Perceiving Different Types of Bad People:

How Moral Person Prototypes Influence Moral Impressions

by

Sarah Dimakis

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Psychology

Dissertation Committee:

Robert Mauro, Chair

Sara Hodges, Core Member

Gerard Saucier, Core Member

Erik Girvan, Institutional Representative

University of Oregon

Summer 2024

© 2024 Sarah Dimakis
This work is openly licensed via CC BY-NC-ND 4.0.

DISSERTATION ABSTRACT

Sarah Dimakis

Doctor of Philosophy in Psychology

Title: Perceiving Different Types of Bad People: How Moral Person Prototypes Influence Moral Impressions

In response to a complex and information dense world, we organize related information into categories (e.g., *furniture*, *animals*) to make it easier to apply existing knowledge to newly encountered objects and situations (Rosch, 1978). Categories are represented in the mind as cognitive prototypes: examples that possess common features of category members (Posner & Keele, 1968; Reed, 1972; Rosch & Mervis, 1975). A key feature of prototype-based categories is that unobserved characteristics are inferred to category members once the prototype is activated (Cantor & Mischel, 1977; Osherson et al., 1990). I propose in this dissertation that we use moral person categories to make inferences about the unobserved or unobservable immoral characteristics of others, which influences our decisions about with whom to interact and how to interact with them. In a series of studies, I demonstrated that people perceive multiple bad person categories, examined their properties, and tested a prototype model of moral character evaluation against dimensional models. This dissertation is the first rigorous investigation of the structure and properties of perceived bad person categories.

In Study 1, fifty prototypical characteristics of a bad person (e.g., lacking empathy, selfishness, racism) were gleaned from previous literature and augmented with additional characteristics generated by participants in a free-listing task. In Study 2, participant sorting data of characteristics from Study 1 revealed that respondents distinguished between multiple types of bad people (e.g., psychopath, abuser, narcissist, bigot). Additionally, categorical and dimensional models of bad person characteristics were generated from the responses in Study 2 and tested with new data in Study 3. Consistent with prototype-based categories, unobserved immoral characteristics that were prototypical of a category were more likely to be attributed to exemplars of that category compared to exemplars of competing *bad person* categories. Further, prototype models outperformed dimensional or multidimensional models, but prototype-unidimensional dual models performed the best. Thus, people infer prototype-consistent immoral characteristics to exemplars of a category, and additionally infer prototype-inconsistent immoral characteristics

that are similar in morality to the exemplar (but not similar on other social dimensions such as intelligence or sociability). Additionally, further providing support for multiple bad person prototype-based categories, large differences were observed in the perception of four bad person prototypes regarding their morality, competence, sociability, gender, race, age, physical appearance, the innateness and permanence of their immoral characteristics, and the ways respondents prefer to interact with them.

Overall, this dissertation provides strong evidence against the unidimensional model of morality commonly used in the moral character evaluation literature and in support of a model of moral character evaluation that includes cognitive prototypes. This research may have significant implications for why we make errors that others are *bad* based on harmless or morally irrelevant characteristics (e.g., physical appearance, harmless behaviors), including in criminal justice contexts. In the future, the methods presented in this paper should be replicated with diverse samples to assess commonalities and differences in the cognitive prototypes of bad people across relevant subgroups (e.g., liberals versus conservatives, experts versus novices).

ACKNOWLEDGMENTS

To start, I wish to express gratitude towards the person who has shaped this project and my academic journey the most, my faculty advisor Dr. Robert Mauro. I am profoundly grateful for your patience, kindness, encouragement, and wisdom over the past six years. I know I am a better researcher and a better person because I got to work with you. I want to thank Annika Zwolinski, Allis McAllister, Ciara Wyatt, and Eva Stojanovic, my extremely dedicated undergraduate research assistants who helped me generate items and code the responses in this paper. I also want to express my sincere gratitude for the members of my committee who have provided very insightful feedback on my research: Dr. Sara Hodges, Dr. Gerard Saucier, and Dr. Erik Girvan.

I want to give a special thanks to Dr. Azriel Grysman, whose research lab I was extremely fortunate to have worked in as an undergraduate that served as a catalyst for my interest in psychology research. I also want to take this opportunity to thank the excellent educators I've been so fortunate to learn from in my life whose courses taught me the academic and life skills I needed to complete this degree, especially Dr. Daniel Anderson, Dr. Courtney Gibbons, Dr. Chinthaka Kuruwita, Dr. Tara McKee, Dr. Cameron Brick, Laurel Webster, Patti Crouch-Cook, Leon Phillips, Ryan Johnson, and Karen Quinn.

Finally, I would like to thank my friends and family for your emotional support throughout this journey. To my friends and cohort mates Liz, Kim, Kyra, Kellyn, Ruby, Moriah, Meg, Andrea, Angela, Cam, and Andrew for being giant sources of happiness and strength. To my parents who sacrificed so much to give me the best education they could and who taught me to value learning over everything else. To my big sister Maria who taught me long division and how to realize you're being conned. And to my partner Raleigh, our cat Amelia, and our dog Zucchini for bringing me so much joy every day. Thank you all.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES 8

LIST OF TABLES 9

I. INTRODUCTION 11

DERIVING A TAXONOMY OF BAD PERSON CONCEPTS 18
Present Study 23
TESTING A PROTOTYPE MODEL AGAINST DIMENSIONAL MODELS OF IMMORAL INFERENCES 23
A Prototype Model 25
A Multidimensional Model 26
A Prototype-Dimensional Dual Model 27
Present Study 28
EXAMINING PROPERTIES OF MORAL PERSON PROTOTYPES 28
Three Dimensions of Person Perception 28
Essentialist Attitudes 30
Social distance 32
Present Study 33

II. STUDY 1: GENERATING CHARACTERISTICS OF A BAD PERSON 34

METHOD 34
Participants 34
Procedure 35
Materials 36
RESULTS 37
Generating a List of the Prototypical Characteristics of a Bad Person 38
Explaining Variance in Conceptions of a Bad Person 39
DISCUSSION 40

III. STUDY 2: EMPRICALLY DRIVEN CATEGORY AND DIMENSIONAL MODELS OF BAD PERSON CONCEPTS 49

METHOD 49
Participants 49
Procedure and Materials 50
RESULTS 53
Number of Groups 53
Group Names 53
Similarity Matrix 54
Deriving a Category Model of Bad Person Concepts 56
Deriving a Dimensional Model of Bad Person Concepts 67
DISCUSSION 69

IV. STUDY 3: TESTING MODELS OF IMMORAL INFERENCES 73

METHOD 75

Participants 75
Research Design 76
Materials 77
Procedure 82
RESULTS 83
Evaluating the Fit of the Empirically Derived Categorical Structure 89
Testing Dimensional Models 91
Evaluating the Fit of the Empirically Derived Dimensional Structure 93
Comparing Prototype, Dimensional, and Dual Models of Immoral Inferences 96
Examining Attributes of Bad Person Prototypes 97
DISCUSSION 109
V. GENERAL DISCUSSION 122
IMPLICATIONS 122
LIMITATIONS AND FUTURE DIRECTIONS 127
Study Good People 127
Study Other Populations 129
Design an Intervention for Moral Perception Errors 132
CONCLUSION 132
APPENDIX A: INITIAL 70 BAD PERSON CHARACTERISTICS 134
APPENDIX B: FINAL 50 BAD PERSON CHARACTERISTICS 136
APPENDIX C: ESSENTALIST BELIEFS SCALE 138
APPENDIX D: SOCIAL DISTANCE SCALE 139
APPENDIX E: MEAN AND STANDARD DEVIATION TABLES FOR LIKELIHOOD RATINGS 140
APPENDIX F: ERRORS FROM PROCRUSTES ANALYSES 145
REFERENCES CITED 148

