Sex and gender differences in moral cognition

David Reilly

A report submitted as a partial requirement for the degree of Bachelor of Psychology (Honours), at Griffith University, Gold Coast campus.

October, 2010

I declare that this written assignment is my own work and does not include:

— material from published sources used without proper acknowledgment; or
— material copied from the work of other students

.................................
Abstract

This study examined the source of individual variation in moral cognition, specifically in moral judgment and moral construal. It was hypothesised that psychological gender (sex-role identity) of participants would show greater differences than those found between males and females; specifically that feminine and androgynous participants would attain higher developmental stages than masculine participants.

Participants (N= 280) were recruited from a student subject pool and the general community to obtain a cross-sectional sample of ages across the lifespan. Participants completed online or paper surveys, containing criterion-judgment tasks on a range of issues to test moral construal, the short form version of the Bem Sex Role Inventory (BSRI; Bem, 1981a) to measure gender, and the Defining Issues Test (DIT; Rest, 1979b) of moral judgment. A subset of participants (n = 84) were also interviewed face-to-face, completing the Ethic of Care Interview (ECI; Skoe, 1993) to measure care-based moral judgment.

As hypothesised, psychological gender was a stronger predictor of moral judgment than biological sex, even after controlling for age and education. Feminine participants were more developed than masculine participants in developmental stage of justice-based moral judgment ($d=.42$) and care-based moral judgment ($d=.64$), while there was no difference between feminine and androgynous participants. Previous inconsistencies in support for, or rejection of sex differences may be partially explained by variation in sex-role identity of samples, and the use of primarily younger student subject pools in recruitment. Additionally, differences were found between males and females on tasks of moral construal. The implications of these findings are discussed, and how measurement of psychological gender might be used to identify those who would benefit from educational interventions aimed at facilitating higher developmental stages of moral judgment.

Author : David Reilly
Supervisor: Heather Green
Submitted : October, 2010
Acknowledgments

I would like to personally thank and acknowledge the advice and guidance of a number of people during the early stages of my project, while researching the literature and structuring my proposal. I would also like to thank the two anonymous reviewers of my proposal, who took exceeding care and detail in providing written guidance as well as much needed praise and encouragement to its merit. I suspect the final thesis would not nearly resemble what it is today without both.

Firstly, Professor Eva E. A. Skoe of University of Oslo, Norway, in providing not only guidance and encouragement, but also a large number of her published journal articles and book chapters that I was unable to access as an undergraduate student, to start me on my way! The care-based moral judgment instrument she has created is used widely to measure the developmental stage of an ethic of care, and was freely granted along with a copy of the Ethic of Care Interview manual. My sincere and profound thanks for all your assistance and advice.

Secondly I would like to thank and acknowledge guidance sought and so generously given in understanding the area of moral cognition from a number of researchers in this area. In alphabetical order, these were: Debora L. Lidell of the University of Iowa, USA for providing a copy of the Measure of Moral Orientation (MMO) instrument; Georg Lind of Konstanz University, Germany for providing a copy of his Moral Judgment Task (MJT) and advice; Steve Sherblom of Lindenwood University, USA; Stephen J. Thoma of the University of Alabama, USA. Last but certainly not least, Professor Larry Walker of the University of British Columbia, Canada, in answering questions and so kindly scanning in a copy of a rather obscure commentary by Carol Gilligan from his personal library, unasked. While I must respectfully disagree with his opinion on gender and morality, his writings have been a pleasure to read and his contributions to the field extensive.
Likewise, the contributions of prior researchers in the area of moral cognition must be acknowledged. The late Professor James R. Rest granted use of his Defining Issues Test (DIT) instrument to all student researchers, which was invaluable for a student struggling to obtain licensing agreements for copyrighted tests. Throughout this project, his ideas and writings have guided me, and his moral judgment tests invaluable. Sandra Lipsitz Bem, whose contributions to gender psychology in advancing the notion of sex-role identity and androgynous, rather than biological sex, inspired me as a first year undergraduate and overcame my cynicism after reading the works of Sigmund Freud. Her work with gender schema theory, and development of the Bem Sex-Role Inventory (BSRI), provides a forceful argument that culture and gender is more potent than biological sex, and that in an ideal world we could seek to combine the best of masculinity and femininity. Carol Gilligan, whose work “In a different voice” spoke to a different type of moral reasoning based on care and compassion, and who challenged our conceptions of the moral domain at a time where women were grossly underrepresented and marginalised in moral research, was the source of inspiration for much of this thesis. She argued, skilfully and persuasively, for the feminine perspective, whose voice I attempted to capture and document with this thesis. While proving elusive and sometimes fleeting in studies comparing men and women directly, it has been most strongly found in the difference between masculine and feminine sex-role identity and I would hope that – in some small way – my thesis would help advance her original arguments. It has been said before that in science we stand on the shoulders of giants. Truly this is so.

I would also like to thank and acknowledge the generous assistance of the following people in helping me recruit participants from the community. Access to highly masculine, and older participants was critical to ensure a representative
sample so as to allow generalisability of these findings, and their kindness and efforts in this regard was essential. Former psychology students Larisa Mesic, Susanne Kimura, my classmate Larraine Peet, and Jamie Mannion of getparticipants.com I can’t thank you enough!

Also, I would especially like to single out both Larraine Peet and Roma Reilly for their assistance with transcription of recorded interviews, and for the last minute assistance of Larraine in coding a batch of interview transcripts so that I could obtain a measure of inter-rater reliability (kappa). I may have been overly ambitious in seeking to interview and then analyse verbal protocols of so many participants, but without their gracious help in transcribing them, it would have been impossible to analyse such a sample size.

Finally I would like to personally thank my project supervisor, Dr Heather Green, without whom this project would not have been possible in its current form. Heather so kindly granted me the freedom and flexibility to attempt an ambitious project in an area slightly tangential to her areas of interest, but which I was passionate about investigating. Other potential supervisors might have been more prescriptive in assignment of topic, but Heather was so supportive and encouraging in her supervision by allowing me such latitude. At the same time, when unforseen hiccups did arise, she was a voice of calm in reassuring me the wheels had not fallen off (completely!) yet. In doing so, Heather provided scaffolding to support my learning curve as a researcher and a safety net just in case. I would particularly like to thank her for her generosity in making additional RA time available to get the transcriptions back on track, and for her exceeding patience in marking and editing my writing. The detail, and care, she took in providing comments, not just corrections, has improved the scholarly quality of this thesis considerably 😊
Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Theoretical approaches to moral cognition</td>
<td>2</td>
</tr>
<tr>
<td> Piaget’s cognitive-developmental view of morality</td>
<td>2</td>
</tr>
<tr>
<td> Kohlberg’s cognitive developmental model of morality</td>
<td>4</td>
</tr>
<tr>
<td> Methodology employed in assessing moral judgment</td>
<td>8</td>
</tr>
<tr>
<td> Criticisms and limitations of Kohlberg’s theories</td>
<td>10</td>
</tr>
<tr>
<td> Gilligan’s ethic of care</td>
<td>12</td>
</tr>
<tr>
<td> Methodology employed in assessing care-based moral judgment</td>
<td>13</td>
</tr>
<tr>
<td> Gilligan’s ethic of care developmental model</td>
<td>14</td>
</tr>
<tr>
<td> Moral construal</td>
<td>16</td>
</tr>
<tr>
<td> Sex differences in moral cognition</td>
<td>17</td>
</tr>
<tr>
<td> Meta-analyses of sex differences</td>
<td>17</td>
</tr>
<tr>
<td> Sex-Roles and Gender Identity</td>
<td>20</td>
</tr>
<tr>
<td>The Present Study</td>
<td>21</td>
</tr>
<tr>
<td>Method</td>
<td>22</td>
</tr>
<tr>
<td> Participants</td>
<td>22</td>
</tr>
<tr>
<td> Materials</td>
<td>23</td>
</tr>
<tr>
<td> Procedure</td>
<td>31</td>
</tr>
<tr>
<td> Analyses</td>
<td>32</td>
</tr>
<tr>
<td>Results</td>
<td>32</td>
</tr>
<tr>
<td> Bem Sex-Role Inventory (BSRI)</td>
<td>32</td>
</tr>
<tr>
<td> Missing Data</td>
<td>33</td>
</tr>
<tr>
<td> Justice-based Moral Judgment using the Defining Issues Test (DIT)</td>
<td>33</td>
</tr>
<tr>
<td> Care-based Moral Judgment using the Ethic of Care Interview (ECI)</td>
<td>36</td>
</tr>
<tr>
<td> Moral Construal</td>
<td>38</td>
</tr>
<tr>
<td>Discussion</td>
<td>40</td>
</tr>
<tr>
<td> Justice-based moral judgment</td>
<td>40</td>
</tr>
<tr>
<td> Care based moral judgment</td>
<td>43</td>
</tr>
<tr>
<td> Moral construal</td>
<td>46</td>
</tr>
<tr>
<td> Implications of the study</td>
<td>48</td>
</tr>
<tr>
<td> Limitations of the study</td>
<td>49</td>
</tr>
<tr>
<td>Summary</td>
<td>50</td>
</tr>
<tr>
<td>References</td>
<td>51</td>
</tr>
<tr>
<td>Appendix A - Bem Sex-Role Inventory</td>
<td>A1</td>
</tr>
<tr>
<td>Appendix B – Defining Issues Test</td>
<td>B1</td>
</tr>
<tr>
<td>Appendix C – Criterion-judgment tasks</td>
<td>C1</td>
</tr>
<tr>
<td>Appendix D – Ethic of care developmental stages</td>
<td>D1</td>
</tr>
</tbody>
</table>
Introduction

The study of moral cognition seeks to understand the way in which people perceive and think about moral issues. Over recent decades, conceptualisation of moral cognition has been expanded, challenged, and refined through the contributions of pioneers such as Piaget, Kohlberg, Rest and Gilligan. Yet great controversy remains over the source and nature of individual differences in moral cognition, and how factors such as culture and gender roles interact with and guide moral development (Nucci, 2007; Nucci & Gingo, 2010). Thus, there remains a need for research that illuminates our understanding of the basic processes of moral cognition and individual differences (Rest, 1994).

The purpose of this study was to investigate individual variation in two key areas of moral cognition: (a) moral construal in the way that individuals identify and construe an action or an issue as involving moral elements, and (b) moral judgment or the way people evaluate and make choices about the right course of action to take. The specific focus was the influence of sex and gender on such processes.

The issue of purported sex and gender differences in morality is contentious and some would argue divisive (Hyde, 2005), because of a chequered history in scientific research of denigrating the cognitive capabilities of women (Fine, 2010). Historically, it has been argued that women are less developed than men in moral capacity (Freud, 1927; Kohlberg & Kramer, 1969; Piaget, 1965), that they think in qualitatively different ways (Gilligan, 1982), that they are better and more noble, or that they think identically to men (Walker, 1984, 2005; Walker & Frimer, 2009). This study was conducted from a perspective that documenting and affirming the diversity of different gender perspectives has merit (Hare-Mustin & Marecek, 1988), when approached with sensitivity.
No single study or methodology can settle such a debate, but this study aims to add to a growing body of literature that suggests individual differences in moral cognition are related to, but not dictated by, gender and cultural identity. Inconsistency in research supporting or rejecting differences between males and females on moral reasoning tasks might be explained by the confounding effect of psychological gender, or sex-roles. To clarify, sex refers to the biological makeup of an individual as male or female, while gender is the socio-cultural identity and gender schema of an individual (Unger, 1979).

This thesis investigated whether psychological gender, or sex-role identity, is more influential than biological sex as a predictor of moral cognition. Before detailing the present study, relevant theories and findings of past researchers are reviewed.

Theoretical approaches to moral cognition

Freud's (1961a) psychoanalytic exploration of morality, posited failure to act in a moral manner as being the result of a weak ego and the individuation process. Freud noted that biological differences and the differential path of boys and girls explain why women “show less sense of justice than men” (Freud, 1927, pp. 141-142). This view was challenged by other psychoanalytic writers (Chodorow, 1978; Horney, 1967; Jung, 1957), but the lack of empirical support and explanatory power regarding individual variations in morality left open room for more concrete theoretical approaches.

Piaget's cognitive-developmental view of morality

Piaget (1965) examined the development of morality from a cognitive perspective, basing his theory of morality on detailed naturalistic observation of the behaviour of boys and girls, especially in their play behaviour. Piaget believed that
children learned moral values not just from moral instruction from adults, but from peer interaction with other children during rule-bounded games of competition and play (Lever, 1976).

Here, in play activity, Piaget noted great differences between boys and girls. Boys, Piaget (1965) found, created on their own an elaborate system of rules for arbitrating disputes and conflicts, and relied on a strict code of behaviour to govern their actions. It was here that conflicts and disputes were argued and resolved; giving rise to notions of what was “right”, and what was “fair”, and the basis for construction and adherence to moral codes.

