a partially-immersive, simulated social situation, in which they witness a target character violating a harmless social rule, and are subsequently placed in a position where they must decide whether or not the target is punished for their misdeeds, each SATEST scenario begins with a direct measure of the participants' sympathetic feelings towards the target. The ambiguities and unaccountability of each scenario were specially written to eliminate all

salient confounding social influences on this decision, while leaving the participant with little more than their positive or negative evaluative intuitions to base their judgment on (though in accordance with insights offered by Greene & Haidt, 2002, and Paxton, Ungar & Greene, 2011, the measure is sensitive to the time taken to make the decision and any explicit deliberations entertained while doing so, which are factored into the scoring). Following this initial decision, the participant is led through non-judgmental conversational cues, prompting a justification for their chosen course of action, and an attribution as to why the target character had performed their social transgression in the first place. The unique value of this vignette design lies in its utilisation of the motive of moral justifications (particularly those relating to moralistic punishment), to provide participants with a socially acceptable and potentially reputation-enhancing channel through which they can express a distinctly non-sympathetic response towards a target character who, strictly speaking, has harmed no one, and appears to be in emotional distress. The present theory predicts that, if the participants' empathetic mechanisms are activated by the target character, they will reliably give the target the benefit of the doubt and spare them their comeuppance, employing the ambiguity of the situation to attribute exonerating explanations for poor conduct to targets with whom they sympathise. However, unlike in many other measures of both prejudice and moral judgment, which require participants to embrace the potential social penalties of assuming a cruel, punishing or vindictive stance towards a prospective victim (see Crandall, Eshleman & O'Brien, 2002), the SATEST allows participants who experience no activation of sympathetic feeling towards the target in the initial presentation to express their negative evaluations in a socially defensible (and potentially self-enhancing) manner, by appealing to the violated rule as grounds for moralistic punishment, and perhaps even as evidence of the target's unsavoury or immoral character. The SATEST approach is thus designed to measure the three aspects of social and moral cognition most influential in the expression prejudice,

while exploiting the contrived conflict between sympathetic feeling and adherence to social rules to overcome the problem of honesty and impression-management that typically plagues both moral judgment and intergroup prejudice research (Greenwald, McGhee, & Schwarz, 1998; Richman et al, 1999; Christensen & Gomila, 2012).

The three studies detailed in this journal article represent three sequential phases testing the hypothesised properties and measurement capacities of the SATEST task, each of which builds upon the study preceding it. It was necessary that these three studies be evaluated together in this manner (despite the resulting great length of the journal article), as the utility of the SATEST as a moral psychology measure could only be established if it demonstrated internally consistent psychometric properties, and the capacity of the SATEST to detect prejudicial differences in the activation of moral processes naturally depends on its baseline ability to measure moral processes at all. In addition to the data reported in the article (total $N = 720$), the two major variants of the SATEST (the original form and the skin-colour varied form) each underwent preliminary pilot-testing, primarily for the purpose of bug-testing the implementation of the program in Adobe Flash, with each pilot-test comprised of 20 undergraduate participants (total $N = 40$). As these pilot-studies did not test the relationship between the SATEST indices and any other measures (for the sake of brevity), they were mostly omitted from the publication manuscript of this paper.

The following article was submitted for publication to the *Journal of Experimental Social Psychology*, under the title ‘Sympathy vs. Social Rule Adherence: A New Measure of Interpersonal Empathy’.

Declaration for Thesis Chapter 7

In the case of journal article featured in Chapter 7, the nature and extent of my contribution to the work, and the contributions of the other listed co-authors is as follows:

<i>Name</i>	<i>Nature of Contribution</i>	<i>Contribution</i>
Tim Marsh	Decision concerning the topic of the paper	90%
	Search and review of the literature	
	Design and programming of measurement tools	
	Administration of study and data collection	
	Analysis and interpretation of data	
	Principle writing and editing of the manuscript	
Simon Boag	Advice on topic and approach	10%
	Assistance with editing and cutting	
	Suggestions for the refinement of the manuscript	

Sympathy vs. Social Rule Adherence:

A New Measure of Interpersonal Empathy

Running title: *Sympathy vs. Social Rules*

Tim Marsh
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: timothy.marsh@mq.edu.au

Simon Boag
Department of Psychology
Macquarie University
Sydney, NSW, 2109
Australia
Email: simon.boag@mq.edu.au

ABSTRACT

In recent decades, research concerning the mechanisms of intergroup prejudice and investigations into bases of human moral judgments have converged on the conceptual common ground of empathy. The growing understanding of the evolutionary origins and neurological underpinnings of empathetic emotions and motivations have revealed functional characteristics of the cognitive processes underwriting sympathetic social behaviour that have yet to be employed in the ongoing tasks of psychometric testing. Drawing on data from three studies (each with a university undergraduate sample, total $N = 720$), the authors introduce a new measure of sympathetic concern, which relies upon eliciting conflicting sensations of interpersonal empathy and deontological rule-adherence in 12 computer-simulated social interactions. The psychometric properties and construct validity of the measure are explored, and the potential applications of this new measurement methodology in both social and moral psychology are discussed.

Keywords: evolutionary psychology; moral judgement; prejudice; empathy; psychometrics

Sympathy vs. Social Rule Adherence:

A New Measure of Interpersonal Empathy

In humans, as with all mammals and most birds, the range of sensations and motivations associated with the concept of ‘empathy’ are often considered the core of the psychology of altruism (de Waal, 2008). While definitions of what precisely empathy is have varied considerably since the early investigations of the philosophy of ethics (Hoffman, 2000; Eisenberg & Morris, 2001; Preston & de Waal, 2002), a prevailing conception in contemporary psychology is that a mental act is ‘empathetic’ if it is directed towards, or deeply inclusive of, the perceived psychological states of another, particularly their emotions (see Decety, Norman, Berntson & Cacioppo, 2012, for a neurobehavioural review of empathetic cognitive processes). The nuances of precisely what psychological processes a researcher may be referring to when speaking of empathy are well-captured in de Waal’s 2008 review of the subject, which emphasises three broad tiers of empathetic cognition: (1) *Emotion Contagion*: the automatic perception and vicarious adoption of basic emotional states displayed by others; (2) *Sympathetic Concern*: experiencing motivations regarding the emotional states of others independent of managing one’s own experience of emotion contagion (for example, seeking to console a recently wronged peer), and; (3) *Empathic Perspective-Taking*: cognitively adopting the perspective of an other (as in the simulation variant of Theory-of-Mind; see Preston & de Waal, 2002), such that one’s perspective is inclusive of their emotional, sensory and motivational states. These tiers map well onto the forms of empathy that can be elicited from various non-human animals. For example, great apes can be prompted to exhibit sympathetic concern, whereas tested species of monkeys only exhibit signs of emotion contagion (O’Connell, 1995; Watts et al, 2000; Schino et al., 2004). Beyond this, these three tiers of empathetic function also describe the relative

automaticity and neurological interdependence of these processes in humans (Decety & Grézes, 2006; de Vignemont & Singer, 2006).

Typically, altruistic, cooperative, or generally pro-social behaviour is conceptualised as requiring at least the cognitions of sympathetic concern, for simple emotion contagion alone is known to motivate distinctly selfish or socially disadvantageous behaviours, both in humans and non-human animals (see de Waal, 1996 for a comparative account). In social psychology, much of the early work investigating altruistic and pro-social interactions focused specifically on the circumstances and interpersonal pressures that elicit, enable, or impede the expression of helping behaviours (Latané & Rodin, 1969; Turiel, 1983). While these lines of research have yielded invaluable insights into phenomena such as bystander effects (Darley & Latané, 1968; Watts et al, 2000), the influence of stereotypes (Brewer, 1988; Devine, 1989; Hewstone, Hantzi & Johnston, 1991; Stangor, Lynch, Duan & Glass, 1992), and the fundamentally distinct responses elicited by ingroup- and outgroup-members (Brewer, 1979; Krebs & Van Hesteren, 1994; Hart et al, 2000), the *intrapersonal* mechanisms of empathetic and sympathetic motivations have been relatively neglected (de Vignemont & Singer, 2006).

Conversely, in the field of moral psychology, many of the earliest accounts of altruism and social helping (such as those of Freud, 1913; 1930) were deeply rooted in the motivational role of empathetic and sympathetic emotions. However, during the ‘cognitive revolution’ of the 1950s and 60s, inspired by the developmental insights of Piaget (1965) and his contemporaries, moral psychology—and the area of moral judgment, in particular—came to be dominated by a stage-based cognitivist paradigm, which in large part sought to ground

moral behaviour in unemotional processes of reasoning (see Rest et al., 2000 for an overview). While producing valuable insights into the characteristics of interpersonal situations most salient to effortful reasoning on moral topics, this approach—dominated primarily by Kohlberg (1964; 1969) and his descendent schools of thought (Rest et al., 1999)—disregarded the influence of affect and emotion *a priori*, taking abstract ethical principles to be the proper domain of moral motivation (reviewed in Greene & Haidt, 2002). As with social psychology in general, however, recent decades have seen a renewed appreciation for the influence of emotion and motivation in the moral domain, prompted primarily by innovations in neuroimaging research (reviewed in Decety, Norman, Berntson & Cacioppo, 2012), and a distillation of phylogenetic consistencies in the biopsychological apparatus that humans share with our evolutionary next-of-kin (reviewed in Krebs, 2008). Consequently, contemporary moral and social psychology have converged on the common territory of empathetic motivation in social interactions (Haidt, 2007), and now stand to benefit from a joint exploration of the evolved underlying mechanisms of interpersonal sympathy (Cushman, 2011). However, one key impediment, as some theorists have observed (Haidt, 2007; Christensen & Gomila, 2012), lies in the methodological history of both fields, where measurement tools are either tailored to the demonstration of particular effects (see Chapman & Anderson, 2011), or are built upon the now-questioned conceptual assumptions of decades past (see Haidt & Graham, 2007).

In this paper, the authors seek to propose and validate a new means of measuring interpersonal sympathetic concern, in a manner that is compatible with the evolutionary analysis of underlying cognitive mechanisms, while also overcoming a wide range of methodological limitations that have been identified in the prior literature (Haidt & Graham, 2007; Christensen & Gomila, 2012). Over the course of three studies and two re-test follow-

ups, this paper explores the development, internal psychometric properties and evidence for convergent and divergent construct validity of the *Sympathetic Attributions Towards Emotive Social Transgressors task* (SATEST).

The Evolutionary Approach to Empathy and Sympathy

As with all analyses undertaken through the *adaptationist* approach, evolutionary psychology regards the phenomena of empathy in humans as a developmentally refined set of responses and behaviours, which are the output of a variety of neurocomputational mechanisms that are, themselves, the product of natural selective forces across the history of our species (Tooby & Cosmides, 2005; Buss, 2005). Since the behavioural manifestations of empathy are species-typical in humans (Kruger, 2003) and appear to act upon specific subsets of information in consistent ways (Krebs, 2005a; 2008), as well as showing gradual partial forms in our phylogenetic relatives (de Waal, 2008), it is important to investigate the possibility that the mechanisms underlying empathy are adaptations, which originally evolved in response to some recurring fitness challenges faced by our distant common ancestors (Kurzban & Leary, 2001). Whether the mechanisms of empathy are best understood as an adaptation (or set of adaptations), or whether they are byproducts of some other adaptations (or perhaps even a form of preserved phylogenetic ‘noise’), is best determined by searching for evidence of functional ‘good design’ with regards to recurring ancestral survival or reproductive problems. As Tooby and Cosmides (2005) explain, the positive criteria of ‘good design’ allow one to identify possible adaptations via their improbable functional shaping by selective pressures, whereas traits may only be identified as byproducts or ‘noise’ in the conspicuous absence of such evidence of selection.

With regards to the possible functions of empathy in animals generally, and humans in particular, there have been a variety of proposed problems that cognitive routines such as these may address (Kruger, 2003; Krebs, 2005a; 2008). The two most prevailing and likely mutually-inclusive functions of empathetic cognition are both summarised by de Vignemont and Singer (2006): (1) *An Epistemological Function*, wherein an organism automatically mirrors the perceived psychological states of others through their own sensations and emotions, in order to better understand and predict their motives and behaviours, and; (2) *A Social Cohesion Function*, wherein other-directed feelings serve to foster and reinforce pro-social and cooperative behaviour with appropriate others, such as kin, coalitional allies, and ingroup members. While significant evidence has accumulated in favour of the first of these two functions (as outlined in Decety & Jackson, 2004), in conjunction with a range of other, less strictly empathetic cognitive mechanisms investigated by multiple schools of Theory-of-Mind researchers (see Ferstl et al., 2008), it is the latter of these two functions that bears greater relevance to the study of interpersonal sympathetic concern, and which has proved influential in evolutionary investigations of social and moral psychology (de Waal, 2008; Krebs, 2008).

Social-Intuitionist Theory

Evolutionary psychology generally regards emotions as clusters of functionally-related motivational states and cognitive biases that organisms adopt and shift between in reaction to important changes of circumstance (Haidt, 2001; 2007; Cushman, 2011). For example, the emotions commonly labelled as ‘anger’ are aroused most often by the perception of being wronged (physically or socially), and physiologically prepare an individual for retributive or defensive physical action, while also fostering perceptual biases adaptive for potentially

violent confrontations (see Krebs, 2003). In this vein, the pro-social emotional processes associated with empathy are a subset of several social emotions, which collectively serve to motivate an organism to selectively cultivate and defend social outcomes that contribute to their ultimate reproductive fitness (Nowak & Sigmund, 2005). Through the effects of the selective forces of *kin-selection* and *reciprocal altruism* (Haidt, 2007), many animals (including humans) exhibit other-oriented motivated behaviours that increase reproductive outcomes through shared genetic inheritance (Krebs & Denton, 1997), or enhance proximal survival outcomes through reliable cooperation (Tooby & Cosmides, 1996; Krebs & Janicki, 2004). Evidence from comparative psychology and primatology (de Waal, 2008) suggests that the mechanisms specifically underlying empathy and sympathetic concern in humans are functional expansions of the more rudimentary mechanisms that facilitate *affective communication*, *parental nurturance*, and *social attachment* in other, phylogenetically similar mammals (see Decety, Norman, Berntson & Cacioppo, 2012, for an in-depth review). Furthermore, disruptions to social behaviours corresponding with neurological damage to brain regions functionally associated with interpersonal empathy (reviewed in Koenigs et al., 2007) provide additional support for the essential motivational role of empathetic emotions in pro-social behaviour (see Batson, 2011, for further discussion).

Following a range of studies spanning the 1990s, which demonstrated the relative primacy of emotional influences in many social and moral decisions (reviewed in Greene & Haidt, 2002), Haidt (2001) and colleagues formulated and refined the *social-intuitionist* theory of moral psychology (Haidt, 2007; Haidt & Graham, 2007). Contrary to much of the moral judgement and social decision-making literature of preceding decades, the social-intuitionist approach identified the central role that affect-laden intuitions play in the majority of contentious real-world moral and interpersonal dilemmas. In this account, the majority of

moral-domain social decisions are made rapidly, in the form of impulsive rejections made on the basis of emotional feelings of ‘wrongness’, which predictably emerge when a proposed course of action violates an unspoken intuitive standard (Haidt, 2001). Moral reasoning, under most circumstances, actually takes the form of post-hoc justifications for decisions made intuitively, a relationship well-demonstrated in the literature on ‘moral dumbfounding’ (see Greene et al., 2004; Greene, 2007). Haidt (2007) later expanded the social-intuitionist approach, so as to account for the specific instances in which traditional moral reasoning is employed to overcome or mediate between intuitive reactions (Greene et al., 2001; Greene et al., 2004; Greene et al., 2009), and also to extend the scope of the evolutionary analysis beyond the strictly harm- and justice-based framing of earlier moral judgment approaches (Haidt & Graham, 2007).

Through analysis of the functional characteristics of the social judgments commonly identified as eliciting a ‘moral’ response, Haidt and colleagues (Haidt, 2007; Haidt & Graham, 2007) identified five foundational classes of affect-laden intuitions, which define the emotional and motivational parameters that lead the majority of tested participants to feel that a moral-domain social transgression has taken place (though some more recent analyses include a sixth foundation, as is discussed below). The first two foundational moral intuitions are concerns for *harm*, and concerns for *fairness*. As Haidt (2007) identifies, these are the two domains of moral transgression that have received the most attention in the cognitivist history of moral psychology (see also, Greene & Haidt, 2002), and are generally considered the most influential sensitivities in an individual’s perception of an act as moral or immoral. Given the relational sensitivity of these first two intuitions, social-intuitionist theory suggests that concern for harm (which ties closely to the activation of empathy), is largely a manifestation of kin-selection pressures since both similarity and interpersonal closeness are crucial

mediating factors (Krebs, 2003; 2005b). Similarly, concern for fairness is most likely a product of reciprocal altruism, as judgments of fairness remain largely divorced from affiliation and closeness, and conversely appear to inform desire for closeness and cooperation with non-kin (Krebs & Van Hesteren, 1994; Krebs & Janicki, 2004; Nowak & Sigmund, 2005). The three additional moral intuitions are concerns for *authority*, concerns for *purity*, and concerns for *ingroup-loyalty* (Haidt & Graham, 2007). Our intuitions concerning authority are thought to exist primarily due to the pressures of social dominance and hierarchy, the patterns of which are well-documented in both humans (Sidanius & Pratto, 1999) and other mammals (Harcourt & de Waal, 1992). The moral intuitions that regard the maintenance of purity (in physical, social, and often religious conceptions) is understood to be a predictable product of the mammalian disgust response, which serves as an adaptation for toxin and disease avoidance (Moll et al., 2005, see also Kurzban & Leary, 2001 for the relation of disgust to social stigmatisation). The final concern of ingroup-loyalty echoes the fundamental group-based biases extolled throughout much of social psychology, and reflects coalition-based evolutionary concerns which likely emerged in the ancestral context of intergroup conflict (Miller, Zielaskowski, Maner & Plant, 2012; McDonald, Navarrette & Van Vugt, 2012). When studying political views, Haidt and associates have successfully employed a sixth moral foundation, *liberty*, which unites elements of harm, fairness and authority to best describe many intuitions relating sparing vulnerable parties the oppression of authorities on a large scale (see Haidt, 2012, chapter 7, for an overview). Liberty, thus construed, has limited applicability outside of the political context, as its primary utility appears to be as a means to disambiguate the loadings of major political philosophies (i.e., conservatism, liberalism and libertarianism) onto the other foundations (for example, the intuitions of fairness can equally support seeking equality of treatment and equality of

outcomes, which map onto the opposing philosophies of conservatism and liberalism respectively).

Evolved Intuitions and Patterns of Prejudice

As Haidt & Graham (2007) explore, there is a cohesive adaptationist case to be made for each of the 5 intuitive domains, though both the literature on moral judgment, and the promoted social standards of many western industrialised nations generally de-emphasise the latter three. In fact, their analysis indicates that within western nations, a significant difference can be identified between politically liberal and politically conservative individuals, wherein conservatives tend to value all 5 moral intuitions as similarly important considerations (with harm and fairness concerns as the consistent top-two), in sharp contrast to liberals who regard harm and fairness concerns highly, but place little-to-no value in authority, purity, or ingroup-membership (Haidt & Graham, 2007).

Considering that political conservatism is often a reliable predictor of prejudicial beliefs (see Roets & Van Hiel, 2011, for an overview), it comes as no surprise that the social and evolutionary literature on various forms of societal discrimination has identified several trends connecting prejudice to specific evolved intuitions (Haidt & Graham, 2007). For example, evidence from evolutionary and comparative psychology suggests that most (if not all) mammals can make essential computational use of details signalling the sex and life-history/developmental stage of their peers (Kurzban & Leary, 2001). This fits appropriately with evidence from social psychology suggesting that the apparent gender and age of other people are intuitively fundamental characteristics encoded immediately and automatically

upon exposure to new individuals (Brewer, 1988; Messick & Mackie, 1989; Stangor, Lynch, Duan & Glass, 1992). The seeming intractability of such intuitive distinctions goes some way towards understanding the ease with which age- and gender-based discriminations seem to form in most human populations (Fiske, Cuddy, Glick & Xu, 2002). Similarly, the psychologically ‘essential’ nature of the intuitive gender-binary offers some insight into the characteristic rejections and discomforts associated with prejudice against persons who identify as transgender or intersex (Clements-Nolle, Marx & Katz, 2006), in addition to prejudice directed towards homosexuals (Haslam & Levy, 2006, see Drescher & Merlino, 2007, for further discussion). Furthermore, the aforementioned intuition concerning purity, and its wider correlates in the disgust literature, also inform some of the distinctive characteristics of prejudice towards persons who are chronically ill (Scambler, 2004; Paterson & Hopwood, 2010), physically-disabled (Rohmer & Louvet, 2012), suffer from mental illness (Kendell, 2004; see Corrigan & Wassel, 2008, for further detail), and perhaps even the obese (Lieberman, Tybur & Latner, 2012). Beyond direct intuitive effects such as these, as the ingroup-loyalty intuition suggests, many forms of prejudice may be exacerbated by the general outgrouping of dissimilar others, in response to even minor perceived conflicts of interest between demographic groups (Kurzban & Leary, 2001; Kurzban, Tooby & Cosmides, 2001; Cosmides, Tooby, & Kurzban, 2003).

The Evolution of Outgroup Categorisation

As the social-intuitionist theory and a wealth of diverse social psychology findings suggest, the tendency to divide the social world into favoured ingroups, and a range of disfavoured outgroups, appears deeply engrained in cognitive architecture of humans and other social mammals (Keeley, 1996; see Smuts et al, 1987, for non-human examples). Perhaps more so

than any of the other above-discussed social intuitions, the inclination to treat ingroup- and outgroup-members differentially is crucial to understanding both the appearance, and the conspicuous absences, of human empathy in various situations. Much of the literature on the behavioural presentation and neurological correlates of empathetic cognition has focussed on its reliable automaticity (Hoffman, 2000). However, as de Vignemont & Singer (2006) discuss at length, empathetic feeling does not reliably emerge under all social contexts, and the patterns of its emergence or absence are far from random. Several recent studies (Singer et al, 2006; de Vignemont & Singer, 2006; Lamm et al, 2007) have indicated that even on the level on neural activation, empathetic engagement with a suffering confederate is moderated by details of social context, most notably, the degree of empathy experienced by participants towards a confederate is drastically reduced by diminished perceptions of fairness and personal affiliation, either as sympathetic concern, or via simple emotion contagion. While the mechanisms underlying the human experience of empathy are regarded as species-typical (de Waal, 2008), part of their design appears to include a means of ensuring that empathetic feeling is not directed towards inappropriate targets.

Unlike some other social animals, anthropological and comparative psychological evidence suggests that our evolutionary ancestors did not regard all other hominids they encountered as peers, but rather, that they engaged in considerable (though perhaps not ubiquitous) intergroup conflict (see McDonald, Navarrete, & van Vugt, 2012, for a recent review). When some members of one's species are treasured allies, and others are potentially dangerous rivals, there is a significant evolutionary onus on being capable of identifying worthwhile allies, distinguishing them from rivals, and reliably tracking which of the two categories any given person is known (or likely) to fall into (Cosmides, Tooby, & Kurzban, 2003). In their analysis of the phenomenon of racial encoding, Kurzban, Tooby and Cosmides (2001)

identified that it is highly unlikely that humans evolved a set of adaptations specifically to notice, track, and attribute expectations on the basis of racial characteristics. The reason for this is that what we now understand as ‘racial’ differences between groups of people are the result of extended periods of breeding isolation (Cosmides, Tooby & Kurzban, 2003). This is generally a product of geographic distance, and as such, during a time in human history when the primary mode of travel was on foot, it is highly unlikely that our ancestors could encounter sufficiently isolated populations as to display racial differences, with sufficient frequency that some benefit may come from the ability to distinguish race easily (Cosmides, Tooby, & Kurzban, 2003). Rather, Kurzban, Tooby and Cosmides (2001) found strong evidence to suggest that humans are equipped with a refined means of tracking coalitional-alliances and ally-groups, and that the general prevalence of racial encoding is a byproduct of these coalition-tracking mechanisms, when participants live in historical circumstances that suggest that race-features may predict social allegiance. In line with their predictions, racial encoding was significantly reduced when superior cues of allegiance were available, in sharp contrast to other intuitively fundamental categories (such as age and sex), which could not be influenced in the same manner.

The coalition-tracking mechanisms inferred by Kurzban, Tooby and Cosmides (2001) serve as an ideal model for the general process of distinguishing ingroup and outgroup. In line with the ingroup-loyalty intuitions of the social-intuitionist model (Haidt, 2007), and fully compatible with Kurzban and Leary’s (2001) analysis of the motivations of social exclusion and stigmatisation, the coalition-tracking account of outgroup categorization specifies that assessments of ingroup- and outgroup-membership should not only sort individuals with whom one has experience, but attempt to generalise predictions about group-membership and personal characteristics to strangers as well. While our hunter-gatherer ancestors were

thought to cohere mostly to groups of kin and close-friends, early humans likely clustered around resources in temperate areas, necessitating cooperative behaviour and favourable relationships with large groups of people, some who will be seen rarely, and others who are implicit ingroup members, but with whom one may have never shared direct contact (Cohen, 2012). In such circumstances, easily identifiable cultural markers that signify group membership would be invaluable, and there is reason to suspect that much of the outgroup-driven prejudice and discrimination experienced today is due to the sensitivity of these cognitive mechanisms to false-positives generated by benign differences of background and circumstance (Van Bavel & Cunningham, 2009; see also Cohen, 2012, for a discussion of spoken accent as just such a signal). Moreover, this coalition-tracking mechanism of ingroup and outgroup categorisations corresponds to the competitive, zero-sum fitness pressures necessary to inform strategic empathy-modulation on the basis of meaningful social appraisals (see Kurzban & Leary, 2001, for an in-depth discussion of adaptive motivations in intergroup conflict). These mechanisms of coalition-management allow for adaptive empathy-modulation in environments where both ingroup cohesion and outgroup mistrust or hostility afford a competitive, selective advantage.

Measure Development

The SATEST measurement tool was designed to target the evolutionary conception of empathy and sympathetic concern outlined above, with specific sensitivity to the aforementioned modulation of empathetic engagement on the basis of coalitional outgroup categorisation. In this regard, the authors sought to establish a testing methodology that could be readily adapted to investigations of sympathetic concern and empathy-motivated helping behaviours from both moral judgment and intergroup social psychology perspectives.