LIST OF FIGURES

Figure Page

1. Computerized Similarity-Sorting Task 52
2. Distribution of Groups Created in Initial Sort 54
3. Elbow Plot of Within-Cluster Variance for Initial Sort 57
4. Plot of Average Silhouette Width for Initial Sort 57
5. Dendrogram of Hierarchical Cluster Analysis Ward's Method (k = 3)..... 59
6. Dendrogram of Hierarchical Cluster Analysis Ward's Method (k = 7)..... 60
7. Dendrogram of Hierarchical Cluster Analysis Ward's Method (k = 10)..... 61
8. Sorting Agreement for Seven-cluster Solution 65
9. Three-Dimension Solution from MDS (Morality, Competence Dimensions)..... 70
10. Three-Dimension Solution from MDS (Morality, Sociability Dimensions) 71
11. Example Instructions for Introducing an Exemplar..... 76
12. Likelihood Exemplars Have Unobserved Characteristics 85
13. Recreating of Category Structure Derived in Study 2 89
14. Morality and Competence Ratings for Exemplars and Characteristics 94
15. Morality and Sociability Ratings for Exemplars and Characteristics..... 95
16. Imagined Gender for Each Prototype 99
17. Imagined Race for Each Prototype 100
18. Imagined Age for Each Prototype..... 101
19. Essentialism Ratings by Prototype..... 106
20. Desired Social Distance by Prototype..... 107

LIST OF TABLES

Table Page

1. Bad Person Characteristics Unconstrained Listing Ranked by Frequency..... 42
2. Bad Person Characteristics Unconstrained Listing Ranked by Mean Badness 44
3. Descriptives and Zero-Order Correlations of Bad Person Characteristics 47
4. Most Frequent Labels Given to Describe Initial Groupings..... 55
5. Membership Values from C-Means Analysis..... 66
6. Research Design Demonstrating How Exemplars Are Defined..... 77
7. Pool of Characteristics Describing Exemplars 80
8. Factor Loadings of Judgments of Exemplars and Characteristics..... 82
9. Model Predicting Likelihood of Psychopath from Category Membership..... 86
10. Model Predicting Likelihood of Abuser from Category Membership..... 87
11. Model Predicting Likelihood of Bigot from Category Membership 88
12. Model Predicting Likelihood of Narcissist from Category Membership 88
13. Model Fit Metrics for Time of Responding to Inferences..... 91
14. Model Fit Metrics for Models Likelihood of Inference..... 98
15. Means (SDs) for Differences Between Cognitive Prototypes 103
16. Participant Generated Unobserved Characteristics of Bad Person Exemplar 111
17. Participant Generated Unobserved Characteristics of Good Person Exemplar..... 112
18. Participant Generated Unobserved Characteristics of Psychopath Exemplar 113
19. Participant Generated Unobserved Characteristics of Abuser Exemplar 114
20. Participant Generated Unobserved Characteristics of Bigot Exemplar..... 115
21. Participant Generated Unobserved Characteristics of Narcissist Exemplar..... 116

22. AI Images from Appearance Descriptions of Basic Prototypes 117

23. AI Images from Appearance Descriptions of Superordinate Prototypes..... 118

I. INTRODUCTION

The label *bad person* is used to describe a wide range of individuals who violate moral norms. This includes those in positions of power who organize and benefit from suffering on a massive scale, such as genocidal dictators, terrorist organization leaders, cult figures, sex traffickers, and drug lords. It includes common criminals directly responsible for inflicting harm onto others, such as serial killers, mass shooters, sex offenders, batterers, stalkers, scammers, and burglars. Even those who have not broken any laws are sometimes referred to as *bad people*, including political adversaries, members of marginalized communities, billionaires, atheists, 'climate criminals,' schoolyard bullies, and self-absorbed popular 'mean girls.'

The people listed above differ in a lot of important ways. Some are violent, while others are too focused on themselves. Some have high intelligence that makes it easy to deceive others, while some individuals' immorality stems from their perceived uninformed or unintelligent beliefs. Some get close to their targets through their warmth and charm, while others are reclusive and disagreeable. Some are seen as capable of reform, while others are perceived to have immoral characteristics that are hardwired in their brain or DNA. Given the wealth of important information to attend to, how do we know what to expect and how to act when we encounter people like these?

In response to an information dense world, we organize related information into categories (e.g., furniture, animals) to help us know how to interact with and what to expect from new stimuli in our environment (Rosch, 1978). Categories are represented in the mind as

cognitive prototypes: average or typical examples with features abstracted from experiences[1] with category members (Posner & Keele, 1968; Reed, 1972; Rosch & Mervis, 1975) and as exemplars: specific category members that a person has interacted with or learned about (Medin & Schaffer, 1978; Medin & Smith, 1981). A new stimulus is granted category membership according to the degree of similarity between the new stimulus and the mental representation of the category (Dopkins & Gleason, 1997; Rosch & Mervis, 1975). For example, when we encounter a chair we have never seen before, we recognize it as a *chair* because it is similar to our mental representation of a chair. Once we have categorized it as a *chair*, we immediately know how to interact with it. A chair can be sat on.

In the social domain, we categorize people into groups defined by shared attributes, such as *women*, *Canadians*, *extroverts, runners,* and *accountants*. These social categories and their mental representations (e.g., prototypes, stereotypes[2]) enable us to make quick inferences and

[1] Experiences with category members does not need to be direct. For categories where it is rare to encounter a category member, the cognitive prototype may be entirely constructed from indirect experiences. For example, we are likely to receive more information about serial killers from the media (e.g., news, true crime documentaries) than from direct experiences because of how rare it is to encounter a category member.

[2] Both stereotypes and prototypes shape how we perceive and interact with strangers. While stereotypes are (usually negative) beliefs and attitudes about members of social categories (e.g., women, Asians), prototypes refer to the mental representations of concepts within and outside the social domain (Hilton & Hipple, 1996; Rosch, 1976). In this research, I use the term

interaction decisions about people we have limited knowledge of (Bodenhausen et al., 2012; Cantor & Mischel, 1977; Macrae & Bodenhausen, 2000). For example, upon learning that your neighbor is an accountant, you might infer that he has the stereotypical characteristics of accountant -- for instance, that he is *boring* (Tilburg et al., 2022) -- and interact with him accordingly. You might assume that he is trustworthy enough to approach at the block party, although a boring conversationalist.[3]

An important social interaction decision is to decide whether to approach or to avoid someone. Moral psychologists have theorized that people make this distinction when we divide people into *good people* and *bad people*, where *good people* are typically those we want to form relationships with and *bad people* are typically those we want to avoid (Hartman et al., 2022; Heiphetz, 2020; Helzer & Critcher, 2018; McHugh et al., 2022). Just like we categorize a new chair based on its resemblance to the *chair* prototype, I propose in this dissertation that we evaluate moral character based on how closely a target resembles our mental representation of a *good person* or *bad person*. Once a categorization is made, I hypothesize that we are more inclined to approach, praise, and form relationships with those we perceive as good and to avoid and punish those we perceive to be bad.

In addition, I propose that there are more than two categories in the moral person domain. In general, categories representing a concept are hierarchically structured, with a *basic level*

---

prototypes instead of stereotypes as I am referring to the mental representations of social categories rather than beliefs or attitudes about social categories.

[3] I mean no offense to accountants. Your job is very valuable, and I'm sure you are really fun to be around.

where categories are highly informative and distinct from each other, a *superordinate level* where categories are more abstract, more different from each other, and have more members, and a *subordinate level* where categories are more concrete, more similar to each other, and have fewer members (Rosch, 1978). For instance, if *furniture* is at the superordinate level, then *chair* would be a category at the basic level, and *desk chair* would be a category at the subordinate level.

We should expect a similar type of hierarchy for *bad people*. Given the diversity of members in the *bad person* category, the mental representation of a *bad person* is not sufficiently detailed to make nuanced inferences and interaction decisions. For example, just as describing a chair as *furniture* would not provide specific advice on how to interact with the object, categorizing someone as a *bad person* would lead to uncertainty surrounding how to interact with and what to expect from them. For instance, while some *bad people* pose an immediate physical threat to our safety, others are merely untrustworthy or unreliable friends or business partners.

Therefore, I propose that we also perceive moral person basic categories that are less inclusive than the superordinate *bad person* category, such as *psychopath*, *bigot*, *pedophile, dictator*, *terrorist*, and *narcissist*. Lower-level moral person categories have richer mental representations compared to superordinate categories because they contain additional, category specific information. And since basic categories optimize *cue validity*, or the ability for a given piece of information to predict membership into a category (Rosch et al., 1976), I propose that the formation and maintenance of basic categories plays a crucial role in moral character evaluation.

