Gendered Morality

By

Elena Antonucci

Millersville University of Pennsylvania, 2022

A Senior Thesis Submitted to the Department of Psychology & The University Honors College

In Partial Fulfilment of the Requirements For the University Honors Baccalaureate

In the

College of Education and Human Services

Faculty of the Psychology Department

MILLERSVILLE UNIVERSITY

Spring 2022

This is a placeholder page and not part of the original document submitted for approval.

The original page of this document containing the signatures of the parties involved has been extracted to protect their privacy.

Please contact the

Millersville University Archives & Special Collections

with any questions.

Placeholder version 1.0

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work:

a. CITI training in Social & Behavioral Research, RCR, Essentials of Research Administration, and the entire COVID-19: Public Training Series including Mental Health for Higher Ed and Healthcare, Participating in Vaccine Research, Remote Contact Tracing, What You Need to Know About COVID-19 Vaccine, COVID-19: Insights for Higher Ed Learners, and COVID-19: Back to Campus.
b. Approval for the present study from the Millersville University Institutional Review Board for conducting research on humans.

Abstract

Many studies have assessed how the gender of a person taking a moral dilemma measurement effects the way in which they respond. When using the moral reasoning framework of Kohlberg, results about potential differences in the ways males and females respond to moral dilemmas vary from test to test (Eagly et al., 2000; Barriga et al., 2001). However, when using Gilligan's Ethics of Care Interview (ECI), results about gender are much more consistent (McLeod-Sordjan, 2014). This study takes a new approach, looking at the gender of a character in a moral dilemma rather than the character of participant, and measuring how responses change. The ECI has a male and female version, in which wording differences exist to highlight gender expectations (Skoe, 2014). This study uses these two conditions as controls, and additionally utilizes two experimental conditions where the gender and pronouns in each moral dilemma are swapped. A series of *t* tests are then completed to search for significant differences in responses between conditions. No significant differences were found. This may be due to limitations related to the ECI.

Keywords: gender; morality; moral reasoning

Dedication

This study and its results are dedicated to the many supporters of Elena Antonucci, who creation of this thesis or completion of this degree would not be possible without.

Acknowledgements

Firstly, I would like to acknowledge my thesis advisor, Dr. Richard Joseph Behun. The guidance and education he has given me throughout this process has been infinitely valuable and will continue to serve me in my future career. His resources and advice have helped to shape this thesis and to shape me. This thesis would not have been possible without him helping me go from a vague idea to a completed study.

I would also like to acknowledge my other thesis committee members. Dr. Elizabeth Thyrum has acted as an academic advisor to me throughout my college career and has helped shape me into the professional I am becoming. Dr. Hannah Osborn took a chance on helping me despite being new to the university, I am forever grateful that she decided to work with me. Her suggestions have helped to point me in the right direction. I am indebted to both of these wonderful women who provided me with amazing advice throughout this process.

Additionally, I would like to thank my friends and family for keeping me grounded during this process and giving me reasons to take breaks. You all have been particularly supportive and have given me the opportunity to explain my study in depth to people from numerous different backgrounds both in and outside of psychology.

Table of Contents

Declaration of Committee……..ii
Ethics Statement……..iii
Abstract……...iv
Dedication……...v
Table of Contents……vii

Chapter 1. Introduction……1
Chapter 2. Literature review……...5
Chapter 3. Methods……20
Chapter 4. Data Analysis……30
Chapter 5. Discussion……..39

References……..……51

CHAPTER 1

Introduction

Moral reasoning is a very multifaceted aspect of psychology, and many different elements contribute to it. This thesis focuses on gender, and how the societal politics of gender effect moral reasoning. To be more specific, this thesis and the study that was completed for it examine the effects of the gender of the character in the moral dilemma on moral reasoning. This is in contrast to the vast majority of studies on this topic, which tend to focus on the gender of the participant responding to the moral dilemmas rather than the gender of the character in the moral dilemma.

Background

Before beginning, it is important to define what is meant by gender. The most modern view of gender is that it is a spectrum where not everyone identifies with the gender they were assigned at birth, or any gender at all. This perspective is very recent, and still not accepted by everyone today. Thus, even well-meaning scientists have not had time to publish many studies to date which include this expanded view of gender. The result of this is a body of data which only views gender in terms of binary, cisgender individuals. Although this thesis will make every attempt to be inclusive, it should be assumed unless specified otherwise that the term gender refers to this older way of thinking. This is not an attempt by the author to exclude or simplify, but rather the end result of a need to make scientific inferences based solely on the available data. Biological sex is similar to the more rigid view of gender. This term is associated with reproductive processes present at the time of birth, and not with changes to those body parts that may be made later in one's life or any changes concerning identity and expression. In further

effort to be inclusive and acknowledge how this thesis is limited by the time in which it was written, it will utilize the terms associated with biological sex (i.e., male and female) rather than those associated with gender (i.e., man and woman) whenever possible.

The present study uses gender as an independent variable (see an important note on this below). Moral reasoning, as measured by Carol Gilligan's Ethics of Care Interview (ECI), is this study's dependent variable. Gender, as defined above, has two levels in this study: male and female. It is important to note that the gender of the participant is not relevant, but rather the gender of the character in the moral dilemma that is manipulated. Each level has a control condition and an experimental one. Condition A, which will now be referred to as Control Group A, refers to the original ECI questions as designed for a female character. Similarly, condition B, which will be referred to as Control Group B, refers to the original ECI questions as designed for a male character. In condition C, which will be called Experimental Group A, participants see the questions from Control Group A, although the gender and pronouns of the characters has been switched. Similarly, in condition D, which will now be called Experimental Group B, participants see questions from Control Group B, although the gender and pronouns of the characters has been switched.

Statement of Problem

This study serves the purpose of filling a gap in the literature on gendered morality. There are many studies that examine how males and females respond to moral dilemmas. After some debate in the field, it is generally agreed that both genders understand morality at the same level and that females tend to use higher levels of moral reasoning than males do. The levels in question follow the ideas of Lawrence Kohlberg and Carol Gilligan. Other studies have examined societal expectations based on gender roles and found significant differences. These

points are discussed in detail below. However, to date, fewer studies have combined moral reasoning as a function of gender and gender roles. The studies that have successfully combined these ideas give an implication of moral expectations based on gender roles, but do not use moral dilemma tests (i.e., either the ECI or Kohlberg's moral dilemmas) in their methodology. Thus, nothing more than inferences can be drawn from those data sets. The present study analyzes differences in participants' responses to the ECI as a function of the gender of the character in the moral dilemma. While the ECI has both a male and female version of each dilemma, they are worded differently to fit traditional male and female gender roles. The present study examines responses when the dilemmas are worded the same for both males and females. Based on the relevant literature, this study hypothesizes that participants' responses to the ECI will differ depending on if the character in the moral dilemma is a male or a female. The following research questions will also be addressed:

1. Are there significant differences between what type of behaviors are viewed as moral for males and females?
2. Are females expected to consider the needs of others over their own needs?
3. How do responses change when a moral dilemma that is specifically worded to incorporate male gender roles has a female character instead?
4. How do responses change when a moral dilemma that is specifically worded to incorporate female gender roles has a male character instead?

Purpose

The purpose of this study is to examine if the gender of a character in a moral dilemma question effects the way participants respond. Specifically, this study hopes to examine if gender roles

affect how participants answer moral dilemmas. In addition, this study aims to apply any found differences to more generalized characteristics associated with gender.

CHAPTER 2

Literature Review

Morality in psychology

Piaget

Gendered morality is a specific subfield that falls within the domains of psychology, philosophy, and gender studies. This thesis focuses on the field of psychology. Gendered morality as a field of study evolved over the course of many years. It began with Jean Piaget's stages of cognitive development. These stages span from a child's birth to adolescence and outline the development of thought and reasoning (Babakr, 2019). Piaget outlined four stages, beginning with the Sensorimotor stage. The Sensorimotor stage lasts from the time a child is born until they are around two years of age. As the name suggests, children in this stage of development take in the world and attempt to understand it though means of the stimuli in their surroundings. During this time, children develop and master the skills of causality and object permanence. Causality deals with causes and immediately acting effects. For example, a child may begin to connect spilling a bottle of milk on themself with the sensation of being covered in said milk. Object permanence refers to the awareness that an object continues to exist even when it is not in one's direct sight (Malik & Marwaha, 2021). It is important to note at this point that the logic behind causality cannot be applied to abstract thought or instances where there is a large time delay between the causes and the effect. For example, a child under the age of two cannot comprehend the sociopolitical causes of an event such as World War II.

Piaget's second stage is Preoperational. The Preoperational stage picks up directly following the end of the Sensorimotor stage and lasts until a child is around 7-years-old. This stage of development deals with the introduction of representational or symbolic thought and is

most easily understood through the tendency of children to “play pretend”; which involves using an object in the physical environment as a symbol for a mental image of something else (Scott & Cogburn, 2021). For example, a child “playing restaurant” may use a stick as a spatula or a leaf as money. In this scenario, the stick is a symbol that represents the spatula, and the leaf is a symbol representing money. Children in this stage see the world in these terms. Cognitively, they are aware that the stick and leaf are not truly a spatula and money but rather symbols of those things.

As the preoperational stage ends, the Concrete Operational stage begins. Children will remain in this stage until around 12-years-old. During this time, children gain the ability to think using more than one dimension. For example, when noticing the amount of water poured into a glass, the child can also note the relative circumference of the glass and apply this logic to understand that the same amount of water in a different glass will fill up a different amount of space. Children in this stage also gain more logic-related operations and thus can understand numerical concepts. They gain the ability to group things based on a similar characteristic and to rank the things they have characterized based on various criteria such as height, weight, length, etc. However, children still struggle with applying these newfound logical abilities to abstract situations (Ojose, 2008).

The final stage is Formal Operational. As the Concrete Operational stage ends, the Formal Operational stage begins and continues throughout adulthood. During this time, children gain the important ability to deal with abstraction, and categorize things in terms of concepts (Ahmad et al., 2016). As the Concrete Operational and Formal Operational stages are both operational in nature, they are interrelated and build on one another. The final stage takes all the

logical skills learned during the Concrete Operational period and applies them to more abstract ideas and events.

The work of Piaget became the basis for the work of many later researchers who applied his ideas to morality. One such person is Lawrence Kohlberg, who sought to expand on the work of Piaget's model to address the development of moral reasoning.