Specifically, the model that the SATEST operationalises, and thus ultimately tests, predicts that empathetic responses to the apparent plight of others will be reliably modulated by cues taken to signal membership in coalitional groups. By measuring expressions of sympathetic concern, while controlling for extraneous factors that can influence responding, one can compute a profile of a subject's typical empathetic reaction to presumed ingroup members. In addition to serving as a measure of individual differences in a morally-relevant domain, this response pattern can be manipulated through the presentation of presumed outgroup members in similar plights, generating difference scores directly indicative of empathy-modulation triggered by coalitional cues (as in Study 3), which this model theorises to be an influential affective component in modern prejudice.

As such, the development of the SATEST has called for a balancing act between measurement efficacy and ease of administration, so as to address the lack of appropriate tools in the literature that are broad-purpose, as opposed to designed-from-scratch to test a single effect (see Christensen & Gomila, 2012, p.1251, for a discussion of the difficulties in comparing dilemma-based measures). The SATEST methodology is therefore designed to rely only on standard and widely available computer hardware and software (most browser applications), with a manageable administration time (approximately 15 minutes) and secure options for online testing. The following sections outline the behavioural indicators selected to signal sympathetic concern, the design decisions intended to control for impression management, affect-elicitation, and ecological realism, and finally the variables of influence controlled and manipulated in the writing of each SATEST scenario.

Behavioural Expressions of Sympathetic Concern

Like many measurement tools employed in the study of social and particularly moral decision-making (Christensen & Gomila, 2012), the SATEST relies upon the presentation of scenario vignettes which frame a hypothetical situation for the participant, in order to pose questions concerning the evaluations, attitudes, and prospective courses of action that this situation would elicit if true. The challenge in developing this kind of task comes not only from framing a situation that appropriately elicits the reactions under investigation. Questions and response options that will be interpreted consistently by multiple participants also need consideration (Christensen & Gomila, 2012), and which will prompt overt behavioural reactions that may be considered viable correlates with the hypothesised processes of interest (Borsboom, Mellenberg & Van Heerden, 2003; 2004; see also Markus & Borsboom, 2011, for a review of this kind of item response approach). This challenge is magnified in the attempts to measure feelings and motivational states, as without the use of costly neuroimaging equipment one must make greater inferences concerning how reliably any given responses signals the presence or magnitude of the proposed state (Markus & Borsboom, 2011). Since a tool such as the SATEST must rely upon the overt response choices of participants, it cannot hope to realistically capture direct variance in a participant's true subjective experience of sympathetic concern. With this in mind, the design strategy of the SATEST instead seeks to measure three probable behavioural manifestations of experiencing sympathetic concern (as indicated by the existing literature), within the context of presented scenarios specially constructed to control and minimise alternative potential influences on the responses in question.

The first indicator used was *helping behaviour*, defined here as committing to a course of action within the scenario which will result in a sympathy-target experiencing an appreciable reduction in distress. Helping, in this sense, is the response most extensively studied in literature employing social and moral dilemmas (Greene & Haidt, 2002), but is likely also the response most easily influenced by alternative motivations, including simple impression-management (Christensen & Gomila, 2012).

The second indicator used, as explored extensively in the moral reasoning (Rest et al., 1999) and moral ‘dumbfounding’ literature (Haidt, 2001), was *personal vs. deontological justification* for one’s actions (see Kurzban, DeScioli & Fein, 2012). Specifically, when asked to account for why a preceding helping decision was or was not made, a justification may be described as ‘personal’ when it identifies motivating factors oriented towards interpersonal feeling and affect, particularly concerning social closeness or engagement (aligned with the inclusion conceptions described in Haidt, 2007). Conversely, a justification may be described as ‘deontological’ when the motivations expressed concern adherence to social rules, principles or standards, regarded independently from the personal circumstances of those involved (Greene et al., 2009). Research into the character of justifications offered for social choices (Haidt, 2001), particularly those related to harm in the context of punishment (Haidt & Graham, 2007), suggest that the level of empathetic sympathy experienced by the participant reliably predicts personal-domain justifications, whereas rule-based judgments remain distinctly dispassionate (Greene & Paxton, 2009; Paxton & Greene, 2010; Paxton, Ungar & Greene, 2011).

The final behavioural indicator employed by the SATEST is the *attribution of the target's motivation*. Drawing upon the distinctions first outlined in the study of fundamental attribution error (Ross, 1977), the participant is asked to ascribe a likely motivation for an ambiguous undesirable act performed by the scenario's sympathy-target. Under such circumstances, 'positive' attributions (those which suggest exonerating details, temporary or coerced reasons for poor behaviour; see Linke, 2012) are identified by some moral judgment researchers (Krebs & Van Hesteren, 1994; Krebs & Denton, 1997; Krebs & Janicki, 2004; Nichols & Knobe, 2007) as signifying both empathetic sympathy, and self-similarity (which corresponds to the social-intuitionist conceptions of ingroup, Haidt & Graham, 2007; Haidt, 2007). Conversely, negative attributions (emphasising global and stable self-determination of poor behaviour; Linke, 2012) demonstrate the opposite effect, predicting both unsympathetic regard and a sense of social distance (Miller, Zielaskowski, Maner & Plant, 2012). Through these three indicators, SATEST results are afforded a high degree of nuance in profiling the likelihood that a participant has experienced sympathetic concern towards the targeted character within each scenario, provided that other sources of influence on these responses (identified below) are sufficiently anticipated and controlled.

Impression-Management

A key concern for any measure that relies upon social dilemmas is the possibility that participants will intentionally bias their responses in an attempt to present themselves in a more socially desirable light to the researcher (Richman et al, 1999). This problem is especially pronounced in the moral judgment and prejudice literatures (Christensen & Gomila, 2012), where the ease with which a participant may identify the socially 'wrong' answers makes explicit response data notoriously unreliable (Greenwald, McGhee, &

Schwarz, 1998). Rather than relying on the inclusion of items intended to identify self-enhancing response tendencies (as in Rest, 1975), the SATEST methodology is instead designed to exploit a recurring conflict identified in the literature between participants' perceptions of harm and fairness (as identified in Haidt & Graham, 2007). The basic design of a moral dilemma is to present the participant with a situation that expresses a conflict between two values that they are presumed to hold (Christensen & Gomila, 2012). In traditional cognitivist moral dilemmas, this was most often a conflict between the personal interests of a protagonist character, and a 'right' course of alternative action as dictated by deontological conceptions of fairness (Rest et al., 1999). However, as both Haidt (2001) and Greene (2001) explored, dilemmatic conflicts between rival social and moral intuitions can elicit a sense of equivalence in the participant, reducing or removing the overt conception that one course of action is the 'right' one (Greene & Haidt, 2002).

To capitalise on this, the design of each SATEST scenario specifically frames an equivalence-promoting conflict between the most primary moral intuitions of harm and fairness (Haidt, 2007), by presenting a scene in which the sympathy-target character is in personal distress, due to the impending comeuppance of having violated a social rule. The moral intuition of fairness interacts with, and is strengthened by, assessments of harm and other intuitions (such as authority and purity), but is none-the-less elicited by the violation of any explicit social rule (Haidt & Graham, 2007). To this end, the SATEST presents the participant with situations in which a target is due to experience considerable, but socially-endorsed, retribution for violating a social rule, but where the participant is in a position to spare the target their punishment. Through this contrivance, the SATEST provides the participant with an option to act upon their sense of sympathetic concern towards the target, experiencing only minor conflict over the rule-violation (as each transgression is a typical 'victimless crime'). However, should the participant not feel sympathetically motivated to

help the target, they are granted a socially justifiable reason for letting them suffer some adverse consequences (a viable social strategy endorsed by Nowak & Sigmund's (2005) analysis of motivated reputation-management). By presenting the participant with no definitively 'wrong' alternative, but providing more intense cues for the activation of the harm rather than fairness intuitions, the SATEST should be able to significantly reduce the instance of wilful impression-management that might otherwise inflate the behavioural indicators of sympathetic concern.

Facial Affect Signalling and Ecological Validity

A near-universal limitation of traditional moral dilemmas is their reliance on purely text-supported vignettes (Christensen & Gomila, 2012). While most participants respond well to written narratives, variations in the verbal comprehension (Sanders, Lubinski & Benbow, 1995) and visual imaginations (Amit & Greene, 2012) of participants introduce a range of potentially problematic inconsistencies. Of the highest importance to dilemmas focusing on the social intuition of harm (Haidt, 2007), is the role that facial affect perception plays in the real world elicitation of sympathetic concern (Iacoboni, 2005; Cannon, Schnall & White, 2011). As the analysis performed by Nichols and Knobe (2007) suggests, the more abstract and emotionally aloof a written vignette is, the less a participant's relevant moral intuitions are likely to activate, whereas, conversely, dilemmas that rely upon emotive language unduly skew participants' reactions in the opposite direction. Beyond this, the explicit wording of written vignettes can conspicuously signal the variables of interest (such as race) to participant, inflating the chance that they will perceive the implied 'wrong' valence of possible answers and engage in impression-management (Christensen & Gomila, 2012).

To address these issues, the SATEST presents the participant with a transitioning first-person perspective of the simulated social scenario, which provides appropriate visual cues that accompany the real-time narrated text. Although researchers have recently found some promising social effects through the use of fully-immersive virtual reality simulations (Wilcox et al., 2006; Gillath, McCall, Shaver & Blascovich, 2008; Llobera et al., 2010), such simulations offer limited situational context and response possibilities relative to written vignettes (Slater et al., 2013). As such the SATEST has been designed with a hybrid approach, wherein the participant's on-screen perspective is semi-immersive, guided through a simulated social interaction by both the first-person visual presentations, and a vignette narrative that is both visually presented at the bottom of the screen, and narrated in real-time via voice-acting. In each scenario, the participant views a situation from a controlled perspective while interacting with a visually and audibly represented 'friend' character (akin to the programmed cohorts in many virtual reality studies; Levine et al., 2002; Levine et al., 2005), and all questions and answers are as presented as conversational with this 'friend'. This presentation, supplemented by the visual cues provided, serves to not only ground the material details of the dilemma without excessive writing, but allows for elements of the scenario to be introduced subtly, where they may be noticed by the participant without the text drawing their overt attention (e.g., racial characteristics, age, apparent illness, etc.).

Each SATEST scenario is therefore capable of presenting the sympathy-target on-screen, where their facial-affect (expressing distress, by default) is clearly visible to the participant, without their emotional state needing to be inferred or explicitly described in the text. This allows for a more ecologically valid elicitation of the participants' empathetic reactions towards the target, without signalling to the participant that measuring their sympathy towards the target is a goal of the study. Furthermore (as was explored below in Study 3), the

visual characteristics of the target characters may be varied subtly, without explicitly signalling to the participant that the manipulation of some feature (such as race) is a part of the study. While present versions of the SATEST have relied exclusively on illustrations to provide the visual cues within the studies, care was taken to ensure that all facial affect representations prominently featured the key details (most notably brow-contour and mouth shape) that humans rely upon to interpret emotional presentations on abstract, illustrated, or animated faces (Stevens, Charman & Blair, 2001; Carr & Lutjemeier, 2005; Creed & Beale, 2008; Chen, Russell & Nakayama, 2010).

As part of the ecological considerations concerning the visual cues and conversational format of the interface, each SATEST scenario must balance two concerns previously identified in the literature between participant's rejection of non-relatable situations (Casebeer, 2003), and the potentially biasing influence of well-established response patterns for choices participants have faced before (Borg et al., 2006). To address this, the initial twelve SATEST scenarios are set in environments and situations that are familiar and commonplace in western industrialised countries (such as offices, suburban neighbourhoods, and cinemas; see Table 1 for the complete list), but present the participants with decisions that, while plausible, are distinctly uncommon (such as being in a position to intervene on the issuing of a parking ticket). These scenarios are a far cry from the life-or-death situations described in many moral dilemmas, specifically because the SATEST aims to elicit the patterns and degrees of sympathetic concern that shape the majority of social interactions in a participant's life (see Greene & Haidt, 2002, for a discussion of common versus extreme social reactions).

Table 1*List of the 12 SATEST Scenarios with Situational Variables*

Target Is About To Be Caught...	Helping Target Requires...	Intervention Performed By...
Parking in a restricted space	Action	Friend
Using restricted work emails for personal chats	Inaction	Participant
Sneaking contraband food into a cinema	Action	Participant
Borrowing library books despite outstanding fees	Inaction	Friend
Bringing a small pet into a no-pets-allowed building	Inaction	Participant
Overloading garbage into inappropriate bins	Action	Participant
Misusing a department-store employee discount	Inaction	Friend
Taking stationary in an office without permission	Action	Friend
Entering an executive break-room without permission	Inaction	Participant
Exceeding lawn water use during a restricted period	Action	Participant
Sneaking into a drive-in movie without paying	Inaction	Friend

Commonly Neglected Methodological Variables

Beyond the major issues described above, in initially developing the SATEST the authors sought to control for as many confounding variables as possible. Recently, Christensen and Gomila (2012) compiled an extensive review of the use of moral dilemmas in moral and social psychology, and identified 19 distinct variables that have been empirically demonstrated to influence participants' responses, but are rarely fully acknowledged and controlled by researchers. Employing Christensen and Gomila's three primary categories (*dilemma formulation, participant characteristics/related of characters, and dilemma conceptualisation*), what follows is a brief explanation of each variable and how it has been addressed, controlled, or manipulated in designing the SATEST.

Within dilemma formation, Christensen and Gomila (2012) identify that decisions of *presentation format* between pen-and-paper and computerised presentation affect some aspects of participant response-consistency. In line with their recommendations, the SATEST is computerised, and proceeds through screen-presentations with controlled time-limits, which regulate when and for how-long participants are exposed to each question. Consistent with their recommendations concerning *expression style*, *word-framing*, and *word number count*, each SATEST scenario distinctly avoids the use of emotive or strongly quantifying words, and maintains a consistently short word count (can be read from start-to-finish in less than 90 seconds). The controlled first-person presentation of the SATEST aligns with their suggestions for *participant perspective*, as does the consistent unveiling of details between SATEST scenarios for *situational antecedent* and *order of presentation*. *Type of question* is also fully standardised between scenarios, and as per Christensen and Gomila's recommendation, the SATEST permits participants to express *justifications*.

With regards to participant characteristics and those of related characters, all three studies below detail the *demographic details of participants*, as recommended, and standardise character features so as to control *ingroup/outgroup* influences, except in Study 3 where this is specifically manipulated. Christensen and Gomila (2012) also suggest controlling for the predictable influences of *kinship/friendship* and *speciesism*, which the SATEST addresses by depicting all sympathy-targets as humans that are strangers to the participant.

Lastly, concerning dilemma conceptualisation, Christensen and Gomila (2012) advise that the *intentionality* behind any assessed acts is of key concern to participants. The intentions of the transgressing targets is standardised as being ambiguous in the SATEST, and attribution of

their intention is operationalised as a measurement variable. As per recommendations, the *kind of transgression* presented between scenarios are all violations of social rules which pose no immediate or distant harm to any victim, but incur a reasonable penalty for the sake of deterrence (see Table 1). As such, the *directedness of harm* is standardised (as none), which also controls for the influence of the harm's *trade-off* and the *normality of harm*. Lastly, beyond the intentional ambiguity of the target's motivation for social transgression, the dialogue with the friend character is standardised so as to assure the participant that the apparent outcome of either decision is essentially guaranteed, satisfying Christensen and Gomila's pragmatic advice concerning the *certainty of events*.

Overview of Present Studies

A series of studies were designed to explore both the internal psychometric properties and construct validity of the SATEST methodology. Short of neuroimaging correlations with the brain structures identified in previous research (see Decety, Norman, Berntson & Cacioppo, 2012, for an overview), evidence to support the SATEST's efficacy in targeting key behavioural expressions of sympathetic concern was provided by the discovery of theoretically appropriate patterns of interrelations with related measurement tools. Due to the aforementioned lack of broad-purpose measurement methodologies specifically designed to track signs of sympathetic concern in commonplace circumstances, predictions of construct convergence and divergence were divided between the two primary research goals of empathy studies. The first goal being the measurement of individual variance in moral judgment style, as is explored in Studies 1 and 2 (see Bartels, 2008, and Christensen & Gomila, 2012, for overviews of this literature), and the second goal of measuring the

variations in empathetic motivation underlying group-based discriminations, as is explored in Study 3 (see Cikara, Bruneau & Saxe, 2011).

Study 1

The first study, following the initial development of the initial 12 scenarios and interface tool, sought primarily to investigate the psychometric properties of the SATEST with regard to factorial loadings of the three anticipated behavioural indicators of sympathetic concern, in addition to a fourth behavioural indicator intended to control for individual variations in activity and social engagement. In addition to this, SATEST was compared to two popular measurement tools from the moral judgment literature, the classic Foot (1967) variant of the Trolley Dilemma, and the cognitive-developmentally based Defining Issues Test (DIT, original version; Rest, 1975), widely employed in study of moral reasoning. The Trolley Dilemma was selected for its key use in the study of utilitarian reasoning, and its well-established vulnerability to the participant's disposition towards activity in hypothetical scenarios (Cikara, Farnsworth, Harris & Fiske, 2010). The DIT was selected for its robust history of use, and clear relevance to the construct of deontological reasoning in social decisions (see Rest et al, 1999, for an overview).

It was predicted that the behaviour scores generated by the SATEST scenarios would cluster into 4 largely distinct factors, corresponding with the participant's disposition towards activity, and the three aforementioned indicators. Scores indicating activity and inactivity irrespective of altruistic outcome are expected to align as positive and negative valences of one factor. *Helping behaviour* would draw from the valence and expressed enthusiasm of

participants' initial decision in each scenario. *Deontological justifications* would align with the rule-focused explanations for one's decision, in addition to the number of slow considerations (those taking longer than 6 seconds, and thus more likely to rely upon deliberate conscious reasoning; see Greene, 2013), whereas those justifications oriented around pro-social and dismissive affects will align with activity and attributions, respectively. Finally, expressed positive and negative *attributions* would predictably align in opposing directions (see Table 2 for a summary of the behavioural indicators drawn from the SATEST scenarios). In the interest of simplicity, each behavioural indicator was scored additively, with each decision adding 1 point to the relevant index (e.g. selecting a 'Rightness or wrongness of act' option adds 1 point to one's Deontological Index), with opposing indicators (e.g. 1 in Positive Attributions and 1 in Negative Attributions) mitigating each other when the final Indices are calculated (as when subtracting Negative Attributions from Positive Attributions to obtain the final Attributions Index). The Activity and Helping Indices also incorporated the magnitudes indicated by the participant's initial decision, with 'Guess so' worded choices contributing 1 point, 'Probably should' worded choices contributing 2 points, and 'Definitely' worded choices contributing 3 points.

Table 2*List of types of SATEST Considerations, Justification and Attribution Options*

Participant Indicates...	Personal (+)	Personal (-)	Deont	Attr (+)	Attr (-)
Considerations					
Rules of this situation			+		
Rightness or wrongness of act			+		
Target's position/feelings	+				
Justifications					
Felt like being nice	+				
Target would be grateful	+				
Target won't do so again	+			+	
Person out-values the rules	+				
Didn't feel like helping		+			
I have limited responsibilities			+		
Target requires punishment		+	+		
Rules are more important		+	+		
Attributions					
Target probably inconsiderate					+
Target may be unaware of rules				+	
Target doesn't care about rules					+
Target may have good excuse				+	
Target is 'that kind of person'					+

Due to the noted susceptibility of the Trolley Dilemma to participants' action-orientations and dispositions (Cikara, Farnsworth, Harris & Fiske, 2010), it was hypothesised that participant decisions to throw the switch and exchange the life of one bystander to save 5 others, would show a significant, but weak, positive correlation with the SATEST's activity index. As for the DIT, the literature has established that pro-social sympathies are measured to some degree in its computed P-Score, but that the majority of influence on P-Score derives

from evidence of principled and dispassionate moral reasoning (see Crowson & DeBacker, 2008, for a critical review). As such, it was hypothesised that DIT P-Score would demonstrate a significant but weak positive correlation with the SATEST's helping-behaviour index, but would show a significant moderate-to-strong positive correlation with the SATEST's deontological-reasoning index. Both DIT relations were predicted to remain significant, even controlling for participant's activity index level, and participant age (also known to be a P-score correlate; Rest et al, 1999).

Method

Participants and procedure. Of an initial two-hundred and thirty-two undergraduate psychology students, recruited from a university subject-pool in exchange for psychology course-credit, eight were discounted from the study due to unacceptable scores in false-response m-scores of the DIT (as per the directions of the measure). The remaining two-hundred and twenty-six participants (156 female, 68 male) completed the three measures via a secure browser-based computer interface, constructed in *Adobe Flash*, on a PC of their choosing. The interface was designed to detect sufficient processing speed in the host machine, and participants were instructed that completion of the study would only be possible on a machine possessing an active internet connection, a keyboard and mouse, a colour monitor, and either speakers or headphones for sound output. Participant age was typical for an undergraduate sample ($M = 20.05$ years, $SD = 4.17$), with a minimum age of 18 and a maximum age of 50.

Measures. The study interface presented the three subsections to the participants in a randomised order. The Trolley Dilemma was presented in its standard textual format (Foot, 1967), and participants indicated both their binary decision to throw or not throw the switch, in addition to two 7-point Likert-scale items asking they indicate how ‘appropriate’ they believed either course of action to be (ranging from *Very Appropriate* to *Very Inappropriate*).

The Defining Issues Test (original version, Rest, 1975) was converted to an online format, preserving the original paragraph structure and layout of response-boxes on each page-view. The wording of three of the vignettes was modified slightly so as to better reflect the expectations of modern participants (for example, the descriptor ‘Oriental’ was changed to ‘Asian’; references to the ‘recent Vietnam War’ were replaced with the more contemporary Iraq War, etc.). The percentage P-score (indicating ‘Post-Conventional Moral Reasoning’) was computed by the interface in accordance with the specifications of the scoring manual, in addition to the percentage M-Score, which indicated non-serious response tendencies in the participants that were used to exclude the data of eight participants. Participants were presented with 6 vignettes, in random order, followed by questions which included a list of principled considerations concerning the morally right or wrong nature of the dilemma. The task involved ranking the four most important considerations, in addition to giving Likert-style responses concerning the significance of each consideration. An algorithm was also employed to locate response patterns which demonstrated inconsistencies between the rankings of items and their associated Likert-scores, as per the scoring instructions, but no participants with sufficiently low M-Scores presented sufficient discrepancies in their scores to warrant additional exclusions.

The 12 scenarios of the Sympathetic Attributions Toward Emotive Social Transgressors (SATEST) task were presented in random order, with their original illustrations and voice-acting tracks. After each situation was framed to the participant, the friend-character would conversationally prompt the participant as to whether or not they wished to assist the sympathy-target presented, or whether they wished to see them receive comeuppance. The 12 scenarios were counterbalanced, so as to systematically vary whether helping the sympathy-target would demand action or inaction on the part of the participant, and whether the act itself would be performed by the participant directly, or through the friend-character as a proxy. Participants have a third option to request more time from their friend to consider their course of action, during which time they may elect one or more of several optional considerations, before returning to the decision to either help or not help. Those who take more than 6 seconds to respond are interpreted by the interface as already engaging in considerations (as supported by the findings of Greene et al., 2001; Greene & Haidt, 2002), and are conversationally moved into the consideration options automatically. The decision is initially presented as 3 options, 'Yes', 'No', and 'Let me think', but clicking either the yes or no options expands the presented buttons, to give the participant 3 levels on which to express their commitment, ranging from slight to definite. This was done to minimise the mandatory reading time of the participant, given the SATEST's timed presentation. While selecting either 'yes' or 'no' resets the count-down, participants can still press any of the initial 3 buttons before making up their mind. As such, while initially presented as a binary choice with a third 'more time' option, the initial decision is encoded closer to a 6-point Likert-scale, ranging from 'no, definitely' to 'yes, definitely'. Following their initial decision, the friend-character asks the participant to provide a justification for why they made the decision they did (from 4 multiple-choice options matched to the valance of the decision), followed by a final question asking the participant to attribute why they

thought the target committed their social transgression in the first place (from 5 multiple-choice options matched to the valence of the initial decision). It should be noted that this pilot testing of the SATEST measure was not designed to invite back a subset of the participants for follow-up testing at a later date, and as such could not provide test-retest data that can be compared to the corresponding data in Studies 2 and 3, both of which examine test-retest reliability directly.

Analyses. To investigate the psychometric properties of the SATEST, principle components factor-analysis was performed on the behavioural output scores, to confirm their adherence to the 4-factored model predicted by the above-discussed theory (oblique, rather than orthogonal, rotation was selected, as factors were expected to correlate given the nature of the items). The internal consistency of the measure was also examined via Cronbach's alpha test of interrelatedness. The hypothesised relations between the three measures were investigated first by Bonferroni-adjusted Pearson's correlation coefficients, followed by a hierarchical multiple regression analysis of the predictive relations between the SATEST variables and the DIT P-Score, so as to control for the influence of participant activity-level on the two construct-relevant relationships.