Such practices, Piaget argued, differed greatly from play of girls, whose games often included few if any rules, and whose rules were not always adhered to in the presence of conflict. Games were less competitive, more cooperative, and group harmony was emphasised as more important than winning or losing. Indeed, Piaget noted, many games such as skipping did not have an outcome in terms of win or loss, remarking that for a girl “She enjoys losing quite as much as winning” (Piaget, 1965, p. 79).

Observational studies by other researchers confirmed such differences (Lever, 1976; Maccoby & Jacklin, 1974), with girls often suspending games in the presence of conflict, rather than placing relationships at risk (Gilligan, 1982, p. 10). This sensitivity to the feelings of others, and maintenance of social harmony, was qualitatively different to the playmaking of boys. Boys, on the other hand, would continue to quarrel over rules of the game (Lever, 1976). Piaget argued that these different gender perspectives form the basis for the fundamental differences in the moral perspectives taken by boys and girls. Piaget concluded that girls are lacking in
a sense of justice, concluding “in the main, the legal sense is far less developed in girls than in boys” (Piaget, 1965, p. 77).

Piaget also described a transition between two modes of morality: from awareness and obedience of rules and external authority, termed heteronomy, towards a respect for rules as a mutually agreed upon social contract which is freely chosen but may be waived in some circumstances, termed autonomy (Piaget, 1965, p. 65). These two perspectives of justice represent a transition which occurs in conjunction with cognitive development. At first, what is “right” is externally determined, but as autonomy develops, what is “right” includes the welfare of others, and the wider society at large.

Piaget (1968) advocated a developmental approach to cognitive and moral development, introducing the concept of four qualitatively different stages (sensorimotor, preoperational, concrete operational, formal operational). Children go through each stage to progress to the next. This notion of a cognitive-developmental sequence of stages played a strong part in later theories of moral development by Kohlberg. Briefly, Piaget made two major contributions—1) a concept of distinct, qualitatively different, stages of cognitive and moral development; 2) the idea that an individual must progress through each stage in sequence without regressing to an earlier period, and that attainment of higher stages represents growth and greater development of cognitive structures.

Kohlberg’s cognitive developmental model of morality

Research into moral psychology, before the contribution of Kohlberg, was sparse (Rest, 1989). Psychoanalytic approaches fell out of favour, and morality was perceived in psychology to be largely a product of values and beliefs (Kohlberg,
assimilated through prior learning experiences early in childhood (Bandura, 1991).

Lawrence Kohlberg advanced a different perspective, however, and saw morality as being the product of internal processes of reasoning and very much a cognitive endeavour. Inspired by the cognitive-developmental approach of Piaget, Kohlberg sought to extend moral psychology research to include a cognitive perspective. Kohlberg (1958) argued that general theories of personality and learning were inadequate. Kohlberg employed a novel methodology, that of studying the ways in which people engaged in tasks of hypothetical moral judgment from a cognitive perspective, by asking them to describe and talk about hypothetical moral dilemmas. From his interviews with a sample of young boys, Kohlberg conceived a cognitive developmental model (CMD) of moral judgment (see Figure 1).

1.) preconventional level (Stages 1,2) which is egocentric and self-oriented;
2.) conventional level (Stages 3,4) which is other-oriented and takes rights into account; and
3.) postconventional level (Stages 5,6) where moral judgment takes a more principled societal view.

Kohlberg (1973, 1976) argued that these developmental stages possessed several important properties. Firstly they were sequential, in that one passed from lower stages to higher stages in order as cognitive and moral maturity developed. Secondly they were invariant, in that one should not regress from a higher stage to a lower one. Thirdly, they were universal in that they applied to all peoples. Chief evidence for these stages and properties was a longitudinal study of approximately fifty boys and their sequential progression through Kohlberg’ stages (Colby et al., 1983).
Figure 1. Kohlberg’s three levels of moral judgment, preconventional, conventional, and post-conventional.
Though stage development is sequential, development and attainment of successively higher stages by all persons is not guaranteed. Cognitive development of the child is a prerequisite for more sophisticated moral development, and while it is common to reach a more advanced cognitive level than moral, the opposite is rarely if ever found (Walker, 1980). Pre-conventional moral development is found primarily in children under the age of 9, some adolescents and in criminal offenders (Kohlberg, 1976). Conventional morality represents the norm for adolescents and adults in society. Cultivation of post-conventional morality through education and life experiences requires formal operational thinking as a prerequisite (Kohlberg & Gilligan, 1972; Kuhn, Langer, Kohlberg, & Haan, 1977), but represents a minority of adults (Kohlberg, 1976; Rest, Narvaez, Bebeau, & Thoma, 1999). Of key interest to researchers are which aspects of personality and development contribute to the cultivation of post-conventional moral judgment (Kohlberg, 1984, pp. 170-172).

A full understanding of the subtle nuances between each of the six stages of Kohlberg’s model is not required; however the definition of Stage 3 and the differences between conventional and post-conventional thinking is helpful in understanding the controversies in his research, and alternate points of view. For this reason, a brief guide is offered to the reader, and supplementary material can be found in Appendix B.

Stage 3 morality coincides with a growth in perspective-taking ability, and a concern for relationships with and well-being of others. A change in focus from the self and towards others takes place, as fulfillment of social roles, responsibilities and expectations of others supersede shallow individual interests. Kohlberg (1976) suggests that the “Golden Rule” typified Stage 3 cognition, and that to be seen to be good in the eyes of others is important. An emphasis on relationships with others is
the primary concern, taking precedence over personal interests and even over rules themselves. Kohlberg (1981) termed this the “Good boy, nice girl” orientation. Being good is living up to the expectations of others and caring for their interests even over one’s own.

In choosing to place “care for others” at a relatively low level of moral development, Kohlberg received strong criticism. Holstein (1976) notes that the stages were defined as ordinal, rather than nominal, with each stage being cognitively and morally superior to the previous. Since Stage 3 thinking care for others represents a traditionally feminine perspective, and a fixation at this stage would suggest developmental arrestment, by implication this suggests moral inferiority in femininity (Gilligan, 1982, 1983). Bandura argued that this placed a “value” judgment in Kohlberg’s system, questioning “By what logical reasoning is a morality rooted in law and order (Stage 4) morally superior to one relying on social regard and concern for others (Stage 3)?” (Bandura, 1991, p. 47).

Stage 4 represents a shift away from personal relationships and towards a differentiation between a “societal point of view and interpersonal agreement”, and Kohlberg recognised this as a point of growth. Except in the most extreme of cases, one’s duty is to obey these laws and to work within that framework. Sometimes this will even come at the expense of personal relationships. For example, a decision to report an escaped prisoner (a classic Kohlberg dilemma) who happens to be a friend would demonstrate growth from Stage 3 to Stage 4.

Post-conventional reasoning occurs less seldom, and involves greater cognitive complexity than at previous levels. Stage 5 recognises that the purpose of laws is to protect individual rights and well-being, and that they represent a philosophical “social contract”. Here, Kohlberg takes a utilitarian view of the “greatest good for the
Sex and gender differences in moral cognition

Kohlberg also distinguished between two fundamentally different moral orientations, which mirrors the distinction Piaget (1965) noted between a heteronomous and autonomous morality. Moral dilemmas may be viewed from a perspective of rules, rights and authority (Type A reasoning), or from a justice and human welfare perspective (Type B reasoning). Post-conventional reasoning generally focuses on a human welfare perspective.

Methodology employed in assessing moral judgment

The task of operationally defining a construct such as moral judgment is a complex one. Direct observation is inadequate, as many different moral justifications can be given for the same action; behaviour demonstrates the property of equifinality, in that one may choose to act, or not to act, for a multitude of reasons that a neutral observer is not privy to. Instead, Kohlberg (1958) relied on qualitative sources of information, primarily interviews, to formulate and evaluate his cognitive-developmental model.

By providing standardised hypothetical situations, expressed as moral vignettes, Kohlberg applied a structured interview approach with standardised probe questions. While the moral decision chosen was important, of greater importance was the reasoning behind a course of action. In explaining the reasons why they would do
such a thing, and providing a rationale, Kohlberg attempted to elicit the cognitive structure behind moral decisions (Rest, 1986; Turieel, 1976). This method was applied first to children, and then later to adolescents and adults to determine the developmental stage of the participant (Kohlberg, 1958, 1984).

While discussions of abstract hypothetical dilemmas provide a rich source of information to analyse, the process was subjective and not easily replicated by other researchers. Kohlberg was challenged on these grounds (Rest, 1989). In response to such criticisms, Kohlberg formalised the interview method and published the Moral Judgment Interview (MJI; Colby & Kohlberg, 1987).

Kohlberg’s MJI instrument represented the first formal interview instrument for assessing the moral judgment construct. Kohlberg published extensive manuals of the coding scheme used to rate participant’s verbal protocols, which were revised many times throughout the course of Kohlberg’s career. Partly this was in response to criticism by other researchers, as flaws in the coding scheme and empirical evidence of sex-bias (see below) were identified (Kohlberg, Levine, & Hewer, 1983). Furthermore, as Kohlberg believed the sequential and invariant nature of his model to be absolute, the scoring system was revised to fit disconfirming evidence and empirical results so that participants who regressed or skipped no longer did so (Rest, 1986).

The MJI allowed others to evaluate and test Kohlberg’s CMD with new samples drawn from the wider community, and to test for influences of gender and culture. It also allowed researchers to retrospectively analyse previous protocols, and to apply the coding scheme to new vignettes designed with less abstract and more emotionally pulling issues. However the multi-volume coding system Kohlberg produced requires extensive training and time to score, limiting sample size of
research. Additionally, scoring decisions based on transcripts are subjective, and may not accurately reflect all internal cognitions. Alternate instruments, such as Rest’s Defining Issues Test (DIT; Rest, 1979a), have become more widely used in research for this reason.

**Criticisms and limitations of Kohlberg’s theories**

Kohlberg stimulated empirical research into the construct of moral judgment in the decades that followed with mixed endorsement and criticism - for a more detailed coverage see Kohlberg, et. al. (1983) and Rest (1989, 1994). Only those issues associated with sex differences in moral judgment are discussed herein.

Kohlberg conceived his moral judgment stages on the Socratic tradition that moral judgment is based on principles of justice and governed by rules and rights (Kohlberg, 1981; Kohlberg, et al., 1983; Rest, 1989). This is a philosophical stance, and fails to reflect all modes of moral decision making. Based on interviews with solely male respondents (standard methodological practice at the time), Kohlberg produced a scoring system that he claimed was universal to all peoples, genders and cultures. This presented a serious threat to generalisability of his theories.

Kohlberg and Kramer (1969) provided empirical findings that moral judgment, in women, was less developed. Many studies (Alker & Poppen, 1973; Bussey & Maughan, 1982; Fishkin, Keniston, & MacKinnon, 1973; Haan, 1978; Haan, Langer, & Kohlberg, 1976; Haan, Smith, & Block, 1968; Langdale, 1983; Poppen, 1974; Saltzstein, Diamond, & Belenky, 1972) replicated Kohlberg’s (1969) finding of “lower” moral reasoning in females. Such findings led researchers to questions of gender bias in Kohlberg’s conception of morality, rather than the egregious assertion initially claimed that women’s morality was developmentally arrested, with Stage 3
reasoning representing the “functional morality for housewives and mothers” (Kohlberg & Kramer, 1969, p. 108).

In memorializing Kohlberg’s moral development legacy, Rest (1989) points out that the chief evidence he offered for his stage theory was a longitudinal study of fifty boys and their sequential progression through Kohlberg stages (Colby, et al., 1983). Holstein (1976) refuted the construct validity of this model with a longitudinal study including both boys and girls, finding that regression to early stages was common, and sex differences were present. Holstein found the modal stage of girls to be Stage 3, while boys were Stage 4, but dismissed the likelihood of women being morally inferior as questionable (Holstein, 1976).

Kohlberg took such charges of gender bias seriously and acknowledged this by revising his detailed scoring manuals for the MJI many times until the form which we have today (Colby & Kohlberg, 1987). Thus the definition of Stage 3 thinking is not as strict as in earlier versions, and many participants are credited with having achieved higher stages than previously. This makes it hard, for the modern reader, to distinguish between criticisms that were valid at the time, and those that remain valid today.

Walker (1984) argued that under a revised scoring system, there existed no gender differences in the attainment of Kohlbergian moral stages, and that when they did occur they were largely the product of confounding influences like occupation and education. Changes in sex roles in the decade since merit further investigation of whether such a claim still holds.

That gender bias did exist in early versions of Kohlberg’s work is an historical footnote, but what is more important to take away from this is the alternative lines of inquiry these early findings stimulated. Justice-based frameworks of moral reasoning
have an important role to play, but they fail to encompass all moral orientations in
society (Gilligan, 1982, 1983). At best, it is an incomplete picture of the moral
domain (Nucci, 2007; Nucci & Gingo, 2010; Walker, 2004; Walker, Pitts, Hennig, &
Matsuba, 1995). A more comprehensive view of the moral domain is called for
which encompasses wider moral perspectives (Flanagan & Jackson, 1987),
something that Kohlberg ultimately came to acknowledge as necessary (Kohlberg, et
al., 1983).