Kohlberg

Lawrence Kohlberg's new model linked the cognitive developmental ideas of Piaget to the development of morality. Thus, the stages created by Kohlberg are meant to accompany and progress alongside the stages created by Piaget (Zhang & Zhao, 2017). Kohlberg ended up creating three widely accepted sequential levels of moral reasoning, each split into two sublevels. As detailed by Zhang and Zhao (2017), the first of Kohlberg's levels is *preconventional morality*. Preconventional morality begins at birth and continues until a child is around the age of nine, roughly aligning it with Piaget's sensorimotor and preoperational stages of development. Overall, this stage deals with the concepts of good and bad. Kohlberg detailed that children start in the first substage and progress sequentially, beginning with the obedience and punishment orientation. At this point, children will base their decisions to moral dilemmas on whether or not they will be punished. Cognitively, this shows that children view rules as always being correct and are unable to comprehend morality beyond existing rules. The second substage is the self-interest orientation. This substage is largely self-explanatory; with the main idea being that rules which actively benefit the individual are good while rules which have no benefit to the individual are bad. Interestingly, Kohlberg believed that most people who engage in criminal activity never progress past this stage.

The second level of Kohlberg's' theory is *conventional morality*, which lasts until about 15-years-old, aligning it with Piaget's Concrete Operational stage. Although, Kohlberg believed that the majority of people stop progression through the stages once the conventional level is reached, hence the naming of it as conventional. Overall, the conventional level deals with choosing moral actions to gain favor in social relationships. The third substage is known as the "good boy/good girl orientation", which encompasses the spirit of this overall level. In this substage, people will answer moral dilemmas based on what their peers will approve of, regardless of whether it is truly right or wrong. The fourth substage deals with the law and order orientation. During this time, people begin to move beyond a desire for social approval and begin to act morally based on the concept of duty. People who are in the law and order substage generally feel that they have an obligation to follow rules because they have a duty to (Barger, 2000). Cognitively, this implies that between substages three and four people shift from a desire to please their peers to a desire to be respected by them.

The final level is *postconventional morality*. It is associated with Piaget's final stage of cognitive development. Just as Piaget believed that not everyone could reach his final level, Kohlberg believed that some people will never operate under postconventional morality. Overall, the postconventional level deals with responding to moral dilemmas with individually defined but not self-serving moral reasoning. Substage five encompasses the social contract orientation. For those who reach this substage, rules go from being firmly followed because of duty to being more flexible for the same reason. Rules during this time are seen as guidelines in place to serve the rather abstract concept of the greater good. This point of view allows for outdated or harmful laws to be changed or updated for reasons other than selfish desire. The sixth and final substage is home to people who deal with moral dilemmas in terms of universal ethical principles. This

stage requires thinking outside of oneself and a constant, active choice to see problems as universal instead of personal matters (Boom et al., 2018). Although Kohlberg veered away from Piaget's theory by stating that not everyone reaches this level, it is clear through his inclusion of the abstract concepts characteristic of substages five and six that postconventional morality is still based on Piaget's formal operational stage of development. The ideas of Piaget and Kohlberg are widely taught in psychology classes to this day, despite some recent accusations of biases.

Gender in psychological morality

Gilligan

The introduction of gender to this field began with Carol Gilligan's critique of Kohlberg's model. Gilligan noted that measurements of morality were created by males and favor a stereotypically male way of viewing moral dilemmas. She demonstrated that females inherently have more care-centered traits that males may develop later in life, and that morality measurements thus must include more elements of connections with others instead of solely focusing on justice as they had in the past (Gilligan, 1982). This idea of a gender dichotomy in moral orientation has been looked at following this study and found to be questionable. While there are studies which find that females do tend to focus on care and males tend to focus on justice when viewing Kohlbergian moral dilemmas (Gilligan & Attanucci, 1998; Gilligan & Wiggins, 1987; Eagly et al., 2000), the differences in answers to these moral dilemmas between males and females are often too similar to imply any inherent difference in the way these two genders view morality (Pratt et al., 1991; Jaffee & Hyde, 2000; Barriga et al., 2001). Some argue that society is what accounts for this supposed dividing of thought processes, instead of the real

internal workings of individuals (Gerson, 2002). Regardless of the proposed reasons, the studies responding to Gilligan's critique showed very mixed results.

While this may seem like Gilligan's original ideas were disproven by the subsequent data, her main point is not the existence of an across-genders difference in moral orientation but that the existing framework for taking psychological measures is where the flaws lie. She published a response to the conversation around her work in which she reiterated that females were excluded from framework and theory building in psychology, and the use of all male samples were common in the early days of the field. Thus, the only moral measurement that can be taken on females is how far they deviate from a male-defined standard (Gilligan, 1986). In this response, she criticizes those who critiqued her work for not seeing a connection between the experiences and exclusion of females and the resulting cognitions and behavior they present in various psychological tests. Despite her criticism, this male-defined standard has repeatedly shown females significantly outscoring males (Bernardi, 1997; Wark & Krebs, 1996) to the point where a meta-analysis of 56 studies were able to support that females outscore males on Kohlberg's moral levels at every age and level of education (Thoma, 1986). To solve the problem of not having female-oriented moral dilemma tests, Gilligan created the Ethic of Care Interview, which is both an interview assessment to measure moral reasoning and a framework for evaluating what level of moral reasoning a person utilizes when responding to moral dilemmas.

Gilligan's ECI

The use of other frameworks such as Gilligan's ECI (Ethic of Care Interview) has been used more in recent years in some studies in place of the traditional Kohlbergian framework. Similarly to Kohlberg, Gilligan utilizes sequential levels in her breakdown of care-centered

moral thinking. Another similarity is that both methods of moral testing do not consider the way a person responds to the moral dilemma but instead focuses on why they chose to respond that way. Each level depicts stronger understanding of a differentiation between oneself and others, and of the human balance between codependence and independence (Skoe, 2014). The overall format is that of a semi-structured interview. This means that the majority of the questions in the interview are predetermined, and participants must give open ended responses. Skoe (2014) explained that the structure in the ECI comes from three situations which are meant to highlight common moral dilemmas that are interpersonal in nature. These situations highlight conflicts about an unplanned pregnancy, marital infidelity, and caring for a parent. For each of these dilemmas a male and female version of the scenario exist and vary in the way they are worded as to further explain a situation that may not be assumed due to traditional gender roles, or change the context of a dilemma to better fit with traditional gender roles. In addition to these, a fourth dilemma is created by the participant so that at least one of the moral dilemma inducing situations will apply directly to them. Answers to these questions are analyzed and participants are placed in one of three levels, or one of two transitional levels.

Level 1 is the lowest of Gilligan's stages of moral reasoning. People who use this form of moral reasoning are said to be mainly focused on survival. Interpersonal relationships are thought about in self-protective terms, where the thing being protected is one's own happiness. Also characteristic of people in this level is an avoidance of pain. Ultimately, the result of this line of thinking is that the feelings and needs of other people are neglected or denied (Skoe, 2014). From this first level, Gilligan's work as an expansion of Kohlberg's is clear, as the reasoning behind decisions is very concrete and mimic Kohlberg's idea of preconventional morality.

Skoe (2014) goes on to explain that Level 1 is followed by a transitional level, which Gilligan labelled as Level 1.5. This in-between level transitions people from thinking about themselves to thinking about others. People in this stage struggle between a realization that solely searching for one's own interests is selfish with a reluctance to fully commit to the idea of responsibility in relationships. People in this stage share an awareness of the needs of others, but still choose to care for themselves more than the other person. Once an individual reaches Level 2 however, this changes. Level 2 embraces the idea of responsibility to the point where one's own needs are excluded. An important distinction appears here between what qualifies as good and what qualifies as right. Being good summarizes an internal concern for other people where personal desires are sacrificed, while being right summarizes how a person's actions are externally defined by other people. There is a desire to be liked and approved of by other people, which may act as the motivation for people during this stage who think in terms of what is right instead of what is good. Again, the expansion from Kohlberg's framework is clear during these stages. There are elements of conventional morality in peer acceptance as a primary motivator for acting morally. Elements of duty are also called into consideration and expanded upon by Gilligan in these levels.

Level 2 is followed by the second transitional level, Level 2.5. During this stage, people go from thinking about goodness or rightness to a much higher concern with personal honesty in relationships. In this transitional level more complexities appear, and individuals tend to acknowledge more moral gray areas. This transitional level leads to the final level: Level 3. People in this level are fully aware of the interconnected nature of human relationships. Thus, they understand the importance of balancing care for the self with care for others and see that by caring for one you can also care for the other. Answers to moral dilemmas that fall into this level

take all involved parties into consideration, including the self, and attempt to minimize harm to all of them (Skoe, 2014). In other words, the Kohlbergian and Piagetian concept of abstraction appears in Level 2.5 and grows in Level 3 to mirror the Kohlbergian idea of universal ethical principles, but specifies it for relationships with other people; again, expanding on the ideas set forth by Kohlberg.

Mixed Results

The ECI specifically measures a care-based moral orientation and is meant to complement Kohlberg's moral reasoning measurement, which are largely based in justice-oriented morality. As mentioned above, it has been shown that males and females think in terms of both care and justice when responding to morality. Generally, the higher that a person scores on the care-based orientation, the higher they will score on the justice-based orientation; However, the care-centered form of moral reasoning is more relevant to the overall identity of females than to males (Skoe & Diessner, 1994; Skoe & Marcia, 1991). The positive relationship between these two moral orientations has been shown many times (Donleavy, 2008; Conley, Jadack & Hyde, 1997), with females generally outscoring males in both (Arvizu, 1995; McLeod-Sordjan, 2014). Because care based moral reasoning increases as justice based moral reasoning does, and because these findings are consistent between sources, the argument that there is no gender dichotomy in the use of justice or care based moral reasoning is supported when using the ECI despite being questionable when using Kohlberg's moral dilemmas test.

In summary, research on the field of gendered morality has reported mixed results up until recently. When the Kohlbergian method is used, gender differences may be found or may not be. When the ECI is used, results are consistently on the side of no gender differences in

moral reasoning, but a gender difference in scores. Because the ECI has a more consistent data, it will be the basis for the moral measurement used in the rest of this thesis.

What each side can agree on is that development through the moral stages, both Kohlbergian and Gilliganian, varies by age but not by gender (Kohlberg & Gilligan, 1971). This means that the age of a person is a more significant factor in predicting their placement in either Kohlberg or Gilligan's levels than their gender. The other claim from the previously mentioned literature which has not been contradicted is that both males and females utilize both the care and justice moral orientations when responding to moral dilemmas. It seems that with the development of the ECI, the big question of a gender dichotomy in moral reasoning has been answered. However, there may be some confounding variables still left to explore.