Table 3*Factor Loadings of SATEST Behavioural Indicators in Studies 1 and 2*

Indicator	Study 1				Study 2			
	Action	Help	Deont	Attrib	Action	Help	Deont	Attrib
All Choices to Act	.68	-.21	-.02	.20	.77	-.27	-.01	.16
All Choices to Not Act	-.79	.15	.09	-.24	-.85	.19	.14	-.17
Choices to Help								
Self + Action	.21	.82	.08	.09	.26	.90	.05	.03
Self + Inaction	.15	.84	.07	.06	.11	.87	.03	.02
Other + Action	.28	.78	.02	.04	.25	.89	.05	.05
Other + Inaction	.22	.74	.05	.05	.14	.83	.05	.07
Choices to Not Help								
Self + Action	-.15	-.88	.07	-.08	-.17	-.95	.05	-.05
Self + Inaction	-.03	-.87	.06	-.06	-.09	-.89	.05	-.08
Other + Action	-.12	-.85	.01	-.09	-.19	-.96	.07	-.03
Other + Inaction	-.01	-.73	.19	-.04	-.06	-.93	.08	-.05
Pos. Personal Justifications								
Self + Action	.79	-.11	.07	-.07	.82	-.13	.01	-.09
Self + Inaction	.67	-.17	.02	-.01	.79	-.19	.01	-.07
Other + Action	.74	-.12	.04	-.16	.74	-.15	.02	-.11
Other + Inaction	.67	-.03	.05	-.06	.77	-.12	.01	-.09
Neg. Personal Justifications								
Self + Action	-.11	.25	-.25	-.76	-.15	.22	-.23	-.81
Self + Inaction	-.14	.21	-.23	-.72	-.13	.24	-.28	-.78
Other + Action	-.12	.16	-.27	-.78	-.17	.19	-.26	-.83
Other + Inaction	-.20	.19	-.25	-.69	-.15	.21	-.25	-.76
Deontological Justifications								
Self + Action	-.19	-.10	.77	.13	-.24	-.06	.87	.09
Self + Inaction	-.22	-.04	.73	.16	-.21	-.07	.90	.07
Other + Action	-.14	-.06	.81	.07	-.18	-.01	.83	.04

Other + Inaction	-.17	-.12	.75	.04	-.27	-.03	.85	.07
Positive Attributions								
Self + Action	-.03	.15	.07	.78	-.03	.17	.05	.87
Self + Inaction	-.08	.12	.06	.74	-.05	.15	.03	.83
Other + Action	-.19	.15	.05	.69	-.07	.16	.05	.89
Other + Inaction	-.09	.12	.12	.78	-.01	.12	.06	.86
Negative Attributions								
Self + Action	-.18	-.16	-.06	-.89	-.11	-.09	-.05	-.96
Self + Inaction	-.12	-.18	-.05	-.75	-.13	-.10	-.07	-.92
Other + Action	-.23	-.11	-.05	-.83	-.08	-.12	-.05	-.94
Other + Inaction	-.24	-.15	-.04	-.75	-.04	-.07	-.05	-.89
Considerations								
Self + Action	.16	.15	.77	.18	.18	.05	.88	.13
Self + Inaction	.18	.16	.68	.16	.13	.02	.85	.17
Other + Action	.13	.22	.73	.17	.11	.06	.91	.12
Other + Inaction	.11	.09	.62	.13	.17	.08	.87	.15

Note: Study 1 $N = 226$, Study 2 $N = 248$; rotated to oblimin criteria ($\delta=0$): loadings $> .40$ in bold type.

Table 4*Pearson Correlations Between SATEST Indices, DIT P-Score, and Trolley Dilemma Scores*

	DIT P-Score	Trolley Approval	Trolley Disapproval
SATEST Activity Index	-.028	.202**	-.126*
SATEST Helping Index	.151**	.106	-.028
SATEST Deontology Index	.314***	-.027	.018
SATEST Attribution Index	.029	.038	.007
DIT P-Score	-	-.094	.035

Note: $N = 226$ in all samples; Bonferroni correction for these comparisons is $\alpha = .05/6 = .008$.

* $p < .05$; ** $p < 0.008$; *** $p < 0.001$

Table 5*Hierarchical Multiple Regression Predicting DIT P-Score*

Step	Predictors	β	ΔR	ΔR^2_{Adj}
1	Age	.211***	.227***	.045
	Gender	-.085		
2	SATEST Activity Index	-.014	.227	.042
3	SATEST Helping Index	.179**	.370***	.091
	SATEST Deontology Index	.249***		
4	SATEST Attribution Index	-.046	.373	.090
	Trolley Approval	.088		
	Trolley Disapproval	-.020		

Note: $N = 226$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$

Results and Discussion

With the initial decision responses taken as a 6-point Likert-style scale (recoded to orient towards helping or not-helping the sympathy-target), the 12 SATEST scenarios demonstrated an acceptable internal consistency (Cronbach's $\alpha = .743$). Also, as Table 3 summarises, the factor structure revealed by principle components analysis under standard oblimin rotation criteria supports the psychometric veracity of the SATEST in two ways. Firstly, the 10 domains of behavioural indicators adhered to a 4-factor structure resembling the 4 dimensions predicted by the theory, with each correlating in the appropriate positive or negative valance. The selection of this number of factors was supported by the distinct flattening of the associated scree-plots when additional factors were proposed, indicating poor reduction in explanatory eigenvalues for any subsequent factors that could be added (with a cut-off below an eigenvalue of 1, in accordance with the Kaiser-Guttman guideline). Second, the principle components analysis confirms that the behavioural indicators drawn from the 4 combinations of action/inaction and self/other scenario variables perform near-uniformly, with sufficient similarities for them to be computed as common response indices.

The bivariate correlations between the DIT P-Score, the Trolley Likert-responses, and the four SATEST indices demonstrated the theorised relationships predicted by the theory, supporting the convergent validity of the SATEST as a measure of sympathetic concern. Both the Helping and Deontological indices significantly positively correlated with DIT P-Scores, displaying the weak and moderate effect sizes anticipated. Furthermore, participant responses concerning the appropriateness or inappropriateness of throwing the switch significantly correlated positively and negatively, respectively, with the SATEST Activity index, supporting the hypothesised role of the index as a measure of tendency to endorse proactive

behaviour in dilemmas. Lastly, the hierarchical regression analysis provided both convergent and divergent evidence for construct-supported relations between the SATEST and the DIT, by confirming that the predicted relations remained uniquely significant predictors of P-Scores even when controlling for the hypothetical general activity levels of the participants.

Study 2

Following from the results of Study 1, study two sought to replicate the initial convergent and divergent construct relations observed between the SATEST and the DIT, utilising a marginally updated version of the SATEST that employed more visually distinct illustrations (better line and colour quality) and more audible voice-acting (with clearer pronunciation and no recording feedback). A long-standing criticism of cognitivist moral judgment tasks is that they are confounded with cognitive ability. Several critics, most notably Sanders, Lubinski and Benbow (1995), have observed that the lengthy vignettes, complicated scoring, and lingual cues in the framing of the questions, bias the P-Score of the DIT so as to favour participants with superior verbal cognitive abilities. To address this, Study 2 endeavoured to replicate prior evidence for the cognitive ability confound in cognitivist moral judgment tasks (the DIT in particular) by simultaneously employing a set of measures of cognitive ability, while also testing to ensure that the SATEST methodology does not possess similar vulnerabilities.

It was therefore predicted that of the four GFGC ability subscales used in this study (Stankov, 1997, developed on the ‘fluid and crystallised’ ability distinctions defined by Cattell, 1971; 1987, and refined by Horn & Noll, 1994; 1997), the Vocabulary measure would demonstrate

a significant, moderate-to-strong positive correlation with the DIT P-Score, and that this relationship would remain significant after controlling for age and gender, which have demonstrated significant relationships with both variables in other samples (Rest et al., 1999). Conversely, it was predicted that no SATEST indices would demonstrate significant correlations with any GFGC ability subscales (it must be noted, however, that this prediction cannot be fairly assessed via null-hypothesis testing, and is merely intended to specify a condition that, if violated, would strongly contradict the theorised properties of the SATEST). Furthermore, it is hypothesised that the relationship between DIT P-Score and SATEST helping-behaviour index will remain significant after controlling for all GFGC subscales. Beyond this, it was hypothesised that the same properties of factor-loading and internal consistency observed in the Study 1 would be replicated in this updated presentation of the SATEST.

Method

Participants and procedure. Of an initial two-hundred and fifty-five participants who were recruited from a university subject-pool in exchange for psychology course-credit, seven were discounted from the study due to unacceptably high scores on the false-response M-Scores of the DIT (as per the directions of the measure). The remaining two-hundred and forty-eight participants (190 female, 58 male) completed, in order, the SATEST, DIT and GFGC subscales, in a secure browser-based computer interface, constructed in *Adobe Flash*, on a PC of their choosing (subject to the same limitations outlined in Study 1). The age range of participants met the typical characteristics of an undergraduate sample ($M = 20.93$ years, $SD = 5.91$), with a minimum age of 18 and a maximum age of 49.

Measures. Beyond the above-listed enhancements to the visual and audio quality of the SATEST scenarios, both the SATEST and the DIT were presented in a manner identical to conditions described in Study 1. The GFGC measure (Stankov, 1997) was employed four subscales, each dedicated to a discrete cognitive testing ability, namely *Vocabulary*, *Linguistic Associations*, *Esoteric Analogies*, and processing of *Partially-Concealed Words* (for further details, see Stankov, 2000). Each subscale was adapted for presentation on a computer screen, providing separate screen-views for each task, and digitally enforcing the requisite time-limits and pacing cues prescribed in each subscale's directions.

Analyses. In addition to reproducing the psychometric validations (internal consistency and principle components analysis) and correlational analyses outlined in Study 1, the authors performed an additional hierarchical regression analysis in order to re-examine the predictive relationships between the helping-behaviour and deontological-justification indices of the SATEST, and the P-Score of the DIT, by first controlling for the variance explained by the GFGC subscales.

Table 6*Pearson Correlations Between SATEST Indices, DIT P-Score, and GFGC Subscales*

	DIT P-Score	Vocabulary	Associations	Esoteric	Concealed
DIT P-Score	-	.169**	.089	.135*	.009
SATEST					
Activity Index	.034	.089	.090	.086	.056
Helping Index	.174**	.025	.106	.019	.113
Deontology Index	.328***	.083	.044	.090	.078
Attribution Index	.041	.098	.100	.092	.063

Note: $N = 248$ in all samples; Bonferroni correction for these comparisons is $\alpha = .05/6 = .008$.

* $p < .05$; ** $p < 0.008$; *** $p < 0.001$

Table 7*Hierarchical Multiple Regression Predicting DIT P-Score*

Step	Predictors	β	ΔR	ΔR^2_{Adj}
1	Age	.255***	.255***	.061
2	GFGC Vocabulary	.149	.315	.077
	GFGC Associations	-.066		
	GFGC Esoteric Analogies	.156		
	GFGC Concealed Words	-.157		
	SATEST Activity Index	.044		
3	SATEST Helping Index	.238***	.442***	.161
	SATEST Deontology Index	.187***		
4	SATEST Attribution Index	-.046	.442	.158

Note: $N = 248$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$

Results and Discussion

When regarding the initial decision responses of the SATEST as a 6-point Likert-style scale (as in Study 1), the SATEST again demonstrated an acceptable internal consistency (Cronbach's $\alpha = .765$). By virtue of a sub-sample of 43 participants who took the SATEST a second time two weeks after the initial testing, a set of test-retest reliability correlations were conducted for each SATEST index, all of which demonstrated strong effect-sizes (the lowest being $r = .689$, $p < .0001$ for the Attribution Index). Also, as is summarised in Table 3, principle components analysis under standard oblimin rotation criteria once again supported the psychometric properties of the SATEST, as established in Study 1.

The bivariate correlations between the DIT P-Score, GFGC subscales, and the four SATEST indices supported and replicated the theorised relationships observed in Study 1, in addition to demonstrating the verbal ability vulnerabilities attributed to the DIT in the literature (Sanders, Lubinski & Benbow, 1995; Rest et al., 1999), vulnerabilities which were not shared by the SATEST. The hierarchical regression analysis replicated the convergent and divergent evidence of construct validity explored in Study 1, and confirmed that both Helping and Deontological SATEST Indices remain significant predictors of DIT P-Score, even when controlling for the influence of verbal skills and general cognitive ability. Unexpectedly, the regression analyses revealed that the relationships observed between the DIT P-Score and the implicated GFGC subscales did not retain independent predictive significance after controlling for the age of the participant. This suggests that the shared component of these variables is related closely to participant age (the most likely possibilities being some approximation of education level and life experience, as suggested by Sanders, Lubinski & Benbow, 1995). Conversely, while controlling for the influence of age and cognitive ability

reduced the predictive significance of the SATEST Deontological Index, relative to the trends observed in Study 1, both Deontological and Helping Indices remained independently significant predictive variables throughout the analysis. This demonstrates that these SATEST variables not only measure both of the key theorised elements targeted by the DIT (that is, signs of interpersonal compassion, and evidence of contemplative moral reasoning), but do so in a more methodologically robust manner that shows less reliance on indications of participant age and verbal cognitive ability than the DIT.

Study 3

Having established both convergent and divergent evidence for the construct validity of the SATEST with regards to sympathetic concern in the moral judgment domain, Study 3 sought to generalise beyond this to explore the efficacy of theoretically appropriate SATEST indices in predicting patterns of prejudicial and discriminative behaviour between different demographic group-members. To this end, the SATEST was modified so as to selectively manipulate the apparent race of the sympathy-target in half of the presented scenarios (the Skin-Colour Manipulation, or SC-SATEST), so as to compute not only separate indices of the above-studied behaviours, but to compute patterns of differences between the contrasted race-conditions.

In order to explore the construct validity of these demographically-motivated variations in sympathetic concern, Study 3 also included two distinct means of measuring racial discrimination. The first, as is commonly employed in surveys studying racial attitudes, was an explicit measure of racial discrimination known as the Modern Racism Scale (MRS). In

contrast, the second measure was designed to measure subtle cognitive attitude-differentials that have been described in the literature as a form of ‘implicit’ racial discrimination (Cunningham, Preacher, & Banaji, 2001; Levin & Banaji, 2006), utilising the Implicit Association Test methodology developed by Greenwald, McGhee and Schwarz (1998; refined in Greenwald, Nosek & Banaji, 2003; Nosek, Greenwald & Banaji, 2005; hereafter referred to as the Racial-Attitudes IAT or RA-IAT).

In line with previous research involving similar constructs, in a sample population that was unlikely to host large and pervasive racial biases (that is, university undergraduates; see Levin & Banaji, 2006), it was predicted that there would be no significant correlations between the implicit and explicit measures of racial prejudice. As with the null-prediction raised in Study 2, it must be noted that these parametric statistics do not provide a reliable means to ascribe further meaning to the absence of statistically significant correlations. This null-prediction is mentioned purely to remind readers that, contrary to the intuitions of many, correlations between implicit and explicit measures of related are considered plausible, but atypical, and that the presence of such a correlation may indicate an unusual degree of prejudice or candour in the sample.

The authors hypothesised that any racially-motivated differences in the SATEST’s Helping index (expressed as a light-minus-dark difference score) would significantly positively predict implicit prejudice towards dark-skinned individuals. Also, it was predicted that racially-motivated differences in the SATEST’s Attributions index would significantly positively predict explicit prejudice towards dark-skinned individuals. It was further hypothesised that these two predictive relations would remain statistically significant when

controlling for the variance explained by gender, age, ethnicity and other relevant operationalisations of racial prejudice.

Method

Participants and procedure. Two-hundred and forty-six participants (188 female, 58 male) were recruited from a university subject-pool in exchange for psychology course credit. Via a secure browser-based computer interface, constructed in *Adobe Flash*, on a PC of their choosing, each participant completed the SC-SATEST, followed by the RA-IAT and the MRS. Participants self-identified as a wide (though skewed) range of racial ethnicities, with 142 (37%) identifying as Caucasian of European descent, 30 (12%) identifying as of Mediterranean descent, 24 (10%) of Middle-Eastern descent, 30 (12%) of East Asian descent, 18 (7.5%) of South Asian descent, and 2 (1%) of Australian Indigenous descent. The age range of participants met the typical characteristics of an undergraduate sample ($M = 21.05$ years, $SD = 5.8$), with a minimum age of 18 and a maximum age of 48.

Measures. The SC-SATEST varied the visual stimuli of the previous SATEST design by randomly assigning the participant to one of two racial-manipulation conditions, in which opposite halves of the 12 SATEST scenarios were presented with a sympathy-target that possessed dark-brown coloured skin, dark eyes and black hair. This was to ensure that even proportions of the four potentially-varying subsets of the SATEST outlined in Study 1 always possessed proportionate members of light- and dark-skinned sympathy-targets so as to compute a valid set of comparisons. Due to the illustrated abstraction of the present SATEST visual stimuli, it was not possible to depict the highly subtle and specific facial features

commonly employed to visually distinguish the race of a face (Levin & Banaji, 2006). As such, there are a number of possible interpretations of the race of the dark-skinned sympathy-targets (for example, of African descent, of South-Asian descent, etc.), which conservatively limits this study's ability to specifically pair racial stimuli for comparison.

The RA-IAT (Nosek, Greenwald & Banaji, 2005) uses racial stimuli facial profile photographs of adults from 5 racial groups in order to produce more generalised results comparable with the outputs of the SC-SATEST. These five racial groups, though possible to analyse separately, are clustered in the following score-computations into Light-Skinned stimuli (featuring the faces of Caucasian and East-Asian adults), and Dark-Skinned stimuli (featuring the faces of adults of African, South-Asian, and Australian Indigenous decent). The faces presented in all images possessed only minor variations in age and weight, and were counterbalanced to present even gender proportions. Unlike typical racial-comparison IATs, racial categories each consisted of multiple racial groups, delimited by light and dark skin-tones (this method was first explored in Marsh & Boag, in development). The associated positive and negative attitude words were taken from the standard inventory used in attitude IATs (Greenwald, McGhee & Schwarz, 1998), and standardised RA-IAT D-Scores were calculated in accordance with the guidelines specified by Nosek et al. (2005), which have shown strong internal consistency and test-retest reliability in prior studies (Nosek, Greenwald & Banaji, 2005), though are potentially susceptible to stereotype priming (see Dasgupta & Greenwald, 2001, for details).

The MRS (McConahay, 1986) poses 10 statements concerning the role and influence of a named racial or ethnic group in the stated national culture of the participant, to which the

participant responds on a 7-point Likert-scale of agreement (ranging from Strongly Disagree to Strongly Agree, where higher scores represent greater degrees of racism in normally coded items). Three iterations of the MRS were used in this study, specifying the racial demonyms of three Australian racial minority groups, people of African descent, people of South-Asian descent, and people of East Asian descent (with provided national examples to ensure participants understood each category; for example, that Sri Lanka, India and Pakistan are South-Asian countries). The items were presented together, in random order, with responses to each subscale proving highly internally consistent (each Cronbach's $\alpha > .950$), and also highly consistent between the subscales (with a collective Cronbach's α of .793).

Analyses. Two sets of internal consistency and principle components analyses were conducted on the behavioural outputs of both the Light-Skin and Dark-Skin subsets of the SC-SATEST, to confirm that each subsection maintained sufficiently similar psychometric structure to the full inventory to warrant subsequent comparisons. Bonferroni-adjusted correlation coefficients were calculated between the SC-SATEST difference variables, the RA-IAT D-Scores, and the three MRS scores as a preliminary test of the relationships predicted by the theory. In order to test these predictions while controlling for the mutual influence of the two forms of prejudice (in addition to the demographic variables), two hierarchical regression analyses were conducted. The first explored the predictive relationships between implicit negative attitudes towards dark-skinned targets and the SC-SATEST difference in Helping. The second targeted the relationship between the SC-SATEST difference in Attributions, and the prominent MRS-African scores.

Table 8*Factor Loadings of Established Indices for Light- and Dark-Skin Subsets of the SC-SATEST*

Indicator	Light-Skin				Dark-Skin			
	Action	Help	Deont	Attr	Action	Help	Deont	Attr
All Choices to Act	.83	-.08	-.14	.19	.86	-.17	-.13	.17
All Choices to Not Act	-.86	.02	.15	-.17	-.78	.03	.17	-.12
Choices to Help	.07	.89	-.14	.11	.07	.83	-.17	.18
Choices to Not Help	-.08	-.91	.15	-.09	-.11	-.93	.19	-.04
Positive Personal Justification	.80	.07	.26	.16	.71	.09	.28	.13
Negative Personal Justification	-.10	.06	-.17	-.69	.07	.09	-.11	-.64
Deontological Justifications	-.27	-.22	.79	.04	-.23	-.28	.73	.07
Positive Attributions	.20	.20	-.01	.89	.15	.23	.14	.76
Negative Attributions	-.19	-.17	-.08	-.94	-.14	-.15	-.05	-.83
Considerations	.17	.14	.91	.16	.11	.26	.86	.22

Note: $N = 246$; rotated to oblimin criteria ($\delta=0$): loadings $> .40$ in bold type.

Table 9*Pearson Correlations Between SC-SATEST Differences, RA-IAT D-Scores, and MRS Subscales*

	RA-IAT D-Scores		MRS Subscales		
	Light-Skin	Dark-Skin	African	South-Asian	East-Asian
SC-SATEST					
Activity Diff	-.096	.079	-.047	-.070	-.075
Helping Diff	-.171**	.169**	.051	.046	.087
Deontology Diff	-.041	-.034	-.028	-.008	.020
Attribution Diff	.069	.019	.170**	.129*	-.084
RA-IAT D-Scores					
Light-Skin	-	-.646***	-.167**	-.131*	-.091
Dark-Skin	-.646***	-	.173**	.079	.052

Note: $N = 246$ in all samples; Bonferroni correction for these comparisons is $\alpha = .05/6 = .008$.

* $p < .05$; ** $p < 0.008$; *** $p < 0.001$

Table 10*Hierarchical Multiple Regression Predicting Implicit and Explicit Racial Attitudes*

Dependent Variable	Step	Predictors	β	ΔR	ΔR^2_{Adj}
Dark-Skin D-Score	1	Age	.121	.135	.010
		Gender	-.074		
		Ethnicity	.022		
	2	MRS-African	.167**	.214**	.032
	3	Helping Difference	.151*	.260*	.054
	4	MRS-East-Asian	-.123	.291	.054
		Activity Difference	.049		
		Deontology Difference	-.029		
		Attribution Difference	.050		
MRS-African	1	Age	.073	.010	.002
		Gender	.026		
		Ethnicity	-.077		
	2	Dark-Skin D-Score	.169**	.185**	.023
	3	Attribution Difference	.162*	.228*	.036
	4	Activity Difference	-.063	.224	.026
		Helping Difference	-.017		
		Deontology Difference	-.027		

Note: $N = 246$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$ *Results and Discussion*

As Table 8 shows, principle components factor analysis performed upon the behavioural indicators draw from the Light- and Dark-Skin subsets of the SC-SATEST confirmed the presence in each subsection of the factorial loadings observed in the full SATEST, as demonstrated in Studies 1 and 2. Furthermore, the uniform factorial structure of the two subsets provides evidence that they can be meaningfully compared in order to compute

difference variables along each of the 4 factor-supported indices of the SATEST methodology. Evidence for the consistent psychometric properties of the SC-SATEST was also obtained via internal consistency analyses of the two Helping subsections, both of which demonstrated adequate Cronbach's alpha levels ($\alpha = .761$ in the Light-Skin subset, and $\alpha = .710$ in the Dark-Skin subset). The test-retest reliability of the computed indices of each subset were also assessed to be adequate, via a subsample of 22 participants who completed the SC-SATEST a second time two weeks after the initial testing (Deontological Index from the Dark-Skin subset demonstrated the weakest correlation, with $r = .631, p < .001$).

The correlation matrix, outlined in Table 9, revealed the expected significant positive correlation between the difference variable of the SC-SATEST Helping Index and the RA-IAT D-Score signifying implicit prejudice towards Dark-Skinned stimuli. As hypothesised, there was also a positive correlation between the SC-SATEST Attribution Index difference variable and the MRS explicit prejudice measure directed towards South-Asian and (even more so) African targets. Beyond this, significant positive correlations were observed between the measures of implicit and explicit prejudice towards those with dark skin, which were not anticipated in this sample. Results such as these are consistent with the partial conceptual overlap theorised between explicit and implicit measures of racial prejudice (Hofmann et al., 2005), but are none-the-less atypical, raising the possibility that sample participants may have demonstrated uncommonly candid explicit racial sentiments. Despite this, the hierarchical regression analyses outlined in Table 10 reveal that the SC-SATEST difference scores for the Helping Index and the Attribution Index remained statistically significant predictors (if only marginally) of implicit and explicit prejudice towards targets with dark skin, respectively, even when controlling for the mutual influence of implicit and

explicit prejudices observed in this study. This demonstrates that that these SATEST difference variables offer unique incremental predictive power in understanding the racially motivated down-regulation of sympathetic emotions, in addition to the forms of attributions typically employed to justify unfavourable decisions.

General Discussion

The three studies outlined above were designed to explore the psychometric properties of the new SATEST methodology in three key respects: Firstly, as a coherent measurement tool that can reliably and unintrusively generate scores corresponding to four related but conceptually distinct constructs that were predicted in the social and moral psychology literature. Second, as a moral judgment measure capable of detecting meaningful individual differences in expressions of sympathy, justifications and attributions when presented with social transgressions. Lastly, as a measure of feature-driven interpersonal prejudice (specifically, in this case, concerning race) based on detectable differences in moral judgment variables between scenarios targeting transgressors from visibly different demographic groups. These three applications of the SATEST are interrelated, with the conceptual plausibility of each application depending the demonstrable success of the application preceding it. As such, the tenability of each proposed application, and the methodological limitations therein, are addressed below in order, concluding with a summary discussion of how this preliminary data on the SATEST approach reflects upon the underlying moral, social, and evolutionary theory that shaped it.

Psychometric Properties of the SATEST

As with many psychometrics designed to study social and moral decision-making, the SATEST measurement tool is, in essence, a series of questions posed in response to a set of framing dilemmas. In accordance with the recommendations of Christensen and Gomila (2012, as explored above), meticulous efforts were taken to avoid a wide range of subtle confounding influences in dilemma formation and response-phrasing in crafting the 12 SATEST scenarios, but one of the key design goals of the SATEST measure was to maximise the ecological validity of the vignettes. While a certain degree of artifice is implicit when responding to prompts on a screen via keyboard and mouse controls, and the narration of vignettes was a practical necessity in establishing context, the SATEST scenarios were carefully designed to represent plausible, everyday situations that a typical citizen of any industrialised Western nation may conceivably find themselves in. This stands in contrast to many dilemma-based measures, particularly in the moral judgment literature, which often pose exceptional circumstances as a means of enticing the motivation of the participant, and ensuring that the dilemmas are sufficiently unfamiliar to the participant to render it unlikely that they possess an already well-deliberated response (the DIT itself serves as a strong example; see Rest et al., 1999). Though the coincidences and encounters of the SATEST scenarios are uncommon enough to ensure participants are unlikely to have experienced similar situations first-hand, they each remain sufficiently commonplace that a typical participant would not struggle to imagine encountering such a setting in their real lives. This feature allows the SATEST scenarios to be plausibly phrased with direct personal pronouns (e.g. “you are”) in a manner that would appear incongruous with such outlandish situations as choosing to flip a switch on a runaway trolley, or advising a man on whether or not to steal treatment drugs for his dying wife. Furthermore, as Christensen and Gomila (2012) explore, many dilemma-based measures encounter difficulties when posing response questions to the

participant, as the phrasing of the questions (e.g., “What would be the right thing to do here?”) introduces a degree of abstraction that causes the participant’s immersion and engagement with the vignette to suffer. As such, the SATEST scenarios each feature a ‘friend’ cohort character who poses contextually plausible questions to the participant conversationally and in near real-time, in an attempt to minimise the abstraction from the dilemma required when responding. This conversational format also allows for an inconspicuous means of discerning when the participant is engaged in lengthy deliberation, relying on Greene’s (2013) insights into the timed nature of declarative reasoning to provide measurable interactive points of deliberation, which triggered automatically after several seconds of participant consideration.