_Gilligan’s ethic of care_

Carol Gilligan (1977, 1982, 1983) argued that by interviewing only male
participants during the development of his moral judgment theory, Kohlberg had
missed hearing the feminine voice. In her book “In a different voice” she argued that
this voice was equally as sophisticated, but different, from the justice-based view of
Kohlberg. Gilligan advocated an alternate moral orientation rooted in an ethic of care
and responsibility which was qualitatively different to an ethic of justice. This thesis
was not completely new, echoing the voices of moral philosophers such as Aristotle,
but Gilligan went beyond what might just be dismissed as a feminist condemnation
of Kohlberg’s “masculine” model of justice. Gilligan offered a compelling
psychological framework for an alternate moral orientation of care, as well as
empirical research to support such a claim.

Kohlberg (1976) and Piaget’s (1965) theories of moral development based their
definitions of morality on a sense of justice and individual rights. Gilligan argued
(1977, 1982, 1983) that such a position may be representative of men, but concerns
of care for the feelings and interests of others were more representative of women.
Gilligan proposed that there were actually two moral orientations: 1) an _ethic of_
justice and rights, and 2) an ethic of care. Though not universally a female perspective, an ethic of care was found more strongly in women (Gilligan, 1982).

The mistake, Gilligan and colleagues argued, was in assuming that justice represented a complete picture of the moral domain, and that any concerns relating to an ethic of care were relegated to Stage 3 thinking as developmental arrestment. More comprehensive assessment of the construct of moral reasoning using a developmentally appropriate measure might well remove such differences between men and women, if women were assessed from an ethic of justice and an ethic of care (Gilligan, 1982).

Methodology employed in assessing care-based moral judgment

While qualitative interviews were still used, Gilligan differed from Kohlberg by conducting interviews about real-life moral dilemmas faced by participants, as well as abstract hypothetical ones. Gilligan included older adolescents and young adults (Gilligan & Murphy, 1979), women considering abortion or adoption in unplanned pregnancy (Gilligan & Belenky, 1980), and a longitudinal study of college students, where participants were interviewed initially and then five years later after graduation (Gilligan, 1981).

Gilligan (1982, 1983) did not specify a formal framework for analysing the content of interviews, but researchers have proffered a number of coding schemes for analysing interviews from an ethic of care perspective (Johnston, 1988; Lyons, 1983; Skoe, 1993). In challenging the justice-only moral framework of Kohlberg, Gilligan extended research into moral decision making to include an ethic of care, and towards the inclusion of real-life dilemmas. Ultimately, Kohlberg came to realise that the moral domain was broader than his earlier research had suggested (Kohlberg, et
al., 1983; Rest, 1989), and believed that Gilligan’s work had highlighted a valid criticism.

Kohlberg (1983) offered:

It is for this reason that we make the following proposal: i.e., that there is a dimension along which various moral dilemmas and orientations can be placed. Personal moral dilemmas and orientations of special obligation… represent one end of this dimension and the standard hypothetical justice dilemmas and justice orientations represent the other end. (pp. 24-25)

Though the tension between Kohlberg and Gilligan’s work is often portrayed as a duel (Jorgensen, 2006), the two moral orientations are both valid aspects of a larger latent construct of moral judgement and complementary. Strong empirical evidence reveals that both care and justice perspectives are present in moral cognition (Brabeck & Weisgerber, 1988; Flanagan & Jackson, 1987; Gilligan & Antanucci, 1988; Jaffee & Hyde, 2000; Pratt & Royer, 1982; Skoe, Cumberland, Eisenberg, Hansen, & Perry, 2002). Gender differences in both must be evaluated to answer the question of whether males and females think differently about moral issues (Thoma, 1986).

**Gilligan’s ethic of care developmental model**

Gilligan (1982) proposed three qualitatively different stages of care. These have since been expanded on by other researchers, documenting the developmental progression of an ethic of care (see Appendix D for supplementary materials).

*Stage 1* represents a position of self-concern, where immediate needs and interests take priority over the needs of others, and responsibilities or duties. It is a strongly ego-centric perspective, based on survival and what is best for the individual. In this fashion, it is similar to pre-conventional moral development. Here,
however, the emphasis is less on specific rights and laws, and more the discretionary personal decisions one can make about caring for others within the framework of social and legal rules. A transitional stage, between survival and personal responsibility, marks the emergence of concepts such as responsibility to care for others and a conception of selfishness (Skoe, 1995). Care for oneself becomes seen as “selfish” when in conflict with responsibilities towards others.

Stage 2 represents a morality of self-sacrifice (Skoe, 1987), where responsibility and the needs of others outweigh the needs of the self. “Goodness” is equated with fulfilling socially expected roles, caring for others and putting their needs ahead of one’s own. Maintaining social harmony, avoiding conflict, and supporting dependents becomes the primary focus of concern, with care for self relegated to lesser importance. Gilligan argued that this was the “traditional” feminine voice (Gilligan, 1977, 1982; Skoe, 1987). The merit of this position is that it involves strong care for others, even at the expense of rules and duties – to care and help others is more important. A cost, however, is imposed on the individual, in that he or she lacks autonomy and self-expression. To do what is “expected” is the “right” thing to do, even if it neglects one’s own rights and needs. Self-expression and assertiveness have the power to hurt relationships, and can be seen as selfish. A transitional stage also exists, where a man or woman re-evaluates the relationship between self and other, beginning to see it as unequal. This can be a time of confusion, where the logic of self-sacrifice comes into question.

Stage 3’s hallmark is a balance between care for self and care for others. One recognises that in order to care for others effectively, one must also care for oneself. This is no longer seen as a “selfish” act or as irresponsible, and recognition of the interconnectedness between self and other is achieved. Care becomes the primary
Sex and gender differences in moral cognition

psychological concern, and the consequences of action and inaction for all parties are taken into account. One’s obligation is to care for all parties, to consider their views and perspectives, but to achieve a balance between self-interest and concern for others. In this fashion, it best represents the androgynous personality, achieving feminine ideals of care and masculine ideals of autonomy and self-assertion (Berzins, 1979; Kaplan, 1979).

Cross-cultural and longitudinal studies have validated the presence of these three developmental stages (Juujärvi, 2006a, 2006b; Pratt, Skoe, & Arnold, 2004; Skoe et al., 1999; Skoe & von der Lippe, 2002), and there is some evidence to suggest that it may be gender-related (Skoe, 1995; Skoe, et al., 2002). However the majority of research focuses on sex differences, rather than considering sex-roles and gender-identity.

**Moral construal**

Another domain pertinent to potential sex and gender effects in moral cognition is the way in which people perceive various social and personal issues, and their construal of the moral domain. Turiel (1983, 1989, 1994, 2006) argues that schemata about which issues fall into the moral domain, as distinct from matters of social custom and personal choices, is pertinent to moral cognition (Nucci, 1981; Nucci, Guerra, & Lee, 1991; Nucci & Turiel, 2000; Smetana, 1995; Turiel, 1983, 2006). Social customs “do not carry a moral quality as their rightness or wrongness is determined by social consensus” (Kuyel, 2002, p. 40), yet they may often be considered bad or wrong, and rated as such.

For an issue or an action to be interpreted within a psychological framework of moral judgment and reasoning, it must first be identified as such by those involved, and assessed as important. Criterion judgments assess only moral construal.
(Saltzstein, 1994), and not the valence of the emotion (for example, contraception may be seen as “moral” issue, regardless of whether one is for or against, or how strongly). Strong gender differences have been found in the self-reported importance of moral issues (Wark, 2000; Wark & Krebs, 1997), so it is proposed that classification of issues to the personal or moral spheres of influence might also yield gender differences. Previous work has shown differences in moral acceptance of behaviours (Henry, Morrissey, & Sullivan, 2006; Katz, Santman, & Lonero, 1994) but to date research of gender differences in moral construal remains limited.

**Sex differences in moral cognition**

Researchers have been mixed in their reactions to Gilligan’s hypothesised sex differences in moral orientations (Galotti, Kozberg, & Farmer, 1991; Gilligan & Antanucci, 1988; Pratt, Golding, & Hunter, 1984; Pratt, Golding, Hunter, & Norris, 1988; Pratt & Royer, 1982; Rothbart, Hanley, & Albert, 1986; Walker, 1989; Walker, et al., 1995; Walker, Vries, & Trevethan, 1987). Where differences between the sexes appear, it has been argued that it is primarily a function of problem content, with women more likely than men to focus on personal relationship problems when asked to recall real-life dilemmas (Pratt, Golding, Hunter, & Sampson, 1988; Walker, 1989; Wark & Krebs, 1996, 2000). Another argument put forward is that education is a confounding factor explaining sex differences (Walker, 1984, 2005). However the issue is still controversial, and subject to vigorous debate, e.g. Sherblom (2008, 2009) and rebuttal by Walker (2009). Generally, though, research into moral cognition focuses on moral orientation, and ignores the development of stages of moral judgment.

**Meta-analyses of sex differences**

Walker (1984) conducted the first published meta-analysis examining the question of sex differences in moral judgment. Walker examined studies using
Kohlberg’s Moral Judgment Interview (MJI), and found that at least in childhood and early adolescence there were few sex differences, and these were largely attributable to age and education confounds (Walker, 1984). Baumrind (1986) heavily criticised this review on statistical grounds, and heavy handed exclusion of disconfirming studies, while Thoma (1986) points out that Walker’s analysis was subject to overestimation in null finding studies where no effect size was present, and that substantial changes in the scoring system between earlier and later studies makes the picture less clear.

Thoma’s (1986) meta analysis into studies using the Defining Issues Test (DIT) found support for a difference between genders, but in the opposite direction to that expected. Contrary to the findings of Walker, across a combined sample of more than 6,000 participants, females scored significantly higher in development of moral judgment with an average effect size of $d = -.21$ across all ages, and of $d = -.28$ in adult samples. Thoma (1986) argues that this be considered a small “but stable gender effect size which consistently favours women across the lifespan”. Thus, when using an objective instrument of moral reasoning, women are no less capable of justice-based moral judgment and may actually show a small advantage.

As consensus grew for a distinct ethic of care and justice, many researchers examined the question of gender differences, and whether care was more strongly associated with women. Jaffee and Hyde (2000) sought to refute such differences, in a review article and meta-analysis of sex differences in moral orientation (preference for either justice-based or care-based arguments for moral decision-making). The authors did not seek to examine in their review whether stage differences - the focus of this present study - were present, instead citing Walker (1984) and Thoma (1986). Moral orientation and development of moral stages are related but distinct topics –
though the two are often conflated in literature with Jaffee and Hyde (2000) being cited as debunking all sex differences (Walker, 2005). Nonetheless, their work remains important as evidence of qualitative differences in moral perspectives of men and women.

Examining the question of preference for moral orientation, Jaffee and Hyde (2000) found in their meta-analysis an overall effect size for care of -.28, favouring females, and a justice effect size of .19 favouring males. Additionally, it should be noted that the magnitude of effect size was substantially higher across the lifespan. Samples from young adults (20-49) showing an effect of $d = -.33$ for care and $d = .40$ for justice, while older participants (50 or over) showed an effect size of $d = -.63$ for care and $d = .63$ for justice reasoning. That an interaction between age and gender might exist often goes unremarked, but shows the importance of surveying across the lifespan when investigating sex differences.

Jaffee and Hyde’s (2000) study has been interpreted by others both as definitive evidence of an absence of sex differences (Walker, 2005), and compelling evidence of strong sex differences in care and justice reasoning (Sherblom, 2008). Arguments typically revolve around whether these are “small” effect sizes using Cohen’s (1988) conventions, often dismissing them as not being practically meaningful. In doing so, however, such arguments overlook the much stronger differences present in adulthood.

It is indeed unfortunate that misinterpretation of Jaffee and Hyde’s (2000) meta-analysis prevails, and that it is evidenced as proof of uniformity between the moral cognition of men and women (Walker, 1991, 2005). Furthermore, in generalising the conclusion of “minimal differences” in moral orientation preference to all aspects of moral cognition, inquiry into other areas has been thwarted. Thus
further inquiry into possible sex differences in attainment of moral stages and moral construal is merited.

Sex-Roles and Gender Identity

In the absence of consensus in the Kohlberg-Gilligan debate over moral judgment, and a comprehensive unifying model of moral cognition, further research is required. One promising line appears to be that moral reasoning is gender-related, but not gender-specific (Gilligan, 1986). Biological sex, as a predictor, may be less of a factor than sex-role concepts and psychological gender. Children acquire sex-role concepts early in life that shape how they perceive themselves and others (Kohlberg, 1966; Kohlberg & Ullian, 1974), and sex-roles may dictate morality (Kohlberg & Kramer, 1969). Gender schema theory argues that these form the basis for cognitive schemata about gender identity (Bem, 1981b). Highly sex-typed individuals have a restricted behavioural repertoire (Bem, 1975; Bem & Lenney, 1976), as well as cognitive perspectives and self-schema (Kaplan, 1979; Ryan, David, & Reynolds, 2004). Additionally, emotional differences in level of empathy (Karniol, Gabay, Ochion, & Harari, 1988; Laurent & Hodges, 2009; Myyrya, Juujärvi, & Pesso, 2010; Skoe, 2010), and pro-social traits (Karniol, Grosz, & Schorr, 2003; Skoe, et al., 2002) have been associated with gender identity.