More recently, Gilligan's ideas have continued to be upheld outside the topic of gender. Her idea of embracing balance instead of dichotomy between the self and others in moral responsibility has been shown to more accurately represent the moral reasoning used by nurses (Nistelrooij and Carlo, 2017) and non-professionals tracking the health of themselves and loves ones during the COVID-19 pandemic (Branicki, 2020). Additionally, Ethics of Care has been used to understand the motivating forces behind creative industry workers (Alacovska & Bissonnette, 2021) and maternal healthcare workers (Tuyisenge, Crooks, & Berry, 2020).

Within the topic of gender, there have been recent attempts to make care ethics a more masculine topic. When Father's Rights Groups implement these Gilliganian ideas however, there is a tendency to also implement problematic, perhaps hierarchical thinking that is sometimes used in inappropriate contexts (Jordan, 2020). In terms of friendships examined by the ethics of care, males tend to offer less physical help in caring for a friend's loved ones than female friends do. Additionally, more males than females refuse any type of care from both family and friends

(Policarpo, 2019). Some social welfare researchers believe that males are beginning to further embrace the ethics of care in the modern day, a claim that is backed up by all the previous research cited in this thesis, but larger societal changes surrounding workplace policies and expectations would be required for this way of thinking to translate into behavioral changes (Tarrant, 2018).

Gender Roles

The Nature of Gender Roles

Gender roles are generally thought to be the roles in society and in interpersonal relationships that males and females are expected to fill, solely based on their gender. Like everything else, the exact nature of these roles has shifted over time. The societal needs of the modern human differ greatly with the needs of the pilgrims, after all. One reason for this change is changing technology. Technological inventions, like that of the internet, have created a vast network of public spaces where elements such as gender can be studied. When Internet users are observed in contexts where one's gender remains anonymous, those who identify as men and those who identify as women abandon their gender roles, thus implying that they are not internal truths but instead social constructs (Haraway, 2016). However, when one's gender is revealed in an online platform, the result is that gender roles are reinstated (Miller et al., 2016), thus implying a human desire to categorize and even stereotype by gender.

Gender Differences in Moral Behavior

Previously mentioned literature states that females tend to outscore males on moral dilemma tests, meaning that females are generally placed into higher levels of reasoning than males are. Despite this, these two genders have the same level of moral understanding when matched on other relevant factors such as age or education (Barriga et al., 2001; Nunner-Winkler

et al., 2007). This means that males and females have the same capacity to understand morality and suggests that females tend to have an easier time applying that understanding to current situations. However, as moral reasoning tests are largely hypothetical in nature people may answer differently in a test than how they would really behave when placed in the same situation in real life. Data from real life does, however, add support to the trend of females outscoring males. Analyses of crime rates in various places around the world support the theory that males partake in more antisocial behavior than females, although because this result was received through archival data this is only a correlational relationship (Barriga et al.,2001; Steffensmeier & Allan, 1996; Esmail et al., 2013; Sharma, 2015; Steffensmeier et al., 2013); thus, other factors such as judge and jury gender biases and moral reasoning assessments were not looked at when reaching this conclusion in the studies cited above. Based on crime rate data, it appears that men commit more crimes than women despite having the same level of moral understanding. However, because their reasoning for committing those crimes in not known, these data do not support or reject the argument that males exhibit lower levels of moral reasoning than females.

Gender Roles and Moral Behavior

Social Role Theory argues that gender roles arise out of observation and generalization, first in physical differences between genders and then through social roles. A main idea of this theory is that physical differences between genders led to a division of labor, and over the centuries people have observed the roles that people of both genders hold in society, which in turn leads them to make generalizations about what internal character traits made those males and females fit into their roles. These characteristics are then generalized to similar people, who in this case tend to be people of the same gender (Eagly & Wood, 2012). Geology presents an interesting example of how Social Role Theory can lead people to incorrect conclusions. For

example, males' physical strength and size over females lead many to believe that they were hunters while women gathered food in early civilization; but recent geological findings have shown many female skeletons alongside hunting materials that were originally incorrectly assumed to be other tools (Haas et al., 2020). Based on physical differences between gender, society made an assumption about what roles physically stronger individuals should fill versus what roles should fall to weaker individuals. Once the stereotype that females must be child-bearers and not hunters was formed, it was used to inform the observations from all other geological finds to the point that hunting materials found by female skeletons were misidentified as various other types of tools. Thus, females have incorrectly been assumed to be less prepared for wilderness survival than males for many years (Khajavei, 2017).

Gender stereotypes are shown to children every day through advertisements and have a significant effect on whether children think a traditionally gendered toy (i.e., a remote-control car) is acceptable for a child of the opposing gender to play with (Pike & Jennings, 2005). Although this example of an expectation based on one's gender is rather insignificant to the larger workings of the adult world, it shows how Social Role Theory begins early in the lifespan. Seeing only young male children playing with a remote-control car tells all young children that this is a male toy, and from the content of those commercials stereotypes can begin to form, as implied by the study's findings that children find the toy unacceptable for the opposite gender. Alternatively, seeing young female children as well as young male children playing with a remote-control car does not give young children the opportunity to form the observational basis that could grow into a stereotype.

These expectations can also be applied to morality. Again, this begins at a young age. By the time children are in middle school, females are already very likely to assign negative

character traits to males (Nunner-Winkler et al., 2007), implying based on Social Role Theory that females have witnessed males holding less moral roles in society and generalized that to inform their ideas about what characteristics males have. In adolescents a similar trend continues with young females expecting other females to uphold higher moral standards than males. Young males, conversely, expect females to have the same level of moral reasoning that they do (Salgado, 2018). This interestingly implies that females at this age have been exposed to other females in roles that require higher-level moral reasoning, but men have not. By the time adulthood is reached, the gendered expectations of what one must do to be seen as moral get much more specific. Females are expected to be communal and avoid being dominant to be seen as morally upstanding, while males are expected to be agentic, independent, unemotional, and not shy to receive the same label of morally upstanding (Koenig, 2018). When responding to morality questions as one assumes a person of the opposite gender would, it was found that females tended to think that males are not as sensitive as they truly are to the pain of others, and that they do not care as much as they truly do about equality. Males, on the other hand, assumed females did not care as much for power as they truly do (Niazi et al., 2020).

Gender Roles in Morality Measurements

Clearly, there are some gendered differences when it comes to morality. Females and males have the same level of understanding about morality, but females tend to outscore males in moral reasoning measurement. In addition to this, people tend to expect different behavior from a person based on their previously held beliefs, whether or not they are correct, about how morally different genders behave. Thus, when gender becomes a relevant factor in a moral dilemma measurement, the participants' responses should change as a result of gender-biased views. More specifically: a participant's response to ECI-based questions should change, that is, be scored at a

different level of moral reasoning, based on whether those questions include a male or female character.

CHAPTER 3

Methods

Participants

This study included a sample generated of psychology students at Millersville University of Pennsylvania. Originally, four psychology classes were chosen to provide the sample. The idea behind this was that two professors would give access to this study to willing students from two sections of the same class. Because it is unlikely that a student would be taking two different sections of the same class during the same semester, the author of this study was able to ensure as little student overlap between conditions as possible. The only overlap this set up would allow for is if a student was in one of the two sections of the used class taught by the first professor, as well as one of the two sections of the used class taught by the second professor. To eliminate this possibility, the only eligibility requirement for participation, other than being a legal adult, was that the student had not already participated in this survey as part of another class. These were the ideal parameters for finding participants; however, in order to get an adequate sample size for Test C, the experimental group A test, a fifth class from a different psychology professor had to be included.

As part of the informed consent that participants read before deciding to participate in this study, they were told that they may withdrawal their consent at any time without penalty. To further ensure that the only data analyzed in this thesis came from willing participants, the final page of each survey reminded participants of this fact and prompted them to hit submit if they understand and agree that their data may be used for this study. If the test was unfinished, this author took it as a sign that participants exited the webpage they were taking the study on and thus did not see this final warning. Due to this precautionary criteria, 8 tests were not used. This

left the final sample size at N=87. This total number breaks down unevenly across groups with test A receiving n=28 test submissions, test B receiving n= 19 submissions, test C receiving n=15, and test D receiving n=25 submissions.

The ages of participants ranged from 18 to 38-years-old, meaning that this demographic has a range of 20 years. The mode of these data is also qual to 20. The psychology majors at Millersville University consists mainly of 'women' (College Factual, 2022), although it is not clear if this statistic is referring to those who identify as female or those who were deemed to be female at birth. The random sample taken of psychology students in this school were consistent with College Factual's stastic, with 72 of the total 87 responses coming from individuals who identify as female. A total of 12 of the responses came from those who identify as male, and 3 of the responses came from individuals who identify as nonbinary. If sex at birth is to be used instead, this data set includes 75 responses from individuals who were deemed to be females at birth, and still 12 responses from individuals who were deemed to be males at birth.

Due to lack of time, access, and funding this study used a convenience sample. This sample does not accurately represent the population at large, or the population of other college majors. This complication should be kept in mind when reviewing the results of this study.

Instrument and Materials

As mentioned in the literature review, this study used the Ethics of Care Interview as a basis for the moral reasoning measurement. Fortunately, information on the ECI is public, allowing this study to use a recreation of the actual measurement instead of creating a new one from scratch. Control Groups A and B are the questions published by Skoe (2014) as they are originally worded. Experimental Groups A and B are a recreation of the ECI which only changed the names and pronouns of the characters in the moral dilemmas. Thus, Control Group A and

Experimental Group A utilized the same questions, but the gender and pronouns of the characters in the dilemmas is switched in Experimental Group A. Control Group B and Experimental Group B functioned in the same way. Content changes in the experimental groups are limited to the changing of names and pronouns, and thus the recreation has not been subjected to any reliability or validity studies. The original version of the ECI has been tested for reliability as well as validity and found to be both reliable, with an inter-rater reliability score generally ranging from .85 to .95, and valid (Skoe, 2012). All information on the content of the moral dilemmas which will be presented to participants along with instructions on how to properly score the ECI has been provided by author Eva E. A. Skoe (2014).