By this design, unlike many other psychometric tools, the SATEST approach relies heavily on ecological validity and an intuitive conversational interface to endow a relatively small set of response items with a great contextual value. This was a risky strategy, in contrast with the more common approach of generating a large pool of similarly phrased response items, which are aggregated into common constructs discovered or confirmed via factor analysis, since a smaller pool of items is more easily disrupted and rendered uninterpretable by inconsistent responding should participants fail to feel immersed in the vignettes. As such, the three phases of factor analysis performed on the various SATEST versions in the present studies must be interpreted as indications of how reliably the appropriate response-items cluster into the four distinct indices predicted, and also as indirect indications of how well the SATEST scenarios fostered a meaningful sense of immersion in participants. To this end, the authors decided against employing confirmatory factor analysis of the four predicted indices, instead relying upon the flattening of scree-plots and the Kaiser-Guttman cut-off, in order to more easily detect whether variations from the expected four factor solution could emerge as a

result of unanticipated response styles. As a further conservative measure, each factor analysis relied upon an oblique rotation method (direct oblimin), for it was anticipated that the four indices may be highly correlated. Though none of the three factor analyses yielded correlations between factors higher than -0.20 (between the Action and Deontological Indices; all other correlations below 0.10), suggesting that this four factor structure was compatible with an orthogonally rotated solution, the authors have elected to preserve the oblique rotation for simplicity of interpretation. Also, while a stronger result may have been possible using an exploratory factor analysis method that assessed common variance between items, the authors chose principle components analysis to obtain factors that account for total variance observed. The decision was theoretically motivated, as factor analyses based upon common variance between items are best suited when searching for latent variables presumed to be a direct underlying cause of the variance observed. This approach is ill-justified by the theory underpinning the design of the SATEST, as no discrete latent variables have yet been proposed to directly cause the behaviours of interest. The SATEST approach assumes only a functionally coordinated set of behaviours consistent with the theorised evolutionary rationales discussed above (particularly those of de Waal, 2008, and Haidt, 2007) and does not yet propose to test any structural details of the mechanisms making these behaviours possible. As such, the SATEST indices are proposed as aggregated descriptive constructs drawn from the behaviour indicators tested, rather than underlying latent variables thought to cause the behaviours measured, and thus are modelled most appropriately via principle component analyses of total variance observed. At this stage of development, and given the intended ecological validity of the SATEST design, there is no meaningful way to distinguish what components of the total variance could be defensibly considered ‘measurement error’.

With these limitations and interpretative constraints in mind, the highly differentiated and highly consistent pattern matrix results obtained for the three factor analyses provide strong evidence in favour of the four theorised output indices of the SATEST. Despite the oblique rotation and lack of confirmatory specification for four factors, each set of behavioural indicators loaded onto the anticipated index, valenced in the anticipated direction.

Furthermore, the magnitude of each expected loading did not fall below 0.62 in any of the three analyses, while no items in any analysis loaded onto any secondary factor to a magnitude higher than 0.28. These distinct loadings also provide indirect evidence for the efficacy of the meticulously controlled confounding elements and high ecological validity of the SATEST scenarios, as even moderate degrees of meaningless or inconsistent responding by participants would be expected to introduce considerable ‘noise’ into these results, given the range of behavioural indicators that were generated from relatively few participant responses. The lack of any significant differences in loadings between those scenarios wherein helping required action and those requiring inaction, as well as the absence of such differences between scenarios where action was undertaken through the ‘friend’ character as opposed to directly, suggests that all 12 scenarios are sufficiently interchangeable to allow for split conditions (as in Study 3) or perhaps even shortened versions of the measure. It is particularly encouraging that the split subsets of the 12 scenarios employed in the skin-colour manipulation of the SATEST each independently retained the desired factorial structure, though further testing would be required to determine whether smaller comparative splits (such as 3-3-3-3 rather than 6-6) would perform just as well.

With regards to the future development of the SATEST, it remains to be seen whether further changes to increase the immersion and ecological validity of the scenarios would enhance the psychometric properties thus observed, or hinder them. Promising directions include the use

of photorealistic images, or perhaps even full-motion video, but while such innovations would likely increase participant immersion (Gillath et al., 2008; Llobera et al., 2010), it is possible that certain degrees of ambiguity that are present in the current illustrated versions of the SATEST are key to the observed success. For example, the ambiguities of somewhat underspecified illustrations may be necessary for participants to not find it jarring to be presented with a stranger on-screen and be told that this is their ‘friend’. Similarly, the ambiguities of the current illustrations allow many elements of physical appearance unrelated to characteristics of interest (e.g., facial attractiveness, weight, age) to remain unspecified (or serve as artificially strong elicitors; see Sherman & Haidt, 2011), whereas employing concrete photo-realistic images would demand that each confounding variable be present, and somehow controlled (be it through standardisation or many parallel testing conditions). Lastly, while the present studies provide evidence that the key psychological trade-off underpinning the SATEST measure (the conflict between sympathetic affect towards a troubled target, and the social license to punish rule-breakers) produces effective indicators of sympathy, justification and attribution in this current suite of scenarios, it remains to be established how well this finding may perform in less familiar, or culturally non-Western contexts.

Measuring Moral Judgment

Given the evidence above, supporting the psychometric reliability and theoretical validity of the SATEST methodology, its value as a measure of moral judgment can be assessed in terms of its convergent and divergent relationships with other tools in the moral literature.

Theoretically speaking, the adherence of the SATEST response items to the four indices predicted in the literature suggests at least face validity with regards to the simple expression

of sympathetic feeling (captured by the Helping Index) and baseline activity level in hypothetical scenarios (captured by the Activity Index). For these elements, in addition to simple positive expressions of affect (which also load onto the Activity Index, in the SATEST), such face validity may be the only evidence obtainable within the constraints of online survey methods, as direct social and behavioural indicators (such as physiological fluctuations signalling affect or engagement) were unavailable within the scope of this study. There is, however, precedence in the moral and social literature (notably Haidt, 2001; Greene & Haidt, 2002) to accepting such face validity as the basis of simple indicators of affect, provided the moral and social context of the questions do not provide confounding means of expressing conflicting affect states (for example, harming someone to put them out of their misery, appearing sympathetic only to cultivate reputation, etc.). These concerns are not only avoided in the SATEST design, but employed knowingly and strategically to elicit plausible conflicts between affective desires to behave sympathetically, and justified motivations to righteously punish a social-rule transgressor. Beyond these intentional dual concerns (a conflict between intuitions of care and justice), one of the key features of the SATEST design was ensuring that each vignette was morally equivalent in terms of the other three core moral intuitions specified in the Social Intuitionist model (Haidt, 2007). Also, in addition to this face validity, the SATEST Activity Index received evidence for convergent validity in the form of its significant correlations (positive and negative, where predicted) with the basic Trolley Problem responses in Study 1, which are in part shaped by participants' general tendency 'to act' in hypothetical questions (Greene et al., 2009). The SATEST Helping Index also approached statistical significance in its correlations with the Trolley Problem responses, though it is at this point unclear whether this non-significance represents a problem for the SATEST's validity, given the harshly utilitarian calculus of the Trolley Problem and its tendency to elicit negative affect in many participants.

Evidence for convergent validity concerning the Helping and Deontological Indices was obtained in the correlations and hierarchical regression analysis between the SATEST and the DIT. Although it is designed to focus primarily on deliberate moral reasoning while eschewing simple affect-driven reactions, the DIT's P-Score (and its later edition replacements) has also long served as a general indicator of morally compassionate, non-selfish tendencies in adults (reviewed in Rest et al., 1999). In both Studies 1 and 2, P-Scores showed significant positive relationships with both the Helping and Deontological Indices of the SATEST, despite the two indices not correlating with each other. This suggests that each index correlates separately with one of the two aspects of moral judgement captured by the DIT's P-Score. This suggestion was further supported by the regression analyses, wherein both indices were shown to be significant incremental predictors of DIT P-Score, even after controlling for the influence of the other.

As with many measurement methods based in the strictly-cognitive neo-Kohlbergian tradition of moral psychology, the DIT's P-Score has been criticised for the degree to which its measurement goals are confounded with other factors that typically grow with age and maturity, most notably (given its complex scoring and vignettes) verbal cognitive ability (Sanders, Lubinski & Benbow, 1995). This methodological and conceptual shortcoming of the DIT was replicated in Study 2, wherein the SATEST showed no such vulnerabilities. Furthermore, the predictive value of the SATEST Helping and Deontological Indices both remained statistically significant predictors of DIT P-Score, even when controlling for verbal and cognitive ability measures and participant age. This provides strong evidence for the claim that the SATEST Helping Index meaningfully measures morally benevolent

dispositions and sympathetic concern, whereas the Deontological Index meaningfully measures the degree of deliberate moral reasoning engaged during the vignettes. Beyond the evidence provided by comparisons with the DIT, it would be instructive to compare the SATEST Indices with other measures of moral judgement that rely more on the influence of affect and intuition. However, at present affect-driven moral judgment measures are typically only employed in highly specific testing contexts (see Haidt, 2007), distinctly limiting their utility in assessing measures designed to be broadly applicable, such as the DIT and the SATEST.

Measuring Prejudice

By virtue of convergent and divergent validity demonstrated in Studies 1 and 2, the SATEST can be regarded as a reliable enough measure of key moral judgment variables to potentially serve as a measure of demographically motivated prejudice. As was explored above, the reliable and theoretically supported factorial structure of the SATEST was successfully replicated when the measure was divided into two counterbalanced sets of six scenarios, on retaining the light-skinned target illustrations used in Study 2, and the other modified these illustrations so as to depict targets belonging to dark-skinned racial groups. This skin-colour manipulation of the SATEST allowed difference scores to be generated for each of the four Indices, and thus was theorised to be a potentially viable measure of racial prejudice by conceiving of racist attitudes as demographically motivated decreases in sympathetic concern (measured by the Helping Index) and increases in negative attributions for transgression behaviours (measured by the Attributions Index).

The value of the SC-SATEST as a measure of racial prejudice received support via significant correlations with measures of implicit racism (for the Helping Index difference score) and explicit racism (for the Attributions Index difference score). These relations were further supported in the multiple regression analyses, wherein both SATEST Indices' difference scores remained statistically significant predictors of their respective forms of racism, even when the other form of racism, gender, age, and ethnicity were controlled for. While it is encouraging that the SC-SATEST was able to provide significant predictors for both of the primary forms of racial attitudes measured in the literature, these results should be interpreted cautiously, however. Firstly, because the observed effect sizes in the regression analyses turned out unimpressively small after adjustment, and secondly, because the sample used in Study 3 showed a somewhat atypical significant correlation between explicit and implicit measures of racism, suggesting that some aspect of participant's racial attitudes (potentially explicit racist attitudes) were unusually candid, limiting how comparable these findings may be to previous samples.

As with the moral judgment analyses, these results at the very least provide preliminary evidence that there is merit to measuring racial prejudice, and perhaps other forms of prejudice and discrimination, as feature-motivated biases in one's affective sympathy and styles of attribution. It would likely prove instructive for future research to continue testing the forms of demographically motivated moral biases that the SATEST methodology is capable of detecting, (e.g., biases related to gender, age, or apparent illness). Some forms of discrimination, such as discrimination on the basis of sexual orientation, would require some significant revisions to the current suite of SATEST scenarios if it were to be manipulated in the same unobtrusive manner as was employed in the SC-SATEST. It remains unclear at this point, however, precisely how much the present set of scenarios can be modified to include

additional characters or more diverse circumstances, before the established psychometric properties begin to suffer.

Underlying Mechanisms

Although the core social, moral, and intuitive conflicts that were synthesised to make the SATEST approach possible are grounded in the converging evolutionary literature discussed above, the relevant theory is not yet sufficiently informed by experimental data to meaningfully hypothesise about what kinds of psychological processes and mechanisms causally underlie the behavioural tendencies observed. As with most analyses based in the adaptationist approach, the theory underpinning the SATEST methodology must begin with an identification of the general fitness problems thought to shape the mechanisms in question (see the *Evolutionary Approach to Sympathy* section above), and proceed to preliminary predictions concerning the most likely behavioural manifestations of these mechanisms if they were functionally ‘well-designed’ (Tooby & Cosmides, 2005; Buss, 2005). The most problematic conceptual stumbling-block at this stage is the possibility of evolutionary mismatch, for if our social and environmental context has changed sufficiently from the circumstances in which these psychological mechanisms became (presumably) species-typical, we cannot guarantee that their observable behavioural manifestations will appear functional or adaptive under our present conditions. The hypotheses proposed and explored in this paper were predicated on the conditional assumption that, if indeed we possess the moral interpersonal mechanisms predicted by the literature, that they will behave in a generally functional manner within our present cultural and technological context. These assumptions were considered reasonable, and worthy of some empirical exploration, simply because any set of psychological mechanisms so potentially crucial to social, hierarchical and

reproductive success, are unlikely to propagate to species-ubiquity in a form that cannot functionally adapt to a wide range of situational and cultural variants (Tooby & Cosmides, 1989). Given the supporting evidence revealed in the results of these three studies, we can interpret these findings as also tentatively supporting the theoretical assumptions made in generating these hypotheses, although at this early stage it would be truly unwise to commit to these assumptions prematurely. Since the primary adaptive function of many moral psychological mechanisms is thought to concern reputation-management, conflict resolution and ally-recruitment (DeScoli & Kurzban, 2013), it remains a possibility that the mechanisms in question may only need to appear coherent to others under key social circumstances, while providing little consistent guidance to our true interpersonal attitudes and behaviours. The soundness of the assumptions underlying the SATEST measure and its specific hypotheses can only be fairly assessed with a far greater critical exploration of when justice-care conflicts occur, and how other fitness-related variables (such as disgust-sensitivity, mortality-salience, priming of intergroup conflict, etc.) affect their expression.

Conclusion

The three studies explored in this paper provide preliminary convergent and divergent evidence for the construct validity of the SATEST as a means of measuring multiple behavioural manifestations of sympathetic concern. The evidence drawn from comparisons to other measures of moral judgment offer strong indications that the standard form of the SATEST can capably identify variation in action-oriented dispositions of participants, their empathetically-motivated decisions to help others, and their tendency to address such dilemmas with explicit, contemplative moral reasoning. The evidence provided for the efficacy of the Skin-Colour manipulation of the SATEST in detecting meaningful,

demographically-motivated variations in sympathetic concern is less conclusive, yet still significant given the demonstrated relationships between helping-behaviour differences and implicit racial prejudice, and attribution-differences and explicit racial prejudice. It remains to be seen whether the measurement efficacy of the SATEST methodology is limited by its present reliance on illustrations rather than photo-realistic visual cues, or whether the results observed here rely on the relative abstractness of these stimuli to elicit these response patterns from participants. Future research employing the SATEST must explore these potential limitations, in addition to testing additional permutations of dilemma scenarios which utilise the same sympathy vs. social-rules conflict that proved effective in the present studies.

References

- Bartels, D.M. (2008). Principled Moral Sentiment and the Flexibility of Moral Judgment and Decision Making. *Cognition*, 108, 381-417.
- Batson, D.C. (2011). What's wrong with morality? *Emotion Review*, 3, 230-36.
- Borg, J.S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation. *Journal of Cognitive Neuroscience*, 18(5), 803–817.
- Borsboom, D., Mellenbergh, G.J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203-219.
- Borsboom, D., Mellenbergh, G.J., & Van Heerden (2004). The concept of validity. *Psychological Review*, 111, 1061-1071.
- Brewer, M.B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86, 307-324.
- Brewer, M.B. (1988). A dual process model of impression formation. In T.K. Srull & R.S. Wyer (Eds.), *Advances in social cognition* (pp. 1-36). Erlbaum, New Jersey.
- Buss, D.M. (2005). *The Handbook of Evolutionary Psychology*. Wiley, New Jersey.
- Cannon, P.R., Schnall, S., & White, M. (2011). Transgressions and Expressions: Affective Facial Muscle Activity Predicts Moral Judgments. *Social Psychological and Personality Science*, 2(3), 325-331.
- Carr, M.B., & Lutfemeier, J.A. (2005). The relation of facial affect recognition and empathy to delinquency in youth offenders. *Adolescence*, 40, 601–619.

- Casebeer, W. D. (2003). Moral cognition and its neural constituents. *Nature Review of Neuroscience*, 4, 841–847.
- Cattell, R.B. (1971). *Abilities: Their structure, growth, and action*. Houghton Mifflin, New York.
- Cattell, R.B. (1987). *Intelligence: Its structure, growth, and action*. Elsevier Science, New York.
- Chapman, H.A. & Anderson, A.K. (2011). Varieties of moral emotional experience. *Emotion Review*, 3, 255-257.
- Chen, H., Russell, R., Nakayama, K., & Livingstone, M. (2010). Crossing the 'uncanny valley': adaptation to cartoon faces can influence perception of human faces. *Perception*, 39(3), 378-386.
- Christensen, J.F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience & Biobehavioral Reviews*, 36(4), 1249-1264.
- Cikara M., Bruneau E., & Saxe R. (2011). Us and Them: Intergroup failures of empathy. *Current Directions in Psychological Science*, 20(3), 149-153.
- Cikara, M., Farnsworth, R.A., Harris, L.T., & Fiske, S.T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social Cognitive and Affective Neuroscience*, 5, 404-413.
- Clements-Nolle, K., Marx, R., & Katz, M. (2006). Attempted Suicide Among Transgender Persons. *Journal of Homosexuality*, 51(3), 53-69.

- Cohen, E. (2012). The evolution of tag-based cooperation in humans: The case for accent. *Current Anthropology*, 53(5), 588-616.
- Corrigan, P.W., & Wassel, A. (2008). Understanding and influencing the stigma of mental illness. *Journal of Psychosocial Nursing and Mental Health Services*, 46, 42-48.
- Cosmides, L., Tooby, J. & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences* 7(4), 173-179.
- Creed, C., & Beale, R. (2008). Emotional intelligence: Giving computers effective emotional skills to aid interaction. *Studies in Computational Intelligence*, 115, 185-230.
- Crowson, H.M., & DeBacker, T.K. (2008). Political identification and the DIT: Re-evaluating an old hypothesis. *Journal of Social Psychology*, 148, 43-60.
- Cunningham, W.A., Preacher, K.J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163-170.
- Cushman, F.A. (2011). Morality from the frog's eye view. *Emotion Reviews*, 3(3), 261-263.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800-814.
- Decety, J., & Grézes, J. (2006). The power of simulation: Imagining one's own and other's behaviour. *Cognitive Brain Research*, 1079, 4-14.
- Decety, J., & Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3, 71– 100.

- Decety, J., Norman, G.J., Berntson, G.G., & Cacioppo, J.T. (2012). A neurobehavioral evolutionary perspective on the mechanisms underlying empathy. *Progress in Neurobiology*, 98, 38-48.
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477-496.
- de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10(10), 435-441.
- Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- de Waal, F.B.M. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Harvard University Press, Cambridge.
- de Waal, F.B.M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, 59, 279-300.
- Drescher, J., & Merlino, J.P. (2007). *American Psychiatry and Homosexuality: An Oral History*. Harrington Park Press, New York.
- Eisenberg, N., & Morris, A.S. (2001). The origins and social significance of empathy-related responding. A review of empathy and moral development: implications for caring and justice by M.L. Hoffman. *Social Justice Research*, 14, 95-120.
- Ferstl, E.C., Neumann, J., Bogler, C., & von Cramon, D.Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, 29, 581-593.

- Fiske, S.T., Cuddy, A.J.C., Glick, P. & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.
- Freud, S. (1913/1918). *Totem and Taboo*. Translated by A.A. Brill. Moffat, Yard & Co., New York.
- Freud, S. (1930/1962). *Civilization and its discontents*. Translated by J. Riviere. Leonard & Virginia Woolf at the Hogarth Press, London.
- Gillath, O., McCall, C.A., Shaver, P., & Blascovich, J. (2008). Reactions to a needy virtual person: Using an immersive virtual environment to measure prosocial tendencies. *Media Psychology*, 11, 259-282.
- Greene, J.D. (2007) Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, 11(8), 322-323.
- Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Press, New York.
- Greene, J.D., Cushman, F.A., Stewart, L.E., Lowenberg, K., Nystrom, L.E., & Cohen, J.D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364-371.
- Greene, J., & Haidt, J. (2002) How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517-523.

- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., & Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389-400.
- Greene, J.D., & Paxton, J.M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences USA*, 106(30), 12506-12511.
- Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral Judgment. *Science*, 293, 2105-2108.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1002.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon, New York.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116.

- Harcourt, A.H., & de Waal, F.B.M. (1992). *Coalitions and Alliances in Humans and Other Animals*. Oxford University Press, Oxford.
- Hart, A.J., Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., & Rauch, S.L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *NeuroReport*, 11, 2351–2355.
- Haslam, N., & Levy, S. R. (2006). Essentialist beliefs about homosexuality: Structure and implications for prejudice. *Personality and Social Psychology Bulletin*, 32, 471- 485.
- Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology*, 21, 517-528.
- Hoffman, M.L. (2000). *Empathy and moral development: Implications for caring and justice*. New York, Cambridge University Press.
- Hofmann, W., Gschwendner, T., Nosek, B.A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology*, 16(10), 335-390.
- Horn, J.L., & Noll, J. (1994). A system for understanding cognitive capabilities. In D.K. Detterman (Ed.). *Current Topics In Human Intelligence* (pp. 151-203). Ablex, New Jersey.
- Horn, J.L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J.L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). Guilford Press, New York.

- Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15, 632-637.
- Keeley, L.H. (1996). *War before civilization: The myth of the peaceful savage*. Oxford University, New York.
- Kendell, R.E. (2004). The Myth of Mental Illness. In J.A. Schaler (Ed.) *Szasz under Fire: The Psychiatric Abolitionist Faces His Critics*. Open Court, Illinois.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138), 908-911.
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M.L. Hoffman & L.W. Hoffman, (Eds.), *Review of Child Development Research, Vol. I* (pp. 381-431). Russel Sage Foundation, New York.
- Kohlberg, L. (1969). Stage and sequence: the cognitive developmental approach to socialization. In D.A. Goslin (Ed.) *Handbook of Socialization Theory* (pp. 347-480). Rand McNally, Chicago.
- Krebs, D.L. (2005a). The evolution of morality. In D. Buss (Ed.) *The Handbook of Evolutionary Psychology*, (pp. 747-771). John Wiley & Sons, New Jersey.
- Krebs, D.L. (2005b). An evolutionary reconceptualization of Kohlberg's model of moral development. In R. Burgess & K. MacDonald (Eds.) *Evolutionary Perspectives on Human Development*, (pp. 243-274). Sage Publications, California.
- Krebs, D.L. (2008). Morality: An evolutionary account. *Perspectives on Psychological Science*, 3, 149-172.

- Krebs, D.L., & Denton, K. (1997). Social illusions and self-deception: The evolution of biases in person perception. In J. A. Simpson & D. T. Kenrick (Eds.) *Evolutionary Social Psychology* (pp. 21-47). Erlbaum, New Jersey.
- Krebs, D.L., & Janicki, M. (2004). Biological foundations of moral norms. In M. Schaller & C. Crandall (Eds.) *Psychological Foundations of Culture* (pp. 125-148). Erlbaum, New Jersey.
- Krebs, D.L., & Van Hesteren, F. (1994). The development of altruism: Toward an integrative model. *Developmental Review*, 14, 1-56.
- Kruger, D.J. (2003). Evolution and altruism: Combining psychological mediators with naturally selected tendencies. *Evolution and Human Behavior*, 24, 118-125.
- Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, 33, 323-333.
- Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*, 127(2), 187-208.
- Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, 98(26), 15387-15392.
- Lamm, C., Batson, C.D., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19, 42-58.
- Latane, B., & Darley, J. (1969). Bystander "Apathy". *American Scientist*, 57, 244-268.

- Latané, B., & Rodin, J. (1969). A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Experimental Social Psychology*, 5, 189-201.
- Levin, D.T., & Banaji, M.R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology*, 135, 501-512.
- Levine, M., Cassidy, C., Brazier, G., & Reicher, S. (2002) Self-categorization and bystander non-intervention: Two experimental studies. *Journal of Applied Social Psychology*, 32, 1452–1463.
- Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, 31, 443–453.
- Lieberman, D.L., Tybur, J.M., & Latner, J.D. (2012). Disgust sensitivity, obesity stigma, and gender: contamination psychology predicts weight bias for women, not men. *Obesity*, 20(9), 1803-1814.
- Linke, L.H. (2012). Social Closeness and Decision Making: Moral, Attributive and Emotional Reactions to Third Party Transgressions. *Current Psychology*, 31(3), 291-312.
- Llobera, J., Spanlang, B., Ruffini, G., & Slater, M. (2010). Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception*, 8(1), 3-12.
- Markus, K. A., & Borsboom, D. (2011). Reflective measurement models, behavior domains, and common causes. *New Ideas in Psychology*, 31(1), 54-64.

- Marsh, T., & Boag, S. (*In Press*). Evaluative Attitudes and Identification with Light- and Dark-Skinned Racial Groups.
- McConahay, J.B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). Academic Press, Orlando.
- McDonald, M.M., Navarrete, C.D., & Van Vugt, M. (2012). Evolution and the Psychology of Intergroup Conflict: The Male Warrior Hypothesis. *Philosophical Transactions of the Royal Society-Biological Sciences*, 367(1589), 670-679.
- Messick, D.M., & Mackie, D.M. (1989). Intergroup relations. *Annual Review of Psychology*, 40, 45-81.
- Miller, S.L., Zielaskowski, K., Maner, J.K., & Plant, E.A. (2012). Self-protective motivation and avoidance of heuristically threatening outgroups. *Evolution and Human Behavior*, 33, 726-735.
- Moll, J., de Oliveira-Souza, R., Moll, F.T., Ignácio, F.A., Bramati, I.E., Caparelli-Dáquer, E.M., & Eslinger, P.J. (2005). The moral affiliations of disgust: a functional MRI study. *Cognitive and Behavioural Neurology*, 18(1), 68-78.
- Nichols, S., & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Noûs*, 41, 663-685.
- Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166-180.