While thousands of published studies into moral judgment have been published that look directly, or as an afterthought, at whether differences exist between males and females, relatively few have examined sex-roles and gender identity (Jaffee & Hyde, 2000). Mixed support for this line of research has been found (e.g. Arbuthnot, 1975; Elm, Kennedy, & Lawton, 2001; Ford & Lowery, 1986; Leahy & Eiter, 1980; Lifton, 1985; Pratt & Royer, 1982), but like sex differences in moral reasoning such research has been plagued by methodological flaws. These include unreliable or ad-
hoc operational definitions of gender identity, samples drawn solely from young adult student populations, failure to rule out education level as a contributing factor, and failure to control for the type of dilemma reported by participants. Furthermore, gender differences in other aspects of moral cognition are largely unexplored, as most research focuses solely on judgment, ignoring other aspects like moral construal, motivation and character (Rest, 1994), which may influence decision making.

The Present Study

In this study, the influence of sex and gender on moral cognition in student and community samples is examined. Using a mixture of quantitative and qualitative methods, three aspects of moral cognition are examined:

1) *moral construal*, assessed by criterion-judgments about social issues
2) *justice-based moral judgment stage*, assessed by Rest’s Defining Issues Test of post-conventional reasoning
3) *care-based moral judgment stage*, assessed by Skoe’s Ethic of Care Interview

Based on prior research, and seeking to rule out the confounding variables of age and education, the following hypotheses were made:

1. Participants with highly feminine sex roles (feminine and androgynous) would show greater post-conventional moral judgment (*an ethic of justice*) than highly masculine participants.

2. Participants with highly feminine sex roles (feminine and androgynous) would show more well-developed care-based moral reasoning than highly masculine participants. Feminine sex roles would be associated with predominance of Stage
2 reasoning, while androgynous participants would exhibit predominantly Stage 3 reasoning.

3. Gender would be a stronger predictor of justice-based moral stage than biological sex, after controlling for age and education.

4. Gender would be a stronger predictor of care-based moral stage than biological sex, after controlling for age and education.

5. Gender differences in moral construal would be present. Specifically issues related to harm to others (lying, accepting bribes, etc.) would be more frequently seen as moral by female and feminine participants than male and masculine participants, while issues related to autonomy (sexual and reproductive decisions) would be more frequently seen as personal.

Method

Participants

Two hundred and eighty participants (110 male, 170 female) were recruited for this study, drawn from a student subject pool and Australian/New Zealand community samples. Ages ranged from seventeen to eighty one (M = 35.29, SD= 15.65), and included a good mix of educational backgrounds from early secondary through to postgraduate tertiary. A priori, a sample size of approximately 90 participants was specified for a minimum cell size of 15-20 in each of the four gender categories, as some sex by gender combinations (for example, feminine males) are less frequently found in the population. Recruitment of the community sample was expected to assist with including participants who encompassed a breadth of gender roles (strongly gender typed, and weakly), as well as a range of life ages and experiences, as previous research has been limited in sampling only from young adult subject pools. Additionally, the gender skew in psychology student
subject pools meant that community sampling was required to recruit additional male participants.

Participants completed written or online survey instruments, and a subset of participants \( n = 84 \) were selected by the researcher for face-to-face interviews. These were drawn from the student sample and from the community, comprising 26 males and 58 females, representing 30% of the full sample. Though desirable to interview all participants, recruitment from online community samples and the need for anonymity made this impractical. Missing values analysis (MVA) was conducted to verify that moral judgment did not differ between survey-only and survey with interview participants.

Materials

Gender role identity. Psychological gender was assessed using the 30 item short form of Bem’s Sex-Role Inventory (BSRI; Bem, 1974, 1981a). At face value, the measure is a general personality inventory that incorporates neutral filler items and socially desirable items so that the true intent of the instrument is not readily apparent. The instrument is based on cultural definitions of socially desirable sex-typed personality traits (Bem, 1981a), and the premise that highly sex-typed individuals chose to endorse traits that are consistent with a masculine or feminine self-image, while androgynous individuals chose a mixture of both (Bem, 1981a; Schmitt & Millard, 1988).

The short form of the instrument (see Appendix A) contains ten masculine, ten feminine, and ten neutral or socially desirable traits alternated to disguise the intent to differentiate participants according to sex-role identity. Personality traits are rated on a 7-point Likert scale (1 = “Never or almost never true”, 2 = “Usually not true”, 3 = “Sometimes but infrequently”, 4 = “Occasionally true”, 5 = “Often true”, 6 =
“Usually true”, \(7=\text{“Always or almost always true”}\). A separate masculine and feminine scale is produced by averaging responses across scale items, resulting in a continuous scale.

Previous attempts at operationalising psychological gender were based on the difference between masculine and feminine attributes; if one score was higher than the other, this indicated a masculine or feminine sex-typing and a difference approximating zero being androgynous where gender roles are equal. Spence, Helmreich and Stapp (1975) noted that participants who scored low on both masculine and feminine represented an entirely different category than participants who scored highly on both, and advocated classification into four distinct categories. Accordingly this has become standard practice. Participants can then, on the basis of a population-based median split (Bem, 1975), be assigned into a high or low masculinity group, and high or low femininity, to produce four categories of participant – masculine, feminine, undifferentiated (low on both masculine and feminine identification) and androgynous.

Despite concerns that cultural gender stereotypes may have drifted over time (Auster & Ohm, 2000; Twenge, 1997), confirmatory factor analysis studies show that the personality adjectives within the BSRI still reliably load against a distinct masculine and feminine factor, and possess strong validity (Holt & Ellis, 1998). The BSRI remains the most widely used androgyny instrument (Wang, 2007), is simple to administer and score electronically, and previous work has found that it has strong predictive validity and reliability in Australian samples (Reilly & Mesic, 2009).

Selection of the short form of the BSRI instrument was chosen over the longer form both for practicality (so as to not unduly burden participants who elected to complete surveys and interviews), and for its robust psychometric properties which
are chiefly the same as the longer form version. Internal consistency alpha coefficients for the short form BSRI masculine and feminine scales in normative samples are reported .84 and .87 respectively (Bem, 1981a); alpha coefficients with the sampled participants were .85 and .88 respectively.

Justice-based moral reasoning stage. Justice-based moral reasoning was assessed using Rest’s Defining Issues Test (DIT) instrument, which is an objective written instrument for assessing moral stage in participants (Rest, 1979a; Rest, Narvaez, Bebeau, et al., 1999). The instrument is designed to measure the development of moral stages attained by participants, and is particularly sensitive to the preference for post-conventional moral judgment (Stages 5 and 6 of Kohlberg’s moral stages), but also measures judgment using lower stages.

Rest (1979a), and colleagues developed a written instrument of moral judgment in contrast to the interview methodology of Kohlberg which used transcription and analysis of verbal discourse. Drawbacks of the interview approach include the considerable time required to collect such interviews sequentially, that analysis of transcripts requires extensive training to interpret transcripts using an extensive multi-volume scoring manual (Colby & Kohlberg, 1987), and that sample sizes in research were typically small due to the one on one nature of interviewing. Though interviews have merit, and provide a rich source of qualitative data, the advantage of the DIT is its efficiency, and that it provides quantitative data that is objectively scored and calculated by computer.

As an alternative to production measures, DIT is a “recognition” task which is believed to represent more accurately the internal cognitive processes and thoughts involved in moral judgment (Rest, 1979a, 1994). Interviews are “production” tasks, relying on explicit recall of moral arguments that may be actually be implicit and
difficult to verbalise (Rest, Narvaez, Bebeau, et al., 1999). Hypothetical moral dilemmas are provided as stimulus to activate moral schemas, and then specific arguments are offered to tap specific moral stages (Thoma, 2002). Previous research identified that people express a preference for moral arguments prototypical of their moral stage of development and reject those at higher or lower levels (Rest, Turiel, & Kohlberg, 1969).

Each section of the DIT contains a hypothetical moral vignette, and asks the participant to make a moral decision (to act, not to act, or cannot decide). They are then asked to evaluate twelve different thoughts and issues that may be relevant to the decision. Participants rate on a 5-point Likert scale how important that statement was to their decision and thinking (Great, much, some, little, none), and statements represent a wide range of moral viewpoints and Kohlbergian stages. Participants are then asked to rank the four most important issues in their thinking. Rating data can be further analysed for consistency with ranking data, but is not part of the actual outcome of the DIT and often discarded. Forcing participants to rate each statement, however, ensures they have read and considered them.

The main scale of the DIT is comprised of the participant’s ranking of statements, and the associated moral stage they represent. Participants may endorse statements from a range of moral stages when rating items, but the act of ranking the four most important decisions shows a preference for a certain type of moral schema (Rest, Douglas, Richard, JoAnna, & Douglas, 1974).

The DIT instrument offers a number of scales which can be used to examine moral judgment based on rating and ranking information, such as preference scores for each Kohlbergian stage. The most commonly used scale is the P-score, which reflects the preference for principled thinking (Stages 5 and 6) over conventional
moral judgment (Stages 2, 3, and 4), and discriminates principled from conventional moral judgment. Scores from each dilemma are summed and then converted to a percentage preference across all dilemmas, the P-Index. This is used as the primary outcome of the test for analysis of group differences.

A number of safeguards are built into the DIT to ensure that a considered and reflective response is given, such as the inclusion of “meaningless” statements which might be endorsed by a participant who “fakes good” and selects statements that are lofty and pretentious (Rest, 1979a, 1979b). These are used to validate legitimate attempts to answer truthfully and thoughtfully from socially desirability bias. Additional processing of responses was undertaken to validate responses (such as inconsistencies between rating endorsements of statements and subsequent ranking of those statements), using procedures outlined in the DIT manual (Rest, 1979b).

The three version short form of the original DIT was selected for this study (see Appendix B), as it strikes a balance between the time required to complete, reliability and validity, as well as being freely available to administer and score. A more recent version (DIT2) of the instrument is also available (Rest, Narvaez, Thoma, & Bebeau, 1999), has onerous licensing restrictions that prohibit scoring of responses by researchers and requires paper copies of surveys to be sent away for processing and data sharing. Furthermore, since the original DIT remains the most widely used, it allows for direct comparison with previous studies on gender differences. External validity with other moral instruments (such as the MJI) is high (Rest, Narvaez, Bebeau, et al., 1999), and discriminant validity from verbal ability and measures of general intelligence firmly established (Rest, et al., 1974).
Internal consistency alpha values for the DIT measures was .43, somewhat lower than would be expected from motivated and conscientious participants. Anecdotally, some participants expressed difficulty with the wording of the instrument, and that the language was dated. Although regarded to be at a 9th grade reading level and widely used with adolescents (Rest, 1979b), in practice many participants who did not complete high school found it extremely difficult to complete and understand, and thus may explain lower reliability in the online and community samples. Longer forms of the DIT are also shown to have higher reliability (Rest, 1979a).

Care-based moral reasoning stage. Care-based moral reasoning was assessed using Skoe’s Ethic of Care Interview (1993) and related coding scheme. The Ethic of Care Interview (ECI) presents participants with three standardised vignettes designed to elicit care-based reasoning for self and others (Skoe, 1987; Skoe & Marcia, 1991), as well as asking about a real-life moral dilemma recently faced. Participants are asked to respond to the moral dilemmas and identify the issues and decisions they would make, and standardised probing questions are used to elicit the reasoning and perspectives behind the decision. Interviews typically take between 10 and 20 minutes, and must be administered carefully so as not to influence the interviewee, but also to make sure they have been given sufficient time to explore and elaborate on their thinking process.

Interviews were recorded electronically, and later transcribed by the student researcher for analysis and coding for stage of care level using the criteria detailed in the ECI manual. Initially it was planned to administer a written equivalent of the interview, but the scale’s author was contacted and strongly recommended using the verbal interview (E. E. A. Skoe, personal communication, 27 March, 2010). As a
face-to-face interview, responses are more reflective and detailed, and follow-up questions allow the reasoning process behind a decision to be explored.

Though significantly more time consuming than survey based instruments like the DIT, the quality of information provided by an interview production measure merits its usage and coding for an ethic of care reflects a fundamentally different type of moral and ethical thinking than a justice-based measure. It also provides a counterbalance as a production measure to a recognition measure, and is less likely to be tapping an underlying measure of written reading ability in participants.