Psychologists researching moral character evaluation define moral character as an individual's perceived disposition towards morally relevant thoughts, feelings, choices, traits, and behaviors (Cohen & Morse, 2014; Goodwin et al., 2014; Hartman et al., 2022). Nearly all research on moral character evaluation has neglected to study moral person categories or has concentrated on the superordinate *good person* and *bad person* categories (e.g., Hartman et al., 2022; McHugh et al., 2022). Most researchers use a simple, unidimensional model of moral character where targets can be placed and ranked against each other on a moral dimension from *extremely immoral* to *extremely moral* (e.g., Brandt & Reyna, 2011).

Primarily, researchers have examined whether a single piece of information can influence moral impressions by manipulating one characteristic in a scenario and keeping all other information constant. For example, by having participants indicate on a unidimensional scale of morality where they would place targets who make different decisions in moral dilemmas (Critcher et al., 2020; Uhlmann et al., 2013), who have different faces (Rosa et al., 2021; Todorov et al., 2015), or express different emotions in a moral situation (Anderson et al., 2021; Stearns & Parrott, 2012). When they observe a significant difference between groups, they record the variable as a factor that influences moral impressions.

However, a unidimensional model of moral character evaluation is insufficient because it assumes that members from different moral person categories who score the same on the unidimensional scale are more similar than they are perceived to be. For example, two people perceived to have the same moral character on a unidimensional scale may be expected to violate different moral norms, be assessed to have different capabilities to carry out their intentions, or to be avoided to different extents. Thus, I propose that a more complex model (e.g., multidimensional, multicategory) of moral character evaluation could more powerfully explain

how people infer unobserved immoral characteristics in others and make decisions about with whom to interact and how to interact with them.

In this dissertation project, I tested and compared the unidimensional model to two additional models of moral character evaluation: a prototype model and a multidimensional model. In **a unidimensional model**, a target is inferred to have unobserved characteristics that align with their perceived morality. For example, a moderately bad person would be assumed to have moderately bad characteristics, and less so extremely bad or slightly bad characteristics. In contrast, in **a prototype model**, a target who is similar to a cognitive prototype is inferred to have unobserved immoral characteristics that are prototype consistent. For example, a target who has characteristics similar to the psychopath prototype would be inferred to have unobserved characteristics consistent with the psychopath prototype. In **a multidimensional model**, additional dimensions of person perception (e.g., competence, sociability) can explain why two individuals with the same morality are inferred different immoral attributes. For instance, an immoral, competent, and sociable person would be more likely to be inferred competent and sociable characteristics compared to an immoral, incompetent, and unsociable person.

Fully understanding the process of moral character evaluation is vital because person perception errors in the moral domain can have serious consequences. Failing to recognize someone's capacity for harm can lead one to be the victim of deception, exploitation, or violence. Conversely, falsely accusing someone of poor character can result in a missed opportunity to connect with and benefit from a partnership, or the injustice of wrongful accusations or punishment. Thus, developing a deeper understanding of the mechanisms that underlie moral person judgments can inform interventions aimed at preventing harmful errors

from occurring. In summary, my dissertation had three goals that will deepen our understanding of those mechanisms:

1. **Derive a Taxonomy of Bad Person Concepts**. Expand our understanding of how people store immoral person characteristics by deriving basic categories of the bad person concept. Deriving a taxonomy of bad person concepts from a relatively homogeneous population would establish that individuals within a subculture share mental representations of more than one type of bad person.
2. **Test a Prototype Model Against Dimensional Models of Immoral Inferences**. Explain how unobserved immoral characteristics are inferred about a target from a set of observed immoral characteristics. Testing whether immoral characteristics are inferred to a target because the initial and inferred characteristics are perceived to be prototypical of the same type of person (i.e., prototype model), because of their proximity to the target on a dimension or dimensions (i.e., dimensional models), or because of prototypes and proximity (i.e., a dual model) would improve our understanding of moral character evaluation.
3. **Examine Properties of Moral Person Prototypes**. Evaluate whether we perceive and interact with various types of bad people differently. Testing for differences in perceived morality, competence, sociability, gender, age, race, physical appearance, beliefs regarding the immutability of their immoral characteristics, and desire for social distance of characterizations of exemplars would deepen our understanding of why people make basic category distinctions in the moral domain.

Deriving A Taxonomy of Bad Person Concepts

Knowing what are the basic categories in a domain is useful for understanding how people make everyday distinctions between different members of a superordinate category (Bodenhausen et al., 2012; Rosch et al., 1976). For social categories, identifying the basic level of a taxonomy can inform our understanding of how people naturally describe and understand others. People prefer to describe others at the level of abstraction that is inclusive of the greatest number of people but still predictive of behavior, as opposed to being purely evaluative (John et al., 1991). For example, in the personality hierarchy *stingy*, *selfish*, *unkind*, and *bad*, people prefer to describe others as *unkind*, which is the most abstract word in the hierarchy that still informs what kind of behavior the person is likely to engage in. Similarly, people may prefer to refer to those who violate moral norms with terms that are more predictive of behavior than *bad person,* such as *psychopath*, *narcissist*, or *sex offender*. However, while prior theory has alluded to "an infinite number of categories" in the moral domain beyond the superordinate categories *good* and *bad* (p. 133; McHugh et al., 2022), there have been no sufficiently rigorous empirical investigations to derive a moral person taxonomy.

The very first draft of a moral person taxonomy may go as far back as the 4th century BCE. Scholars have reconstructed writings from one of Aristotle's students, Theophrastus, who penned highly entertaining character sketches of 30 different types of men (Anderson, 1970). Theophrastus depicted each man with a disagreeable characteristic and described how each type of man could be identified across different situations. Many of his characters would be considered a *bad person* by today's standards. For example, the *Man without Moral Feeling* does not feel guilt or remorse when he violates moral norms, the *Insincere Man* compliments others to their face and then speaks ill of them behind their back, the *Arrogant Man* believes he is superior

to others, and the *Shamelessly Greedy Man* hoards more resources than he needs. Other characters possess negative but not necessarily immoral traits, such as the unhygienic *Repulsive Man* who goes to bed without showering and the *Talkative Man* who constantly expresses what he is thinking, even in socially inappropriate situations.

Although the exact purpose (e.g., entertainment, theatre, personal vendetta) of Theophrastus's character sketches is unknown, at some point *The Characters* was used for moral instruction. The preface of most modern translations depicts Theophrastus informing his interlocutor Polycles that the character sketches will "help [our children] to choose the best people to know and be with, and they will have something to aim at" (Anderson, 1970, p. 3). Thus, the highly detailed examples of morally flawed individuals that Theophrastus sketched out many centuries ago may have been used or intended to be used to instruct children on who to avoid.

Additionally, throughout history Christian theologists have sought to compile a list of sins or vices that lead to immoral action to use in their moral instruction. Although the specific vices that are considered sinful have evolved over time, the most influential and well-known set emerged during the Middle Ages: The Seven Deadly Sins – pride, envy, anger, avarice, sloth, gluttony, and lechery (Bloomfield, 1952). Although Christian theology specifies that these are sins that everyone must resist for a virtuous life, it may be that people perceive seven different sinners: a person who believes they are superior to others (pride), a person who is excessively envious of others' success and possessions (envy), a person who cannot control their anger, hatred, and desire for vengeance (anger), a person who hoards wealth and possessions (avarice), a person who lacks motivation and ambition (sloth), a person who overindulges on food and

drink (gluttony), and a person who violates moral norms such as marital fidelity in the pursuit of sexual pleasure (lechery).

In modern social psychology research, Moral Foundations theorists have proposed that there are at least six categories of moral violations, similar to the seven sins from Christian theology: harm, inequality, disproportionality, disloyalty, subversion, and impurity (Atari et al., 2022; Graham et al., 2013). If categories of moral violations map directly onto norm violators, then Moral Foundations Theory would expect six types of immoral agents: a type of person who physically harms others (harm), a type of person who treats some people worse than others (inequality), a type of person who takes more than they deserve (disproportionality), a type of person who is disloyal to their groups and allies (disloyalty), a type of person who disobeys authority and disrespects tradition (subversion), and a type of person who defiles objects, bodies, or institutions that are considered pure (impurity).