- Nowak, M.A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291-1298.
- O'Connell, S.M. (1995). Empathy in chimpanzees: Evidence for theory of mind? *Primates*, 36, 397-410.
- Paterson, B., & Hopwood, M.N. (2010). The relevance of self-management programmes for people with chronic disease at risk for disease-related complications. In D. Krailik, B. Paterson, V. Coates (Ed.), *Translating Chronic Illness Research into Practice, 1st Edition* (pp. 111-142). Wiley-Blackwell, Oxford.
- Paxton, J.M., & Greene, J.D., (2010) Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, 2(3), 511-527.
- Paxton, J.M., Ungar, L., & Greene, J.D. (2011) Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.
- Piaget, J. (1932/1965). *The moral judgment of the child*. Free Press, London.
- Preston, S.D., & de Waal, F.B.M. (2002). Empathy: It's ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1-72.
- Rest, J.R. (1975). Longitudinal Study of the Defining Issues Test of moral judgment: A strategy for analyzing developmental change. *Developmental Psychology*, 11(6), 738-748.
- Rest, J.R., Narvaez, D., Bebeau, M.J., & Thoma, S.J. (1999). *Postconventional Moral Thinking: a neo-Kohlbergian approach*. Lawrence Erlbaum, New Jersey.
- Rest, J.R., Narvaez, D., Thoma, S.J., & Bebeau, M.J. (2000). A Neo-Kohlbergian Approach to Morality Research. *Journal of Moral Education*, 29, 381-396.

- Richman, W.L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires and interviews. *Journal of Applied Psychology*, 84, 754-775.
- Roets, A., & Van Hiel, A. (2011). Allport's Prejudiced personality today: Need for closure as the motivated cognitive basis of prejudice. *Current Directions in Psychological Science*, 20, 349-354.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz, *Advances in experimental social psychology* - 10 (pp. 173–220). Academic Press, New York.
- Rohmer, O., & Louvet, E. (2012). Implicit measures of the stereotype content associated with disability. *British Journal of Social Psychology*, 51(4), 732-740.
- Sanders, C.E., Lubinski, D., & Benbow, C.P. (1995). Does the Defining Issues Test measure psychological phenomena distinct from verbal ability?: An examination of Lykken's query. *Journal of Personality and Social Psychology*, 69, 498-504.
- Scambler, G. (2004) Re-framing stigma: felt and enacted stigma and challenges to the sociology of chronic and disabling conditions. *Social Theory and Health*, 2, 29–46.
- Schino, G., Geminiani, S., Rosati, L., & Aureli, F. (2004). Behavioral and emotional response of Japanese macaque (*Macaca fuscata*) mothers after their offspring receive an aggression. *Journal of Comparative Psychology*, 118, 340–46.
- Sherman, G.D., & Haidt, J. (2011). Cuteness and disgust: The humanizing and dehumanizing effects of emotion. *Emotion Review*, 3, 245–251.

- Sidanius, J., & Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge University Press, New York.
- Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., & Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466–469.
- Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J.J., Campbell, C., & Levine, M. (2013). Bystander Responses to a Violent Incident in an Immersive Virtual Environment. *PLoS ONE*, 8(1), e52766.
- Smuts, B., Cheney, D., Wrangham, R., & Struhsaker, T. (1987). *Primate Societies*. The University of Chicago Press, Illinois.
- Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, 62, 207-218.
- Stankov, L. (1997). *Gf–Gc quickie test battery*. E-ntelligence TestingProducts, Sydney.
- Stankov, L. (2000) Structural extension of a hierarchical view on human cognitive abilities. *Learning and Individual Differences*, 12, 35-51.
- Stevens, D., Charman, T., & Blair, R.J. (2001). Recognition of emotion in facial expressions and vocal tones in children with psychopathic tendencies. *The Journal of Genetic Psychology*, 162, 201–211.
- Tooby, J. & Cosmides, L. (1996). Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith, & R.

- I. M. Dunbar (Eds.), *Evolution of Social Behaviour Patterns in Primates and Man. Proceedings of the British Academy*, 88, 119-143.
- Tooby, J. & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 5-67). Wiley, New Jersey.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, Cambridge University Press.
- Van Bavel, J.J., & Cunningham, W.A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, 35, 321-335.
- Watts, D.P., Colmenares, F., & Arnold, K. (2000). Redirection, consolation, and male policing: how targets of aggression interact with bystanders. In *Natural Conflict Resolution*, (Ed.) F Aureli, FBM de Waal, pp. 281–301. University of California Press, Berkeley.
- Wilcox L.M., Allison R.S., Elfassy S., & Grelik C. (2006). Personal space in virtual reality. *ACM Transactions in Applied Perception*, 3(4), 412-428.

Discussion for Thesis Chapter 7

The empirical aims of the three studies featured in this chapter were to establish the psychometric properties of the SATEST methodology, verify its predicted capacity to measure the differences in empathetic, justification and attribution components of moral judgment, and its utility in measuring the differential patterns of these moral factors when responding to target characters representing light- and dark-skinned racial groups. Evidence from all three studies (explored above) strongly supports the designed sensitivities, reliability and internally-consistent factorial structure of the SATEST measurement tool and its computed outputs, satisfying the necessary minimal criteria for interpreting its validity in the moral psychology and prejudice contexts, and justifying its potential application to similar domains in future studies. The evidence of both convergent and divergent validity explored in the studies attests to the viability of the SATEST methodology, both in measuring moral and prejudicial variance captured by pre-existing scales, and in avoiding or controlling for several key confounding influences observed in earlier measures.

In the domain of phenomena traditionally addressed by moral psychology, studies 1 and 2 demonstrated the SATEST's capacity to simultaneously measure participants' general tendencies towards taking action in hypothetical scenarios (as has been identified in studies employing Trolley Dilemmas; see Cikara et al., 2010), their favourable sympathetic reactions to the plight of described characters (as is the focus of most dilemma-based moral judgment tasks, as reviewed in Christensen & Gomila, 2012), and their degree of engagement in deliberate, conscious, typically deontological reasoning (the central focus of all Neo-Kohlbergian approaches; see Rest et al., 1999—many insights of which are preserved in dual-process Social-Intuitionist approaches; see Greene & Haidt, 2002; Greene, 2007). In addition to these capacities (each element of which showed unique predictive value when modelled together), the SATEST demonstrated strong statistical divergence from several cognitive

ability variables known to confound the results of earlier general-purpose moral judgment measures (notably, the Defining Issues Test; Rest, 1975), including the well-established confounds of participants' verbal ability (see Sanders, Lubinski & Benbow, 1995).

Although the SATEST was designed to measure the activity of more general intrapersonal mechanisms than those typically associated with racial prejudice in the social psychology literature (notably, because the SATEST does not address stereotype-content), the light versus dark Skin-Colour manipulation of the SATEST (SC-SATEST) demonstrated preliminary evidence of expected convergent and divergent validity with both explicit and implicit social cognitive measures of racial attitudes (as conceptualised in Greenwald et al.'s 2002 review). The SC-SATEST divided the 12 original scenarios into two (randomly counterbalanced with regards to activity and agency framings) halves, one set of 6 retaining target characters with features resembling light-skinned racial groups, and the other 6 with characters seemingly belonging to dark-skinned racial groups. Despite this manipulation, each half of the SC-SATEST suitably retained the psychometric properties and factorial structure of the original, which allowed for separate moral responding profiles to be computed for each subset, permitting the calculation of light- versus dark-skinned difference scores for each of the SATEST indices. Significantly, these difference scores demonstrated simultaneous predictive relationships with both explicit and implicit measures of negative racial attitudes. Specifically the difference scores relating to the activation of participant's interpersonal empathy (the Helping Index) significantly and uniquely predicted implicit racial attitudes (as was predicted based on the prevailing social cognition theories of *aversive racism*; see Gaertner & Dovidio, 1986; Greenwald et al., 2003; 2009; Pearson, Dovidio, & Gaertner, 2009), while measured differences in the character of participants' attributions significantly and uniquely predicted explicit racial attitudes (expressed in the form of *modern racism*; see McConahay, 1986; Beal et al., 2000). That said, these relationships must be

interpreted cautiously, for as is common in the social cognitive study of racial attitudes, the total variance explained by any one model remained globally low (Nosek, Greenwald & Banaji, 2005), suggesting that variables not accounted for in these analyses (such as specific stereotype content) may exert a far greater influence on any outcomes of interest.

While the specific empirical hypotheses of each study were supported—thus establishing the core viability of the SATEST methodology as a potentially useful tool in future studies of moral judgment and intergroup prejudice (as was the main goal of this journal article)—it is also crucial to the goals of this thesis that the results of these studies be interpreted in the wider context of the integrative evolutionary theory developed over the course of this thesis. The underlying rationale for the ‘sympathy vs. social rule adherence’ tradeoff at the centre of each SATEST scenario, in addition to the wide range of specific hypotheses explored throughout the preceding three studies, are all predictions generated by the evolutionary synthesis of insights from social, differential, and moral psychology whose conceptual bases are first outlined in Chapter 3, and the core processes of which are described earlier in this chapter.

As the conceptual tools introduced in Chapter 4 specify, provided a theory is oriented towards the common material ontology that grounds modern scientific psychology, any theoretical position can be understood as a set of hierarchically dependent pragmatic assumptions. To render the empirical insights of a field conceptually compatible with the insights of another, it is necessary to identify the assumptions upon which the findings in question were based, and reinterpret the finding’s possible meanings when a key assumption is not shared between the perspectives one seeks to integrate. As a whole, this thesis has adopted the suite of paradigmatic assumptions generally employed by evolutionary psychologists, namely the adaptationist approach (the basics of which are detailed in Chapter 3), with the additional conceptual expansions of individual phenotypic variation explored in

Chapter 5, and strictly those employing appropriately bottom-up explanatory approaches (also described in Chapter 5). As such, the insights into the mechanisms underlying prejudice that could be drawn from the fields of social, differential and moral psychology, were by necessity constricted to those findings that did not depend upon assumptions that are fundamentally rejected by (or incapable of being recast in) the adaptationist approach.

For example, many theories concerning prejudicial personality tendencies in the differential psychology literature (see Roets & Van Hiel, 2011), were omitted from consideration due to their simplistic, top-down explanatory approaches (of the sort criticised in Chapter 5), which are fundamentally inconsistent with the assumptions regarding functional cognitive mechanisms in the adaptationist approach. Other empirical insights, such as those concerning the character of deliberate, conscious moral evaluations at the heart of cognitivist approaches to moral judgment (see Rest et al., 1999), were incorporated following reinterpretation in the light of which grounding assumptions did not stand up to scrutiny. As Greene and Haidt (2002) outline, deliberations of the sort described in this cognitive approach do truly occur with many of their empirically observed characteristics, and the fault merely lies in the (now largely disconfirmed) theoretical assumption that such deliberations are the primary process of moral judgment, rather than a rarely engaged, effortful subsidiary system. Social psychology theories concerning prejudice rarely required such reinterpretation, but rather, were simply expanded with regards to intrapersonal sources of variation that were assumed, in their native fields, to have little influence. These integrations are perhaps best understood via the taxonomy of levels of analyses employed by Duckitt (1992; 1994), introduced in Chapter 1, which relegates all but the most general of categorisation processes to a neglected intrapersonal level, the investigation of which is largely not considered the purview of social psychologists.

The preceding journal article summarises the diverse psychological literature (much of which involves specifically evolutionary and comparative approaches within the three key fields) drawn upon to shape the theory of empathy modulation and coalition management described at the beginning of this chapter. As with all new articulations of theory, this integrated model made a series of novel predictions concerning how participants were expected to behave under controlled (in this case, simulated) circumstances, and the quality of the theory as a whole must be judged not only on its coherent arrangement of previous findings, but on how many of its diverse predictions were supported by empirical evidence. Across the three studies, the responses of participants were consistent with the theory's predictions concerning the primacy of coalitional categorisation of the target characters, and the subsequent differential activation of empathetic feelings, given the semi-anonymous position of the protagonist in each scenario, and the emotive facial stimuli provided. This finding was encouraging, but was also expected with great confidence, as specific demonstrations of similar effects have already been noted in the literature (both moral and social cognitive), rendering the prediction comparatively *safe* within the assumptive context of an evolutionary approach to prejudice. The *riskier* predictions concerned the predicted moralistic motivations, which gave rise to the hypotheses concerning the character of justifications and attributions offered by participants following sympathetic and non-sympathetic initial decisions. These predictions are based upon evolutionary rationales (notably Kurzban, DeScioli & Fein, 2012; DeScioli & Kurzban, 2013) which at the time of the writing of this thesis had been primarily modelled in simulations. The success of these predictions, and their conceptual dependence on the empathy modulation predictions that precede them, offer preliminary empirical support for the complete evolutionary theory of that mechanisms underpinning expressions of prejudice proposed in this thesis. While little more can be said with certainty given the limited evidence contained in these three studies, the same adaptationist reasoning that gave

rise to the current predictions can be employed to tentatively expand the scope of this theory with regards to the character of the mechanisms in question, and how they may vary on the level of individuals. These theoretical speculations, and their standing with relation to the empirical findings of this thesis, are discussed in Chapter 8.

CHAPTER 8

General Discussion and Conclusion

Taken on their own, the 5 publications that comprise the bulk of this thesis have made a range of distinct contributions to multiple areas of psychological science. The theoretical publications, most notably the large journal articles of Chapters 4 and 5, have contributed primarily to the clarification of the ongoing issue of disunity in psychology, providing new directions for future integrative efforts. Chapter 4 introduced, and demonstrated the conceptual efficacy of two new conceptual tools designed to enhance the discussion of theoretical differences between psychological fields, and aid in integrative attempts to dissolve apparent barriers between research traditions. Chapter 5 offered a historical perspective on the apparent rifts between differential and evolutionary psychology, and clarified the resulting differences in explanatory approaches, so as to give context to the recent evolutionary breakthroughs that show promise in bridging the two fields. This analysis demonstrated the conceptual flaws that had crept into the atheoretical traditions of differential psychology, and illustrated how further integrative efforts may capitalise on the explanatory power of evolutionary approaches while retaining many of differential psychology's impressive descriptive methods. The two empirical publications comprising the second half of the thesis made additional contributions to empirical literatures of racial prejudice and moral judgment, each of which also introduced novel research methods approaches. The generalised racial category framings and new methodological approach to racial identity explored in Chapter 6 expanded upon the social cognition literature that relies upon reaction-time measures to infer implicit attitudes. Chapter 7 outlined the majority of the evolutionary approach to empathy-modulation and coalition-management developed during this thesis, while detailing the development and empirical verification of a new measurement tool that

utilises ecologically plausible vignettes with conversational cues and evocative facial stimuli to obtain subtle behavioural indications of moral judgment and intergroup prejudice.

This final chapter focuses on the overall goals of this thesis, particularly newly developed evolutionary theory of the empathy-modulation and coalition-management mechanisms underpinning general prejudicial behaviours and evaluations. The successful predictions of this theory, and the methodological approach it has inspired, outlined in Chapter 7 are far from exhaustive, and this thesis will thus conclude with an exploration of the more speculative elements of this new theory, and how the predictions discussed may be explored in future empirical research.

The Interplay of Theory and Measurement

The overall structure of this thesis was written to demonstrate a progression from theoretical development, to the development of appropriate new research methods, and then to the empirical testing of the new theory's predictions. The SATEST tool was specifically designed to measure: (a) overt behavioural indicators of sympathetic evaluations towards the target character; (b) the degree and character of deliberate conscious considerations; (c) evidence of moralising justifications of initial decisions, and; (d) the attribution of traits to the target character that would invalidate the participant's coalitional obligations. Beyond this, however, many of the careful design decisions of the present SATEST manipulations were designed to control for other anticipated variables predicted by the current theory to influence the intrapersonal mechanisms in question. For example, each scenario was carefully constructed to place the participant in a position of control over the target's immediate fate,

but offering the target no means of identifying the participant or the role of their decision in whatever outcome is delivered. This is because the theory predicts that participants are implicitly motivated to manage their reputations, so as to maximise opportunities for alliances and status and minimise the risk of making enemies whenever possible, and excluding the target's awareness of the participants' actions frees the participant from considerations of social retribution that would otherwise influence their decisions. Each SATEST scenario similarly controls for the instance of noticeable physical or emotional harm, instances of clear unfairness, or violations of socially-accepted hygienic or taboo boundaries (in accordance with the fundamental affect-driven moral intuitions of the Social-Intuitionist approach) so as to preserve the affective character of the central trade-off between sympathy for the target character's plight, and the opportunity to punish a relatively harmless social rule transgression. While these controlled variables were essential to establishing the measurement efficacy of the SATEST's central trade-off in the studies reviewed in Chapter 7, each variable also represents an opportunity for testing the cumulative role of these affective influences in shifting the categorical and coalitional evaluations in future manipulations. Precisely how some of these manipulations may be designed, particularly with regards to the content of specific stereotypes, is reviewed in the next subsection of this chapter.

The results of Chapter 7 outline the preliminary support that the application of the SATEST methodology has provided for the integrated evolutionary theory developed in this thesis, part of which predicts that a participants' profile of responses to comparable SATEST scenarios can be meaningfully expressed by index scores which 'count' relevant behavioural indicators. As such, the 5 index scores generated by the SATEST based on the 12 or 6 scenarios measured (depending on whether it is the original or skin-colour design), can be loosely regarded as representing the participant's 'levels' of moral response to the target stimuli

presented. These index levels have demonstrated predictive relationships with other established moral judgment variables (notably the DIT's P-Score), and more tellingly, differences in these index levels between skin-colour conditions have demonstrated predictive relationships with both implicit and explicit measures of racial attitude preferences. Thus, as with the other variables with which they correlate, the computed 'levels' of SATEST indices can be construed as indirect measurements of some varying intrapersonal feature that differentiates those with typically sympathetic reactions from those with typically unsympathetic reactions. This contention is strengthened by the appreciably high test-retest reliabilities for both SATEST manipulations briefly mentioned in Chapter 7.

However, as Chapter 5 outlines, stable and systematic variation in adaptive, evolved mechanisms, often require an explanation within the adaptationist context, because differentially effective mechanisms tend to be refined by selection into a near-optimal species-typical strategy whenever sufficient heredity and phenotypic certainty are present. Any speculation into how such interpersonal variation comes into being is (by necessity) empirically groundless for a theory so new and under-examined as this one. That said, as Chapter 4 argues, fruitful hypothesis testing often requires the tentative acceptance of risky pragmatic assumptions, in order to generate falsifiable predictions whose empirical testing can help rapidly close off the least promising options. As such, of the options explored in Chapter 5 concerning how selective forces acting upon fitness-influencing mechanisms can maintain systematic phenotypic variation, the theory as it currently stands suggests that the observed individual differences are likely to be the result of ontogenically calibrated strategies, which specify particular response thresholds for different categorically identified groups. Although it remains an open possibility that individuals inherit biases towards more or less sympathetic baseline responses (a contention which only behaviour genetic analyses

could support), the ontogenic calibration of inherited response strategies is considered at least a likely contributing factor for two reasons. Firstly, given the extensive literature (reviewed in Chapters 6 and 7) describing the primacy of ingroup-outgroup categorisations in social cognition, and the evolutionary rationales for this phenomenon concerning the probable incidence of intergroup conflict in early human environments, there are strong grounds to expect that partially pre-specified response suites corresponding to viable coalition members (and those flagged as unsuitable allies or possible enemies), could emerge as an adaptive solution to negotiating social uncertainty and the omnipresent threat of group defection. Second, on a purely pragmatic note, empirical evidence for this variety of calibrated strategy-switching is amenable to straightforward hypothesis testing, both via the longitudinal study of children first developing competitive groupings and divisive world-views, and cross-culturally by searching for the consistency of situational responses of adults from cultural backgrounds with widely varying ideologies of intergroup conflict.

With regards to more risky predictions generated by the current theory, it warrants mentioning that the studies explored in Chapter 7 neither expected, nor found, any significant gender differences in any of the SATEST indices in either the moral or prejudicial contexts. This is not to say that gender, one of the most pervasively influential patterns of genetic differentiation in our species, is anticipated to have no role in moral and prejudicial reactions and evaluations addressed by this theory, but rather, that several of the influences most likely to demonstrate gendered differences were explicitly controlled for in the design of these initial SATEST measures. Most notably, the aforementioned elements concerning the relative anonymity and unaccountability of the participant with regards to the decisions they make about the fate of the target characters were controlled for in all of the present SATEST scenarios. The present theory predicts that, when faced with target characters belonging to

what the participant identifies as a threatening outgroup, gender differences may be expected in the type of affective response triggered by the prospect of the target character learning of the participant's power to intervene. Consistent with the outgroup-male framings discussed briefly in Chapter 7, it is predicted that male participants are more likely to respond aggressively to outgroup threats than female participants, who are predicted to experience more intense fear and desires for avoidance, given the partial niche-selection for male competitive aggression in human evolutionary history. Once again, predictions of this sort are pragmatically valuable not only for their consistency with existing evidence and theory, but for the simplicity with which they could be directly empirically disconfirmed in the event that the predictions are mistaken.

Future Directions

Although the integrated theory outlined in this thesis could, in principle, be studied with a wide range of methodological tools, the complementary design of the SATEST measure to the domains of predictions generated by the current theory suggest that the SATEST approach trialled in Chapter 7 will prove uniquely useful in the exploration of empathy-modulation and coalition-management. As such, this section of the thesis is dedicated to outlining a range of possible manipulations of, and additions to, the SATEST methodology that may shed light on the nuances of human prejudice.

Of perhaps primary concern is the reliance of the present SATEST versions on white male characters as defaults. While the SATEST scenarios are worded with personal pronouns, so as to encourage participants to imagine themselves in the vignettes described, the social plausibility of the 'friend' character potentially varies based on the true social cohorts of the participant. While one might expect the presentation of a Caucasian male friend character in

every scenario to be unremarkable to white male participants, this is likely not the case for female participants and participants of non-white racial backgrounds. Given the background and gender similarities of many peer groups, one possible solution to this issue would be to use demographic information taken from the participant to match the friend characters in each scenario to the gender and racial specifications of the participant. Alternatively, SATEST manipulations could employ questions concerning the typical characteristics of the participants' peers, in the guise of a social skills questionnaire or some similar cover. Such modifications cannot be presumed to serve as unambiguous improvements over the existing SATESTs, however, since Eurocentric tendencies in many Western nations may result in non-white participants none-the-less identifying with white or male characters as less intrusive typical features in a simulated social situation. This concern could perhaps also be addressed via the inclusion of racial identity measures, such as the approach outlined in Chapter 6, which could both inform the friend and target character features shown to participants, and offer additional modifying variables to consider in the subsequent analyses.

Similarly, although the SATEST methodology was designed to be sensitive to the underlying mechanisms theorised to be responsible for many of the shared characteristics of multiple forms of prejudice, only its efficacy in detecting racially-motivated differences in responding has been established. Racial prejudice was selected in the aforementioned studies for both its prevalence in the various prejudice literatures, and for the ease with which target characters with a different skin-tone could be substituted into the existing SATEST framework with only few methodological modifications. That said, the studies reviewed in Chapter 7 provide a proof-of-concept for the efficacy of the central trade-offs of the SATEST in measuring the variables they were designed for, and as such it is likely that variants of the SATEST which modify their scenarios to subtly highlight other target characteristics will prove equally

effective, provided confounding influences are equally well controlled. The obvious candidates for such future studies include other simple visual characteristics of the target character that can be substituted into existing scenarios in much the same manner as the skin-colour manipulation, such as modifications of the target's gender, age, weight, level of wealth as indicated by style of dress, and status as handicapped or able-bodied. It may also be possible to compose comparable vignettes, in which the target character is temporarily depicted in the presence of their romantic partner or other cohort. Manipulations of this sort could be applied to comparisons between singled and married individuals, conceptions of guilt-by-association for targets presented with perceived undesirable affiliates (such as obvious gang-members), and of course, sexual preference or partner-choice prejudices (through the depiction of same-sex couples, interracial couples, couples with large age-gaps, etc.). Provided the conflict of each scenario remains a victimless but easily-understood social rule violation, and the presentation of the target occurs in a context where the participant need not fear discovery by, or retribution from, the target, the presently verified properties of the SATEST scenarios are expected to function as has been observed.

SATEST approaches to the study of moral judgement and prejudicial behaviour could also be modified through the strategic manipulation of elements whose affective influence on participants in the present studies was meticulously controlled, specifically, those relating to the fundamental moral intuitions described by Social-Intuitionist theories. While the manipulation of select scenarios to include elements of overt harm, obvious unfairness, the violation of taboos or antagonism between groups would likely prove instructive in its own right, a greater degree of nuance in participant responses could likely be obtained via the selective inclusion of stereotype content into specific SATEST vignettes (particularly in prejudicial responding). For example, depicting scenarios in which certain racial outgroup

members are framed as being unreasonably violent or entitled (as is a common stereotype associated with socio-economically disadvantaged racial minorities in many Western nations) can be expected to not only negatively impact upon sympathetic decisions and subsequent attributions, but to disproportionately affect these evaluations in those individuals who have internalised such stereotypes to a greater degree. Of particular value, such manipulations may be employed to contrast reactions to both stereotypical and non-stereotypical presentations of outgroup members, perhaps even including single instances of a stereotyped outgroup target as a priming stimulus before later presenting non-stereotyped presentations.

With regards to priming effects, simple SATEST methodologies could also be employed in conjunction with contextual primes, ideally in repeated-measures designs, to measure the effect of particular primes on the typical SATEST response profiles of participants. Priming participants with stimuli pertaining to mortality salience, specific intergroup conflict, and a wide range of negative affect primes such as disgust and anger, may all be reasonably expected to reduce empathetic activation in participants, whereas priming affects such as sadness may have complex effects depending on the initial response profile the participant (for example, highly compassionate individuals may become more sympathetic when upset, whereas low-compassion individuals may disengage even further when sadness is induced to impair their motivations). Given the demonstrably low cognitive load designed into the presentation of the SATEST (as study 2 in Chapter 7 suggests), it is also predicted that participants will lean more towards negative evaluations and increased deliberations if they are instructed to undertake the SATEST with a simultaneous cognitive load task, or when instructed to abstain from food so as to engineer low blood-sugar.