After initial piloting of the instrument, it was found that many participants found it difficult or were unable to bring to mind a recent “real-life” moral dilemma. Typically, real-life moral dilemmas that include issues of self-interest show lower stages of moral schema (Carpendale & Krebs, 1995; Haviv & Leman, 2002; Wark & Krebs, 1996, 1997), and an isolated dilemma on its own introduces additional noise unless several dilemmas are sampled (Walker, et al., 1995). A decision was made, after consultation with the instrument author (E. E. A. Skoe, personal communication, 27 March, 2010) to drop real-life moral dilemmas as outlined in the interview manual and to calculate ethic of care stages using just the hypothetical dilemmas. This represents the upper theoretical maximum of care reasoning, with the caveat that it might not always be used in real-life situations where personal interests are present.

**Criterion-judgment tasks of moral construal.** A series of criterion-judgment rating tasks of ambiguous moral issues was used to measure moral construal (see Appendix C). Participants were presented with a range of social and controversial issues; some of which involved overt moral issues and others a blend of moral and personal decisions.
Participants were asked, regardless of whether they were for or against an issue (such as abortion, or contraception) whether they believed it to be primarily a personal (discretionary) issue or a moral one. Previous studies have examined the valence or severity of moral infraction (Harding & Phillips, 1986; Henry, et al., 2006; Katz, et al., 1994). In these studies, the “wrongness” of actions were reported, participants may have been confounding “moral” issues with “social” taboos – a very subtle but important distinction. Domain theorists such as Turiel (1975, 1978, 1983, 1989) argue that important distinctions exist between social and moral domains. Expansion on this line of inquiry by other researchers has also drawn distinctions between the moral and personal domains (Kelly & Stich, 2008; Nucci, 1981; Nucci, et al., 1991; Nucci & Turiel, 2000). By asking respondents to instead make a criterion-judgment, it was hoped that gender differences in the construal of issues as belonging to the personal or moral domain might be assessed, as well as shedding light on whether general consensus exists on what constitutes a moral issue.

Items for inclusion were selected to represent a broad sample of social and moral issues, and included items from the Morally Debatable Behaviors Scale (Harding & Phillips, 1986; Katz, et al., 1994) which has been previously shown to exhibit some gender differences in endorsement. Additional items were added on the basis of intuition, findings on gender differences in attitudes such as sexual and reproductive issues (Petersen & Hyde, 2010; Whitley & Kite, 1995), as well as several issues that in recent years have been called “moral issues” in the public and political arena, such as climate change, to test how broadly the moral domain is defined by respondents.
Procedure

Participants were recruited from student subject pools to complete either the written surveys (taking approximately thirty to forty minutes) or survey and Ethic of Care Interview. Additionally, community samples were recruited using personal contacts, Facebook, and by placing a notice on GetParticipants.com, a site which assists in contacting people in the community who have registered their willingness to participate in scientific research. An online survey for this purpose, hosted by SurveyMonkey, contained the criterion-judgment tasks, Defining Issues Test, and Bem Sex-Role Inventory. Past research with the DIT has shown equivalency between paper and online versions (Xu, Iran-Nejad, & Thoma, 2007), and this allowed the recruitment of participants without face to face contact, thus broadening our sample. Deemed informed consent was sought in the case of survey-only participants, and additional written consent sought for recording of the ECI. Written consent forms were kept separately from surveys, and a unique participant number was used to link survey and interviews.

Interviews were recorded electronically, and administered one-to-one in a quiet setting either at the university or a place of their choosing. These interviews were transcribed for further analysis and coding. All identifying information, such as age, sex-role identity, student or community sample, was removed so that coding and rating of transcripts was blind to such information.

After giving informed consent, either for the written/online survey, or both for participants recruited face-to-face, participants completed either the written survey measure or the Ethic of Care Interview. Ideally, the order of measures should be counterbalanced, however practical considerations on the availability of a single
interviewer meant that participants would often be filling out a survey while waiting for their interview.

**Analyses**

Data was analysed using SPSS v18, after being imported directly from SurveyMonkey, with written paper responses entered manually into the online survey to assist with data validation. Calculation of DIT index scores is a complex process and involves additional transformation and validation. Though data is often sent to the test authors for processing, calculations may also be performed manually (Rest, 1979b), and SPSS syntax is provided for replication of this process and examination of validation cutoffs (see Appendix A).

**Results**

Data was screened for data-entry errors, univariate and multivariate outliers and normality. All variables for analysis were normally distributed, with the exception of age which showed moderate negative skew (skew = 4.70, \( p < .001 \)). A log transformation was applied to age, and analysis performed with transformed and untransformed age. As there was no change in statistical significance, the untransformed age is reported. Table 1 presents summary statistics for the variables of interest.

*Bem Sex-Role Inventory (BSRI)*

Participants were classified into one of four sex-role categories on the basis of published medians from the BSRI Manual (Bem, 1974). Multiple regression analyses used the masculinity and femininity scores, instead of the sex-role classification, as conversion from a continuous scale to categorical results in a loss of power.

Table 2 presents the distribution of masculine, feminine, androgynous and undifferentiated participants in the sample. Pearson correlations revealed no
Table 1

*Descriptive Statistics for BEM Sex-Role Inventory, Defining Issues Test, and Ethic of Care Interview*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>(SD)</th>
<th>95% CI</th>
<th>Reliability alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEM Masculinity</td>
<td>4.75</td>
<td>.90</td>
<td>4.64 – 4.86</td>
<td>.86</td>
</tr>
<tr>
<td>BEM Femininity</td>
<td>5.23</td>
<td>.87</td>
<td>5.12 – 5.33</td>
<td>.88</td>
</tr>
<tr>
<td>DIT P Index</td>
<td>21.83</td>
<td>13.14</td>
<td>20.09 – 23.58</td>
<td>.45</td>
</tr>
<tr>
<td>Ethic of Care Interview</td>
<td>6.78</td>
<td>1.24</td>
<td>6.50 – 7.06</td>
<td>.63</td>
</tr>
</tbody>
</table>

Table 2

*Distribution of sex-role categories*

<table>
<thead>
<tr>
<th>Sex</th>
<th>Masculine</th>
<th>Feminine</th>
<th>Androgynous</th>
<th>Undifferentiated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>39</td>
<td>13</td>
<td>20</td>
<td>33</td>
</tr>
<tr>
<td>Females</td>
<td>31</td>
<td>34</td>
<td>40</td>
<td>50</td>
</tr>
</tbody>
</table>
significant association between age, education, masculinity and femininity. Chi-square analysis of sex by gender was significant, \( \chi^2(3) = 11.25 \ p = .010 \ \varphi = .21 \). More women than men were in the feminine or androgynous categories, and a greater percentage of women were cross-sex-typed.

**Missing Data**

Missing data can represent a threat to internal validity if the sample including complete measurements differs greatly from those participants with missing data. Since only a subset \( (n = 84) \) of participants were interviewed, missing values analysis (MVA) using SPSS was performed to examine if there were systematic differences between the two samples on demographics of interest (for example, if those completing the interview showed higher moral judgment than non-completers). MVA showed no difference between complete and incomplete ECI participants on sex-role or DIT moral judgment \( (p = .814) \) indicating that the interview sample did not differ in a systematic fashion on gender or morality judgment. Interviewed participants were, however, younger than the online sample \( (p = .001) \) and slightly less educated \( (p = .067) \) reflecting the composition of student subject pool participants. As prior research suggests that sex and gender differences increase throughout the lifespan, an over-reliance on younger interview participants may increase the likelihood of a Type II error, missing differences that might have been found in older participants. Analysis showed the data to be missing at random (MAR), meaning that the interview subgroup did not differ on the dependent variable of morality (Baraldi & Enders, 2010).

**Justice-based Moral Judgment using the Defining Issues Test (DIT)**

Participant responses to the DIT were examined using the exclusion criteria specified in the DIT manual (Rest, 1979b). Computer scored validation checks were
performed, assessing the endorsement of $M$ (meaningless) items, as well as rate/ranking discrepancies. From the initial sample of completed protocols ($n = 253$), thirty-one protocols were excluded, resulting in an exclusion rate of 12.3%. The majority of excluded protocols were from online participants, which may reflect differences between completing the DIT as a paper survey and online. Typical exclusion rates are 5-15% depending on the motivation and reading age of participants (Rest, 1979b), though previous studies at Griffith using the DIT with Australian audiences have shown exclusion rates as high as 18% (Herington & Weaven, 2008). Statistical analyses were performed on the full, and the excluded sample; significance of findings still held, and the validated sample is reported.

An independent samples $t$-test was performed on DIT scores to rule out a selection bias with the choice of sample; differences between males and females as a group should be minimal, if any. No significant difference between males and females was found, $t(221) = -1.45$, $p = .149$, $d = -.19$, though the direction favoured women.

Figure 2 shows mean P-Index scores for the four sex-role categories. A multifactorial sex by gender ANOVA was not reported due to insufficient cell size for feminine males. Instead, a 4x (sex-role) between-groups ANOVA was performed on DIT scores, and all assumptions (independence, normality and homogeneity of variance) held. There was a significant difference for sex-role, $F(3, 219) = 4.44$, $p = .005$, $\eta_p^2 = .06$ which is a small to medium effect size given Cohen’s (1988) conventions.

Three contrasts determined $a$ priori for hypothesized gender differences were performed. The first contrast compared masculine ($M = 17.17$, $SD = 13.03$) participants to feminine ($M = 25.47$, $SD = 12.65$) participants, showing the former
Figure 2. DIT P-Index scores for sex-role identity groups. Error bars represent standard error of the mean (SEM).

Table 3

Zero-order Correlations between DIT Moral Judgment and Predictors

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4a</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DIT</td>
<td>-</td>
<td>.14*</td>
<td>.17**</td>
<td>.10+</td>
<td>-.08</td>
<td>.27***</td>
</tr>
<tr>
<td>2 Age</td>
<td>-</td>
<td>-</td>
<td>.21**</td>
<td>.05</td>
<td>-.01</td>
<td>.08</td>
</tr>
<tr>
<td>3 Education</td>
<td>-</td>
<td>-</td>
<td>-.04</td>
<td>.06</td>
<td>.12*</td>
<td></td>
</tr>
<tr>
<td>4 Sexa</td>
<td>-</td>
<td>-</td>
<td>-.15*</td>
<td>-</td>
<td>.28***</td>
<td></td>
</tr>
<tr>
<td>5 BEM Mascul.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.13*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 BEM Femin.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .10, **p < .05, ***p < .01, ****p < .0001

a Dummy coded variable
group had lower preference for post-conventional reasoning, $t(219) = -3.12, p = .002, d = -.42$. Masculine participants were also compared with the androgynous group ($M = 24.65, SD = 12.67$), and were also found to be lower $t(219) = -2.99, p = .003, d = -.40$. Both of these are considered medium effect sizes. The final contrast showed no significant difference between feminine and androgynous participants ($p = .770$).

An hierarchical multiple regression analysis ($N = 218$) was conducted, in order to assess the hypothesis that the effect of psychological gender is stronger than biological sex, and to rule out potential confounds of age and education. Table 3 presents the zero-order correlations between these variables, and Table 4 presents the hierarchical regression. Sex was dummy coded, with males coded as 0 and females as 1.

Two participants were identified as multivariate outliers, and were investigated as influential scores. Excluding these participants did not greatly increase model fit, and they were retained for analysis using the full sample. All assumptions for regression were met. At Step 1, age and education were entered, accounting for 3.5% of variance in moral judgment, $F_{chg} (2, 216) = 4.31, p = .015$, with education being the stronger of the two predictors. Sex of the participant was entered at Step 2, but did not improve the model fit, $F_{chg} (1, 215) = 2.4-, p = .123$. Finally, masculinity and femininity scores were entered together at Step 3, $F_{chg} (2, 213) = 6.49, p = .002$.

The final model, $F (5, 210) = 4.92, p < .001$, using all of the predictors, explained approximately 10.4% of unique variance in moral judgment scores (95%CI = 2.9% to 17.9%). In the final inspection of regression coefficients shown in Table 4, age and education did not make a significant contribution to moral judgement and can be ruled out as a potential confounds in this research. Furthermore, sex and masculine sex-role identity did not contribute to moral judgement, and only feminine
Table 4

*Hierarchical Multiple Regression of DIT Moral Judgment and Predictors*

(N = 218)

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>(SE B)</th>
<th>β</th>
<th>sr²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.09</td>
<td>.06</td>
<td>.10</td>
<td>.01</td>
<td>.130</td>
</tr>
<tr>
<td>Education</td>
<td>1.24</td>
<td>.58</td>
<td>.15</td>
<td>.02</td>
<td>.034*</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.09</td>
<td>.06</td>
<td>.10</td>
<td>.01</td>
<td>.153</td>
</tr>
<tr>
<td>Education</td>
<td>1.29</td>
<td>.58</td>
<td>.15</td>
<td>.02</td>
<td>.028*</td>
</tr>
<tr>
<td>Sex</td>
<td>2.78</td>
<td>1.80</td>
<td>.10</td>
<td>.01</td>
<td>.123</td>
</tr>
<tr>
<td>BEM Masc.</td>
<td>-.75</td>
<td>.96</td>
<td>-.05</td>
<td>.00</td>
<td>.437</td>
</tr>
<tr>
<td>BEM Femin.</td>
<td>3.48</td>
<td>1.02</td>
<td>.24</td>
<td>.05</td>
<td>.001**</td>
</tr>
</tbody>
</table>

**Note:** R² at Step 1 = .04, p < .05  
R² at Step 2 = .04, p < .05  
R² at Step 3 = .096, p < .001  
+ p < .10  
* p < .05  
** p < .01  
*** p < .001
sex role predicted moral judgment performance. This indicates that poorer justice-
based moral judgment was associated with a lack of feminine sex-role identification,
rather than a masculine identity per se.