The validity of this six-category model of moral violations was tested by asking participants to rate the likelihood that an agent would violate a moral norm, given that they violated a different moral norm (Landy & Bartels, 2018). A cluster analysis revealed that people perceive three types of bad people: a person who violates moral and social norms (e.g., cheating, hurting others, being impolite), a person who does disgusting behaviors, usually of a sexual nature (e.g., watching deviant pornography), and a person who uses power to restrict or deny the freedoms of others (e.g., forcing their daughter to take premed classes). This study gives initial support for the claim that people distinguish between different types of bad people. However, the authors' main objective was to evaluate a taxonomy of moral behaviors, not to derive basic categories of bad people. As a result, they limited their characteristics to behaviors and did not provide any cognitions (e.g., beliefs, intentions) or affect (e.g., guilt, empathy), which are

especially powerful signals of moral character (Critcher et al., 2020; Helzer & Critcher, 2018). Thus, Landy and Bartels (2018) may be underestimating the actual number of perceived basic categories, as they did not test whether people distinguish more finely between categories based on perceived cognitions and affect.

Additionally, personality research may offer insight into a moral person taxonomy insofar that perceived categories in part reflect underlying realities concerning different antagonistic personality traits individuals possess. For example, the Dark Triad proposes three antisocial or harmful personality traits that individuals possess to differing extents: psychopathy, Machiavellianism, and grandiose narcissism (Paulhus & Williams, 2002). Thus, people may perceive three types of bad people: psychopaths, who have emotional deficiencies and inclinations toward impulsive, thrill-seeking behavior, which can lead them to harm others without guilt or remorse (Dolan & Fullam, 2006; Hare et al., 1990; Holmqvist, 2008), Machiavellians, who manipulate and exploit others through careful planning (Jones & Paulhus, 2014; Monaghan et al., 2016; Rauthmann & Kolar, 2012), and grandiose narcissists, whose entitlement and grandiose sense of selves may lead them to become aggressive if their positive self-concepts are threatened (Bushman et al., 2003; Gentile et al., 2013).

Further, others have proposed additional "dark traits" such as sadism, spitefulness, greed, perfectionism, and dependency that could be perceived moral categories (Marcus & Zeigler-Hill, 2015; Mededović & Petrović, 2015). For example, people may perceive the moral category *sadists* who derive pleasure from causing others psychological or physical pain (Baumeister & Campbell, 1999; Buckels et al., 2013) and *spiteful people* who are willing to take some cost to themselves to harm others (Marcus et al., 2014). It is reasonable to expect that any number of these categories may emerge as perceived moral person categories.

Additionally, prior work on personality structure derived from type-nouns may inform the expected structure of perceived categories. For instance, US students indicated the extent to which a large set of type-nouns (i.e., nouns that define a class of people with shared attributes, such as lawbreakers or psychopaths) described themselves and others (Saucier, 2003). A factor analysis revealed that an eight-category solution fit the data reasonability well, revealing a personality structure that included more than one morally relevant factor: an egocentric factor (e.g., snob, gossip, eavesdropper) and a delinquency factor (e.g., lawbreaker, pothead, drunk). Insofar as moral perception in part reflects reality, this work may indicate that US students perceive a type of person that is negative and critical, and another type of person who breaks laws. However, given that participants were young adults who were instructed to describe people they know, there may be additional factors that were not captured by the methodology that are important in criminal contexts (e.g., serial killer, psychopath) or may emerge in other samples. For example, a German study of type-nouns found a factor for men related to cowardice (Henss, 1998).

Recently, Hartman et al. (2022) theorized a taxonomy of moral person types with three superordinate categories *villain*, *hero*, and *victim*, and basic categories *saint*, *striver/sinner*, *demon*, *opportunist*, *virtuous victim*, and *culpable victim*. In this taxonomy, basic categories represent different levels of prototypically to the superordinate category. For instance, a *demon* is evil in all situations, while an *opportunist* is inclined towards evil but whether the inclination is present or not will depend on the situation. This model suggests that the only important differences that people observe between basic moral person categories are valence (i.e., whether a target is perceived to be good or bad) and strength (i.e., whether a target is perceived to be

capable of acting on good or bad intentions). However, no empirical test of this category structure has been conducted.

Present Study

My dissertation project aims to contribute to the literature by first establishing that people distinguish between different types of bad people based on more complete information about their cognitions, affect, behavior, and personal characteristics[4]. This study is necessary, as existing taxonomies of moral character types have been theoretical, derived to understand personality structure rather than person perception, or empirically driven from behaviors only.

Testing a Prototype Model Against Dimensional Models of Immoral Inferences

The unidimensional model of moral character evaluation that is widely used in the moral psychology literature often predicts that two targets with identical moral character scores will be perceived as more similar to each other than they are actually perceived to be. For instance, two

[4] It's essential to clarify that I do not aim to derive a cross-culturally valid taxonomy of moral person types. Just as moral norms differ significantly across cultures (Graham et al., 2009; Haidt et al., 1993; Janoff-Bulman & Carnes, 2013; Rai & Fiske, 2011; Schein & Gray, 2015), I anticipate substantial cross-cultural disagreement regarding which types of bad people are perceived and how they are cognitively represented. To properly test models of moral inferences, it is necessary to understand the taxonomic structure of the participants within the culture being sampled. Therefore, this project derives basic categories from the same pool of participants that were then used to test models of moral judgment, but the taxonomic structure derived should not be interpreted as one that is universally shared. I expand on this point in the general discussion.

individuals who are both considered to be *a moderately bad person* in a unidimensional model may differ in additional ways that are important to accurately predict their behavior, infer unobserved or unobservable characteristics, recommend punishment, and make informed decisions about how and when to interact with them. By measuring and conceptualizing moral character on a unidimensional scale, researchers cannot fully explain how observers make nuanced moral judgments and decisions that are important in real-world contexts.

For example, researchers using a unidimensional measurement of morality have found that we sometimes make moral character judgments that do not reflect the morality of the observed behavior, such as characters in vignettes being rated as more immoral than their behavior warrants (Tannenbaum et al., 2011; Uhlmann & Zhu, 2014) or more moral than their behavior merits (Schaumberg & Mullen, 2017). Of interest here, many studies have found that behaviors or characteristics that are considered relatively harmless can be perceived as strong evidence of a target's poor character. For instance, animal-abusers were rated as having poorer character than wife-abusers, despite wife-abuse being rated as more immoral than animal-abuse (Uhlmann & Zhu, 2014). In another study, job candidates who asked for a $40,000 marble table with their face engraved on it were rated as having poorer integrity than candidates who asked for a $40,000 cash bonus (Tannenbaum et al., 2011).

A unidimensional model of moral character cannot explain these findings. A unidimensional model would predict that a target's morality should be perceived as similar to the morality of the observed characteristics, and targets should be most likely to be inferred characteristics that are similar in morality to the target. However, other models may be better able to explain moral character evaluation: a prototype model, a multidimensional model, and a dual model.

A Prototype Model

According to prototype theory, categories are represented in the mind as a summary of attributes and their relations, abstracted from direct and indirect experiences with members from the category (Posner & Keele, 1968; Reed, 1972; Rosch & Mervis, 1975). Further, membership is granted based on the degree of similarity between the new stimulus and the cognitive representation of the category (Rosch & Mervis, 1975). Thus, if a target's observed characteristics are similar to a perceiver's prototype of a bad person category, then the target would be classified into the category and transferred unobserved or unobservable characteristics that are consistent with the perceiver's mental representation of the category (Cantor & Mischel, 1977; Osherson et al., 1990; Smith & Zárate, 1992). Therefore, a prototype model of moral character evaluation would posit that similarity to a prototype would predict the likelihood that a target is inferred unobserved immoral characteristics.