The ECI consists of four open-ended moral dilemmas. Three of these dilemmas center around socially oriented moral decisions and are written before the participant is given the measurement. A fourth question is always generated by the participant and will be excluded from this study in order to give the experimenter more control. The ECI has a male and female version of each of the three dilemmas, however, they are not worded in the same way. Below is the original ECI in its exact wording along with the recreated version. The control conditions are the original ECI, and the experimental conditions are the recreations.

The Unplanned Pregnancy Dilemma

The first dilemma focuses on conflict surrounding an unplanned pregnancy. The original ECI female condition of this question is written in the following way. For the sake of clarity in this study, it has been labeled denoting condition, version, and question number

> **Control Group 1A:** Lisa is a successful teacher in her late twenties who has always supported herself. Her life has been centered on her work and she has been offered a permanent position for next year. Recently she has been involved in an intense love affair with a married man and now finds that she is pregnant. *What do you think Lisa should do? Why?* (p. 103).

The original ECI male version of that same question is written slightly differently. Again, for clarity in this study, it has been labelled accordingly.

> **Control Group 1B:** Derek is a married, successful teacher in his late twenties. His life has been centered on his work and he has been offered a permanent position for next year. Recently, he has been involved in an intense love affair with a single woman who has just told him that she is pregnant and that it is his child. *What do you think Derek should do? Why?* (p. 104).

There are a few differences between the male and female versions of this dilemma. The male character is married while the female character's relationship status is unknown, and the relationship status of the person the character is having an affair with changes between the two versions of this dilemma as well. The present study keeps these two versions of the dilemma as control conditions and will include another version of each, which is listed below. Lisa's dilemma and all other originally female dilemmas are Control Group A, and Derek's dilemma along with all other originally male dilemmas are Control Group B. Experimental Group A consists of the story in Control Group A, however, the female name and pronouns will be swapped for a male name and pronouns. For Experimental Group A the unplanned pregnancy dilemma uses the name Michael. Finally, Experimental Group B uses the exact wording of Control Group B but with a female name and pronouns. For the unplanned pregnancy dilemma Experimental Group B uses the name Molly. The experimental group questions were taken and created from the ECI.

> **Experimental Group 1A:** Michael is a successful teacher in his late twenties who has always supported himself. His life has been centered on his work and he has been offered a permanent position for next year. Recently he has been involved in an intense love affair with a married woman and now finds that she is pregnant. *What do you think Michael should do? Why?*

> **Experimental Group 1B:** Molly is a married, successful teacher in her late twenties. Her life has been centered on her work and she has been offered a permanent position for next year. Recently, she has been involved in an intense love affair with a single man and has just told him that she is pregnant and that it is his child. *What do you think Molly should do? Why?*

The Marital Fidelity Dilemma

The two original ECI versions of this dilemma are much more similar in wording but still imply a difference between if the character is the one at home with the children. Written below is the original ECI female version of this dilemma.

> **Control Group 2A:** Betty, in her late thirties, has been married to Erik for several years. They have two children, 8 and 10 years old. Throughout the marriage Betty has been at home, looking after the house and the children. For the last few years Betty has felt increasingly unhappy in the marriage relationship. She finds her husband demanding, self-centered and insensitive as well as uninterested in her needs and feelings. Betty has several times tried to communicate her unhappiness and frustration to her husband, but he continually ignores and rejects her attempts. Betty has become very attracted to another man, Steven, a single teacher. Recently, Steven has asked Betty for a more intimate, committed relationship. *What do you think Betty should do? Why?* (p. 103-104).

Written below is the original ECI male version of this same dilemma.

> **Control Group 2B:** Erik, in his late thirties, has been married to Betty for several years. They have two children, 8 and 10 years old. Throughout the marriage Betty has been at home, looking after the house and the children. For the last few years Erik has felt increasingly unhappy in the marriage relationship. He finds his wife demanding, self-centered and insensitive as well as uninterested in his needs and feelings. Erik has several times tried to communicate his unhappiness and frustration to his wife, but she continually ignores and rejects his attempts. Erik has become very attracted to another woman, Carol, a single teacher. Recently, Carol has asked Erik for a more intimate, committed relationship. *What do you think Erik should do? Why?* (p. 104).

Written below are the two experimental conditions, which are not part of the ECI as created by Skoe (2014) but have instead been created for this study.

> **Experimental Group 2A:** Jason, in his late thirties, has been married to Mary for several years. They have two children, 8 and 10 years old. Throughout the marriage Jason has been at home, looking after the house and the children. For the last few years Jason has felt increasingly unhappy in the marriage relationship. He finds his wife demanding, self-centered and insensitive as well as uninterested in his needs and feelings. Jason has several times tried to communicate his unhappiness and frustration to his wife, but she continually ignores and rejects his attempts. Jason has become very attracted to another woman, Joan, a single teacher. Recently, Joan has asked Jason for a more intimate, committed relationship. *What do you think Jason should do? Why?*

> **Experimental Group 2B:** Mary, in her late thirties, has been married to Jason for several years. They have two children, 8 and 10 years old. Throughout the marriage Jason has been at home, looking after the house and the children. For the last few years Mary has felt increasingly unhappy in the marriage relationship. She finds her husband demanding, self-centered and insensitive as well as uninterested in her needs and feelings. Mary has several times tried to communicate her unhappiness and frustration to her husband, but he continually ignores and rejects her attempts. Betty has become very attracted to another man, Daniel, a single teacher. Recently, Daniel has asked Betty for a more intimate, committed relationship. *What do you think Mary should do? Why?*

The Familial Care Dilemma

Written below is the original ECI female version for this dilemma.

> **Control Group 3A:** Kristine, a 26-year-old woman, has decided to live on her own after having shared an apartment with a girlfriend for the last three years. She finds that she is much happier living alone as she now has more privacy and independence and gets more work and studying done. One day her mother, whom she has not seen for a long while as they do not get along too well, arrives at the doorstop with two large suitcases, saying that she is lonely and wants to live with Kristine. *What do you think Kristine should do? Why?* (p. 104).

Written below is the male version of this same dilemma.

> **Control Group 3B:** Chris, a 26-year-old man, has decided to live on his own after having shared an apartment with a friend for the last three years. He finds that he is much happier living alone as he now has more privacy and independence and gets more work and studying done. One day his father, whom he has not seen for a long while as they do not get along too well, arrives at the doorstop with two large suitcases, saying that he is lonely and wants to live with Chris. *What do you think Chris should do? Why?* (p. 104).

Written below are the experimental conditions created for this study,

> **Experimental Group 3A:** Jack, a 26-year-old man, has decided to live on his own after having shared an apartment with a guy friend for the last three years. He finds that he is much happier living alone as he now has more privacy and independence and gets more work and studying done. One day his mother, whom he has not seen for a long while as they do not get along too well, arrives at the doorstop with two large suitcases, saying that she is lonely and wants to live with Jack. *What do you think Jack should do? Why?*
>
> **Experimental Group 3B:** Emily, a 26-year-old woman, has decided to live on her own after having shared an apartment with a friend for the last three years. She finds that she is much happier living alone as she now has more privacy and independence and gets more work and studying done. One day her father, whom she has not seen for a long while as they do not get along too well, arrives at the doorstop with two large suitcases, saying that he is lonely and wants to live with Emily. *What do you think Emily should do? Why?*

Format

This study used four versions of the ECI, each corresponding to one of the conditions. Test A, for example, consists of Control Group A for each of the three moral dilemmas. As previously mentioned, this is the female control group. Participants saw the original wording of each dilemma in the female condition. Control Group B is the male control group, meaning that participants in this group were presented with the original wording of each dilemma in the male condition. Test B utilized these dilemmas. In Test C participants responded to the Experimental

Group A test. In Test D, participants saw the dilemmas from the Experimental Group B test. Each version of the ECI was made into an open-ended survey which was administered virtually and blindly graded by this author as well as two additional psychology students she recruited to be additional scorers.

Procedure

As previously stated, this study had originally planned for four psychology classes to make up this study's sample. Due to lack of student participation on Test C, a fifth class had to be added. One of four links were sent via email to each participating class, with each link granting access to one of the four testing conditions. This email was sent to the first four classes at the same time, and then the surveys were closed one week later. On the final day that the surveys were open, another email containing the link to Test C was sent to a fifth class. This survey remained open for about 24 hours and then was promptly closed. The tests were administered through Qualtrics, and data will be deleted three years after it was collected for the protection of the participants.

The email that granted access to the Qualtrics surveys also contained a short explanation stating that if the student chooses to complete the survey, they are participating in a fellow psychology student's honors thesis regarding how gender effects moral reasoning. In addition to this explanation, attached to each email was an informed consent document with all the proper American Psychological Association protocal. The professors who taught these classes alerted the students of the emails the day they were sent out, and various other reminders were said in class as needed through the week. Completion was expected to take about 10 minutes. Participants were told that they can withdrawal consent to partake in this experiment at any time, and the resulting data will be immediately deleted. As previously noted, 8 participants did not

complete the survey in its entirety. These tests were seen as situations where consent was withdrawn, and the data resulting from these tests were not included in this study. The risks involved in this study were no greater than the risks faced in everyday life.

Participants' answers to the variations of the three pre-generated questions were placed into the proper levels as described in Skoe (2014), which were explained in detail in the literature review of this thesis. The ECI stresses that scorers should not focus on the behaviors which participants suggest for the character in the dilemma, but rather the reasoning behind the suggested behaviors. In other words, the 'what' is not as important as the 'why' when scoring the ECI. Specifically, placement into one of the levels should stem from whose needs are being met in a given scenario. The task of placing each test into its proper levels was undertaken by this author, acting during that moment as a scorer, as well as two other students who she recruited to be additional scorers. The nuance of scoring related to whose needs are being met as well as the focus on the 'why' were stressed to research assistants before they separately scored the tests. Interrater reliability was calculated to be 54% agreement between the original two scorers, but 95% agreement was reached by a majority after the tie-breaker completed their work. Each participant was given a score for each of the moral dilemmas to which they responded. The score corresponds to Gilligan's levels of moral reasoning previously described. For studies such as this one, where self-generated dilemmas are not being used, the mean score for the other three dilemmas was used in its place. All four scores are then added together and divided by four, thus averaging them a second time, and rounded to the nearest 0.5 for level placement (ranging from 1-3).

An independent samples *t* test statistical analysis was then used to compare the mean level of moral reasoning between Control Group A and Experimental Group A, and between

Control Group B and Experimental Group B. Analyses of average level placement between these two groups should isolate the differences in level placement based on how the question is worded. Because the questions are worded differently to reflect gender roles, as previously mentioned, it can be assumed that differences in expected behavior based on gender is the cause of any difference in average level placement.