Care-based Moral Judgment using the Ethic of Care Interview (ECI)

To assess inter-rater reliability using the ECI coding scheme, a randomly
selected sample of 20 transcripts (10 male, 10 female) were classified by a second
rater (another 4th year honours student) not affiliated with the project and blind to
experimental hypotheses. To ensure independence, only the transcripts were
provided rather than the entire data file, so that judgment would also be blind to age,
education, and psychological gender of each participant. Each transcript contained
three stories, for a total of 60 vignette response. A good level of inter-rater reliability
was found, Cohen’s kappa = .684, p < .001 (95%CI + .135).

An independent samples t-test was performed on DIT scores to rule out a
selection bias with the choice of sample; differences between males and females as a
group should be minimal, if any. Women showed slightly higher ECI scores than
men, but the difference was non-significant, t(82) = -1.17, p = .244, d = -.26. This
effect size is consistent with meta-analysis by Jaffee and Hyde (2000) and supports
generalisability of the interview sample.

As only a subset of participants were administered the ECI (n = 84), a
multifactorial ANOVA was impractical due to small cell size for male participants,
especially for cross-sex typed males. Preliminary analysis run with males and
females separately showed the same directional differences between sex-role
categories so results were combined. A between-groups 4x (gender) ANOVA was
conducted on total ECI scores across all three dilemmas (see Figure 3). Homogeneity
of variance was investigated, with a significant Levene’s test; participants in the
Figure 3. Ethic of Care Interview scores by psychological gender. Error bars represent standard error of the mean (SEM).
masculine and undifferentiated group showed slightly wider variance than the feminine and androgynous group. The ratio between the highest and lowest variance groups was low ($F_{\text{Max}} < 3$), however as the cell size for each group was not equal, a more conservative Welch correction to the degrees of freedom was applied. There was a significant difference in ECI total, $F(3, 39) = 3.39, p = .027, \eta_p^2 = .11$.

Planned contrasts determined \textit{a priori} were conducted to test experimental hypotheses, comparing masculine ($M = 6.16$ SD = 1.36), feminine ($M = 7.2$, SD = .82) and androgynous ($M = 7.1$, SD = 1.15) participants. Contrasts showed masculine participants to have attained a lower stage than feminine $t(80) = -2.89, p = .005, d = -.64$ and androgynous $t(80) = -2.30, p = .024, d = -.51$ participants. Stage differences between feminine and androgynous were non-significant $t(80) = .26, p = .797$; the modal stage for both groups was 2.5, and feminine participants were not more likely to represent predominantly Stage 2 reasoning.

Bivariate correlations and scatterplots were examined prior to conducting multiple regression, to determine suitability of predictors. Though it has been argued in the past that associations between sex, gender, and care-based moral judgment have been confounded by age and education, low correlations with these demographic variables were found in the present sample. During piloting of the interview instrument, visual inspection of scatterplots suggested the association between gender and care developmental stage was moderated by sex, with steeper slope coefficients for females than men. When the complete sample was analysed, sex as a moderator failed to reach statistical significance using moderator criteria (Baron & Kenny, 1986), and the interaction term was omitted.

An hierarchical multiple regression analysis ($n = 84$) was conducted, in order to assess the hypothesis that the effect of psychological gender is stronger than
### Table 5

*Hierarchical Multiple Regression of ECI Moral Judgment and Predictors (n = 84)*

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>sr²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.01</td>
<td>.01</td>
<td>.15</td>
<td>.02</td>
<td>.209</td>
</tr>
<tr>
<td>Education</td>
<td>-.05</td>
<td>.10</td>
<td>-.03</td>
<td>.00</td>
<td>.794</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.01</td>
<td>.01</td>
<td>.15</td>
<td>.02</td>
<td>.202</td>
</tr>
<tr>
<td>Education</td>
<td>-.02</td>
<td>.10</td>
<td>-.02</td>
<td>.00</td>
<td>.882</td>
</tr>
<tr>
<td>Sex</td>
<td>.35</td>
<td>.29</td>
<td>.13</td>
<td>.02</td>
<td>.230</td>
</tr>
<tr>
<td>BEM Masc.</td>
<td>-.07</td>
<td>.16</td>
<td>.05</td>
<td>.00</td>
<td>.665</td>
</tr>
<tr>
<td>BEM Femin.</td>
<td>.55</td>
<td>.17</td>
<td>.36</td>
<td>.11</td>
<td>.002**</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.00</td>
<td>.01</td>
<td>.05</td>
<td>.00</td>
<td>.672</td>
</tr>
<tr>
<td>Education</td>
<td>-.04</td>
<td>.10</td>
<td>-.05</td>
<td>.00</td>
<td>.671</td>
</tr>
<tr>
<td>Sex</td>
<td>.21</td>
<td>.29</td>
<td>.08</td>
<td>.01</td>
<td>.471</td>
</tr>
<tr>
<td>BEM Masc.</td>
<td>-.07</td>
<td>.16</td>
<td>.05</td>
<td>.00</td>
<td>.665</td>
</tr>
<tr>
<td>BEM Femin.</td>
<td>.55</td>
<td>.17</td>
<td>.36</td>
<td>.11</td>
<td>.002**</td>
</tr>
</tbody>
</table>

*Note: R² at Step 1 = .03, p = .360   R² at Step 2 = .05, p = .281   R² at Step 3 = .16, p = .017*

* p < .05, **p < .01, ***p < .001
biological sex, and to rule out potential confounds of age and education (see Table 5). Two participants had missing scores for education, and were means-substituted so that they were included in analysis. At Step 1 age and education were entered as predictors, and sex in Step 2; neither of which significantly improved model fit. Finally at Step 3, BSRI masculinity and femininity scores were entered, $F_{chg}(2, 78) = 5.76, p = .005$. Of the two, only femininity ($\beta = .36, p = .002$) made a significant independent contribution as a predictor, with higher femininity scores associated with more advanced stages of care-based moral judgment.

Overall the final model explained 16.1% (95% CI = 2.75 to 29.45%) of the variance in ethic of care stage, $F(5,78) = 3.00, p = .016$. Individually, femininity made a unique contribution of 11.3%, with the remaining variance being shared by other predictors. As hypothesized, sex-role was a stronger predictor than the association with biological sex, even after controlling for age and education.

Moral Construal

Figure 4 presents percentages of respondents seeing an issue as belonging to the moral domain. For the majority of issues, great individual variation exists in whether each is classified as a moral or personal issue. The only issues of strong consensus as being moral issues (greater than 70%) were cheating on taxes, cheating on exams, affairs with a married man or woman, receiving stolen goods and accepting bribes. Additionally, a number of decisions were regarded as inherently personal, and not moral issues (with fewer than 30% of participants endorsing these as moral issues). Several of these have traditionally been regarded as being moral issues, such as taking marijuana, divorce, and contraception. Control measures of obesity as a moral issue (4%) and expressing physical affection in public (5.5%), interestingly, did show some mild endorsement as moral issues.
Figure 4. Percentage of respondents classifying an issue as belonging to the moral domain
Sex-related differences in moral construal were evaluated using chi-square analysis, with exact probabilities calculated as asymptotic and Monte-Carlo estimations lacked sufficient power and accuracy (Mehta, 2008). Per hypotheses, these were grouped into moral transgressions involving social harm, which it was hypothesised would be more frequently construed by women as moral, and issues of personal and sexual autonomy which would be more frequently seen as moral by males. The hypothesised difference in moral construal for psychological gender was investigated, but not supported. There was no difference between masculine, feminine and androgynous participants, contrary to hypotheses.

On issues involving transgressions towards others or social harm, items that were more frequently construed as moral issues by women than men were accepting bribes $\chi^2(1) = 5.02, p = .019, \phi = .14$; cheating on exams, $\chi^2(1) = 10.18, p = .001, \phi = .19$; purchasing stolen goods, $\chi^2(1) = 5.07, p = .018, \phi = .14$; and cheating on tax, $\chi^2(1) = 4.47, p = .024, \phi = .13$. No sex differences were found on lying in one’s own self-interest, divorce, affairs with married partners, or euthanasia as a moral issue.

On items involving personal autonomy and sexual freedom, males more frequently viewed these as moral issues in relation to homosexuality, $\chi^2(1) = 9.23, p = .002, \phi = -.18$; contraception $\chi^2(1) = 2.84, p = .042, \phi = -.12$; taking an antidepressant medication, $\chi^2(1) = 5.91, p = .020, \phi = -.15$; and using pain medication or sleeping tablets, $\chi^2(1) = 6.67, p = .015, \phi = -.16$. One other issue, abortion, approached significance ($p = .062$). In contrast, hypothesised differences in relationships with much younger partners were not found. However women more frequently saw prostitution as a client to be a moral issue, $\chi^2(1) = 4.05, p = .029, \phi = .12$, and for the sex worker themselves though this failed to reach statistical significance; the direct of difference was contrary to hypotheses.
Discussion

The results of this study provide evidence of gender differences in moral cognition, using both quantitative and qualitative methods. Though sex differences in this domain have been widely reported in literature (Jaffee & Hyde, 2000), it has been argued that they are small in nature, and when found could be easily explained by differences in level of education (Walker, 1984, 2005). On the basis of the findings in this study, an alternative argument is made: gender differences in moral cognition exist, but can be found more readily using psychological gender and sex-role identity of participants than their biological sex.

Justice-based moral judgment

Consistent with the first hypothesis, feminine and androgynous participants demonstrated greater preference for post-conventional moral reasoning than masculine participants. As a whole, there were no meaningful differences between males and females which would be reasonably expected if the sample were representative; when found at a population level, stage differences are small (Jaffee & Hyde, 2000; Thoma, 1986). However assessment of psychological gender did differentiate the participants, with masculine participants showing less developed moral judgment performance than feminine and androgynous groups (effect sizes for comparisons $d=-.42$ and $d=-.40$ respectively). To put this in context, educational interventions specifically aimed at improving moral and ethical decision making show extremely modest effect sizes of $d = .18$ on DIT moral judgment against control groups (Schlaefli, Rest, & Thoma, 1985), while meta-analysis by Thoma (1986) showed a difference favouring females of only $d = .28$ in adults. Thus the magnitude of gender differences is quite large in comparison to effect sizes commonly studied in moral cognition.
The direction of stage differences is consistent with some prior research (e.g., Leahy & Eiter, 1980; Pratt, et al., 1984; Pratt & Royer, 1982), but inconsistent with others (e.g., Karniol, Ekbali, & Vashdi, 2007; Lifton, 1985; Wark, 1992). This may be attributable to the increased sample size used in this study, recruitment from community samples, and methodological limitations on defining sex-role identity in previous studies. Classification by the BSRI may have greater predictive validity than other instruments such as the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974). Nonetheless, further replication of these findings is necessary, especially with non-student samples drawn from the wider community.

Prior research has been criticised for failure to control for the effects of age and education when testing for sex differences (Walker, 1984, 2005), and it was necessary to rule these out as confounding factors. It was hypothesised that the influence of psychological gender would be stronger than biological sex, even after controlling for age and education. By including masculinity and femininity as predictors in the final step of an hierarchical regression, we can see that psychological gender has a much greater impact than sex, age or education.

Masculinity showed no association with DIT judgment scores, while femininity did. This helps clarify the differences between masculine, feminine and androgynous participants; it is not a masculine sex-role identity that is associated with lower moral judgment per se, but rather an absence of femininity (i.e. a negative relationship between masculinity and moral judgment is not supported).

Such a finding, though predicted in this study’s hypotheses, may nonetheless seem somewhat counterintuitive, as cultural stereotypes and earlier psychological theorists (Freud, 1927; Gilligan, 1982; Piaget, 1965) associate males and masculinity with a justice orientation. This may have been true, up to a point, for advancement
through earlier stages, but the DIT instrument is particularly sensitive to a preference for post-conventional reasoning (Rest, 1979a; Rest, Narvaez, Bebeau, et al., 1999). A tendency towards Stage 3 thinking on relationships, as asserted by Gilligan (1982), or towards law-and-order justice, Stage 4, might well be seen when using Kohlberg’s MJI. However post-conventional reasoning requires an integration of these perspectives, towards higher societal-level interests. A degree of feminine sex-role identity may be a necessary prerequisite for cognitive-developmental growth away from individual relationships or legal codes, and towards integration of the best interests of society. Doing the “right” thing at this level of reasoning is no longer just obedience to arbitrary rules or personal interests, but a reflective concern for the impacts of a decision on all parties and the wider society.