Reflecting on surprising findings in the moral character evaluation literature, Schein and Gray (2018) noted, "[c]onsider a man who masturbates with a dead chicken. Although this act is 'harmless,' parents likely would not want him babysitting their daughter for fear of the harm he might do to her." Thus, relatively harmless characteristics could lead to severe moral commendation of a target's character when the target is assumed to have immoral characteristics that were not originally observed (i.e., future harmful behavior). For instance, a prototype model would predict that someone who masturbates with a chicken would be categorized as a *sex offender* or *psycho* and inferred other characteristics of their category, and thus receive more condemnation than the behavior warrants.

Additionally, consider the finding that an animal abuser was perceived as having worse moral character than a wife abuser, even though wife abuse was rated as more morally wrong

than animal abuse (Uhlmann & Zhu, 2014). Animal abuse is widely known to be part of the MacDonald Triad, factors that when observed together are theorized to (although not empirically supported) predict a serial killer (Parfitt & Alleyne, 2020). While animal abuse may not be rated as more immoral than spousal abuse, the perceiver may be inferring unobserved characteristics based on perceived similarity to a prototype (e.g., a psychopathic serial killer), and therefore rate the target as more immoral than the action itself would warrant. In another example, when the interviewee requested a marble table with their face on it, it may have activated the perceiver's narcissist prototype, whereas simply asking for more money did not (Tannenbaum et al., 2011). Thus, a prototype model, which posits that targets are inferred unobserved characteristics when they are prototype consistent, may better explain inferences compared to a unidimensional model.

A Multidimensional Model

Another model that may offer additional explanation above a unidimensional model is a multidimensional model. For instance, Hartman et al. (2022) proposed strength as a second dimension of moral character, which concerns differences in the ability to carry out moral or immoral intentions and seemingly captures the *competence dimension* of person perception (e.g., Brambilla et al., 2021; Goodwin, 2015). Given that moral person perception is expected to utilize the same mental processes involved in general person perception, we should also expect that people are attuned to the third dimension of person perception, sociability, which reflects the target's outward extroversion and charm (Brambilla et al., 2021; Goodwin, 2015). Judgments about sociability are relevant to encode because they can predict the target's capacity to integrate into society, persuade others of their moral beliefs, and deceive others into perceiving them as virtuous. Thus, a multidimensional model, which posits that targets would be inferred

unobserved characteristics that are similar to the target on multiple dimensions (e.g., morality, competence, and sociability), may better explain inferences compared to a unidimensional model.

A Prototype-Dimensional Dual Model

A prototype model of moral character evaluation expects that category membership will predict the likelihood that a target will be inferred a characteristic, such that targets are more likely to be inferred characteristics that are consistent with their category and less likely to be inferred characteristics that are inconsistent with their category. Further, if the categorical structure has a strict hierarchy, all category-inconsistent information is equally likely to be inferred. A dimensional model of moral character evaluation offers that the distance from the observed and inferred characteristics on one or more dimensions (e.g., morality, competence, and sociability) predicts the likelihood that a target will be attributed an unobserved characteristic. According to a dimensional model, category membership does not offer any additional explanation.

However, these models may be too simple, as it is possible that both category and distance provide explanation for why characteristics are inferred. For instance, it is possible that certain categories are perceived to be closer in the dimensional space to some categories compared to others. For example, people may perceive a *bigot* and a *narcissist* to be closer in morality, competence, and sociability than they are to a *psychopath*, and therefore infer that a *bigot* has narcissistic traits to a higher degree than psychopathic traits. Thus, a dual model that includes both dimensional and prototype components may best explain moral character evaluation.

Present Study

My dissertation project aimed to contribute to the literature by explaining why, given a set of initial immoral characteristics about a target, additional immoral characteristics are inferred. I hypothesized that the prototype, multidimensional, and dual models would provide a better explanation for immoral inferences than the unidimensional model that is widely used in the literature.

Examining Properties of Moral Person Prototypes

The final goal of my dissertation project was to evaluate whether there are meaningful and systematic differences between how we perceive and interact with different types of bad people. Cataloguing these differences provides valuable insight into why we employ basic categories to make more accurate judgments and decisions. Additionally, by examining the properties of *bad person* prototypes, I have illuminated harmless characteristics (e.g., physical appearance, social identities) that can lead to a false positive categorization that the target intends or desires harm. If I would have failed to find differences between exemplars of various bad person categories, then a parsimonious conclusion would be that there is one category: *bad person*.

Three Dimensions of Person Perception

Impressions of a target can be explained almost entirely by the perceiver's independent assessments of the target's warmth and competence (Abele et al., 2021; Abele & Wojciszke, 2014; Cuddy et al., 2008; Koch et al., 2016). *Warmth* refers to how the target interacts with others, including their friendliness, sociability, honesty, and trustworthiness, and *competence* refers to the target's ability to complete their goals, including their confidence, assertiveness, intelligence, and skill. Recently, researchers have argued for the separation of the warmth

dimension into discrete evaluations of morality and sociability (Brambilla et al., 2021; Brambilla & Leach, 2014; Goodwin, 2015; Leach, 2007). A three-dimensional model better explains the primacy of morality in social judgment and accounts for the reality that morality and sociability are often independent, such as a gregarious thief or a principled introvert. Since morality, competence, and sociability are recognized as the primary dimensions of general person perception, they are expected to be important in our understanding of the distinctions people make about different types of bad people.

Generally, people with positive characteristics are inferred to have unrelated positive characteristics and people with negative characteristics are inferred to have unrelated negative characteristics, i.e., the "halo effect" (Thorndike, 1920). For example, people who violated moral norms like cheating or shoplifting were inferred to have low ability, an effect that was driven by perceptions of low intelligence (Stellar & Willer, 2018). Thus, I hypothesized that people who are described to have prototypical *bad person* characteristics (e.g., lacks empathy) will also be inferred to have lower competence and sociability than people who are described to have prototypical *good person* characteristics. However, I hypothesized that not all basic categories are perceived the same, such that prototypical members of basic *bad person* categories can diverge significantly on morality, competence, and sociability from each other, depending on the characteristics of the prototype.

Morality. The moral dimension is composed of lower-level traits like *unfair*, *dishonest*, *unprincipled*, *untrustworthy,* and *disloyal* (Brambilla et al., 2021; Goodwin, 2015). Perceiving the morality of others is important because it provides perceivers with information about whether the target intends to be harmful or helpful to the perceiver and others (Fiske et al., 2007). Some *bad person* categories (e.g., a psychopath) are expected to be rated as more immoral than other

bad person categories (e.g., a narcissist), based on how dangerous and harmful they are perceived to be.

Ability. The competence dimension consists of lower-level traits like *unintelligent*, *illogical*, *unorganized*, *unathletic*, *uncreative*, and *unclever* (Brambilla et al., 2021; Goodwin, 2015). Perceiving the competence of others is an important moral concern because it provides information about whether the target can fulfill their immoral or moral desires and intentions (Fiske et al., 2007). Further, Hartman et al. (2022) argued that immoral agents who are perceived to have low ability may be perceived more akin to victims than villains, as they are seen primarily as a product of their unfortunate circumstances. Some moral categories (e.g., *a con artist*) may be perceived to be highly intelligent and clever, while other categories may be perceived as irrational and ignorant (e.g., *a bigot*).

Sociability. The sociability dimension includes lower-level traits like *unsociable*, *unhappy*, *disagreeable*, *unfunny*, and *unenthusiastic* (Brambilla et al., 2021; Goodwin, 2015). Judgments of sociability are considered to provide a similar function to judgments of competence because they provide information about the target's ability to attain friends, allies, or supporters that could help them enact their moral or immoral desires or intentions (Landy et al., 2016). Some members of basic categories (e.g., *a psychopath*) may be perceived as solitary individuals on the fringes of society, while others (e.g., *a narcissist, a manipulator*) may come across as very charming and sociable, which allows them to get close to their targets.

Essentialist Attitudes

People, to differing extents, view others as having innate, immutable essences that cannot be observed directly but can explain their behavior and make them different from other types of people (Gelman, 2004; Rhodes & Moty, 2020). In particular, past research has found that people

view others as having "moral essences" that are perceived to be biologically driven, stable, essential to who a person is, differentiate bad people from good people, and can explain why people violate or adhere to moral norms (Heiphetz, 2020; Lynch et al., 2019; Strohminger & Nichols, 2014).