The final concern for the future of the SATEST concerns the possibility of improvements to immersion of the scenarios, which can only be obtained through substantial technical efforts, and at the cost of several presentation ambiguities that may indeed work in the measure's favour. This concern was addressed largely in the journal article featured in Chapter 7, although one aspect that was not addressed was the ease with which the combination of multiple target characteristics may only be possible with higher detail images and settings. For example, the present simple art-style of the SATEST visual stimuli render it difficult, but possible, to depict large variations in the target character's age, or large variations in the target's weight or skin-clarity, but to depict two or more of these conditions (in contrast to the simplified default characters employed at present) at once would prove a remarkable struggle, particularly if trying to maintain the consistent simple lines that allow target expressions to be easily observed from a distance. Though some degree of nuanced depiction may be achieved by adding some details to the text of the narrated vignette, to do so would invalidate the subtlety with which target features are depicted in the SATEST task, and thus ultimately reducing the SATEST's projected ability to overcome impression-management on the part of the participant by obscuring the nature of the decisions the participant is asked to make.

Conclusion

From its inception, the central goal of this thesis was to explore several of the key intrapersonal psychological processes thought to underpin prejudicial behaviours and beliefs in humans. Early reviews of the literature concerning intrapersonal mechanisms of prejudice revealed that several distinct research fields—social, differential, and moral psychology—each offered a range of insights into the psychological phenomena of interest, but conceptualised these insights within highly dissimilar and seemingly incompatible research

traditions. In order to integrate these diverse findings into a coherent whole, this thesis endeavoured to expand upon noteworthy evolutionary psychology studies undertaken within each of the respective fields, as these promising efforts all employed a common conceptual and theoretical methodology: the adaptationist approach. By applying an adaptationist analysis to the psychological phenomena shared by the three fields, this thesis integrated their respective insights into a single evolutionary synthesis, which focuses primarily on the hypothesised mechanisms governing the modulation of empathy and coalition management. From the specific predictions generated by this integrated theory, the final components of this thesis outlined the development of a new psychometric measure, designed to simultaneously measure five classes of behavioural indicators theorised to be directly related to the underlying mechanisms of interest. The testing of the SATEST measurement tool yielded empirical results almost entirely consistent with the predictions of the new theory, offering both evidence for the psychometric veracity of the methodological design, and preliminary support for the theory itself. With the inclusion of the individual theoretical, methodological and empirical achievements of the 5 publications incorporated into this manuscript, the contributions of this thesis to the science of psychology are many and varied, and its overall approach is intended to illustrate the value of robust theoretical and conceptual analyses in a discipline that remains largely divided by uncritically perpetuated research traditions.

References for the Unpublished

Sections of this Thesis

- About, F. E. (2003). The formation of ingroup favoritism and outgroup prejudice in young children. *Developmental Psychology*, 39, 48-60.
- Akrami, N., Ekehammar, B., & Bergh, R. (2011). Generalized prejudice: Common and specific components. *Psychological Science*, 22 (1), 57-59.
- Allport, G.W. (1954). *The Nature of Prejudice*. Addison-Wesley, Cambridge.
- Allport, G.W., & Kramer, B.M. (1946). Some roots of prejudice. *Journal of Psychology*, 22, 9-39.
- Anderson, N. (1996). *A Functional Theory of Cognition*. Hillsdale, N.J: L. Erlbaum Associates.
- Baars, B. (1984). View from a road not taken. *Contemporary Psychology*, 29, 804-805.
- Baars, B. (1985). The logic of unification. *Contemporary Psychology*, 30, 340.
- Barlow, F.K., Louis, W.R., & Hewstone, M.(2009). Rejected! cognitions of rejection and intergroup anxiety as mediators of the impact of cross-group friendships on prejudice. *British Journal of Social Psychology*, 48 (3), 389-405.
- Beal, D.J., O'Neal, E.C.O., Ong, J., & Ruscher, J.B. (2000). The Ways and Means of Interracial Aggression: Modern Racists' Use of Covert Retaliation. *Personality and Social Psychology Bulletin*, 26(10), 1225-1238.
- Bennett, M.R. & Hacker, P.M.S. (2003). *Philosophical Foundations of Neuroscience*. Blackwell Publishing.
- Bernstein, M.J., Sacco, D.F., Young, S.G., Hugenberg, K., & Cook, E. (2010). Being "in" with the in-crowd: The effects of social exclusion and inclusion are

- enhanced by the perceived essentialism of ingroups and outgroups. *Personality and Social Psychology Bulletin*, 36 (8), 999-1009.
- Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., Mummendey, A., Maquil, A., Demoulin, S., & Leyens, J.P. (2009). Does Contact Reduce Prejudice or Does Prejudice Reduce Contact? A Longitudinal Test of the Contact Hypothesis Among Majority and Minority Groups in Three European Countries. *Journal of Personality and Social Psychology*, 96 (4), 843-856.
- Bizman, A. & Yinon, Y. (2001). Intergroup and interpersonal threats as determinants of prejudice: The moderating role of in-group identification. *Basic and Applied Social Psychology*, 23(3), 191-196.
- Blair, I.V., Judd, C.M., Sadler, M.S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality & Social Psychology*, 83(1), 5-25.
- Boag, S. (2011). Explanation in personality research: ‘verbal magic’ and the Five-Factor Model. *Philosophical Psychology*, 24, 223-243.
- Bower, G.H. (1993). The fragmentation of psychology? *American Psychologist*, 48, 905-907.
- Brewer, M.B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86, 307-324.
- Bromley, Y.V. (1987). Anthropology, Ethnology and Ethnic and Racial Prejudice. *International Social Science Journal*, 39(1), 31-43.
- Buss, D.M. (1988). The evolution of human intrasexual competition: tactics of mate attraction. *Journal of Personality and Social Psychology*, 54(4), 616-628.

- Buss, D.M. (2005). *The handbook of evolutionary psychology*. Hoboken, NJ, US: John Wiley & Sons Inc; US.
- Buss, D. M., & Hawley, P. H. (2011). *The evolution of personality and individual differences*. New York: Oxford University Press.
- Butz, D., & Yogeeswaran, K. (2011). A new threat in the air: Macroeconomic threat increases prejudice against Asian Americans. *Journal of Experimental Social Psychology*, 47, 22-27.
- Callahan, M.P., & T.K. Vescio. 2011. Core American values and the structure of antigay prejudice. *Journal of Homosexuality*, 58, 248-262.
- Caplan, D. (2009). Experimental design and interpretation of functional neuroimaging studies of cognitive processes. *Human Brain Mapping*. 30 (1), 59-77.
- Carr, P.B. & Steele, C.M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology*, 45, 853-859.
- Cervone, D. (1999). Bottom-up explanation in personality psychology: The case of cross-situational coherence. In D. Cervone & Y. Shoda (Eds.), *The coherence of personality: Social-cognitive bases of personality consistency, variability, and organization* (pp. 303-341). New York: Guilford Press.
- Cervone, D. (2004). The architecture of personality. *Psychological Review*, 111, 183-204.
- Christ, O., Hewstone, M., Tausch, N., Wagner, U., Voci, A., Hughes, J., & Cairns, E. (2010). Direct contact as a moderator of extended contact effects: Cross-sectional and longitudinal impact on outgroup attitudes, behavioral intentions, and attitude certainty. *Personality and Social Psychology Bulletin*, 36 (12), 1662-1674.

- Christensen, J.F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience & Biobehavioral Reviews*, 36(4), 1249-1264.
- Cikara, M., Farnsworth, R.A., Harris, L.T., & Fiske, S.T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social Cognitive and Affective Neuroscience*, 5, 404-413.
- Cohen, E. (2012). The evolution of tag-based cooperation in humans: The case for accent. *Current Anthropology*, 53(5), 588-616.
- Cohrs, J. C., & Asbrock, F. (2009). Right-wing authoritarianism, social dominance orientation and prejudice against threatening and competitive ethnic groups. *European Journal of Social Psychology*, 32, 270–289.
- Condor, S. & Brown, R. (1988). Psychological processes in intergroup conflict. In W. Stroebe, A. Kruglanski, D. Bar-Tal, & M. Hewstone (Eds.), *The social psychology of intergroup conflict* (p. 3-26). Springer, Berlin.
- Confer, J. C., Easton, J. A., Fleischman, D. S., Goetz, C. D., Lewis, D. M., Perilloux, C., & Buss, D. M. (2010). Evolutionary Psychology: Controversies, Questions, Prospects, and Limitations. *American Psychologist*, 65, 110-126.
- Cosmides, L., Tooby, J. & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences* 7(4), 173-179.
- Cox, W.T.L., Abramson, L.Y., Devine, P.G., & Hollon, S.D. (2012). Stereotypes, Prejudice, and Depression: The Integrated Perspective. *Perspectives on Psychological Science*, 7(5), 427–449.

- Crandall, C.S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, 129(3), 414–446.
- Crandall, C.S., Eshleman, A., & O'Brien, L.T. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82, 359-378.
- Daly, M. & Wilson, M. (2008). Is the "Cinderella effect" controversial?: A case study of evolution-minded research and critiques thereof. In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology*. (pp. 383-400). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.
- Dasgupta, N., & Rivera, L. M. (2006). From automatic anti-gay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control. *Journal of personality and Social Psychology*, 91, 268-280.
- Dawkins, R. (2009). *The Greatest Show on Earth: The Evidence for Evolution*. Free Press.
- de Groot, A.D. (1990). Unifying psychology: A European view. *New Ideas in Psychology*, 8, 309-320.
- De Los Reyes, A. & Kazdin, A.E. (2008). When the evidence says, "Yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science*, 17, 47-51.
- de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10(10), 435-441.

- de Waal, F.B.M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, 59, 279-300.
- Decety, J., & Grézes, J. (2006). The power of simulation: Imagining one's own and other's behaviour. *Cognitive Brain Research*, 1079, 4-14.
- Denton, D.A., McKinley, M.J., Farrell, M., & Egan, G.F. (2009). The role of primordial emotions in the evolutionary origin of consciousness. *Consciousness and Cognition*, 18(2), 500-514.
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477-496.
- Dixon, R.A. (1983). Theoretical proliferation in psychology: A plea for sustained disunity. *The Psychological Record*, 33, 337-340.
- Dovidio, J.F., Glick, P., & Rudman, L.A. (2005). *On the Nature of Prejudice*. Blackwell Publishing, Malden.
- Duckitt, J. (1992). Psychology and prejudice: A historical analysis and integrative framework. *American Psychologist*, 47(10), 1182-1193.
- Duckitt, J. (1994). *The Social Psychology of Prejudice*. Praeger, New York.
- Duckitt, J., & Sibley, C.G. (2010). Personality, Ideology, Prejudice, and Politics: A Dual-Process Motivational Model. *Journal of Personality*, 78(6), 1861-1894.
- Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry*, 6, 267-285.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.

- Eagly, A.H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15, 543-558.
- Ellis, R.D. (2005). The roles of imagery and metaemotion in deliberate choice and moral psychology. *Journal of Consciousness Studies* 12 (8-10), 140-157.
- Fairchild, H., & Gurin, P. (1973). Traditions in the social psychological analysis of race relations. *American Behavioral Scientist*, 21, 757-778.
- Feather, N.T., & Atchison, L. (1998). Reactions to an Offence in Relation to the Status and Perceived Moral Character of the Offender. *Australian Journal of Psychology*, 50 (2), 119-127.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Fiske, S.T., Cuddy, A.J.C., Glick, P. & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902.
- Fitzgerald, C. J., & Whitaker, M. B. (2010). Examining the acceptance of and resistance to evolutionary psychology. *Evolutionary Psychology*, 8(2), 284-296.
- Fodor, J. (1975). *The Language of Thought*, Harvester Press.
- Gabarrot, F., Falomir-Pichastor, J.M., & Mugny, G. (2009). Being similar versus being equal: Intergroup similarity moderates the influence of in-group norms on discrimination and prejudice. *British Journal of Social Psychology*, 48 (2), 253-273.

- Gaertner, S.L., & Dovidio, J.F. (1986). The aversive form of racism. In J.F. Dovidio and S.L. Gaertner (Eds.), *Prejudice, Discrimination and Racism: Theory and Research* (pp. 61-89). Academic Press, Orlando.
- Garth, T.R. (1930). A review of race psychology. *Psychological Bulletin*, 27(5), 329–356.
- Gilgen, A.R. (1987). *The psychological level of organization in nature and interdependencies among major psychological concepts*. In A.W. Staats & L.P. Mos (Eds.), *Annals of theoretical psychology*, Vol. 5. (pp. 179-209). New York, NY, US: Plenum Press; US.
- Gintis, H. (2007). A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences*, 30, 1-61.
- Gladin, L.L. (1961). Toward a unified psychology. *Psychological Record*, 11, 405-421.
- Goertzen, J.R. (2008). On the Possibility of Unification: The Reality and Nature of the Crisis in Psychology. *Theory & Psychology*, 18(6), 829-852.
- Greene, J.D. (2007) Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, 11(8), 322-323.
- Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Press, New York.
- Greene, J., & Haidt, J. (2002) How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517-523.

- Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., & Mellott, D.S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109, 3–25.
- Greenwald, A.G, Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- Greenwald, A.G., Poehlman, T.A., Uhlmann, E., & Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17–41.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A.G, Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- Griskevicius, V., Shiota, M.N., & Neufeld, S.L. (2010). Influence of Different Positive Emotions on Persuasion Processing: A Functional Evolutionary Approach. *Emotion*, 10(2), 190-206.
- Gutsell, J. N., & Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, 46, 841-845.
- 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. (2004). The emotional dog gets mistaken for a possum. *Review of General Psychology*, 8(4), 283-290.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1002.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon, New York.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116.
- Harenski, C.L., Harenski, K.A., Shane, M.S., Kiehl, K.A. (2010). Aberrant neural processing of moral violations in criminal psychopaths. *Journal of Abnormal Psychology* 119 (4), 863-874.
- Hart, A.J., Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., & Rauch, S.L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *NeuroReport*, 11, 2351-2355.
- Haslam, S.A., & Wilson, A. (2000). In what sense are prejudicial beliefs personal? The importance of an in-group's shared stereotypes. *British Journal of Social Psychology*, 39 (1), 45-63.
- Hawley, P.H. (1999). The Ontogenesis of Social Dominance: A Strategy-Based Evolutionary Perspective. *Developmental Review*, 19 (1), 97-132.
- Henriques, G. (2003). The Tree of Knowledge System and the theoretical unification of psychology. *Review of General Psychology*, 7, 150-182.
- Henriques, G.R. (2004). Psychology defined. *Journal of Clinical Psychology*, 60, 1207-1221.

- Henriques, G.R. (2008). The problem of psychology and the integration of human knowledge: Contrasting Wilson's consilience with the Tree of Knowledge System. *Theory & Psychology, 18*, 731-755.
- Huebner, B., Dwyer, S., Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences, 13*(1), 1-6.
- Jaszczolt, K. (1996). Relevance and infinity: Implications for discourse interpretation. *Journal of Pragmatics, 25*(5), 703-722.
- Jayawickreme, E. & Chemero, A. (2008). Ecological Moral Realism. *Review of General Psychology, 12*, 118-126.
- Kteily, Sidanius, & Levin, 2011
- Judd, C.M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100*, 109-128.
- Kantor, J.R. (1979). Psychology: Science or nonscience? *The Psychological Record, 29*, 155-163.
- McGregor, I., Haji, R., & Kang, S.J. (2008). Can ingroup affirmation relieve outgroup derogation? *Journal of Experimental Social Psychology, 44*(5), 1395-1401.
- Kassinove, J.I. (2002). As defined, unification is inevitable. *American Psychologist, 57*, 1127.
- Kelly, R.J. (1998). The crisis in psychology: Trouble in the temple. *Journal of Social Distress and the Homeless, 7*, 211-223.
- Kohlberg, L. (1969). Stage and sequence: the cognitive developmental approach to socialization. In D.A. Goslin (Ed.) *Handbook of Socialization Theory* (pp. 347-480). Rand McNally, Chicago.
- Kimble, G.A. (1996). *Psychology: The hope of science*. MIT Press, Cambridge.

- Klineberg, O. (1968). Prejudice: The Concept. In *International Encyclopaedia of the Social Sciences* (p. 439-448). Macmillan and Free Press, New York.
- Knowles, E.D., & Peng, K. (2005). White Selves: Conceptualizing and Measuring a Dominant-Group Identity. *Journal of Personality and Social Psychology*, 89, 223-241.
- Krebs, D.L. (2008). Morality: An evolutionary account. *Perspectives on Psychological Science*, 3, 149-172.
- Krebs, D. L. & Denton, K. (2006). Explanatory limitations of cognitive-developmental approaches to morality. *Psychological Review*, 113(3), 672-675.
- Krebs, D.L., & Van Hesteren, F. (1994). The development of altruism: Toward an integrative model. *Developmental Review*, 14, 1-56.
- Kreindler, S.A. (2005). A dual group processes model of individual differences in prejudice. *Personality and Social Psychology Review*, 9, 90-107.
- Kristjánsson, K. (2009). Does moral psychology need moral theory?: The case of self-research. *Theory and Psychology* 19(6), 816-836.
- Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, 33, 323-333.
- Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*, 127(2), 187-208.
- Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, 98(26), 15387-15392.

- Maddox, K.B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, 8, 383-401.
- Magnusson, D. (2000). The individual as the organizing principle in psychological inquiry: A holistic approach. In L. R. Bergman, R. B. Cairns, L. G. Nilsson, & L. Nystedt (Eds.), *Developmental science and the holistic approach* (pp. 33–47). Mahwah, NJ: Erlbaum.
- Matarazzo, J.D. (1987). There is only one psychology, no specialties, but many applications. *American Psychologist*, 42, 893-903.
- Matarazzo, J.D. (1992). The unity or diversity of psychology: Concluding remarks. *International Journal of Psychology*, 27, 327-330.
- McCoy, S.K., & Major, B. (2003). Group identification moderates emotional responses to perceived prejudice. *Personality and Social Psychology Bulletin*, 29, 1005-1017.
- McConahay, J.B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). Orlando FL: Academic Press.
- McDonald, M.M., Navarrete, C.D., & Van Vugt, M. (2012). Evolution and the Psychology of Intergroup Conflict: The Male Warrior Hypothesis. *Philosophical Transactions of the Royal Society-Biological Sciences*, 367(1589), 670-679.
- McGrane, J. A., & White, F. A. (2007). Differences in Anglo and Asian Australians' explicit and implicit prejudice and the attenuation of their in-group bias. *Asian Journal of Social Psychology*, 10, 204-210.

- McGregor, I., Haji, R., & Kang, S.J. (2008). Can ingroup affirmation relieve outgroup derogation? *Journal of Experimental Social Psychology*, 44 (5), 1395-1401.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Miller, C.H., Burgoon, J.K., & Hall, J.R. (2007). The effects of implicit theories of moral character on affective reactions to moral transgressions. *Social Cognition*, 25 (6), 819-832.
- Miller, S.L., Zielaskowski, K., Maner, J.K., & Plant, E.A. (2012). Self-protective motivation and avoidance of heuristically threatening outgroups. *Evolution and Human Behavior*, 33, 726-735.
- Milner, D. (1983). *Children and race: Ten years on*. Ward Lock Educational, London.
- Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, 16(3), 696-703.
- Moll, J., de Oliveira-Souza, R., Moll, F.T., Ignácio, F.A., Bramati, I.E., Caparelli-Dáquer, E.M., & Eslinger, P.J. (2005). The moral affiliations of disgust: a functional MRI study. *Cognitive and Behavioural Neurology*, 18(1), 68-78.
- Monin, B. & Miller, D.T. (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology*, 81(1), 33-43.
- Morrow, D. (2009). Moral psychology and the "Mencian creature". *Philosophical Psychology* 22(3), 281-304.
- Nairne, J. S. (1997). *Psychology: The adaptive mind*. Brooks/Cole, California.

- Narvaez, D. (2001). Moral Text Comprehension: implications for education and research. *Journal of Moral Education*, 30(1), 43-54.
- Narvaez, D. & Bock, T. (2002). Moral Schemas and Tacit Judgement or How the Defining Issues Test is Supported by Cognitive Science. *Journal of Moral Education*, 31(3), 297-314.
- Navarrete, C.D., McDonald, M., Molina, L., & Sidanius, J. (2010). Prejudice at the nexus of race and gender: An out-group male target hypothesis. *Journal of Personality & Social Psychology*, 98(6), 933-45.
- Neisser, U. (1995). The unity of psychology and of persons. *International Newsletter of Uninomic Psychology*, 15, 6-12.
- Nesdale, D., Maass, A., Durkin, K & Griffiths, J. (2005). Group norms, threat and children's ethnic prejudice. *Child Development*, 76(3), 1-12.
- Nesdale, D., Maass, A., Kiesner, J., Durkin, K., Griffiths, J., & Ekberg, A. (2007). Effects of peer group rejection, group membership, and group norms, on children's outgroup prejudice. *International Journal of Behavioral Development*, 30, 526–535.
- Nesdale, D., Durkin, K., Maass, A., Kiesner, J., Griffiths, J., Daly, J., & McKenzie, D. (2010). Peer group rejection and children's outgroup prejudice. *Journal of Applied Developmental Psychology*, 31(2), 134-144.
- Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge.
- Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166-180.

- Nowak, M.A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291-1298.
- O'Connell, S.M. (1995). Empathy in chimpanzees: Evidence for theory of mind? *Primates*, 36, 397-410.
- Overmeier, J.B. (1989). Uninomics and learning theory. *International Newsletter of Uninomic Psychology*, 8, 10-14.
- Paxton, J.M., Ungar, L., & Greene, J.D. (2011) Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.
- Pearson, A.R., Dovidio, J.F., & Gaertner, A.L., (2009). The Nature of Contemporary Prejudice: Insights from Aversive Racism. *Social and Personality Psychology Compass*, 3, 1-25.
- Plant, E. A., Devine, P.G., & Peruche, B.M. (2010). Regulatory Concerns for Interracial Interactions: Approaching Egalitarianism versus Avoiding Prejudice. *Personality and Social Psychology Bulletin*, 36, 1135-1147.
- Page-Gould, E., Mendoza-Denton, R., Alegre, J.M., & Siy, J.O. (2010). Understanding the Impact of Cross-Group Friendship on Interactions With Novel Outgroup Members. *Journal of Personality and Social Psychology*, 98(5), 775-793.
- Pratto, F., & Shih, M. (2000). Social dominance orientation and group context in implicit group prejudice. *Psychological Science*, 11(6), 515-518.

- Rai, T.S. & Fiske, A.P. (2011). Moral Psychology is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality. *Psychological Review*, 118, 57-75.
- Rest, J.R. (1975). Longitudinal Study of the Defining Issues Test of moral judgment: A strategy for analyzing developmental change. *Developmental Psychology*, 11(6), 738-748.
- Rest, J.R., Narvaez, D., Bebeau, M.J., & Thoma, S.J. (1999). *Postconventional Moral Thinking: a neo-Kohlbergian approach*. Lawrence Erlbaum, New Jersey.
- Rest, J.R., Narvaez, D., Thoma, S.J., & Bebeau, M.J. (2000). A Neo-Kohlbergian Approach to Morality Research. *Journal of Moral Education*, 29, 381-396.
- Reynolds, K. J., Turner, J. C. Haslam, S. A., & Ryan, M. K. (2001). The role of personality and group factors in explaining prejudice. *Journal of Experimental Social Psychology*, 37, 427–434.
- Richman, W.L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires and interviews. *Journal of Applied Psychology*, 84, 754-775.
- Riek, B.M., Mania, E.W., & Gaertner, S.L. (2006). Intergroup Threat and Outgroup Attitudes: A Meta-Analytic Review. *Personality and Social Psychology Review*, 10(4), 336–353.

- Roets, A., & Van Hiel, A. (2011). Allport's Prejudiced personality today: Need for closure as the motivated cognitive basis of prejudice. *Current Directions in Psychological Science*, 20, 349-354.
- Sanders, C.E., Lubinski, D., & Benbow, C.P. (1995). Does the Defining Issues Test measure psychological phenomena distinct from verbal ability?: An examination of Lykken's query. *Journal of Personality and Social Psychology*, 69, 498-504.
- Schaller, M., Boyd, C., & Yohannes, J.O.M. (1995) The prejudiced personality revisited: personal need for structure and formation of erroneous group stereotypes. *Journal of Personality & Social Psychology*, 68, 544-555.
- Shapiro, J.R., Neuberg, S.L. (2008). When Do the Stigmatized Stigmatize? The Ironic Effects of Being Accountable to (Perceived) Majority Group Prejudice-Expression Norms. *Journal of Personality and Social Psychology*, 95(4), 877-898.
- Sherif, M., Harvey, O. J., White, B.J., Hood, W.R., & Sherif, C.W. (1988). *The Robbers Cave Experiment: Intergroup Conflict and Cooperation*. Wesleyan University Press, Connecticut.
- Sibley, C.G., Harding, J.F., Perry, R., Asbrock, F., & Duckitt, J. (2010). Personality and Prejudice: Extension to the HEXACO Personality Model. *European Journal of Personality*, 24, 515-534.
- Sidanius, J., & Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge University Press, New York.
- Sober, E. (1998). Black box inference: When should intervening variables be postulated? *British Journal for the Philosophy of Science*, 49, 469-498.

- Sober, E. (2000). *Philosophy of Biology*. Westview Press, Colorado.
- Son Hing, L.S., Chung-Yan, G.A., Hamilton, L.K., & Zanna, M.P. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology*, 94(6), 971-987.
- Staats, A.W. (1983). *Psychology's crisis of disunity: Philosophy and method for a unified science*. New York: Praeger.
- Staats, A.W. (1996). *Behavior and personality: Psychological behaviorism*. New York: Springer.
- Staats, A. W. (1999). Unifying psychology requires new infrastructure: Theory, method, and a research agenda. *Review of General Psychology*, 3, 3–13.
- Stam, H.J. (2004). Unifying psychology: Epistemological act or disciplinary maneuver? *Journal of Clinical Psychology*, 60, 1259-1262.
- Sternberg, R. J., & Grigorenko, E. L. (2001). Unified psychology. *American Psychologist*, 56, 1069 –1079.
- Stephan, C.W., Stephan, W.G., Demitrakis, K.M., Yamada, A.M., Clason, D.L. (2000). Women's attitudes toward men: An integrated threat theory approach. *Psychology of Women Quarterly*, 24, 63-73.
- Stevenson, D. & Goldworth, A. (2002). Ethical considerations in neuroimaging and its impact on decision-making for neonates. *Brain and Cognition*, 50(3), 449–454.
- Tajfel, H. (1982). Social psychology of intergroup attitudes. *Annual Review of Psychology*, 33, 1-39.
- Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for outgroup members. *British Journal of Social Psychology*, 48(3), 427-446.