Gender identity and sex-roles, though malleable, are not easily changed as gender-schema is integrated into identity and sense of self (Bem, 1981b). Nor would we want to change sex-role identity by intervention in the hope of facilitating moral development; society is best served by a plurality of perspectives. Nonetheless, sex-role identity may be a useful diagnostic indicator in adolescent and emerging adulthood for educational intervention through ethical training. The combination of high masculinity and low femininity may represent a risk factor for identifying those who may become stalled at lower levels of moral judgment. Since sex differences are small, and other contextual factors such as socio-economic status and education are relatively modest, psychological gender may represent the strongest identifiable risk factor in screening. Furthermore, the educational needs of a masculine subgroup may differ from feminine and androgynous. There is a need for further research to identify which types of moral and ethical training best suit this subgroup, rather than applying generic interventions.
Furthermore, this study suggests that the impact of age and education on justice-based moral judgment is modest at best, at least for this sample, which included participants from the community. Research in this area frequently recruits adolescent and college-age students where age-related effects may be particularly acute, and even a single year might contribute to further development. When sampling across a broad range of ages and educational experiences, their influence is minimal, and this suggests that sampling from across the lifespan is important before attempting to generalise from student subject pools to adult populations.

Care based moral judgment

Consistent with the second hypothesis, participants with a feminine or androgynous sex-role identity demonstrated more advanced stages of an ethic of care, as assessed by the ECI. Prior research (e.g., Juujärvi, 2006b; Pratt, et al., 2004; Skoe & Marcia, 1991) has shown little evidence of meaningful differences between males and females on this measure (though see Skoe, Pratt, Matthews, & Curror, 1996 for contrary evidence in mid to late adulthood). Support for the influence of sex-role identity is much stronger (Skoe, 1995; Skoe, et al., 2002; Söchting, Skoe, & Marcia, 1994). Our findings are consistent with this research, showing a large difference attributable to psychological gender but no evidence for a difference between males and females. This is not to say that a male-female difference does not exist, only that the limited sample size (especially for male interviewees) was insufficient to adequately assess such a claim. Masculine participants were at earlier developmental stages than feminine and androgynous participants ($d = -.64$ and $d = -.51$ respectively). By comparison, effect sizes of $d = .28$ are found for care-based moral orientation (Jaffee & Hyde, 2000) between males and females. Gender
differences in care-based morality were stronger than those found in justice-based moral judgment.

Hierarchical multiple regression was used to control for age and education level, and as with justice-based moral judgment, their influence was minimal with only marginally significant contributions of age which disappear after including sex and sex-role identity. Interestingly, and not hypothesised, it was found that sex may be acting as a moderator between psychological gender and developmental stage of care. There was a strong association between femininity and care, but more greatly in magnitude for female participants, though it failed to meet statistical significance for a moderator. Though an association between femininity and developmental stage of care has generally been found (e.g., Söchting, et al., 1994), two previous studies have found the relationship to be absent or more strongly present only in women (Skoe, 1995; Skoe & Diessener, 1994).

One explanation for these results may be that a feminine sex-role identity is more permissible, and even expected in women, but that men feel more constrained in their behavioural repertoire to respond in caring and nurturing ways. Such an interpretation is consistent with the position of Gilligan and Wiggins (1987) who argued for differential developmental trajectories for morality, and with gender-schema theory (Bem, 1981b). Only psychological gender was examined in this study, and it is possible that attitudes towards public displays of tenderness and concern for others, and endorsement of sex-role stereotypes (such as “only women should express care for others”, or “that to care is to be seen as weak”) may better explain these findings. Though they may self-identify with feminine values and personality traits, cross sex-typed and androgynous males may have been responding in gender-appropriate socially desirable response styles which may not be a true reflection of
their actual feelings. Women, on the other hand, may feel less constrained to describe their feelings of care and concern, and were responding more openly. This position would be consistent with Gilligan (1982)’s thesis of care representing the traditionally female voice, and inconsistent with traditional masculinity. However investigation with a larger sample of male participants is necessary before accepting that sex-role interacts differently with males and females.

Contrary to the hypotheses, feminine and androgynous participants did not differ in stage level. It was hypothesized that feminine participants would be typified by Stage 2 care-reasoning, while androgynous participants would strike a balance between the need to care for others and for themselves (Stage 3). Such a result seems paradoxical, as self-assertion is generally more associated with the androgynous personality (Bem, 1975; Kaplan, 1979) while self-sacrifice and caring for others has been associated with feminine sex roles (Gilligan, 1982). It is also inconsistent with a previous study measuring sex-role identity and developmental stage of care using the ECI (Söchting, et al., 1994). Feminine and androgynous participants showed a mixture between the two stages (modal stage of 2.5), and further investigation is required to determine if this is representative, or if the study merely lacked sufficient statistical power to differentiate the sex-roles.

It seems clear from these findings that an association between sex-role identity and care-based judgment exists. It may therefore be useful as a diagnostic indicator for allocation of limited moral education resources, as well as for better understanding the aetiology of moral judgment. Though only correlational, such findings merit further investigation in a longitudinal study to investigate potential causality, or a bidirectional influence between psychological gender and moral judgment.
Results of the criterion-judgment classification tasks give strong support to differing perceptions and views of the moral domain. Perceptions of what constitutes a “moral” issue, and what remains a personal discretionary decision, are heterogeneous with very few issues seen uniformly. Furthermore these findings may be useful reference points for future investigators examining temporal effects in the moral domain.

Moral judgment is an important aspect in the decision-making process, but for moral judgment processes to be engaged, a decision must first be seen in a morally right or wrong light. It would seem that quite different perceptions of the moral and personal domains are present in society, and this has important implications for public policy efforts and interventions to induce behavioural change. Taking two issues that showed mixed support for example, namely climate change and music/video piracy, there is a strong divide on whether they actually constitute a moral issue or not. Efforts at inducing behavioural change by government and business interests may ultimately prove ineffective without changing the way in which these issues are seen.

On the subject of sex differences, as hypothesised, men and women differed on a number of issues and in the direction expected. These findings are consistent with previous research (Katz, et al., 1994) that men and women see the relative “wrongness” of actions differently, but goes further to show that there are sex differences in whether an action is even perceived as moral in nature. Chi-square analysis generally supported the predicted pattern that women construe issues involving transgressions or harm to others as more frequently involving moral elements than men, and that issues related to medical autonomy and
reproductive/sexual freedom are less frequently seen in a moral light. Of interest though is that the issue of prostitution presented a mixed finding, inconsistent with previous studies (Katz, et al., 1994). Females saw moral issues for clients more frequently than men, but also for the sex worker *themselves* (though the difference was small, and failed to reach statistical significance). Making a distinction between the two, therefore, is important when investigating perceptions of morality.

The hypothesis that psychological gender would show differences in construal of the moral domain was not supported. Given that gender differences were found in moral judgment, one would expect to see the same for construal processes. Though speculative, it might suggest that the aetiology of our perceptions of what constitutes a “moral” issue is different to that of moral judgment, and less malleable to gender-schema. Whether these reflect differences in socialisation, or qualitatively different views of the world between men and women, is difficult to argue but merits further investigation.

Turiel (1978, 1989) identified a number of specific properties that are prerequisite for classification as a moral issue, but this study used a forced-choice response (personal or moral) for criterion-judgments which is a new methodology. Previous research by Nucci (1981) employed cards containing a range of issues and behaviour which were sorted by children into moral, social or personal piles. Convergent validity with other mediums should be established, to identify whether forced choice produces similar results. Anecdotally, participants reported criterion-judgments to be extremely cognitively engaging and thought-provoking, so it is likely that sufficient time was spent reflecting on each issue.

However, it is not clear whether these results reflect implicit values and beliefs held about morality, or whether they reflect only explicit beliefs or a gender-
appropriate response style. For example, one might believe homosexuality and abortion to be immoral but classify these issues as a personal decision to reflect cultural mores and prohibitions against homophobia and respect for a woman’s choice. Alternate methods, such as an implicit association test, might be employed, to strengthen confidence in findings in criterion-judgment tasks.

**Implications of the study**

It goes without saying that the issue of sex and gender differences in moral cognition is contentious (Fine, 2010), and some argue divisive in perpetuating gender stereotypes (Hyde, 2005; Walker, 2005). Therefore one must be cautious in the interpretation of such differences when found, and their practical implications. Historically, findings of sex differences have been used to argue the inferiority of women’s morality (Freud, 1927, 1961b; Gilligan & Wiggins, 1987; Piaget, 1965), so perhaps it is a refreshing change that femininity might be associated with more developed stages of moral cognition, and for women to hold more enlightened beliefs about moral construal. But one must be cautious, however, not to draw the conclusion that any single gender is “better”, “more noble”, or “more virtuous”. To do so is a philosophical perspective, and not a scientific one.

While encouraging greater moral development is an important goal, one must be careful not to disparage or marginalise any group as “less moral”, or to promote another as “more moral”. Rather, moral cognition is fluid and modifiable, and group differences on moral judgment tasks serve only as developmental indicators and risk factors to identify targets for intervention.

Furthermore, these findings suggest that focus on psychological gender rather than biological sex may be a more productive line of research in moral cognition, with effect sizes that are stronger in magnitude than those previously found between
just males and females. Hyde (2005) has argued that general sex differences in psychology are small in nature and of little practical significance, advancing a “gender similarities” hypothesis. Hyde illustrated this powerfully by graphing two normal distributions that differed by an effect size considered “small” by Cohen’s (1988) conventions. By way of comparison, the effect size between males and females for justice-based DIT reasoning is shown in Figure 5a, while Figure 5b shows the effect size and reduced overlap found between masculine and feminine participants in this study. Psychological gender and sex-role identity supports greater explanation of individual differences in developmental stage.

Such an argument also presents an explanation for the varied and contradictory findings of sex differences by some researchers. This may be partially explained by variation in sex-roles between samples. Many studies recruit from psychology student subject pools where individual variation in sex-role identity may be slight (thus supporting null findings), while other subject pools from different regions, faculties, and universities might have greater variation in sex-roles, especially when recruiting from non-psychology and community samples.

Limitations of the study

It is important to acknowledge several limitations of this study. Firstly the sample size for the additional interview measure was small, especially in the number of male participants. This is particularly important in evaluating a possible interaction between sex and gender for care-based moral judgment. Additionally, internal consistency on the DIT instrument, especially with the online sample, was lower than desirable and introduces additional noise to the data. The effect size is quite strong and robust, and therefore unlikely to be a product of measurement error, but may be underestimating the true differences between sex-role categories. Though
Figure 5a. Two normal distributions that are 0.18 standard deviations apart - the difference found between males and females on DIT responses.

Figure 5b. In contrast, two normal distributions that are 0.42 standard deviations apart - the difference between masculine and feminine participants on DIT found in this study.
studies have established discriminant validity of DIT from general intelligence, verbal ability, and political ideology (see for a review Thoma, Barnett, Rest, & Narvaez, 1999) the language of the original DIT test is becoming dated and was confusing for some participants. Replication of these gender differences with the newer version (DIT2) is warranted once psychometric properties and reliability of online administration of the newer form is documented. Additionally, one would want to rule out other potential confounds such as political attitudes towards liberalism or conservatism, or general intelligence.

Summary

This study adds to a growing body of literature on sex and gender differences in moral cognition, and strengthens the argument that, at least in adult populations, individual variation can be better explained by psychological gender than biological sex per se. Conflicting findings of sex differences in prior research may be explained by sample selection bias, and failure to take into account the gender-role identity of participants, as well as an over-reliance on adolescent and young adult participants. Additionally, men and women differ in the way they perceive the moral domain, and whether they construe an issue as involving moral elements. By broadening investigation into morality beyond simply moral judgment, and recruiting from across the lifespan, we may gain a clearer picture of the processes involved in moral cognition and the role that culture and gender schema plays in moral thought and behaviour.
References


*Psychological Review, 88, 354-364.*


*Journal of Personality and Social Psychology, 33, 48-54.*


*Psychology of Women Quarterly, 3, 248-254.*


Appendix A

BEM Sex-Role Inventory (BSRI)

This instrument is used to measure the psychological gender, or sex-role identity, of participants. On the basis of a median split, participants can be assigned to either a masculine, feminine, androgynous or undifferentiated group. The short form (30 item) version was used for this study, and a sample of the instrument is shown overleaf. Neutral filler items are alternated between masculine and feminine items so that the true nature of the test is not readily obvious. Two independent scales are produced representing masculine and feminine self-identity as a continuous variable. As there is a loss in power converting from a continuous to categorical variable, the masculine and feminine subscales are used in analysis.

As this is a copyrighted instrument, only a limited portion of the instrument is reproduced *per the licensing agreement* (included overleaf). A license was purchased to use with up to three hundred participants, which authorizes reproduction of five sample items and the format of the instrument.