Essentialist attitudes are important to study in the moral domain because they predict endorsements of how severely offenders should be punished (Martin et al., 2022; Martin & Heiphetz, 2021; Vel-Palumbo et al., 2019). For instance, in cases where the crime is relatively minor (e.g., theft, property destruction), individuals described as being biologically predisposed to criminal behavior were recommended harsher punishments compared to individuals who committed the same crime but were not described as biologically predisposed (Martin & Heiphetz, 2021). In another study, people who more strongly endorsed essentialist attitudes about criminal behavior were more likely to endorse the death penalty (Martin et al., 2022).

Although previous researchers have not directly examined whether people hold varying essentialist views about different types of bad people, their research has provided indications that we should expect to find them. Mainly, some immoral characteristics seem to be more essentialized than other immoral characteristics. For instance, Martin and Heiphetz (2021)'s manipulation of essentialists attitudes had no effect for major crimes like "murder," which could be explained by an understanding that people already possess highly essentialized attitudes regarding murderers. In another study, respondents indicated that it would be a more dramatic departure from their friend's 'true self' if they were to suddenly become racist or cruel, compared to if they started stealing or cheating on a spouse (Strohminger & Nichols, 2014).

The discovery that some immoral characteristics are perceived to be more biologically driven and stable than others may reflect underlying perceived differences in the types of bad

people believed to possess those characteristics. For example, we might expect some types of bad people (e.g., *a psychopath*, *a pedophile*) to be perceived as having fixed immoral characteristics that are a function of biology, while other types of bad people (e.g., *the narcissist*) may be perceived to have traits that are nurtured and mutable.

Learning which categories are highly regarded as *essential* will further our understanding of how moral judgments of people change or fail to change over time. While members of less essentialized categories could be perceived to change over time and potentially even lose membership to their category, members of highly essentialized categories may be perceived to have a stable categorization. For instance, if the pedophile's sexual attraction towards children is perceived to be biologically determined and fixed, then someone who molests a child may always be perceived as a pedophile, even as decades pass without re-offending. Additionally, members of highly essentialized categories may be more susceptible to harsher punishments, including endorsement for the death penalty, even when the punishment may not be proportional to the offense committed. In contrast, people may perceive members of less essentialized categories as probable or even likely to mature out of their category, and therefore will interpret instances of good behavior as genuine evidence that they no longer belong in the category.

Social distance

Another important distinction between different types of bad people may be how comfortable we are interacting with them. While we may seek to exclude sex offenders from living in our neighborhoods and advocate lifelong imprisonment for psychopaths, we may feel more comfortable coexisting with other types of bad people. People who display strong moral principles are often disparaged by observers who feel morally inferior in comparison (e.g., vegans; Minson & Monin, 2012). In contrast, individuals who exhibit less stringent moral

standards do not threaten the perceiver's positive moral self-concept, and thus the observer may be more willing to form a relationship with such individuals, as long as they do not pose a genuine threat to their physical or mental health. In some cases, people may even benefit from relationships with wrongdoers, as they offer a way to reap the benefits of moral transgressions without direct involvement in the wrongdoing (Melnikoff & Bailey, 2018). Additionally, we often tolerate or in some instances desire narcissistic (Brunell et al., 2008; Grijalva et al., 2015) and psychopathic (Uhlmann et al., 2013) traits in leaders.

Present Study

My dissertation project tested whether there are perceived differences in morality, ability, sociability, essentialist beliefs, desired social distance, gender, race, age, and physical appearance between exemplars of various bad person categories. By cataloguing the attributes of four different prototypes, I established that we have cognitive prototypes in the moral person domain and provided preliminary insights that can inspire and inform future research questions in the field.

II. STUDY 1: GENERATING CHARACTERISTICS OF A BAD PERSON

The purpose of Study 1 was to generate a list of prototypical characteristics (e.g., behaviors, emotions, beliefs, desires, motives, physical appearance) of a *bad person* in the same relatively homogenous population that will be used to derive (Study 2) and test (Study 3) categorical and dimensional models of moral character evaluation. Using a relatively homogeneous population was necessary because people from different cultures or subcultures are expected to have different conceptions of a bad person that would weaken the tests of moral character evaluation in Study 3. For example, if someone does not think that racism or homophobia makes someone a *bad person*, then they may be less likely to infer to that person unrelated immoral characteristics than someone who does think racism or homophobia makes someone a *bad person*. A secondary goal of Study 1 was to conduct an exploratory analysis to test which individual difference variables (e.g., moral values, conservatism, belief in pure evil) can explain differences in conceptions of bad people.

Method

Participants

University of Oregon students were recruited online from the Psychology/Linguistics Human Subjects Pool to take a 30-minute survey and were compensated with course credit. Of the 183 participants aged 18 to 70 (M = 20.14, SD = 4.71), 69.3% identified as a woman, 26.8% identified as a man, and 3.8% identified as non-binary or "other." Additionally, 66.1% identified as White, 12.6% identified as Asian, 11.5% identified as Hispanic, 3.8% identified as Black, and 6.1 % specified an ethnicity other than those listed. Finally, the sample skewed liberal, with 68.3% identifying as extremely or moderately liberal, 25.1% identifying as moderate, and 6.6% identifying as extremely or moderately conservative. According to recent demographics provided

by the university registrar, the sample generally represented the racial breakdown of the typical University of Oregon student but oversampled female students.

Because I did not conduct an inferential test for this study, I did not run a power analysis to determine sample size. I aimed to collect between 100 to 200 participants, which is adequate to survey a broad range of students and is standard across studies that use this methodology to derive prototypical traits (e.g., Tilburg et al., 2022; Walker & Pitts, 1998). Thirty-five participants were removed from the exploratory analysis explaining variance in the conceptions of bad people for failing at least one attention check.

Procedure

In an online study, participants provided descriptions of a *bad person* with the following instructions:

> ***Instructions****: Picture in your mind a "bad person." Think about a real person (not a fictional character), ranging from a person you would consider to be somewhat bad to one you would consider to be extraordinarily bad. Now think of other people, each of whom you would consider to be a "bad person." In the blanks below, describe the characteristics of these bad people, such as their behaviors, emotions, beliefs, desires, motives, intentions, traits, relationships, abilities, and physical appearance.*

The participants were provided 20 blank spaces to write in descriptors. They were told to list as many descriptors as come to mind (one per blank) and that they could click the “add characteristic” button to add more. If they did not add at least 10 characteristics, an error message appeared that said that they needed to have at least 10 characteristics to continue. On the next page, the participants were given a rating matrix table of each characteristic they had written and were told to “rate the extent to which you would consider someone **a bad person** if

they expressed the characteristic" on a scale of 1 (*definitely not a bad person*) to 9 (*definitely a bad person*). On the following page, they rated their free listed characteristics again but with **good person** replaced with bad person, i.e., on a scale of 1 (*definitely not a good person*) to 9 (*definitely a good person*).

After the unconstrained listing and rating task, participants were asked to rate 70 additional characteristics on the two 9-point scales described above. The characteristics were drawn from theory and previous research on which characteristics cue negative moral character judgments (see Appendix A). I included the 70 additional characteristics to ensure that attributes that are perceived to be prototypical of bad people but are not easily accessible (e.g., due to low prevalence in the general population) are considered in the creation of the final list of prototypical characteristics.

Finally, in order to conduct an exploratory analysis to examine how conceptions of what makes a person a *bad person* differ across individual difference variables, participants filled out a short survey to collect their moral values, conservatism, belief in pure evil, gender, age, ethnicity, political orientation, and political affiliation.

Materials

Moral values were measured with the 36-Item MFQ-2 (Moral Foundations Questionnaire-2; Atari et al., 2022). The MFQ-2 measures the extent to which an individual endorses moral principles related to harm (e.g., "caring for people who have suffered is an important virtue"), equality (e.g., "the world would be a better place if everyone made the same amount of money"), proportionality (e.g., "I think people who are more hard-working should end up with more money"), loyalty (e.g., "I think children should be taught to be loyal to their country"), authority (e.g., "I think obedience to parents is an important virtue"), and purity (e.g.,

"I think the human body should be treated like a temple, housing something sacred within") on a scale of 1 (*does not describe me at all*) to 5 (*describes me extremely well*). People who score high in the proportionality, loyalty, authority, and purity domains diverge from the typical liberal undergraduate student and may therefore be expected to have different conceptualizations of a bad person (Graham et al., 2009).