- Tashiro, M. (2004). *Impacts of Neuroimaging on Psycho-Oncology*. *Psycho-Oncology*, 13(7), 486-489.
- Terrizzi, J. A., Jr., Shook, N. J., & Ventis, W. L. (2010). Disgust: A predictor of social conservatism and prejudicial attitudes toward homosexuals. *Personality and Individual Differences*, 49, 587-592.
- Tooby, J. & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 5-67). Hoboken, NJ: Wiley.
- Tooby, J. & Cosmides, L. (2007). Evolutionary psychology, ecological rationality, and the unification of the behavioral sciences. Comment on, A framework for the unification of the behavioral sciences by Gintis. H. *Behavioral and Brain Sciences* 30(1), 42-43.
- Tropp, L.R., & Pettigrew, T.F. (2005). Relationships between inter group contact and prejudice among minority and majority status groups. *Psychological Science*, 16 (12), 951-957.
- Tsang, J. (2002). Moral rationalization and the integration of situational factors and psychological processes in immoral behavior. *Review of General Psychology*, 6, 25-50.
- Viney, W. (1996). Disunity in psychology and other sciences: The network or the block universe? *Journal of Mind and Behavior*, 17, 31-43.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Alfred A. Knopf.
- Yanchar, S.C., & Slife, B.D. (1997). Pursuing unity in a fragmented psychology: Problems and prospects. *Review of General Psychology*, 1, 235-255.

Zuroff, D.C., Fournier, M.A., Patall, E.A., & Leybman, M.J. (2010). Steps toward an evolutionary personality psychology: Individual differences in the social rank domain. *Canadian Psychology*, 51(1), 58-66.

APPENDIX A

The SATEST Scenarios

Scenario 1: Parking Intervention

Condition = Other + Action

Begin

[Scene dark] For this scenario, imagine you work in an office. [Illuminate into street view] You are walking in to work, ahead of schedule, with your friend and co-worker (friend's name) [Rotate view to show friend], who is one of the ground floor security staff. You are discussing with (friend's name) your plans for the weekend...

Friend: It'll be nice to just relax, it's been a busy week.

...when you notice a co-worker from your department [Show target window], (target name), looking out from one of the meeting room windows. You are not friends with (target name), though you have spoken on occasion. [Zoom in on target, looks upset/desperate] (She/He) is looking, with a desperate expression, at the street outside the building. [Closer view of street] In one of the paid parking spaces on the street, which require the purchase of a ticket, you notice (target name)'s car. A parking attendant is standing beside (her/his) car, issuing a parking penalty. [Show target window, close] It's clear that (target name) is in an early meeting, and cannot leave to attend to (her/his) car. (She/he) seems upset, since parking penalties are never cheap. [Turned to friend] (Friend's name) has noticed the situation as well.

Friend: That's rough. Do you know (her/him)? I could speak to the attendant, if you like, get them to overlook this one. It wouldn't be any trouble. Should I?

How will you respond to (friend's name)?

- Yes >
 - "I guess you could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."

- No >
 - “You don’t have to.”
 - “No, don’t worry about it.”
 - “Don’t bother, (she/he) shouldn’t have parked without a ticket.”
- “Let me think about it.”

IF “Let me think about it” or time interval exceeding 8 seconds

Friend: Okay, there’s no big rush. What’s on your mind?

What are you considering?

- “I’m thinking about the importance of parking fines.” [If selected, show car and attendant.]
- “I’m thinking about (target name).” [If selected, show target, close]
- “I’m thinking about whether this is right or not.” [If selected, show broad street view]
- “Nothing, I’ve made up my mind.” [If select, go straight to decision.]

Friend: Right, sure. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess you could.”
 - “Sure, I don’t see why not.”
 - “Definitely, let’s help (her/him) out.”
- No >
 - “You don’t have to.”
 - “No, don’t worry about it.”
 - “Don’t bother, (she/he) shouldn’t have parked without a ticket.”

IF helping

Friend: Okay, sure. This’ll just take a moment.

[View friend talking to attendant] (Friend's name) speaks to the attendant for a minute. They exchange nods, and the attendant takes back the penalty ticket and continues walking down the row of cars. [Facing friend]

Friend: All taken care of. So why did you decide to help (her/him) out?

What motivated your decision?

- "I just felt like being nice."
- "(Target name) would be grateful."
- "I don't think (target name) will park without paying again."
- "Helping people out is more important than obeying parking regulations."

IF not helping

Friend: Okay, no problem then.

[View front of the building, past the window and the car] The two of you continue toward the front entrance of the building. [Turned to friend]

Friend: So why did you decide not to help?

What motivated your decision?

- "I just didn't feel like helping (her/him) out."
- "(Target name)'s parking is none of our business."
- "(Target name) will just keep parking without paying if (she/he) doesn't experience the consequences."
- "Rules like parking regulations are in place for a reason, you can't just make exceptions to them."

Ending

[View building interior, friend in foreground]

Friend: I understand. Why do you think (she/he) parked there in the first place?

How do you answer?

- “(She/He) is probably just inconsiderate.”
- “(She/He) was probably in hurry.”
- “(She/He) is probably just reckless.”
- “(She/He) might not have had any change.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. I’ll talk to you later, okay?

[Empty interior view] You and (friend’s name) part ways, and you head to your office.

Scenario 2: Overlooking Emails

Condition = Self + Inaction

Begin

[Scene dark] For this scenario, imagine you work in an office. [Illuminate into cubicle view] You are just beginning your lunch break, and are being visited by your friend (friend’s name). [Show friend, eating a sandwich] A handful of people in your department, yourself included, have been asked to occasionally look over the department’s email server, on the off chance of discovering someone using their office computers to send personal emails. [Broad cubicle view] It is the company’s policy that sending personal emails from work is not allowed, and there are penalties in place for those caught doing so. You have decided to quickly perform your unofficial duty now, before you and (friend’s name) head out. [Show friend, talking]

Friend: Don’t rush on my account, I’ve already picked up my lunch.

[Show computer] In looking over the lists, you notice that someone has just recently sent an email, which from the address and subject heading appears to be personal. [Modify computer view to highlight one email, text unclear] The sender was (target name), who you don’t know

very well, but have seen many times since (her/his) cubicle is just down the row from yours. [Show row view, target visible at chair, target looks nervous and distracted/worried] You explain what you have discovered to (friend's name). [Show friend, talking]

Friend: It seems like (she/he)'s been caught in the act. Are you going to email the supervisor and report it?

Are you going to report (target name) to your supervisor?

- Yes >
 - "I guess so."
 - "I suppose I should."
 - "Yes, definitely, personal emails aren't allowed."
- No >
 - "I might not."
 - "I don't think I will."
 - "No, I'll let them off the hook this time."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: It's your call. What's on your mind?

What are you considering?

- "I'm thinking about the importance of office policies." [If selected, email list]
- "I'm thinking about (target name)." [If selected, show target, close]
- "I'm thinking about whether this is right or not." [If selected, show broad cubicle view]
- "Nothing, I've made up my mind." [If select, go to decision.]

Friend: I see. Anything else? [Loop back]

What have you decided?

- Yes >
 - "I guess so."
 - "I suppose I should."
 - "Yes, definitely, personal emails aren't allowed."

- No >
 - “I might not.”
 - “I don’t think I will.”
 - “No, I’ll let them off the hook this time.”

IF helping

Friend: All right. What’s the harm, I suppose.

[View target leaving cubicle] You decide to overlook (target name)’s email this time. In a few moments (she/he) seems more at ease and leaves for lunch. Having looked over the email lists for the day, you and (friend’s name) are ready to go. [View friend]

Friend: We’re all set then. By the way, why did you decide to help (her/him) out?

What motivated your decision?

- “I just felt like being nice.”
- “(Target name) would be grateful.”
- “I don’t think (target name) will send more personal emails.”
- “Helping people out is more important than the office email restrictions.”

IF not helping

Friend: All right. You did catch (her/him) pretty much red-handed.

[View computer] Following the procedure you were instructed in, you flag (target name)’s email and forward it to the appropriate supervisor. You can hear down the row that (target name) has left their cubicle, and you and (friend’s name) are prepared to do the same. [View friend]

Friend: So why did you decide did you decide to report it?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “Reporting it is something I was tasked to do, beyond that it’s none of my business.”
- “(Target name) will keep using (her/his) work computer for personal emails if (she/he) doesn’t experience the consequences.”
- “Rules like email policy are in place for a reason, you can’t just make exceptions to them.”

Ending

[View row, friend in foreground]

Friend: I understand. Why do you think (she/he) sent that email in the first place?

How do you answer?

- “(She/He) probably doesn’t care about the email policy.”
- “(She/He) probably forgot about the policy.”
- “(She/He) is probably just reckless.”
- “It might have been something important that couldn’t wait.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Anyway, let’s get you something to eat.

[Empty row view] You and (friend’s name) continue off on your lunch break.

Scenario 3: Movie Usher Warning

Condition = Self + Action

Begin

[Scene dark] In this scenario, imagine you are preparing to see a much-anticipated movie at a cinema. [Illuminate into broad cinema view] [Friend comes into view] You are with your

friend (friend's name) coming away from the front of the line, having just purchased your tickets.

Friend: I can't believe the line is still this long.

[Show line] It is nearly time for the screening to start, but the line is still lengthy, since not only is the movie popular, but this is the last screening for the day. As (friend's name) and yourself make your way down past the line, heading towards the cinema doors, you notice a (woman/man) standing in the line, [Show target, close] carrying a jacket in one hand, and a serving food from the outside food court in the other. (She/he) is speaking to a friend, but also looks somewhat nervous, likely because of the food. You are aware that this cinema specifically does not allow outside food to be brought into movies, and that anyone wishing to do so must sneak their food in without the ushers noticing. [Show line, usher visible] at this point you also notice a cinema employee, walking up from the rear of the line, inspecting those waiting to buy tickets. Since the usher is coming up from behind them, the (woman/man) with the food has not noticed them, and is making no effort to conceal their food. If they are discovered by the usher, the usher will certainly watch them more closely, and not allow them into the theatre with their food, meaning they will either have to give up their food, or miss the movie they want to see. You suspect that the only way (she/he) will not be caught is if you, as you walk past, alert them to the usher's presence, giving (her/him) time to cover their food with their jacket. [Show friend] You point this out to (friend's name), who notices the (woman/man) with food as well.

Friend: Ah, I see. Yeah, they haven't seen the usher at all. Are you going to point them out to (her/him)?

How will you respond?

- Yes >
 - "I guess I could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."
- No >
 - "I might not."
 - "I don't think I will."
 - "No, it's fine. They shouldn't be sneaking food in anyway."
- "Let me think about it."

IF “Let me think about it” or time interval exceeding 8 seconds

Friend: There’s a bit of time before the usher will even pass them. What’s on your mind?

What are you considering?

- “I’m thinking about the cinema’s right to ban outside food.” [If selected, show usher, close.]
- “I’m thinking about the (woman/man) with the food.” [If selected, show target, close]
- “I’m thinking about whether this is right or not.” [If selected, show broad cinema view]
- “Nothing, I’ve made up my mind.” [If select, go straight to decision.]

Friend: Right, sure. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess I could.”
 - “Sure, I don’t see why not.”
 - “Definitely, let’s help (her/him) out.”
- No >
 - “I might not.”
 - “I don’t think I will.”
 - “No, it’s fine. They shouldn’t be sneaking food in anyway.”

IF helping

Friend: Okay, sure. Just be subtle about it.

[View target, close] As you walk past the (woman/man) with the food, you slow down, clearing your throat to grab their attention. [Target looking directly at camera] You gesture towards the rear of the line with your head, [Show usher] causing the (woman/man) to notice the usher, and immediately cover their food [Show target, food covered]. You continue on quickly so as not to draw the usher’s attention. As you make your way towards the theatre doors, (friend’s name) looks over their shoulder, and nods to confirm your intervention’s success. [Show friend]

Friend: That went smoothly. So why did you decide to help (her/him) out?

What motivated your decision?

- “I just felt like being nice.”
- “(She/he) would be grateful.”
- “I don’t think (she/he) makes a habit of sneaking food into movies.”
- “Helping people out is more important than the cinema’s rules about outside food.”

IF not helping

Friend: All right, no problem then.

[Broad view of cinema] You and (friend’s name) continue towards the theatre doors.
(Friend’s name) looks over (her/his) shoulder, and turn back to you. [View friend]

Friend: It looks like the ushers have their eye on (her/him) now. So why did you decide not to say anything?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “Someone else trying to sneak in food is none of our business.”
- “(She/he) will just keep sneaking food in if (she/he) never gets caught for it.”
- “The cinema has a right to make restrictions on what people bring in with them, and people should honour that.”

Ending

[View outside theatre, friend in foreground]

Friend: I understand. Why do you think (she/he) was sneaking in food in the first place?

How do you answer?

- “(She/He) is probably just inconsiderate.”

- “(She/He) might not have known about the rule against outside food.”
- “(She/He) probably just doesn’t care about these kinds of rules.”
- “(She/He) might have been in a hurry and hadn’t eaten yet.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Anyway, let’s find our seats. Don’t forget to switch your phone off.

[Empty view outside theatre] You and (friend’s name) walk into the theatre.

Scenario 4: Overlooking Late Fees

Condition = Other + Inaction

Begin

[Scene dark] For this scenario, imagine you are in a local library, visiting a friend who works at the borrowing counter. [Illuminate into counter view] It has been a fairly slow day at the library, and you have spent the last 20 minutes or so talking casually with your friend (friend’s name). [Distant view, friend and target at counter] (Friend’s name) has just excused (herself/himself) for a moment to attend to a (woman/man) who is at the counter to borrow some books. You can see from where your standing that the (woman/man) has a worried expression on (her/his) face, and appears to be pleading with (friend’s name) about something that has come up on the computer. You notice (friend’s name) excusing (herself/himself), and walking over to your side of the counter with the (woman/man)’s books. [Show friend up close] When you ask what’s going on, (friend’s name) responds:

Friend: I’ve got a bit of an issue over here. (She/he) says (she/he) really needs to borrow these books today, but when I scanned (her/his) card, it showed that (she/he) has some unpaid late fees, and it’s my boss’s policy that we’re supposed to refuse to lend out new books to someone with unpaid fees. The boss isn’t here right now, though, and since the computer logs the fees and the books separately, I could easily just let (her/him) have the books anyway. [Show target at counter] (She/he) says (she/he) would pay the fees, but (she/he) doesn’t have the money on (her/him) just now. I told (her/him) that I’d ask you, my ‘colleague’ about it. [Show friend, close] So I’ll leave it up to you, what do you think I should do? Should I tell them they can’t borrow the books?

How will you answer (friend's name)?

- Yes >
 - "I guess so."
 - "I think you should probably not allow it."
 - "Yes, definitely enforce the rule, and tell them they can't borrow until they pay the fees."
- No >
 - "You could not."
 - "I don't think you have to withhold the books."
 - "No, I really think you should overlook the fees and let (her/him) have the books."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: It's a bit of a delicate matter. What's on your mind?

What are you considering?

- "I'm thinking about the importance of the library's borrowing rules." [If selected, show borrowing desk computer, close]
- "I'm thinking about the (woman/man) who needs the books." [If selected, show target, close]
- "I'm thinking about whether this is right or not." [If selected, show broad library view]
- "Nothing, I've made up my mind." [If select, go to decision.]

Friend: I see. Anything else? [Loop back]

What have you decided?

- Yes >
 - "I guess so."
 - "I think you should probably not allow it."
 - "Definitely enforce the rule, and tell them they can't borrow until they pay the fees."
- No >
 - "You could not."
 - "I don't think you have to withhold the books."
 - "I really think you should overlook the fees and let (her/him) have the books."

IF helping

Friend: All right. There's no harm in letting it slide this time.

[View friend and target at counter] (Friend's name) returns to the counter and scans through the (woman/man)'s books, to their seeming relief. [View friend alone at far counter] (She/he) leaves in a hurry, and (friend's name) returns to your side of the counter. [View friend, close]

Friend: All done. So, why did you decide to help (her/him) out?

What motivated your decision?

- "I just felt like being nice."
- "(She/he) would be grateful."
- "I don't think (she/he) will make a habit of not paying (her/his) late fees."
- "Helping people out is more important than the library late fees."

IF not helping

Friend: Yeah you're right, fees are fees.

[View of target and friend at far counter] (Friend's name) returns to the borrowing counter and speaks to the (woman/man), who offers only brief protest before reclaiming (her/his) library card. [View friend alone at far counter] (She/he) leaves in a hurry, and (friend's name) returns to your side of the counter. [View friend, close]

Friend: Okay, problem solved. What made you decide we should uphold the late fee restriction?

What motivated your decision?

- "I just didn't feel like helping (her/him) out."

- “Your job is really just to act out the library’s rules, (her/his) problems with late fees aren’t really our concern.”
- “(She/he) will just keep avoiding late fees if (she/he) isn’t prevented from borrowing more books as a result.”
- “The rules that say you can’t borrow with outstanding late fees are in place for a reason, you can’t just make exceptions to them.”

Ending

[View row, friend in foreground]

Friend: I understand. Why do you think (she/he) left their fees unpaid in the first place?

How do you answer?

- “(She/He) is probably inconsiderate.”
- “(She/He) forgotten about the outstanding fees.”
- “(She/He) probably doesn’t care about the library’s rules.”
- “(She/He) might have intended to pay them, but was in a rush and didn’t have the right money just now.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Anyway, it’s almost time for my break. My boss should be back soon...

[Empty counter view] You and (friend’s name) continue talk casually while waiting for (her/his) boss to return and man the counter.

Scenario 5: Carpool Exception

Condition = Other + Action

Begin

[Scene dark] For this scenario, imagine you work in an office. [Illuminate into lobby view] It is near the end of the day, and you are preparing to leave. As you walk through the ground-floor lobby of the building, you notice your friend (friend's name) [show friend and group] meeting with a group of people that you recognise as his carpool. (Friend's name) walks over to speak with you. [Friend view, close]

Friend: Hey, good to see you.

(Friend's name) is the driver for the office-arranged carpool that heads through his area. In order to benefit the greatest number of employees, and to reduce the occurrence of co-workers harassing each other for rides, your office has implemented an organised carpool-roster this year. [Show group] Rather than leaving employees to manage the trips themselves, the office asks them to put their names down on the roster, and attempts to organise them into the most convenient group. Because these carpools are rostered ahead of time, it is now against company policy to bother a carpool driver for a ride. [Friend view, close]

Friend: Listen, before you go, I wanted to get your advice on something. (Target name) heard that one of the people in my carpool group didn't show up today, and asked me when we were coming down in the elevator if I could give (her/him) a ride home.

[Show target near group, close] (Target name) is one of your co-workers, though neither you nor (friend's name) know (her/him) very well. (Friend's name) hasn't given (her/him) an answer yet, and (she/he) looks concerned. [Friend view, close]

Friend: Should I take (her/him) along?

How will you advise (friend's name)?

- Yes >
 - "I guess you could."
 - "Sure, I don't see why not."
 - "Definitely, you should help (her/him) out."
- No >
 - "You don't have to."
 - "No, don't worry about it."
 - "Don't bother, (she/he) shouldn't be asking a carpool driver for a ride."
- "Let me think about it."

IF “Let me think about it” or time interval exceeding 8 seconds

Friend: Okay, there’s no big rush. What’s on your mind?

What are you considering?

- “I’m thinking about the importance of the carpool rules.” [If selected, show carpool group]
- “I’m thinking about (target name).” [If selected, show target, close]
- “I’m thinking about whether this is right or not.” [If selected, show broad lobby view]
- “Nothing, I’ve made up my mind.” [If select, go straight to decision.]

Friend: Right, sure. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess you could.”
 - “Sure, I don’t see why not.”
 - “Definitely, you should help (her/him) out.”
- No >
 - “You don’t have to.”
 - “No, don’t worry about it.”
 - “Don’t bother, (she/he) shouldn’t be asking a carpool driver for a ride.”

IF helping

Friend: Okay then, thanks. Wait here while I tell (her/him).

[View friend talking to target] (Friend’s name) informs (target name) that (she/he) will be able to give (her/him) a ride home. (Target name) appears relieved, and walks towards the car park with the other members of the carpool. (Friend’s name) walks back to you. [Friend view, close]

Friend: All set. So why did you decide that we should help (her/him) out?

What motivated your decision?

- “I just felt like being nice.”
- “(Target name) would be grateful.”
- “I don’t think (target name) will make a habit of asking for rides.”
- “Helping people out is more important than the office carpool policies.”

IF not helping

Friend: All right then. Wait here while I tell (her/him).

[View friend talking to target] (Friend’s name) informs (target name) that (she/he) will not be giving (her/him) a ride home. (Target name) appears disappointed by this, and immediately leaves the building while (friend’s name)’s carpool group heads towards the car park. (Friend’s name) walks back to you. [Friend view, close]

Friend: All taken care of. So why did you decide that we shouldn’t help (her/him)?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “(Target name)’s travel arrangements are none of our business.”
- “(Target name) will just keep hassling people for last-minute rides if people don’t turn (her/him) down.”
- “The carpooling policy and rules are in place for a reason, you can’t just make exceptions to them.”

Ending

[View building interior, friend in foreground]

Friend: I understand. Why do you think (she/he) wanted to get a lift in the first place?

How do you answer?

- “(She/He) is probably just inconsiderate.”
- “(She/He) probably needs to be home in hurry today.”
- “(She/He) probably just doesn’t care about the policy.”
- “Something might have interfered with (her/his) usual means of getting home.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Well, I can’t leave my passengers waiting, I’ll see you tomorrow.

[Empty interior view] (Friend’s name) rushes off towards the car park while you exit the building to make your way home.

Scenario 6: Ignoring Pet

Condition = Self + Inaction

Begin

[Scene dark] For this scenario, imagine you live in an apartment building. [Illuminate into front view] You are heading home from an outing with your friend (friend’s name), and have just reached your apartment building. [Show lobby view] You enter the building and go to check your mail box when you hear the sound of a dog barking. [Turn to friend view]

Friend: Did you hear that?

[Show lobby view] It is strange that the barking seems to be coming from inside the building, since your building doesn’t allow pets to be kept in the apartments. [Show lobby view, target with dog walking to elevator] You notice (target name), one of the people who lives on your floor, nervously hurrying to the elevator holding a small bulldog, which you recognise as the source of the barking. You do not see (target name) very often, and do not know (her/him) very well, so it is possible that (she/he) has been keeping a dog secretly and you have just never noticed. [Show friend, close]

Friend: I thought you weren't allowed to keep pets in the building.

You explain to (friend's name) that pets aren't allowed by the building's superintendent, and that residents of the building had been asked to use the anonymous complaints box [show box on wall] to bring violations of the building's rules to the superintendent's attention. [Lobby view] (Target name) doesn't seem to have noticed you, and is still nervously waiting for the elevator. [Friend view]

Friend: I see. So, are you going to write a note to the superintendent to check it out?

Are you going to report (target name)?

- Yes >
 - "I guess so."
 - "I suppose I should."
 - "Yes, definitely. It's strictly no pets allowed."
- No >
 - "I might not."
 - "I don't think I will."
 - "No, I'll let (her/him) off the hook this time."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: It's your call. What's on your mind?

What are you considering?

- "I'm thinking about the importance of the apartment rules." [If selected, show complaint box]
- "I'm thinking about (target name)." [If selected, show target, close]
- "I'm thinking about whether this is right or not." [If selected, show broad lobby view]
- "Nothing, I've made up my mind." [If select, go to decision.]

Friend: I see. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess so.”
 - “I suppose I should.”
 - “Yes, definitely. It’s strictly no pets allowed.”
- No >
 - “I might not.”
 - “I don’t think I will.”
 - “No, I’ll let (her/him) off the hook this time.”

IF helping

Friend: All right. What’s the harm, I suppose.

[View target getting onto elevator] You decide to overlook (target name)’s dog, and not bring it to the superintendent’s attention. When the elevator arrives (she/he) seems hastily steps inside and the doors close. You conclude checking your mail, and continue towards the elevator with (friend’s name). [View friend]

Friend: I’m wondering, why did you decide to help (her/him) out and keep the dog secret?

What motivated your decision?

- “I just felt like being nice.”
- “(Target name) would be grateful.”
- “I don’t think (target name) will is necessarily keeping the dog permanently.”
- “Helping people out is more important than the building’s restrictions on pets.”

IF not helping

Friend: All right. You did catch (her/him) pretty much red-handed.

[View box] You fill out a quick note to the superintendent, explaining that (target name) has been seen taking a dog into the building, and may be keeping it in (her/his) apartment. After sliding the note into the box, you conclude checking your mail, and continue towards the elevator with (friend’s name). [View friend]

Friend: So why did you decide did you decide to report the dog?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “We were all asked to report rule violations in the building, beyond that it’s none of my business.”
- “(Target name) will just continue to secretly keep pets in the building if the superintendent never confronts (her/him) about it.”
- “The rules and codes for the building are in place for a reason, you can’t just make exceptions to them.”

Ending

[Outside elevator view, friend in foreground]

Friend: I understand. Why do you think (she/he) is keeping a dog here in the first place?

How do you answer?

- “(She/He) probably doesn’t care about the building’s rules.”
- “(She/He) might have nowhere else to keep the dog.”
- “(She/He) is probably just reckless.”
- “(She/He) might just be watching it temporarily.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. [Dinging noise] Ah, the elevator has arrived.

[Outside elevator view, empty] You and (friend’s name) walk into the elevator, and you press the button to take you to your floor.

Scenario 7: Bin Misuse Warning

Condition = Self + Action

Self/action - Some employee in shopping complex, dumping goods (cushions?) in a nearby recycling-only bin, you've noticed coming around corner, security is coming up behind you, fines are in place. Warn them or not?

Begin

[Scene dark] In this scenario, imagine you are with a friend, walking through a shopping complex. [Illuminate into mall view] It is the afternoon, and you are out shopping with your friend (friend's name), but so far you have had little luck finding anything of interest in the various stores you have visited. [Show friend]

Friend: Well, the shopping isn't going very well, but at least the air-conditioning is good in here.