CHAPTER 4

Results

Analysis of the Sample

Control group A (n=28 participants) had a mean level placement of 1.86 and a standard deviation of 0.58 levels. Control group B (n= 19 participants) had a mean level placement of 2.05 and a standard deviation of 0.28 levels. Experimental group A (n=15 participants) had a mean level placement of 1.73 and a standard deviation of 0.53 levels. Finally, experimental group B (n= 25 participants) had a mean level placement of 1.9 and a standard deviation of 0.41 levels. The table below expresses these data in a more visual format.

Table 1

Mean and Standard Deviation of Average Level Placement Across Groups

Test	Participants	Mean Score	SD
Control A	28	1.86	0.58
Control B	19	2.05	0.28
Experimental A	15	1.73	0.53
Experimental B	25	1.9	0.41

As previously detailed, when a participant's response to a dilemma indicated a focus on the main character's needs to the exclusion of everyone else's, the response was given a score of 1 or a score of 1.5 if the response indicated an awareness of the selfishness this entails (Skoe, 2014). Conversely, following the directions of Skoe (2014), a score of two was assigned to responses in which the needs of the main character were ignored in favor of the needs of others; and this score was raised to 2.5 when one's own needs came back into the picture. Responses were given a score of 3 if there was a clear demonstration of how each party's needs were addressed and how helping the self can lead to helping others, and vice versa (Skoe, 2014). The scores visible above were calculated by taking the mean of the three moral dilemmas to be a

surrogate fourth score, and then taking the mean of the four scores. This scoring process follows the directions laid out in the ECI (Skoe, 2014). Following a 95% confidence interval, the range of potential means for Control A is from 1.63 to 2.08. This range for Control B spans from 1.91 to 2.18. For Experimental A this range is from 1.43 to 2.02, and for Experimental B the range of potential means in this confidence interval spans from 1.73 to 2.07.

Demographic information collected from participants included gender identity, sex at birth, race, age, and relationship status. Tables displaying this information are below.

Table 2

Gender Identity of Participants by Test

Test	Female Participants	Male Participants	Non-Binary Participants
A	24	2	2
B	17	1	1
C	12	3	0
D	19	6	0

As clear from the table above, many more females than males participated in this study. In addition, very few of the participants had nontraditional gender identities. Participants were free to choose from female, male, transgender female, transgender male, nonbinary, or another gender identity not listed. The table below shows participants by race.

Table 3

Race of Participants

Race	Percent of Participants
American Indian or Alaska Native	2%
Asian	3%
Black or African American	8%
Hispanic or Latino	8%

Native Hawaiian/Pacific Islander	0%
White	77%
Multiracial	1%
Other Race	0%

The majority of participants in this study were white, which will be further discussed in the limitations section. The following table focuses on the age of participants.

Table 3

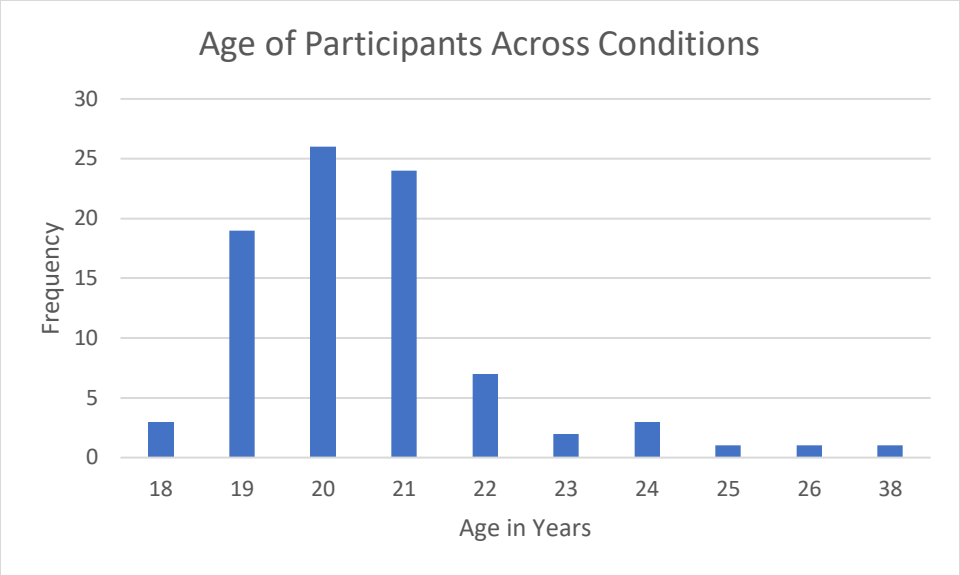
Age of Participants

Age	Frequency
18	3
19	19
20	26
21	24
22	7
23	2
24	3
25	1
26	1
38	1

As evident from the table above, most participants fell into the typical undergraduate age range of 18 to 22, with a few outliers. The outliers are more obviously present in the graph below.

Figure 1

Age of Participants Across Conditions



The final demographic table below presents the relationship status of participants.

Table 4

Relationship Status of Participants

Relationship	Frequency
Single	32
Short-term	17
Long-term	33
Married/engaged (no kids)	4
Married/engaged (kids)	1

It should be noted that a short-term relationship was defined by this study as lasting for any amount of time less than one year. A long-term relationship was defined as any relationship lasting longer than this amount of time. Participants were given a wider range of options when asked about their relationship status, but every person who participated in this study reported one of the categories listed in the table above.

This study used two independent scorers and utilized a third scorer as a tiebreaker when necessary. Interrater reliability was calculated to be 54%, however at least two of the three scorers agreed on 95% of the tests. This issue will be explained in the limitations section.

Research Question 1

Are there significant differences between what types of behaviors are viewed as moral for males and females? To test this question the following hypotheses were created. H0: there is no significant difference between scores for tests with female characters and scores for tests with male characters. H1: there is a significant difference between scores for tests with female characters and scores for tests with male characters. An independent samples *t* test was completed using an averaged group score for control group A and experimental group B as one sample, and an averaged group score for experimental group A and control group B as the other. This put the tests with female characters against the tests with male characters. Analyzed at a significance level of $\alpha=0.05$, there was no significant difference between tests with female characters and tests with male characters on DV, $t\ (77)= -0.34$, $p = 0.73$, despite tests with female characters ($n = 53$, $M = 1.88$, $SD = 0.5$) generally being scored lower than tests with male characters ($n = 34$, $M = 1.91$, $SD= =0.43$). Thus, this data set is unable to reject the null hypothesis and cannot support the claim that there is a significant difference between scores for tests with female characters and tests with male characters. The table below shows this in a visual format.

Table 5

Independent Samples t Test for Research Question 1

	Female Characters	*Male Characters*
Mean	1.877358491	1.911764706
Variance	0.249092888	0.188948307

Observations	53	34
Hypothesized Mean Difference	0	
df	77	
t Stat (t obtained)	-0.33972159	
P(T<=t) one-tail	0.367495559	
t Critical one-tail	1.664884537	
P(T<=t) two-tail	0.734991118	
t Critical two-tail	1.991254395	
-0.34 > -1.99		

A 95% confidence interval for the difference between means ranges from -0.18 to 0.24, with a small effect size ($\hat{d}$=0.06).

Research Question 2

Are females expected to consider the needs of others over their own needs? The ECI level 2 indicates that participants instructed the character in the moral dilemma to ignore their own needs in favor of the needs of others (Skoe, 2014). Tests A and D consist solely of female characters. The proportion of tests which scored at level 2 in these two conditions were 22:53, or 42%. The proportion of tests which scored at level 2 in the male conditions (tests B and C) were 17:34 or 50%, although it should be noted that this sample size is significantly smaller than that of the A conditions. Because this research question does not involve looking for significant differences, a hypothesis test was not used.

Based on these data, there is evidence that females are expected to ignore their needs in favor of the needs of others; however, there is also evidence that more males were expected to give up their own needs than females. This bit of data will be discussed in more detail shortly.

Research Question 3

How do responses change when a moral dilemma that is specifically worded to incorporate male gender roles has a female character instead? To test this question the following

hypotheses were formed. H0: There is no significant difference between control group B scores and experimental group B scores. H1: There is a significant difference between control group B scores and experimental group B scores. To find this answer, another independent samples *t* test was completed between the group's mean score of control group B and experimental group B (tests B and D). Analyzed at a significance level of α=0.05, there was no significant difference between control group B and experimental group B scores on DV, t (42) = 1.46, p = 0.15, despite the control B condition (n = 19, M = 2.05, SD = 0.28) scoring higher than the experimental B condition (n = 25, M = 1.9, SD= 0.41). This data set is unable to support the claim that there is a significant difference between control group B scores and experimental group B scores. The table below lays this information out in a visual format.

Table 6

Independent Sample t test of Research Question 3

	Test B	*Test D*
Mean	2.05263158	1.9
Variance	0.08040936	0.16666667
Observations	19	25
Hypothesized Mean Difference	0	
df	42	
t Stat (t obtained)	1.46203002	
P(T<=t) one-tail	0.07558783	
t Critical one-tail	1.68195236	
P(T<=t) two-tail	0.15117565	
t Critical two-tail	2.0180817	
1.46 < 2.02		

The effect size for these calculations ($\hat{d}$=0.43) is medium. A 95% confidence interval for these data ranges from -0.07 to 0.37.

Research Question 4

How do responses change when a moral dilemma that is specifically worded to incorporate female gender roles has a male character instead? To answer this question, the following hypotheses were formed. H0: there is no significant difference between control group A scores and experimental group A scores. H1: there is a significant difference between control group A scores and experimental group A scores. Following the same format as the previous research question, an independent samples *t* test was completed between the group's mean score of control group A and experimental group A (tests A and C). Analyzed at a significance level of $\alpha=0.05$, there was no significant difference between control A and experimental A scores on DV, $t(31) = 0.71$, $p = 0.48$, despite the control A condition (n = 28, M = 1.86, SD= 0.56) being scored higher than the experimental A condition (n = 15, M = 1.73, SD= 0.53). This data set is unable to reject the null hypothesis and thus unable to support the claim that there is a significant difference between control group A scores and experimental group A scores. The table below displays this information in a visual format.