The extent to which an individual believes that some people are evil was measured with the 22-Item BPE or Belief in Pure Evil Scale (Webster & Saucier, 2013). Participants rated items such as "evil people have an evil essence, like a stain on their souls, which is almost impossible to get rid of" on a scale of 1 *(Disagree Very Strongly)* to 7 (*Agree Very Strongly).* People who score low on this scale may be less likely to condemn others for characteristics that have not caused harm (e.g., having a belief but not acting on it).

Conservatism was measured with the 12-Item Social and Economic Conservatism Scale (SECS; Everett, 2013). The SECS measures the extent to which an individual feels positively or negatively about different social (e.g., religion) and economic (e.g., business) issues on a 100-point scale, where higher scores are considered more conservative. People who score high on this scale are expected to have diverging views from the typical liberal undergraduate student on which behaviors, cognitions, and affect make an individual a bad person (e.g., Graham et al., 2009).

Results

The primary goal of this data analysis was to create a non-redundant list of the most prototypical characteristics of a *bad person* from the unconstrained listing and rating task and the 70-item rating task. A secondary objective was to conduct exploratory analyses to test whether

people have different conceptions of *a bad person* based on their moral values, political orientation, and beliefs regarding the existence of pure evil.

Generating a List of the Prototypical Characteristics of a Bad Person

On average, participants freely listed 15.56 (*SD* = 4.74) descriptors, for a total of 1211 unique descriptors. A team of four researchers (three undergraduate students and the principal investigator) reduced the list based on the following procedural rules for simplifying prototypical characteristics adapted from Walker and Pitts (1998). Disagreements over which characteristics should be considered synonymous were resolved through lengthy discussion over a period of several weeks.

1. Drop modifiers (e.g., *completely lacks empathy* turns into *lacks empathy*).
2. Separate compound phrases (e.g., *lacks empathy and is manipulative* turns into two separate characteristics, *lacks empathy* and *manipulative*).
3. Collapse words and phrases that have synonymous meanings in common use into one (e.g., *lacks empathy* and *lacks sympathy* became *lacks empathy*).
4. Drop idiosyncratic descriptors that cannot be combined with another category (e.g., *Arkansas native*), where *fewer than* 5 participants used the descriptor.

After removing idiosyncratic descriptors, 103 distinct descriptors remained. To determine the final 50 characteristics, I compiled three lists: the top 50 freely recalled descriptors rated by frequency (Table 1), the top 50 freely recalled descriptors rated by mean badness rating, and the top 50 closed list descriptors by mean badness rating (Table 2). To match the closed list descriptors to the freely recalled descriptors, three coders independently matched the closed list items with the free listed items and disagreements were resolved through discussion.

Forty-four characteristics were on at least two lists and were therefore prioritized to be on the final list. Five of the forty-four items were removed for being too multifaceted or too general (e.g., narcissistic, bigoted, toxic, cruel, and harms others). Next, based on high rankings across multiple lists, abuse was split into domestic abuse and child abuse, sadism was split into animal sadism, sexual sadism, and general sadism, and child predator was split into child molestation and sexual attraction towards children. The final seven items were added because they were rated extremely high on a single list (i.e., necrophilia, cannibalism, scamming) or rated close to 50 on more than one list (e.g., exploitative, jealousy, substance abuse, psychotic). None of the final characteristics were rated as strongly indicative of both a good person and a bad person.

To write items for the final list, I used pre-written items from the current study when available; otherwise, new items were created based on discussions with the research team about the meaning of each freely recalled characteristic. A few items were slightly reworded after discussion with the research team raised concerns that they could be misinterpreted. The final list of items can be found in Appendix B.

Explaining Variance in Conceptions of a Bad Person

In an exploratory analysis, I tested whether an individual's moral values, conservatism, and belief in pure evil are related to their conception of a *bad person* (Table 3). Overall, conservatives were less likely to endorse that having racist, transphobic, sexist, or xenophobic beliefs makes someone a bad person, but were more likely to endorse that being an atheist, unpatriotic, being addicted to substances, trading sex for money, and committing petty crimes makes someone a bad person. These findings suggest that conservatives and liberals may have different moral person category structures. For instance, a *bigot* may be more likely to emerge in

a liberal sample, whereas a *welfare queen* may be more likely to emerge in a conservative sample.

Belief in Pure Evil (BPE) positively predicted nearly all items where the person fantasized about but did not act on their immoral desires (e.g., rape, cannibalism, necrophilia, pedophilia, and infidelity), suggesting an individual difference in the perception that a person's badness can be dictated by their desires alone, even in the absence of evidence that the person has or will act on those desires. Further, social conservatism, $r(146) = .30$, $p < .001$ and economic conservativism, $r(146) = .19$, $p = .023$, were both significantly related to BPE, which suggests that conservatives may be more inclined to judge a person's moral character based on their desires alone.

Discussion

In Study 1, participants listed and rated characteristics that are prototypical of a *bad person*. The freely listed descriptors were polished by dropping modifiers, separating compound phrases, collapsing synonyms, dropping idiosyncratic responses ($n < 5$), and resolving discrepancies between coders. Then, the freely listed descriptors were compared to the closed-list ratings to generate a final list of 50 prototypical characteristics of a *bad person*. These 50 characteristics were used in the next study to develop a taxonomy of the *bad person* concept.

Exploratory analyses revealed that conservatives and liberals differed in their judgments of what constitutes a bad person, particularly for judgments concerning racism, substance abuse, and patriotism. Additionally, differences in moral values and the belief that some possess an evil essence also influenced which characteristics were considered prototypical of a *bad person*. These findings support the idea that there will be differences between population subgroups in their conceptions of bad person prototypes, especially across the political spectrum. Thus, as

suspected, investigations of moral person prototypes will have to be made within relatively homogeneous subgroups.

Table 1

Bad Person Characteristics from Unconstrained Listing Ranked by Frequency

		Frequency		Badness Ratings	
Rank	Characteristic	*N*	*Prop*	*M*	*SD*
1	Rude	130	71.0%	6.83	1.25
2	Lying	117	63.9%	6.68	1.36
3	Selfishness	96	52.5%	6.63	1.29
4	Mean	91	49.7%	6.81	1.43
5	Manipulative	82	44.8%	7.59	1.05
6	Narcissistic	65	35.5%	7.09	1.50
7	Easy to anger	64	35.0%	5.86	1.40
8	Egotistical	63	34.4%	6.38	1.33
9	Physically violent	50	27.3%	7.92	1.07
10	Abusive	47	25.7%	8.72	0.62
11	Racist	44	24.0%	8.55	1.30
12	Self-centered	42	23.0%	6.71	1.11
13	Aggressive	42	23.0%	7.40	1.17
14	Lacks empathy	40	21.9%	7.15	1.29
15	Judgmental	38	20.8%	6.50	1.29
16	Harmful	36	19.7%	7.86	1.44
17	Sexist	36	19.7%	8.42	1.40
18	Greedy	35	19.1%	6.94	1.33
19	Hateful	35	19.1%	7.57	1.29
20	Controlling	34	18.6%	6.85	1.33
21	Careless	33	18.0%	6.15	1.54
22	Cruel	33	18.0%	8.00	1.60
23	Bullying	32	17.5%	7.47	1.19
24	Unemotional	29	15.8%	6.62	1.97
25	Ignorant	29	15.8%	5.69	1.20
26	Negative	28	15.3%	5.96	1.73
27	Inauthentic	28	15.3%	5.96	1.40
28	Jealous	28	15.3%	6.50	1.43
29	Distant	26	14.2%	6.23	2.20
30	Close-minded	26	14.2%	6.58	1.60
31	Infidelity	25	13.7%	7.36	1.85
32	Spitefulness	24	13.1%	6.71	1.27
33	Homophobic	24	13.1%	8.17	1.79
34	Stealing	21	11.5%	7.52	1.60
35	Bigoted	21	11.5%	8.14	1.06
36	Freeloader	21	11.5%	7.90	1.14
37	Untrustworthy	20	10.9%	6.10	1.21
38	Entitled	20	10.9%	6.55	1.05
39	Does not care about others	19	10.4%	6.74	1.66