<table>
<thead>
<tr>
<th>Masculinity Score</th>
<th>Below Median</th>
<th>Above Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Median</td>
<td>Undifferentiated</td>
<td>Masculine</td>
</tr>
<tr>
<td></td>
<td>( low-low )</td>
<td>( low fem. - high masc. )</td>
</tr>
<tr>
<td>Above Median</td>
<td>Feminine</td>
<td>Androgynous</td>
</tr>
<tr>
<td></td>
<td>( high fem. - low masc. )</td>
<td>( high - high )</td>
</tr>
</tbody>
</table>

*Figure 1.* Using a median split on masculinity and femininity scores, participants can be classified into one of four categories.
**Personality Inventory**

On this page are listed a number of personality characteristics. We would like you to use those characteristics to describe yourself, by indicating on a scale from 1 to 7, how true of you each of these characteristics is. Individual responses will NOT be analysed or identified; we are looking only for group averages of these personality traits.

**Example: sly**

If I am never sly, then I might record a 1. If it is always true of me a 7. Or I might record somewhere in between, such as 3 or 5.

**SAMPLE ANSWERS**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sly</td>
<td>3</td>
</tr>
<tr>
<td>Carefree</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Never or almost never true</td>
</tr>
<tr>
<td>2</td>
<td>Usually not true</td>
</tr>
<tr>
<td>3</td>
<td>Sometimes but infrequently true</td>
</tr>
<tr>
<td>4</td>
<td>Occasionally true</td>
</tr>
<tr>
<td>5</td>
<td>Often true</td>
</tr>
<tr>
<td>6</td>
<td>Usually true</td>
</tr>
<tr>
<td>7</td>
<td>Always or almost always true</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1. Defend my own beliefs</th>
<th>16. Have leadership abilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Understanding</td>
<td>.....</td>
</tr>
</tbody>
</table>
To whom it may concern,

This letter is to grant permission for the above named person to use the following copyright material:

Instrument: **Bem Sex Role Inventory**

Author: **Sandra Lipsitz Bem**


for his/her thesis research.

Five sample items from this instrument may be reproduced for inclusion in a proposal, thesis, or dissertation.

The entire instrument may not be included or reproduced at any time in any other published material.

Sincerely,

Robert Most
Mind Garden, Inc.
www.mindgarden.com
Appendix B

Defining Issues Test

Rest’s Defining Issues Test (DIT) is a written test of moral judgment which may be group administered or completed online (Rest, Douglas, Richard, JoAnna, & Douglas, 1974). It uses original Kohlbergian dilemmas, such as the famous “Heinz and the Drug”, but is a recognition measure rather than a production one. Instead of verbally asking the participant what is the “right” thing to do and to justify the choice, the participant is presented with the dilemma and a series of twelve items representing different points of view. These represent different stages of moral reasoning, and prior research has shown that participants will endorse judgments at the same or similar stage to their own reasoning, and reject or rate less highly those at higher or lower stages (Rest, Turiel, & Kohlberg, 1969).

An example vignette and rating decision given to participants as part of their instructions is shown. Though non-moral in nature, it gives a good indication of the format of the instrument, without reproducing sensitive materials. As the remainder of the instrument is both sensitive and copyrighted, only a paraphrase of the vignettes used in this particular study are given overleaf. Additionally, stage summaries by Rest are provided and syntax for scoring and validation of responses to allow for independent replication.

Though a copyrighted instrument, the author the late James R. Rest granted a license to use the Defining Issues Test for psychological research, saying “all professional and student researchers affiliated with recognized institutions are encouraged to use it”. The student researcher acknowledges his generosity and the legacy to moral research he contributed.
Instructions and example

In this section of the survey, we are interested in the way you think about moral issues and social problems. Different people often have different opinions about questions of right and wrong. There are no “right” answers in the way that there are right answers to maths problems. We would like you to tell us what you think about several problem stories. These answers will be fed into a computer to find the average for the whole group, and no one will see your individual answers. Some statements might not even make sense, and should be marked as of no importance. A sample story is given below.

Example: Frank Jones has been thinking about buying a car. He is married, has two children, and earns an average income. The car he buys will be his family’s only car. In trying to decide what car to buy, Frank realized that there were a lot of questions to consider. Below is a list of some of these questions. Please rate these questions, to indicate whether you feel they are important factors to consider in making a decision about buying a car.

Instructions: Below you will find a list of issues Frank might view as important. Please place a tick for each issue, rating it from great importance to no importance.

<table>
<thead>
<tr>
<th>IMPORTANCE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Great</td>
</tr>
<tr>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
</tr>
</tbody>
</table>

1. Whether the car dealer was willing to discount for cash. (This issue was of some importance)
2. Would a used car be more economical in the long run than a new car. (This issue was of great importance)
3. Whether the colour was green, Frank’s favourite colour
4. Whether it was good for the environment (Frank feels this is a very important issue)
5. Would a large, roomy car be better for the kids than a compact car.
6. Whether the front connectivity are differential.

(Note that if a statement sounds like gibberish or nonsense to you, mark it “no importance”)

We would now like you to rank these factors in order of importance. Put the number of the most important issue first, followed by the second most important issue, and so on. For example, Frank might consider issue 2 to be critical, followed by issue 5, then issue 3. If you are unsure of a fourth, please leave it blank.

Most important 2
Third most important 1
Second most important 4
Fourth most important 3
Table 1

Summary of moral vignettes used in DIT short form

<table>
<thead>
<tr>
<th>Story Number</th>
<th>Title</th>
<th>Overview of content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Heinz and the drug</td>
<td>Story content from Kohlberg’s moral judgment interview, in which protagonist considers stealing a life-saving cancer medication for his terminally ill wife as he is unable to afford the many thousands required for the cure.</td>
</tr>
<tr>
<td>2</td>
<td>Escaped Prisoner</td>
<td>A man convicted of a crime is sentenced for ten years, and serves out only part of his sentence before escaping. Living under an assumed identity, he devotes his life tirelessly for decades to helping others and atoning for his crime. A good friend discovers that he is actually an escaped prisoner, and ponders what action to take.</td>
</tr>
<tr>
<td>3</td>
<td>School Newspaper</td>
<td>A student who sought permission to publish a school newspaper from his principal about local issues important to the students finds that some parents are angry at the views expressed over issues like uniforms. The principal never thought it would become so popular, and considers banning the newspaper.</td>
</tr>
<tr>
<td>Stage</td>
<td>Morality Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>The morality of obedience: Do what you’re told.</td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>The morality of instrumental egoism and simple exchange: Let’s make a deal</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>The morality of interpersonal concordance: Be considerate, nice, and kind: you’ll make friends.</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>The morality of law and duty to social order: Everyone in society is obligated to and protected by the law.</td>
<td></td>
</tr>
<tr>
<td>Stage 5</td>
<td>The morality of consensus-building procedures: You are obligated by the arrangements that are agreed to by due process procedures.</td>
<td></td>
</tr>
<tr>
<td>Stage 6</td>
<td>The morality of nonarbitrary social cooperation: Morality is defined by how rational and impartial people would ideally organize society.</td>
<td></td>
</tr>
</tbody>
</table>

Source: (Rest, 1994, p. 5)
SPSS Syntax for computer-based scoring of DIT indexes, and validation of responses

For transparency, and independent validation of findings, the explicit exclusion criteria used in this study is given below as SPSS syntax. If three or more items contain rank/rating inconsistencies on ALL vignettes, or eight or more on any SINGLE vignette, they are excluded.

* Inconsistency checks

```spss
COMPUTE HeinzInconsistency = 0. EXECUTE.
Vector vech = Heinz_Q1 TO Heinz_Q12.
LOOP #I = 1 to 12.
+ DO IF vech(#I) < vech(Heinz_Most1).
  + COMPUTE HeinzInconsistency = HeinzInconsistency +1.
+ END IF.
  + DO IF ( vech(#I) < vech(Heinz_Most2) & #I ~= Heinz_Most1).
    /* if there is a rating more important than heinzmost2 and it is NOT actually heinzmost1 */
    + COMPUTE HeinzInconsistency = HeinzInconsistency +1.
  + END IF.
END LOOP. EXECUTE.

COMPUTE PrisonInconsistency = 0. EXECUTE.
Vector vecp = Prison_Q1 TO Prison_Q12.
LOOP #I = 1 to 12.
+ DO IF vecp(#I) < vecp(Prison_Most1).
  + COMPUTE PrisonInconsistency = PrisonInconsistency +1.
+ END IF.
  + DO IF ( vecp(#I) < vecp(Prison_Most2) & #I ~= Prison_Most1).
    /* if there is a rating more important than prisonmost2 and it is NOT actually prisonmost1 */
    + COMPUTE PrisonInconsistency = PrisonInconsistency +1.
  + END IF.
END LOOP. EXECUTE.

COMPUTE NewsInconsistency = 0. EXECUTE.
Vector vecn = News_Q1 TO News_Q12.
LOOP #I = 1 to 12.
+ DO IF vecn(#I) < vecn(News_Most1).
  + COMPUTE NewsInconsistency = NewsInconsistency +1.
+ END IF.
  + DO IF ( vecn(#I) < vecn(News_Most2) & #I ~= News_Most1).
    /* if there is a rating more important than newsmost2 and it is NOT actually newsmost1 */
    + COMPUTE NewsInconsistency = NewsInconsistency +1.
  + END IF.
END LOOP. EXECUTE.

COMPUTE Validresponse=1. EXECUTE.
IF ( Mindex > 8) ValidResponse = 0. EXECUTE.
IF ( HeinzInconsistency >= 3 & PrisonInconsistency >= 3 & NewsInconsistency >=3 ) ValidResponse=0.
EXECUTE.

IF ( HeinzInconsistency > 8 | PrisonInconsistency > 8 | NewsInconsistency > 8) ValidResponse=0.
EXECUTE.
```
Appendix C

Criterion Judgment Tasks

Criterion-judgment tasks on a range of social and potentially moral issues and actions were given. Many of these items were derived from the Revised Morally Debatable Scale used by Katz and colleagues (1994), itself a composition of items from earlier studies. The exact issues and wording of the moral construal task is reproduced overleaf. Though additional items were added by the author, an acknowledgement of the origin and contribution made by previous researchers is made. The instrument is reproduced overleaf.
Public issues and opinions survey

We’d just like to ask you a few demographic questions, to see how different groups of people answer questions. Remember this survey is entirely anonymous – please do NOT write your name on this paper.

1. Gender : Male, Female  2: Year of birth (eg 19xx) : ..............

We’d like to get your opinion on a range of personal, and social issues. We are interested in whether you consider these to be “moral” issues, or more of a social or personal issue. Regardless of whether you are for or against, we are only looking to see whether or not you see them as involving moral issues. Some may involve rather obvious moral elements, and others may involve none.

Please circle one, either ‘p’ersonal issue or ‘m’oral issue.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Personal decision or moral issue?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Accepting a bribe in the course of your work</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>2. Cheating on an exam</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>3. Having a committed sexual relationship, but not being married</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>4. Lying in your own self-interest</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>5. Having a ‘casual’ relationship or one-night stand</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>6. Buying something you know is stolen</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>7. Cheating on taxes if you have a chance to get away with it</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>8. Married men or women having an affair</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>9. Homosexuality</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>10. Contraception</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>11. Divorce</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>12. Euthanasia</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>13. Taking the drug marijuana</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>14. Abortion</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>15. Prostitution (for the sex worker)</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>16. Prostitution (for the client)</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>17. Obesity or heavily overweight</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>18. Climate change</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>19. For a man to have a relationship with a significantly younger woman</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>20. For a woman to have a relationship with a significantly younger man</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>21. Expressing affection in public (eg. kissing)</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>22. Taking an antidepressant medication</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>23. Taking pain medication or sleeping tablets</td>
<td>P ----------- M</td>
</tr>
<tr>
<td>24. Copying and sharing digital music and video (piracy)</td>
<td>P ----------- M</td>
</tr>
</tbody>
</table>
Table D1

*Ethic of care stages, and moral perspective, based on Gilligan’s (1982) conceptions of care*

<table>
<thead>
<tr>
<th>ECI Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>The morality of survival, and caring for oneself (highly egocentric, neglecting needs of others in decision making)</td>
</tr>
<tr>
<td>Level 1.5</td>
<td>Transition from exclusively self-care, to an awareness of the needs of others, but favouring self interest in relationships</td>
</tr>
<tr>
<td>Level 2</td>
<td>The morality of responsibility, and <em>care for others</em>, at the expense of the needs of the self. <em>Goodness</em> is equated with caring for others and putting their needs and wishes first; <em>right</em> is externally defined</td>
</tr>
<tr>
<td>Level 2.5</td>
<td>The morality of reflective care, still putting the needs of others first but less certain (situations are complex, and not ‘black-and-white’)</td>
</tr>
<tr>
<td>Level 3</td>
<td>The morality of <em>care for others balanced by care for self</em>. Needs, wishes and welfare of self and others are balanced, recognising the interconnectedness of relationships and that to care for others one needs to care for oneself as well.</td>
</tr>
</tbody>
</table>

Adapted from Skoe (1993), and Skoe (in press).