Table 7

Independent Samples t test of Research Question 4

	Test A	*Test C*
Mean	1.85714286	1.73333333
Variance	0.33068783	0.28095238
Observations	28	15
Hypothesized Mean Difference	0	
df	31	
t Stat (t obtained)	0.70846177	
P(T<=t) one-tail	0.24197613	
t Critical one-tail	1.69551878	
P(T<=t) two-tail	0.48395225	
t Critical two-tail	2.03951345	
0.71 > 2.04		

These data have a small effect size ($\hat{d}$=0.21), and a 95% confidence interval ranging from -0.23 to 0.50.

Summary

The hypothesis of this thesis, that the gender of the character in a moral dilemma effects how participants respond, cannot be supported by this data set and the four research questions utilized for analyzation purposes. The scores of tests with female characters were not significantly different from the scores of tests with male characters, nor were the experimental conditions significantly different from their controls conditions. A series of *t* tests were used find these answers, and no significant differences were found between any of the four conditions.

CHAPTER 5

Discussion

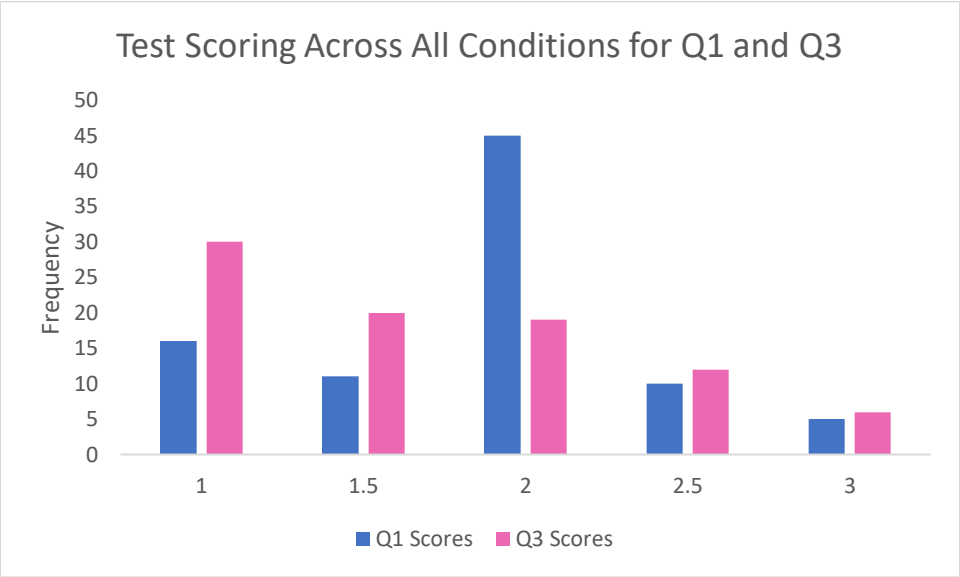
This study did not find any significant differences between group scores, which implies that the gender of the character in a moral dilemma may not effect how people respond. In the first research question it was found that responses to the ECI were not scored significantly differently between tests with female characters and tests with male characters. In the second research question it was found that 42% of tests with female characters were scored at level 2, indicating that females are expected to sacrifice their needs in favor of the needs of others. However, 50% of tests with male characters were also scored at level 2, indicating that gender was not an important factor to placement in level 2. In research questions 3 and 4, control conditions were also not found to be scored at a significantly different level than their respective experimental conditions. This being said, it is the opinion of this author that the use of the ECI ensured results would never have been significant. An argument for this case follows shortly, and more immediately is a discussion of why these results may have been reached based on current research in the field and further analysis of the collected data.

Beyond the dividing topics of male and female, people tend to respond to moral dilemmas in part by considering reciprocity; meaning that they are more likely to sacrifice their own needs in favor of the needs of someone who will likely return this favor at a later date (Zhan et al., 2019). While the moral dilemmas discussed by Zhan and his colleagues are presented to the participant as if they are a secondary character in the situation, this logic was still seen in the present study's data. For example, the mode score for the unplanned pregnancy dilemma (where the emotional closeness of other character's relationship with the main character implies that there may be reciprocity of a good deed) across all tests was 2, whereas the mode score across all

tests for the familial care dilemma (where it was implied that the secondary character would not reciprocate a good deed) was 1. The graph below shows this visually.

Figure 2

Test Scores Across All Conditions for Questions 1 and 3



Recall that a score of 2 indicates that one's own needs are being sacrificed in the proposed solution to the dilemma, and a score of 1 indicates one's own needs are the only priority. This may imply that participants are projecting themselves into the dilemma characters' situations and simply responding with what they think they would do in the same context. Should this be the case, it would change the independent variable of the character's gender to another measured variable of the participant's gender; thus leaving results looking like the experiments discussed in the literature review, with females generally outscoring males. Data from this study shows that, on average, tests taken by those who identify as female were scored at 1.89 and tests taken by those who identify as male were scored at 1.87. Although this is by no means a significant difference, it does not eliminate the possibility of participants projecting themselves into the moral dilemma questions, and a significant difference may have been found in a longer survey

taken by more people. There is nothing from the literature to compare this finding to, but it should be noted that tests taken by nonbinary individuals were scored at an average of 2.00. However, these numbers are misleading, as 72 of the tests were completed by those who identify as female, while only 12 were completed by those who identify as males and only 3 were completed by those who identify as nonbinary.

Other Findings

This study found numerous things it did not set out to find.

Safety

By virtue of having open-ended responses, this study was able to analyze more qualitative data. Safety was one of these topics. In the martial fidelity dilemma, the main character is in an unhappy marriage and has piqued the interest of another potential lover. Despite the dilemma never implying the unhappy relationship is abusive, there were numerous participants that factored the main character and children's safety into their responses. This pattern showed up in 21% of test A responses, 0 test B responses, 13% of test C responses (although it should be noted that this is only 2 tests, as test C has an exceptionally small sample size), and 16% of test D responses. To combine these numbers in reference to gender breakdown: abuse was assumed in 10 tests where the main character was a female married to a male. All of these tests mentioned the words safety or security, violent or violence, or otherwise implied abuse. Responses that did not meet those specific wording criteria but otherwise imply abuse follow. Someone taking test A suggested Betty have "some kind of third party in the discussion so if Erik gets mad or anything someone can step in". Another participant noted that "Betty also has the responsibility of her kids especially since she is aware of how awful Erik can be".

Contrasting this, only 2 participants assumed abuse in situations where a male character is married to a female character. Even in these situations, the responses looked very different. Whereas a participant responding to a female version of this question noted that she "should let people know what's going on in case her husband becomes violent" both responses for male conditions questioned the abuse. A participant taking test C said that the male in his question "may also want to look into Domestic Violence counseling, because while not violent this still can qualify for DV". This response at the same time notes that some level of protection may be needed while also rejecting any possibility that the male is being physically abused. The only other response in a male condition to mention abuse states that "the situation could also become very abusive if the wife were to find out about the affair especially if Joan was a teacher of one of their kids". This is the only statement which truly considers abuse in a male condition.

Interestingly, the concern for the safety of the female and her children was spread relatively equally across female conditions. 6 participants assumed abuse in the condition where the female was the sole parent providing care, and 4 assumed abuse in the condition where her husband was the sole parent providing care. This means that merely the mention of an unhappy marriage where a female's spouse does not listen leads some to ponder her safety. Test B, where the male was not providing any home or childcare, did not have any instances assumed abuse. 2 participants questioned abuse when the male was the sole care provider in the family.

There are multiple larger 'truths' that this study can imply, and without more in-depth answers it would not be possible to figure out which from this data set is correct. These data could possibly imply that males are seen as more likely to abuse a spouse than females in heterosexual relationships. This takeaway is backed up by the fact that 1 in ever 4 women will at some point in their lives experience violence at the hands of an intimate partner, while 1 in 9

men will have the same experience at some point in their lives (National Coalition Against Domestic Violence, n.d.). However, it could also be possible that we are socialized to believe that females are more likely to face abuse than males.

Relationship Status Breakdown

One of the demographics collected from each participant was their relationship status. The majority of participants were either single (n=32) or in a long-term relationship, which was defined as being longer than 1 year (n=33). A significant amount of participants were in a short-term relationship, defined as being less than 1 year (n=17). Very few participants were either married or engaged (n=4), and even fewer were married or engaged with children (n=1). Interestingly, each of these groups had similar average scores. The mean score for single participants was 1.88, this happens to be the same mean score for participants who are in a short-term relationship. Those in a long-term relationship scored the highest on average, with a mean of 1.95. While these numbers are all closer together, the mean scores for married or engaged individuals is much lower. Participants who were married or engaged without children had an average score of 1.63, while the single participant who was married or engaged with children scored 1.

This finding might indicate that individuals enter into relationships relatively willing to sacrifice for the other partner, and this generosity peaks in long-term relationships; however, marriage correlates with more selfish behavior. This author would not be comfortable making any claim based on this observation due to the small sample size of married or engaged individuals (n=5) compared to the participants who are not married or engaged (n=82).

Study Limitations

Due to time, access, and financial constraints, this study has multiple limitations.

Limitations Related to Participants and Participation

This study utilized a convenience sample. Four psychology classes at Millersville University cannot be expected to accurately represent other majors at the same university, the same major at other universities, or the population at large. Participants were 82.8% female, 13.8% male, and 3.4% nonbinary. This gender breakdown is not representative of the population at large. Additionally, the racial breakdown of participants in this study is not representative of the population at large. In fact, 77% of participants were White, making this racial group the majority represented in this sample. The next largest racial demographic of participants was 8%, representing participants that were Hispanic or Latino, and an additional 8% of participants that were Black or African American. Additionally, 3% of participants were Asian, 2% were Native Americans of the contiguous United States or Alaska, and 1% of participants were multiracial.

The control group for the A condition received 13 more tests than its experimental counterpart, almost doubling in group sample size. Because the group scores are given based on averages, this size variation may be a limitation. Likewise, the experimental group for the B condition received 6 more tests than its control counterpart. While this difference in size is not as extreme, this study's results should still be viewed with caution for this reason.

Limitations Related to the ECI

The ECI is a wonderful source that combines gender and moral reasoning; however, it is not a measurement that is meant to be used in the style of the present study. This measurement was chosen for the present study due to the fact that it is available to the public without any cost. This study did not have the funding to pay for a longer measurement, nor the time required to create and pre-test a truly original one. The choice this author made to use the ECI meant that

she had no control over the number of dilemmas presented or the way in which scores would be given.