[Show security guard, from behind] You are walking at a brisk pace, and overtake a security guard who is walking slowly in the same direction as you. [Show mall view] As you approach and round the next corner, [Show bin view, with target] you notice a (woman/man) wearing a shirt identifying (her/him) as an employee of the nearby home-and-hardware store, standing beside a garbage bin with a small trolley. [Show target, close] (She/He) is quickly taking what appears to be damaged or poorly cut pieces of foam and wood out of the trolley, and is discarding them. However, [Show signs] the garbage bin is clearly labelled "Paper recycling only", with other notices claiming "Fines apply for misuse of garbage bins". (Friend's name) taps you on the shoulder. [Friend view]

Friend: They really should install a security camera here, but all the same, (she/he) is going to be in trouble when that security guard behind us comes around the corner. We need to keep going, since we don't want to be around when (she/he) gets caught. Are you thinking of saying something to warn (her/him)?

How will you respond?

- Yes >
 - "I guess I could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."
- No >
 - "I might not."

- “I don’t think I will.”
- “No, it’s fine. (She/He) shouldn’t be misusing the recycling bin.”
- “Let me think about it.”

IF “Let me think about it” or time interval exceeding 8 seconds

Friend: Well, the security guard is walking very slowly. What’s on your mind?

What are you considering?

- “I’m thinking about the importance of the paper-only recycling bin.” [If selected, show signs, close.]
- “I’m thinking about the (woman/man) dumping the garbage.” [If selected, show target, close]
- “I’m thinking about whether this is right or not.” [If selected, show mall view]
- “Nothing, I’ve made up my mind.” [If select, go straight to decision.]

Friend: Right, sure. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess I could.”
 - “Sure, I don’t see why not.”
 - “Definitely, let’s help (her/him) out.”
- No >
 - “I might not.”
 - “I don’t think I will.”
 - “No, it’s fine. (She/He) shouldn’t be misusing the recycling bin.”

IF helping

Friend: Okay, sure. But we’d better hurry.

[View target, close] As you walk past the recycling bin and the (woman/man) with the trolley, you clear your throat loudly to get (her/his) attention. [View target, looking at

camera] You quickly gesture over your shoulder with your thumb, mouthing the words ‘Security’s coming’ as you continue past. [View mall corridor] As you pass (her/him), you hear the hasty sounds of the recycling bin being closed, and the trolley wheeling away in the opposite direction. (Friend’s name) glances over (her/his) shoulder. [Friend view]

Friend: Looks like (she/he) made it. That was close. So why did you decide to help (her/him) out?

What motivated your decision?

- “I just felt like being nice.”
- “(She/he) would be grateful.”
- “I don’t think (she/he) is going make a habit of misusing recycling bins.”
- “Helping people out is more important than the strict content of garbage bins.”

IF not helping

Friend: All right, no problem then.

[View target, close] You and (friend’s name) continue past the (woman/man) with the trolley, receiving a quick sideways glance as (she/he) continues to unload (her/his) garbage into the bin. [View mall corridor] As you pass (her/him), you hear the calling out of the security guard. (Friend’s name) glances over (her/his) shoulder. [Friend view]

Friend: It looks like (she/he)’s been caught, it was a whole trolley full of garbage. So why did you decide not to say anything?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “Someone misusing a recycling bin is none of our business.”
- “(She/he) will just keep misusing if (she/he) never experiences the consequences for doing so.”
- “That bin is labelled paper recycling for a reason, you can’t just make exceptions to that.”

Ending

[View friend, side-on]

Friend: I understand. Why do you think (she/he) was dumping that garbage there in the first place?

How do you answer?

- “(She/He) is probably just inconsiderate.”
- “(She/He) might have not noticed the signs.”
- “(She/He) probably just didn’t think (she/he) would get caught.”
- “(She/He) might have really needed to get rid of the garbage in a hurry.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Anyway, let’s see if we have any better luck in these stores.

[Empty view mall corridor] You and (friend’s name) walk towards your next destination to continue shopping.

Scenario 8: Overlooking Employee Discount

Condition = Other + Inaction

Begin

[Scene dark] For this scenario, imagine you are in a department store, visiting a friend who works at one of the retail counters. [Illuminate into counter view] Your friend (friend’s name) [Show friend, close] has been working a late afternoon shift, but since business is slow this time of day, the two of you have had ample time to simply stand around and talk. [Target and friends appear at counter] A (woman/man), wearing a nametag identifying them as another employee of this store, and a number of (her/his) friends, have approached (friend’s name)’s counter. [Friend view, close]

Friend: Ah, looks like (target name) has some purchases to make. Please excuse me a moment.

[Counter view, friend, target and crowd present] (Friend's name) returns to the service side of the counter to greet (target name), and begins to scan (her/his) items through. There appear to be a large number of items, [target view, close] and (target name) looks somewhat nervous as (she/he) speaks to (friend's name). (Friend's name) appears to excuse (herself/himself) and walks back over towards you. [Friend view, close]

Friend: I need your advice on something. (Target name) is purchasing a lot of items, and it's clear that only some of them are for (her/him), while the others are for (her/his) friends. The problem is that (target name) wants to use (her/his) employee discount to buy all of this. Its company policy that we're only allowed to use our employee discount on items purchased for ourselves, not for gifts. I'm just not sure whether I should deny (her/him) the discount because (she/he) is misusing, or if I should just let it slide this time. What do you think, should I deny (her/him) the employee discount on (her/his) friend's purchases?

How will you answer (friend's name)?

- Yes >
 - "I guess so."
 - "I think you should probably not allow it."
 - "Yes, definitely enforce the policy, and tell them they can't use their discount for other people's purchases."
- No >
 - "You could not."
 - "I don't think you have to deny (her/him) the discount."
 - "No, I really think you should help (her/him) out and let (her/him) have the discount."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: There's no real rush, they're just talking amongst themselves. What's on your mind?

What are you considering?

- “I’m thinking about the importance of the store’s policies.” [If selected, show friend]
- “I’m thinking about the (target name).” [If selected, show target, close]
- “I’m thinking about whether this is right or not.” [If selected, show broad counter view]
- “Nothing, I’ve made up my mind.” [If select, go to decision.]

Friend: I see. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess so.”
 - “I think you should probably not allow it.”
 - “Yes, definitely enforce the policy, and tell them they can’t use their discount for other people’s purchases.”
- No >
 - “You could not.”
 - “I don’t think you have to deny (her/him) the discount.”
 - “No, I really think you should help (her/him) out, and let (her/him) have the discount.”

IF helping

Friend: Okay. There’s no harm in helping (her/him) out this time.

[View friend and target/crowd at counter] (Friend’s name) returns to the counter and punches in the keys at the cash register needed to give (target name) (her/his) employee discount. (Target name) and (her/his) friends seem somewhat relieved as they bid (friend’s name) goodbye as they leave the store. (Friend’s name) walks back to your side of the counter. [View friend, close]

Friend: All done. So, why did you decide to help (her/him)?

What motivated your decision?

- “I just felt like being nice.”
- “(She/he) would be grateful.”
- “I don’t think (she/he) will make a habit of misusing (her/his) employee discount.”
- “Helping people out is more important than the store’s discount policy.”

IF not helping

Friend: Yeah you're right, rules are rules.

[View of target, crowd, and friend at far counter] (Friend's name) returns to the counter and speaks to (target name), who offers only brief protest before apologising to (her/his) friends, and paying for the goods. [View counter with friend] The group leaves quickly, and (friend's name) returns to your side of the counter. [View friend, close]

Friend: Okay, it's taken care of. What made you decide we should uphold the store policy?

What motivated your decision?

- "I just didn't feel like helping (her/him) out."
- "Your job is just to act in accordance with store policies, (her/his) desire to over-extend (her/his) discount isn't really our concern."
- "(She/he) will just keep misusing (her/his) discount if (she/he) isn't prevented from doing so."
- "These company policies are in place for a reason, you can't just make exceptions to them."

Ending

[View row, friend in foreground]

Friend: I understand. Why do you think (she/he) tried to over-use (her/his) discount in the first place?

How do you answer?

- "(She/He) is probably inconsiderate."
- "(She/He) probably just wanted to save money for (her/his) friends."
- "(She/He) probably doesn't take the discount restriction seriously."
- "(She/He) might not have realised that you can't use your employee discount on gifts."
- "(She/He)'s just that kind of person."

Friend: Yeah, you're probably right. Anyway, my shift is almost over, my replacement should be here to take over in a few minutes.

[Empty counter view] You and (friend's name) continue talk casually while waiting for (her/his) replacement to arrive and take over the counter.

Scenario 9: Stationary Permission Intervention

Condition = Other + Action

Begin

[Scene dark] For this scenario, imagine you work in an office. [Illuminate into office view] You are returning from your lunch break with your friend (friend's name), and are walking with (her/him) through (her/his) department on your way back to your own. As you approach (friend's name)'s office, you hear the sound of someone being halted by security. [Show cupboard view, target and security] You round the corner, and see a member of security stopping a (woman/man) who is walking out of the stationary cupboard, carrying a seemingly large amount of stationary. [Friend view]

Friend: Oh, that's (target name), I've seen (her/him) around the department. It looks like security suspects (her/him) of stealing stationary.

[Show cupboard view, security and target] You can't make out most of what they are saying, but from what you and (friend's name) can gather, security is asking (target name) [show target, close] for the name of the supervisor who gave (her/him) access to the stationary cupboard, and (target name) is claiming to not remember the name of the supervisor, saying that (she/he) had been given permission via email, after asking a handful of supervisors. [Show cupboard view, security and target] The office restricts access to stationary by locking the cupboard doors with 4-digit PIN codes, which are changed weekly and only sent to the supervisors in a department. (Friend's name) is one of this department's supervisors, [friend view, close] and appears to doubt (target name)'s story.

Friend: Not a very likely story, but then again, (target name) may get in some serious trouble if (she/he) is caught actually stealing stationary. I could just step in quickly and say I gave

(her/him) permission to access the stationary cupboard, security would be satisfied with that.
Do you think I should?

How will you respond to (friend's name)?

- Yes >
 - "I guess you could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."
- No >
 - "You don't have to."
 - "No, don't worry about it."
 - "Don't bother, (she/he) shouldn't be stealing stationary anyway."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: Okay, there's no big rush. What's on your mind?

What are you considering?

- "I'm thinking about the importance of keeping our office supplies safe." [If selected, show security, close.]
- "I'm thinking about (target name)." [If selected, show target, close]
- "I'm thinking about whether this is right or not." [If selected, show broad cupboard view]
- "Nothing, I've made up my mind." [If select, go straight to decision.]

Friend: Right, sure. Anything else? [Loop back]

What have you decided?

- Yes >
 - "I guess you could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."
- No >
 - "You don't have to."
 - "No, don't worry about it."
 - "Don't bother, (she/he) shouldn't be stealing stationary anyway."

IF helping

Friend: Right, okay. Just give me a moment.

[Show cupboard view, friend, target and security] (Friend's name) walks up to the two and speaks to (target name), pretending that (she/he) had requested the stationary that (she/he) is carrying. [Show cupboard view, friend and target] After a quick exchange of words, the member of security promptly moves on, after which (friend's name) gives a reassuring nod to the target, and walks back to you. The two of you continue towards (friend's name)'s office. [Friend view, close]

Friend: All taken care of. So why did you decide that we should help (her/him) out?

What motivated your decision?

- "I just felt like being nice."
- "(Target name) would be grateful."
- "I don't think (target name) is likely to steal stationary again, if this really was stealing."
- "Helping people out is more important than the regulation of office supplies."

IF not helping

Friend: Okay. Well, no problem then.

[Show cupboard view, target and security] The two of you continue towards (friend's name)'s office. As you walk by, you can hear the sound the stationary cupboard opening again, as the member of security forces (target name) to return the stationary until (she/he) can verify that (she/he) had permission to take any. [Friend view, close]

Friend: So why did you decide not to help?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “(Target name)’s taking of stationary is none of our business.”
- “(Target name) will just keep stealing stationary if (she/he) doesn’t experience the consequences.”
- “Access to office supplies is restricted for a reason, you can’t just make exceptions.”

Ending

[View office, friend in foreground]

Friend: I understand. Why do you think (she/he) was taking stationary in the first place?

How do you answer?

- “(She/He) is probably just greedy.”
- “(She/He) may have really needed the supplies.”
- “(She/He) probably thought (she/he) wouldn’t get caught.”
- “(She/He) might have been telling the truth about the email story.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Anyway, I’ll talk to you after work, okay?

[Empty office view] You and (friend’s name) part ways, and you head back to your office.

Scenario 10: Overlooking Break-Room Access

Condition = Self + Inaction

Begin

[Scene dark] For this scenario, imagine you work in an office. [Illuminate into cubicle view] You are just beginning your lunch break, and are being visited by your friend (friend’s name). [Show friend] Your floor has quickly become less populated, [show row view] with most people heading out to lunch rather than electing to stay in their offices or cubicles. [Show friend]

Friend: It still surprises me how quickly everyone clears out. [Sideways glance] Say, isn't that your boss's office?

[Show office, target entering] You follow where (friend's name) is looking, and see a co-worker who you are not very familiar with, (target name), looking around nervously while stepping into your boss's office. You recall that your boss usually leaves early for lunch, and always leaves the building to do so, and you wonder what business (target name) has in there. [Show target leaving office] As (target name) emerges, you recognise that (she/he) is carrying a door-access card, which from the level of clearance suggested by the coloured stripe on it, is most likely a spare belonging to your boss. [Friend view]

Friend: Is (she/he) actually is sneaking away with your boss's door I.D. card?

[Show elevator view] You watch (target name) quickly makes (her/his) way towards the elevator, and swipes the card in (her/his) hand, gaining access to the restricted floors higher in the building, and pressing the 'up' button on the elevator. [Show target at elevator view] (She/He) looks nervous as (she/he) waits for the elevator to arrive. You wonder what business (target name) may have on the higher floors, but the only option you can think of is that (she/he) intends on accessing the executive break room, and the superior coffee and stacks available there. It is possible that (she/he) could get away with this, as very few executives actually stay in the building for their lunch break. It is, however, strictly against the office security policies to use anyone's access card but your own. (Friend's name) has been watching with you, and is now turned to face you. [Friend view]

Friend: I doubt that (she/he) is allowed to do what (she/he)'s doing. Do you think we should email security, and have them check the access records for inconsistencies?

Are you going to report (target name)'s activities to security?

- Yes >
 - "I guess so."
 - "I suppose I should."
 - "Yes, definitely, accessing areas with someone else's card can't be allowed."
- No >
 - "I might not."
 - "I don't think I will."
 - "No, let's let them off the hook this time."

- “Let me think about it.”

IF “Let me think about it” or time interval exceeding 8 seconds

Friend: Well, I’ll leave it up to you. What’s on your mind?

What are you considering?

- “I’m thinking about the importance of the office security policies.” [If selected, show office view]
- “I’m thinking about (target name).” [If selected, show target, close]
- “I’m thinking about whether this is right or not.” [If selected, show broad row view]
- “Nothing, I’ve made up my mind.” [If select, go to decision.]

Friend: I see. Anything else? [Loop back]

What have you decided?

- Yes >
 - “I guess so.”
 - “I suppose I should.”
 - “Yes, definitely, accessing areas with someone else’s card can’t be allowed.”
- No >
 - “I might not.”
 - “I don’t think I will.”
 - “No, let’s let them off the hook this time.”

IF helping

Friend: Yeah, okay. What’s the harm, I suppose.

[View target entering] You decide to overlook (target name)’s activities this time. In a few moments (she/he) steps onto the elevator and disappears from view. You and (friend’s name) resume your plans to head out to lunch. [View friend]

Friend: We should be off, then. By the way, why did you decide to not to report (her/him)?

What motivated your decision?

- “I just felt like being nice.”
- “(Target name) would be grateful.”
- “I don’t think (target name) will make a habit of accessing areas restricted to (her/him).”
- “Helping people out is more important than the office access policies.”

IF not helping

Friend: All right. We did pretty much see (her/him) taking it.

[View computer] Before leaving, you quickly write an email to the head of security, detailing what you saw, and suggesting that they check the door-access records if they seek to pursue the matter. [Show elevator view] After sending the email, you notice (target name) has stepped into the elevator and disappeared from view. You and (friend’s name) resume your plans to head out to lunch. [View friend]

Friend: Well that’s over and done with. So why did you decide did you decide to report (her/him)?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “We had an obligation to at least bring it to security’s attention, beyond that it’s none of our business.”
- “(Target name) will keep accessing restricted areas with other people’s cards if (she/he) isn’t reprimanded for it.”
- “The office’s security and access protocols are in place for a reason, you can’t just make exceptions to them.”

Ending

[View row, friend in foreground]

Friend: I understand. Why do you think (she/he) sneaking around with the boss's access-card anyway?

How do you answer?

- “(She/He) probably doesn't care about the office's security restrictions.”
- “(She/He) might have some important business on the higher floors.”
- “(She/He) is probably just reckless.”
- “It's possible (she/he) had the boss's permission to use the card, and (she/he) was acting nervously for some other reason.”
- “(She/He)'s just that kind of person.”

Friend: Yeah, you're probably right. Anyway, let's go get something to eat.

[Empty row view] You and (friend's name) continue off on your lunch break.

Scenario 11: Water Restriction Cover-up

Condition = Self + Action

Begin

[Scene dark] In this scenario, imagine you live in a house in a suburban neighbourhood.

[Illuminate into street view] It's a Saturday morning, and you are walking home from a nearby corner store with your visiting friend, (friend's name). [Friend view]

Friend: The houses around here are really starting to show signs from the heat.

[Zoom in front lawns] Some of the lawns in the area are beginning to visibly die. There has been a problem in the area for the last few weeks, involving the water supply. Some underground pipe damage had cut off water to your area, meaning that to get any water at all it must be siphoned off from adjacent areas. As a result, the local authorities have put in place some restrictions on water use, the most notable being expensive fines for anyone who uses water on their lawns, pavement or swimming pools during the warm daylight hours. [Show

target house, with target and inspector] As you near the corner turning onto your street, you notice one of the water restriction inspectors standing on the pavement, speaking to one of the residents on your street, who you recognise as owning the house on the corner plot. [Show inspector, close] The two of them seem to be having a conversation, [Show target, close], though (she/he) looks as if (she/he) is nervous or troubled. [Show target house, with target and inspector] From what you can guess, the inspector is talking to (her/him) because of (her/his) lawn, which despite its large size seems to be surviving the water restrictions well. The situation also catch's (friend's name)'s eye as the two of you reach the corner. [Friend view]

Friend: It must seem suspicious, (she/he) must do a lot of watering at night. [Friend head turn] On the other hand...

[Show lawn side 2 and sprinkler] You look at where (friend's name) is gesturing, on the second half of lawn, concealed from the view of the house owner and the inspector, is a small sprinkler, quietly spraying water onto the lawn. Although the inspector doesn't seem to have noticed it, this is likely what the owner is so nervous about, and is perhaps trying to distract the inspector from. [Show faucet] Following the hose with your eyes, you notice that the faucet supplying water to the sprinkler is immediately at hand. (Friend's name) has noticed it too.

Friend (quietly): It looks like (she/he) is going to be fined when the inspector notices this. The facet is right at hand though. Are you thinking of turning it off for (her/him)?

How will you respond?

- Yes >
 - "I guess I could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."
- No >
 - "I might not."
 - "I don't think I will."
 - "No, it's fine. (She/He) shouldn't be using (her/his) sprinkler during the day."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: Well, they're still in the middle of talking about something. What's on your mind?

What are you considering?

- "I'm thinking about the importance of the water restrictions." [If selected, show sprinkler.]
- "I'm thinking about the (woman/man) who lives here." [If selected, show target, close]
- "I'm thinking about whether this is right or not." [If selected, show broad street view]
- "Nothing, I've made up my mind." [If select, go straight to decision.]

Friend: Right, sure. Anything else? [Loop back]

What have you decided?

- Yes >
 - "I guess I could."
 - "Sure, I don't see why not."
 - "Definitely, let's help (her/him) out."
- No >
 - "I might not."
 - "I don't think I will."
 - "No, it's fine. (She/He) shouldn't be using (her/his) sprinkler during the day."

IF helping

Friend: Okay, sure. But be quiet about it.

[Show lawn side 2 and sprinkler, turned off] You quickly step over to the facet and turn the handle, watching as the spray of water from the sprinkler dies down. You continue down the street with (friend's name) while (she/he) looks over (her/his) shoulder. [Show friend]

Friend: It looks like (she/he) isn't going to get a fine issued. So why did you decide to help (her/him) out?

What motivated your decision?

- “I just felt like being nice.”
- “(She/he) would be grateful.”
- “I don’t think (she/he) is going to keep watering (her/his) lawn during the day.”
- “Helping people out is more important than these water restrictions.”

IF not helping

Friend: Okay. Well, no problem then.

[Street view] You continue down the street towards your house, with (friend’s name) looking casually over (her/his) shoulder as you gain distance from the house on the corner. [Friend view]

Friend: Well, it seems that (she/he) has some explaining to do, the inspector has found the running sprinkler. So why did you decide not to intervene?

What motivated your decision?

- “I just didn’t feel like helping (her/him) out.”
- “Someone else’s water usage is none of our business.”
- “(She/he) will just keep wasting water on (her/his) lawn if (she/he) never experiences any consequences for doing so.”
- “These water restrictions are being enforced for a reason, you can’t just make exceptions to the rules.”

Ending

[View friend, side-on in street]

Friend: I understand. Why do you think (she/he) was watering (her/his) lawn during the day in the first place?

How do you answer?

- “(She/He) is probably just inconsiderate.”
- “(She/He) might have forgotten about the restrictions.”
- “(She/He) probably didn’t think (she/he) would get caught.”

- “(She/He) might have some important reason from preserving (her/his) lawn.”
- “(She/He)’s just that kind of person.”

Friend: Yeah, you’re probably right. Anyway, let’s get out of this heat.

[Empty street view] You and (friend’s name) continue up the street and return to your house.

Scenario 12: Overlooking Snuck-in Patrons

Condition = Other + Inaction

Begin

[Scene dark] For this scenario, imagine you are spending the evening visiting a friend who works in a Drive-In theatre. [Illuminate into view of screen and cars] Your friend (friend’s name) [Show friend, at table] works here on weekends, and has obtained permission for you to keep (her/him) company while (she/he) attends one of the external snack bars.

Friend: Be glad we both brought jackets, it gets colder and colder as the night goes on.

[View screen and cars] Ahead of you, cars are driving in and finding spaces in front of the screen. Even though the movie is about to start, the cue of cars at the ticket booth is still long. [Show ticket booth] At this theatre, for a car to gain admission, a ticket must be purchased for each passenger in the car, so line must move slowly enough for the booth workers to look closely. [Show target car, close] One of the cars recently allowed in catches your attention, when its trunk is opened a (woman/man) climbs out [Show target car, target climbing out], re-adjusting a blanket that (she/he) was seemingly hidden beneath when the car drove past the ticket booth. [Show target, close] Without interacting with the other two people in the car, the (woman/man) has started nervously walking up towards the central building, which houses the snack bars and bathrooms. (Friend’s name), following where you’re looking, has noticed the (woman/man) too. [Friend view]

Friend [looking sideways]: Uh-oh, looks like we have someone sneaking in. That's bad news for them, because the Drive-In has a no-tolerance policy on people sneaking in, they'll be forced to leave. [Looking at camera] We're supposed to report people sneaking in to the manager, if we see any, but I doubt anyone else has noticed. What do you think, should I turn (her/him) in?

How will you answer (friend's name)?

- Yes >
 - "I guess so."
 - "I think you should probably report it."
 - "Yes, definitely enforce the policy and have them ejected from the theatre."
- No >
 - "You could not."
 - "I don't think you have to, if no one would even know."
 - "No, I really think you should help (her/him) out and just overlook it."
- "Let me think about it."

IF "Let me think about it" or time interval exceeding 8 seconds

Friend: Well there's no rush, the movie hasn't even started yet. What's on your mind?

What are you considering?

- "I'm thinking about the importance of the ticket policies." [If selected, show ticket booth]
- "I'm thinking about the (woman/man) who snuck in." [If selected, show target, close]
- "I'm thinking about whether this is right or not." [If selected, show screen and cars]
- "Nothing, I've made up my mind." [If select, go to decision.]

Friend: I see. Anything else? [Loop back]

What have you decided?

- Yes >
 - "I guess so."
 - "I think you should probably report it."
 - "Yes, definitely enforce the policy and have them ejected from the theatre."
- No >
 - "You could not."

- “I don’t think you have to, if no one would even know.”
- “No, I really think you should help (her/him) out and just overlook it.”

IF helping

Friend: Okay. There’s no harm in letting it slide.

[View screen and cars] The (woman/man) continues into the building, most likely the bathroom, and several minutes later return to (her/his) car without incident, this time stepping into the back seat. Just as (friend’s name) had suspected, no other employees had noticed (her/him) climbing out of the trunk. [View friend, close]

Friend: Well, no harm done. So, why did you decide to help (her/him)?

What motivated your decision?

- “I just felt like being nice.”
- “(She/he) would be grateful.”
- “I don’t think (she/he) will make a habit of sneaking into movies.”
- “Helping people out is more important than upholding the theatre’s ticket policies.”

IF not helping

Friend: Yeah you’re right, that is the policy after all.

[View of screen and cars] (Friend’s name) excuses (herself/himself) for a few minutes, and then returns to the snack bar. [Friend view]

Friend: Well, I’ve pointed the car out to the manager, it’s up to them now.

[Show target car] A few minutes after the (woman/man) from the trunk returns to (her/his) car, [Show target car, with security] the Drive-In security personnel approach the car. [Show

screen and cars] After a short conversation, the car backs up out of its parking space, and drives away towards the Drive-In exit. [Friend view, close]

Friend: Looks like that's taken care of. So what made you decide that we should turn them in?

What motivated your decision?

- "I just didn't feel like helping (her/him) out."
- "Reporting people who're sneaking in is part of your job, beyond that (her/him) trying to sneak into the movie is none of our business."
- "(She/he) will just keep sneaking into Drive-In movies if (she/he) never gets caught for doing so."
- "The Drive-In theatre has a right to charge people for entry, you can't just make exceptions to that."

Ending

[Friend view]

Friend: I understand. Why do you think (she/he) tried to sneak in at all?

How do you answer?

- "(She/He) is probably just inconsiderate."
- "(Her/His) friends might have been trying to save (her/him) money."
- "(She/He) probably didn't think (she/he) would be caught."
- "(She/He) might not have been able to afford a ticket."
- "(She/He)'s just that kind of person."

Friend: Yeah, you're probably right. We'd better prepare for a bit of a rush, people always try to buy their popcorn just before the movie starts.

[Show screen and cars] (Friend's name) continues to sell food and drinks to the Drive-In patrons as you wait for the movie to start.

Appendix B of this thesis has been removed as it may contain sensitive/confidential content