The ECI is meant to be taken in a one-on-one, face-to-face format. This allows for a fourth dilemma to be created by the participant (Skoe, 2014). Instead, due to the online nature of the surveys used in the present study, Skoe (2014) instructs that an averaged score of the first three questions be used in place of a fourth question; and those four scores be averaged together to grade each test. It is the opinion of this author that using mean scores for tests with only three real questions may make for an unrealistic result. This restriction of range may be the most troubling limitation that the present study faced as it eliminates any room for complexity in responses. For example, if in a given test a participant's responses to the first two questions are given a score of 2, the last question could be given a score of 1, indicating that the participant expects the character to be sacrificial in romantic/sexual relationships, but more selfish in familial ones. Even in this case, the overall score given to this test would be 1.5. Even though the majority of the answers indicate the participant expects the characters to ignore their own needs, the end result reads as if the participant expects the characters to proceed selfishly albeit aware of their selfishness. The complexities of this line of thinking are completely lost by this one summary number. If this test was longer, with 20-30 questions about romantic/sexual relationships and 20-30 about familial ones, an overall score may more accurately represent the participant's line of thinking.

In addition to this, the score for the averaged fourth question and the overall test score both had to be rounded to the nearest 0.5 level, to fit with Skoe's (2014) framework, leaving even less room for nuance. In the example above, the averaged score did not come out to 1.5, but rather to 1.67; however, it could not be recorded as such. This left very few opportunities for

different scores in the test and may have ultimately contributed to this study's incapability to find a significant difference between any of the scores. Overall, if the ECI had less range restriction this study may have found different results.

Possible Limitations Related to Interrater Reliability

The interrater reliability was expected to be measured around 0.80 (Skoe, 2014), in this study it was calculated to be 0.54 between the first two scorers. Despite Skoe advertising her average interrater reliability to be around 0.80, it is unclear if this refers to majority consensus before or after the tie-breaker is brought in. Should it refer to after this third scorer is recruited, than this study's calculated interrater reliability, 0.95, is well within the range of validity.

Should Skoe's calculation refer to the percentage of agreement before the tie-breaker does their work, than this study's low score on this matter may come from a lack of adequate training. This author trained all scorers by explaining the purpose of the study, providing them with Skoe's (2014) instructional handbook on this matter, and answering any question that arose from their reading. Generally, participants seem to have understood; however, the tie-breaker did not follow the rounding to the nearest 0.5 level rule. This made calculations more difficult and required some of their overall test scores to be rounded yet again. The 5% of tests, 4 total tests, which did not reach majority agreement account for situations where the final rounded score sat perfectly in between final test scores assigned by the other two scorers. An example of this situation would be a hypothetical test where scorer A gives a test the score of 2 and scorer B gives the same hypothetical test a score of 2.5, and the tie-breaker decides the same test should be scored at 2.25. It should be noted that Skoe does allow for 'half scores' of .25 or .75, but only as a last resort which should be used sparingly (Skoe, 2014).

This author acknowledges that more could have been done to solve this problem. For example, the tie-breaker could have been asked to fix this problem. This author did not want to risk any effects from having the same person grade the same tests twice and trying to appeal to what they believe this author wants to see after being asked to complete their work again. While another tie-breaking scorer could have been found, this study had a harsh completion deadline that was quickly approaching; leaving no time to recruit and train another individual.

Limitations Related to Data Analysis

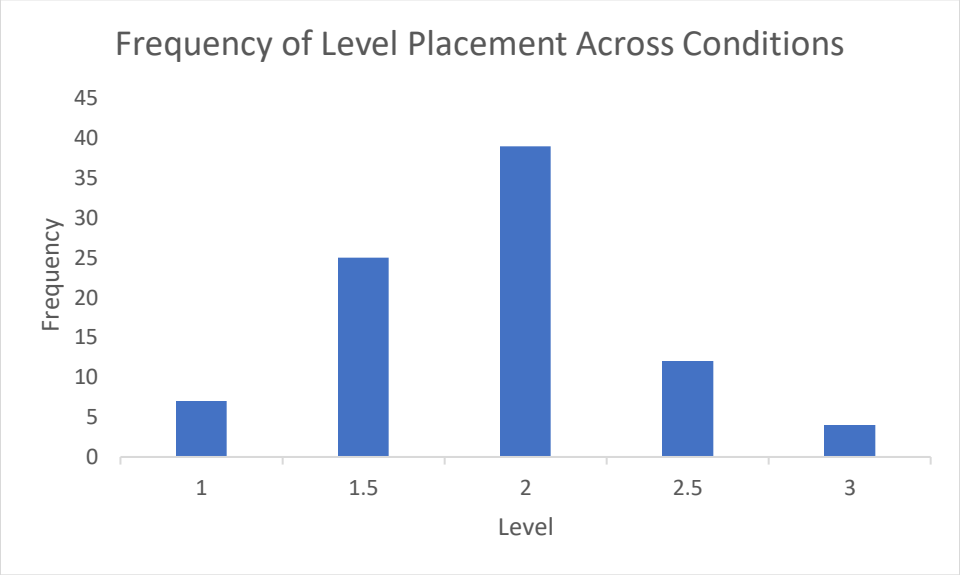
This study was unable to run any tests to determine ideal sample size before the participants were brought in-- another victim of this study's limited timeline for completion. It was the opinion of this author and the committee she created that 30 people per test would be ideal. This number was not reached in any condition. A post-hoc power analysis revealed that for the first research question (regarding the difference in scores between tests with female characters and tests with male characters) had 4.8% power. The second research question (regarding the proportion of tests with female characters which scored at level 2) also had 4.8% power, as the same groups were involved in both of these research questions. The third research question (regarding the difference in scores between the control and experimental B conditions, which were the original male condition created by Gilligan and its experimental, female counterpart) had 30.2% power. The final research question (regarding the difference in scores between Gilligan's original female condition, control A, and its experimental counterpart) had 11.2% power. All of these post-hoc analyses used a significance level of $\alpha=0.05$. Needless to say, this study did not achieve high power in any condition.

Additionally, a *t* test was used to test each hypothesis. Each research question required two 2 groups to have their mean scores tested against each other, these groups were designed to

be independent of each other; however, these data are slightly skewed. This skew, shown below, caused some debate as to if a *t* test or a Mann Whitney U Test should be used.

Figure 3

Frequency of Level Placement Across All Conditions



However, after completing a Shapiro-Wilk test, this data set can be classified as approaching normality and thus the standard *t* test can be used. This data set has a skewness value of 0.2 and a kurtosis value of 0.07, putting them well within the range of parametric assumptions. The present study used the independent version of this hypothesis test, although it is possible that some participants may have participated in two different conditions. There were steps in place to avoid this scenario: multiple sections of one course were chosen where possible as the same student would likely not be in both sections of the same course. As a backup to this, students were informed that they were not allowed to complete the study if they already had participated in it as part of another class. Still, to ensure zero overlap between groups would have required asking the participants' names, which would breach their promised anonymity.

Study Implications

These data imply that the gender of a character in a moral dilemma does not have an effect on how people respond to the moral dilemma. This would also indicate that what is seen as moral for men is no different than what is seen for moral as women, contrary to the articles cited in the literature review section of this thesis. However, the limitations of this study are numerous, these data should only be referenced with an understanding of these limitations.

Although this study could not support the claim, it is possible that participants answer moral dilemmas by projecting themselves into the character's situation. This may explain the lack of difference found, and how this set of data does not fit with the current data that is published. This is an area of research without much study. Obviously, participants project themselves into moral dilemmas where they get to be the character (i.e. "would you do ___"); however, not much research has been done on projection of the self in moral dilemmas where the self is not a character. This is an area that would benefit from research in the future.

Directions for Future Research

Testing the claim that gender politics are trumped by projection of the self in moral dilemmas which feature characters instead of the self would add a lot of understanding to this study and to the field at large. There are no current studies on this topic, leaving room for lots of future research and experimentation. This study should also be improved upon. The same hypotheses tested by a measure other than the ECI may yield different, more reliable results which would bring increased understanding to the field. Researchers who take on this task would need adequate funding and time to effectively complete this proposed study. To avoid the pitfalls of other moral dilemma measurements, which were discussed in detail in the literature review, future research would be best served by creating an expansion of the ECI. A study of this nature may yield a more accurate test score but would still be subject to range limitation as Gilligan

only created 5 levels of moral reasoning. To circumvent this particular issue without taking on the problem of mixed results in the literature for other moral dilemma measurements, an entirely new test could be created, piloted, and then used in future studies. While this may be more tedious and require more time and funding, the literature review of this paper details how it may be necessary, as consistent data on a problem-free measurement does not yet exist for the study of gender and morality.

Additionally, modern understanding of the wider view of gender identity should be applied to this and other topics. This particular study only had 3 participants with nontraditional gender identities, which is not enough to draw any valid conclusions. However, a study similar to this with a focus on nontraditional gender identities (i.e. transgender, nonbinary) may paint a more compete picture of human moral reasoning and how its effected by biological sex versus socialized ideas of gender.

Conclusion

This is one of the first studies to focus on characters in moral dilemmas rather than the participant taking the measurement. Despite its numerous limitations, this research acts as a precedent for examining the further reaching impacts of gender politics which may be more subtle than blatant sexism. More research should be done in this area to see if a significant difference can be found when the circumstances better allow for nuanced results.

References

Ahmad, S., Ch, A. H., Batool, A., Sittar, K., & Malik, M. (2016). *Play and cognitive development: Formal operational perspective of Piaget's theory*. *Journal of Education and Practice*, 7(28). Retrieved from https://files.eric.ed.gov/fulltext/EJ1118552.pdf

Alacovska, A., & Bissonnette, J. (2021). Care-ful work: An Ethics of Care approach to contingent labour in the creative industries. *Journal of Business Ethics*, *169*(1), 135-151. https://doi.org/10.1007/s10551-019-04316-3

Arvizu, D. R. (1995). The care voice and American Indian college students: An alternative perspective for student development professionals. *Journal of American Indian Education*, *34*(3), 1–17. https://doi.org/https://www.jstor.org/stable/24398307

Babakr, Z. H., Mohamedamin, P., & Kakamad, K. (2019, August 15). *Piaget's cognitive developmental theory: Critical review*. Asian Institute of Research, 517-524. DOI:10.31014/aior.1993.02.03.84

Barriga, A. Q., Morrison, E. M., Liau, A. K., & Gibbs, J. C. (2001). Moral cognition: Explaining the gender difference in antisocial behavior. *Merrill-Palmer Quarterly*, *47*(4), 532–562. https://doi.org/ https://www.jstor.org/stable/23093698

Barger, R. N. (2000). A summary of Lawrence Kohlberg's stages of moral development. *Chapter 8 Ethics*. Retrieved from https://www.qcc.cuny.edu/socialsciences/ppecorino/medical_ethics_text/Chapter_2_Ethical_Traditions/Reading-Barger-on-Kohlberg.htm.

Bernardi, R.A & Arnold, D.F., S.R. (1997), An examination of moral development within public accounting by gender, staff level, and firm. *Contemporary Accounting Research*, 14: 653-668. https://doi.org/10.1111/j.1911-3846.1997.tb00545.x

Boom, J., Brugman, D., & Thoma, S. (2018). Stages of moral judgment development: Applying item response theory to Defining Issues Test data. *Journal of Moral Education*, *48*(4), 423–438. https://doi.org/https://doi.org/10.1080/03057240.2018.1540973

Branicki, LJ. COVID-19, ethics of care and feminist crisis management. *Gender Work Organ*. 2020; 27: 872– 883. https://doi.org/10.1111/gwao.12491

Conley, T. D., Jadack, R. A., & Hyde, J. S. (1997). Moral dilemmas, moral reasoning, and genital herpes. *Journal of Sex Research, 34*(3), 256–266. https://doi.org/10.1080/00224499709551892

Donleavy, G. D. (2008). No man's land: Exploring the space between Gilligan and Kohlberg. *Journal of Business Ethics*, *80*(4), 807–822. https://doi.org/http://www.jstor.com/stable/25482183

Eagly, Alice & Wood, Wendy. (2012). Social role theory. Handbook of theories in social psychology. 2. 458-476. 10.4135/9781446249222.n49.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). Lawrence Erlbaum Associates Publishers.

Esmail, A. M., Penny, J., & Eargle, L. A. (2013). the Impact of culture on crime. *Race, Gender & Class*, *20*(1/2), 326–343. http://www.jstor.org/stable/43496922

Gerson, K. (2002). Moral dilemmas, moral strategies, and the transformation of gender: Lessons from two generations of work and family change. *Gender and Society*, *16*(1), 8–28. https://doi.org/https://www.jstor.org/stable/3081874

Giligan, C. (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Gilligan, C. (1986). Reply. *Signs*, *11*(2), 324–333. https://doi.org/https://www.jstor.org/stable/3174055

Gilligan, C., & Attanucci, J. (1988). Two moral orientations: Gender differences and similarities. *Merril-Palmer Quarterly*, *34*(3), 223–237. https://doi.org/ https://www.jstor.org/stable/23086381

Gilligan, C., & Wiggins, G. (1988). The origins of morality in early childhood relationships. In C. Gilligan, J. V. Ward, J. M. Taylor, & B. Bardige (Eds.), *Mapping the moral domain: A contribution of women's thinking to psychological theory and education* (pp. 111–138). Harvard University Press. (A slightly different version of this chapter has appeared in J. Kagan and S. Lamb (Eds.) "The Emergence of Morality in Young Children," University of Chicago Press, 1987, 277–306)

Haas, R., Watson, J., Buonasera, T., Southon, J., Chen, J. C., Noe, S., Smith, K., Llave, C. V., Eerkens, J., & Parker, G. (2020). Female hunters of the early Americas. *Science advances*, *6*(45), eabd0310. https://doi.org/10.1126/sciadv.abd0310

Haraway, D. J. (2016). *Cyborg manifesto*. University of Minnesota Press.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: a meta-analysis. *Psychological bulletin*, *126*(5), 703–726. https://doi.org/10.1037/0033-2909.126.5.703

Jordan, A. (2020). Masculinizing care? Gender, Ethics of Care, and fathers' rights groups. *Men and Masculinities, 23*, 20 - 41.

Khajavei, Naiyear, "Women and the wilderness: a review of barriers to participation, current coping strategies, and guidance for future programs" (2017). University Honors Theses. Paper 462. https://doi.org/10.15760/honors.459

Koenig A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in psychology*, *9*, 1086. https://doi.org/10.3389/fpsyg.2018.01086

Kohlberg, L., & Gilligan, C. (1971). The Adolescent as a philosopher: The discovery of the self in a postconventional world. *Daedalus, 100*, 1054-1087. https://www.semanticscholar.org/paper/The-Adolescent-as-a-Philosopher%3A-The-Discovery-of-a-Kohlberg-Gilligan/e5743ca3aff8551ee0baf87057de78e69d00f03b#paper-header

Malik, F., & Marwaha, R. (2021). Cognitive development. In *StatPearls*. StatPearls Publishing.

McLeod-Sordjan R. (2014). Evaluating moral reasoning in nursing education. *Nursing ethics*, *21*(4), 473–483. https://doi.org/10.1177/0969733013505309

Miller, D., Costa, E., Haynes, N., McDonald, T., Nicolescu, R., Sinanan, J., Spyer, J., Venkatraman, S., & Wang, X. (2016). Gender. In *How the World Changed Social Media* (1st ed., Vol. 1, pp. 114–127). UCL Press. https://doi.org/10.2307/j.ctt1g69z35.15

NCADV. (n.d.). *NCADV: National Coalition Against Domestic Violence*. NCADV. Retrieved April 10, 2022, from https://ncadv.org/STATISTICS

Niazi F, Inam A, Akhtar Z (2020) Accuracy of consensual stereotypes in moral foundations: A gender analysis. PLoS ONE 15(3): e0229926. https://doi.org/10.1371/journal.pone.0229926

Nunner-Winkler, G., Meyer-Nikele, M., & Wohlrab, D. (2007). Gender differences in moral motivation. *Merrill-Palmer Quarterly*, *53*(1), 26–52. http://www.jstor.org/stable/23096290

Ojose, B. (2008). *Applying Piaget's Theory of Cognitive Development to mathematics instruction*. The Mathematics Educator, *18*(1), 26-30. Retrieved from: https://files.eric.ed.gov/fulltext/EJ841568.pdf

Policarpo, V. (2019). Friendship and care: Gendered practices within personal communities in Portugal. *Journal of Gender Studies*, *8*(1), 57–69. https://doi.org/https://doi.org/10.1080/09589236.2017.1394820

Pike, Jennifer & Jennings, Nancy. (2005). The effects of commercials on children's perceptions of gender appropriate toy use. Sex Roles. 52. 10.1007/s11199-005-1195-6.

Pratt, M. W., Diessner, R., Hunsberger, B., Pancer, S. M., & Savoy, K. (1991). Four pathways in the analysis of adult development and aging: Comparing analyses of reasoning about personal-life dilemmas. *Psychology and aging*, *6*(4), 666–675. https://doi.org/10.1037//0882-7974.6.4.666

Salgado M (2018) Gender-Biased expectations of altruism in adolescents. *Front. Psychol*. 9:484. doi: 10.3389/fpsyg.2018.00484

Scott, H. K., & Cogburn, M. (2021, April 30). *Piaget*. StatPearls [Internet]. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK448206/

Sharma, P. (2015). Gender-based crime and gender inequality in India: A subnational analysis. *Economic and Political Weekly*, *50*(45), 48–55. http://www.jstor.org/stable/44003994

Skoe, Eva. (2014). Measuring care-based moral development: The Ethic of Care Interview.. Behavioral Development Bulletin. 19. 95-104. 10.1037/h0100594.

Skoe, E. E. A. (2014). Measuring care-based moral development: The Ethic of Care Interview. *Behavioral Development Bulletin, 19*(3), 95-104. http://dx.doi.org/10.1037/h0100594

Skoe, E. E., & Diessner, R. (1994). Ethic of Care, justice, identity, and gender: An extension and replication. *Merrill-Palmer Quarterly*, *40*(2), 272–289. https://doi.org/https://www.jstor.org/stable/23087865

Skoe, E. E., & Marcia, J. E. (1991). A Measure of care-based morality and its relation to ego identity. *Merrill-Palmer Quarterly*, *37*(2), 289–304. https://doi.org/https://www.jstor.org/stable/23087367

Steffensmeier, D., & Allan, E. (1996). Gender and crime: Toward a gendered theory of female offending. *Annual Review of Sociology*, *22*, 459–487. http://www.jstor.org/stable/2083439

Steffensmeier, Darrell J., Jennifer Schwartz, and Michael Rochea. "Gender and twenty-first century corporate crime: Female involvement and the gender gap in enron-era corporate frauds." *American Sociological Review* 78, no. 3 (2013): 448–76. http://www.jstor.org/stable/23469220.

Tarrant, A. (2018) Care in an age of austerity: Men's care responsibilities in low-income families, *Ethics and Social Welfare*, 12:1, 34-48, DOI: 10.1080/17496535.2017.1413581

The psychology major at Millersville University of Pennsylvania. The psychology major at Millersville University of Pennsylvania - College Factual. (2022, February 2). Retrieved April 2, 2022, from https://www.collegefactual.com/colleges/millersville-university-of-pennsylvania/academic-life/academic-majors/psychology/

Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. Developmental Review, 6(2), 165–180. https://doi.org/10.1016/0273-2297(86)90010-9

Tuyisenge, G., Crooks, V.A., & Berry, N.S. (2020). Using an ethics of care lens to understand the place of community health workers in Rwanda's maternal healthcare system. *Social science & medicine, 264*, 113297 .

van Nistelrooij, I., & Leget, C. (2017). Against dichotomies: On mature care and self-sacrifice in care ethics. *Nursing ethics*, *24*(6), 694–703. https://doi.org/10.1177/0969733015624475

Wark, G. R., & Krebs, D. L. (1996). Gender and dilemma differences in real-life moral judgment. Developmental Psychology, 32(2), 220–230. https://doi.org/10.1037/0012-1649.32.2.220

Zhang, Q. and Zhao, H.H. (2017) An Analytical overview of Kohlberg's Theory of Moral Development in college moral education in main- land China. *Open Journal of Social Sciences*, 5, 151-160. https://doi.org/10.4236/jss.2017.58012

Zhan, Y., Xiao, X., Tan, Q., Zhang, S., Ou, Y., Zhou, H., Li, J., & Zhong, Y. (2019). Influence of self-relevance and reputational concerns on altruistic moral decision making. *Frontiers In Psychology*. https://doi.org/https://doi.org/10.3389/fpsyg.2019.02194