		Frequency		Badness Ratings	
Rank	Characteristic	*N*	*Prop*	*M*	*SD*
40	Disloyal	19	10.4%	6.83	1.25
41	Toxic	18	9.8%	7.33	1.57
42	Psychotic	18	9.8%	6.22	1.48
43	Power hungry	18	9.8%	7.22	1.52
44	Two-faced	17	9.3%	7.88	0.99
45	Few relationships	17	9.3%	5.24	1.68
46	Annoying	17	9.3%	5.88	1.54
47	Deceitful	16	8.7%	7.13	1.36
48	Unforgiving	16	8.7%	6.44	1.09
49	Murder	16	8.7%	8.75	0.58
50	Loud	16	8.7%	4.31	2.02

Table 2

Bad Person Characteristics from Unconstrained Listing Ranked by Mean Badness

Rank	Characteristic	Badness ratings		Frequency	
		M	*SD*	*N*	*Prop*
1	Child predator	9.00	0.00	7	3.8%
2	Transphobic	8.77	0.60	13	7.1%
3	Murder	8.75	0.58	16	8.7%
4	Rape	8.73	0.59	15	8.2%
5	Abusive	8.72	0.62	47	25.7%
6	Sadistic	8.69	0.63	13	7.1%
7	Racist	8.55	1.30	44	24.0%
8	Sexist	8.42	1.40	36	19.7%
9	Bad intentions	8.40	0.55	5	2.7%
10	Homophobic	8.17	1.79	24	13.1%
11	Bigot	8.14	1.06	21	11.5%
12	Cruel	8.00	1.60	33	18.0%
13	Physically violent	7.92	1.07	50	27.3%
14	Freeloader	7.90	1.14	21	11.5%
15	Two-faced	7.88	0.99	17	9.3%
16	Harmful	7.86	1.44	36	19.7%
17	Opposing political beliefs	7.85	0.99	13	7.1%
18	Lawbreaker	7.83	1.75	12	6.6%
19	Superiority complex	7.80	1.40	10	5.5%
20	Emotional abuse	7.80	1.30	5	2.7%
21	No guilt or remorse	7.80	1.10	5	2.7%
22	Xenophobic	7.67	3.27	6	3.3%
23	Gaslighting	7.63	1.19	8	4.4%
24	Manipulation	7.59	1.05	82	44.8%
25	Hateful	7.57	1.29	35	19.1%
26	Stealing	7.52	1.60	21	11.5%
27	Gossiping	7.50	1.08	10	5.5%
28	Dominance	7.50	1.38	6	3.3%
29	Bullying	7.47	1.19	32	17.5%
30	Aggressive	7.40	1.17	42	23.0%
31	Religious intolerance	7.40	3.58	5	2.7%
32	Possessive	7.40	0.89	5	2.7%
33	Cheats	7.36	1.85	25	13.7%
34	Toxic	7.33	1.57	18	9.8%
35	Condescending	7.29	1.44	14	7.7%
36	Humiliates others	7.25	1.16	8	4.4%
37	No accountability	7.25	1.22	12	6.6%
38	Power hungry	7.22	1.52	18	9.8%
39	Lacks empathy	7.15	1.29	40	21.9%

Rank	Characteristic	Badness ratings		Frequency	
		M	*SD*	*N*	*Prop*
40	Intrusive	7.14	1.46	7	3.8%
41	Deceitful	7.13	1.36	16	8.7%
42	Narcissistic	7.09	1.50	65	35.5%
43	Troubled Upbringing	7.00	2.00	6	3.3%
44	Greedy	6.94	1.33	35	19.1%
45	Disloyal	6.89	1.82	19	10.4%
46	Controlling	6.85	1.33	34	18.6%
47	Rude	6.83	1.25	130	71.0%
48	Mean	6.81	1.43	91	49.7%
49	Irrational	6.80	0.84	5	2.7%
50	Does not care about others	6.74	1.66	19	10.4%

Table 3

Descriptives and Zero-Order Correlations Between Individual Variables and Badness Ratings of Bad Person Characteristics

	Badness		Moral Foundations						Conservatism		
Item	*M*	*SD*	*Care*	*Equality*	*Prop*	*Loyalty*	*Authority*	*Purity*	*Social*	*Economic*	*BPE*
Child molestation	8.99	.12	.15	.05	.05	.13	.12	.08	-.02	.11	.09
Murder	8.97	.16	.16*	.10	.09	.10	.07	.09	-.15	.01	.05
Rape	8.89	.39	.25**	.29***	.03	-.04	-.05	-.04	-.14	-.22**	.06
Child abuse	8.87	.43	.26**	.18*	-.04	.05	.05	.03	-.06	-.14	.01
Partner abuse	8.69	.63	.30***	.21**	.04	-.01	.01	.02	-.10	-.15	.11
Hitman	8.67	.79	.22**	.01	.00	.11	.06	.15	-.02	-.05	.13
Animal sadism	8.56	1.06	.07	.31***	-.13	-.12	-.13	-.12	-.19*	-.28***	.01
Rape fantasies	8.52	1.03	.20*	.23**	-.01	.08	.05	.07	.09	-.05	.17*
Cannibalism	8.51	1.05	.01	.00	.10	-.01	.09	.06	.08	.08	.21*
Sadism	8.47	.88	.31***	.29***	.02	.03	-.04	.00	-.16	-.20*	.11
Homophobic behavior	8.47	.84	.25**	.27***	-.01	-.11	-.16	-.12	-.30***	-.17*	.00
Racist behavior	8.45	.95	.23**	.31***	.00	-.11	-.13	-.15	-.30***	-.17*	-.04
Racist beliefs	8.40	1.04	.15	.23**	-.08	-.29***	-.30***	-.30***	-.45***	-.18*	.07
Scammer	8.30	1.02	.14	.07	.09	.08	-.05	.02	-.12	-.09	.14
Necrophilia	8.24	1.28	.18*	.20*	.07	.06	-.04	.03	-.07	-.04	.17*
Sexist behavior	8.18	1.18	.16	.24**	-.07	-.20*	-.22**	-.32***	-.43***	-.09	-.03
No guilt	8.14	1.13	.25**	0.18*	.03	.00	-.06	-.03	-.15	-.16	.26*
Homophobic beliefs	8.11	1.29	.19*	.35***	-.08	-.19*	-.22**	-.34***	-.47***	-.20*	.00
Terrorism	8.11	1.20	.14	.03	.01	.05	.06	-.06	.03	.05	.18
Transphobic beliefs	8.09	1.28	.26**	.43***	-.19*	-.27***	-.31***	-.34***	-.51***	-.38***	.05
Pedophilia	8.09	1.36	.14	.11	.11	.06	.05	.06	.08	.01	.29***
Sexist beliefs	8.03	1.26	.35***	.37***	-.04	-.08	-.16*	-.20*	-.41***	-.31***	.03

	Badness		Moral Foundations						Conservatism		
Item	*M*	*SD*	*Care*	*Equality*	*Prop*	*Loyalty*	*Authority*	*Purity*	*Social*	*Economic*	*BPE*
Exploitative	8.01	1.09	.23**	.27***	-.10	.00	-.09	-.09	-.19*	-.26**	-.02
Transphobic behavior	7.97	1.46	.31***	.42***	-.15	-.26**	-.30***	-.33***	-.48***	-.40***	.01
Stalking	7.95	1.11	.18*	.20*	-.06	-.10	-.19*	-.13	-.25**	-.15	.04
Bullying	7.90	1.00	.27***	.22**	.11	.03	-.01	-.04	-.09	-.13	.17*
Money seeking	7.89	1.41	.18*	.30***	-.16*	-.12	-.29***	-.15	-.32***	-.23**	.01
Infidelity	7.71	1.24	.20*	.16*	.10	.13	.07	-.03	-.06	-.09	.01
Controlling	7.65	1.11	.36***	.11	.03	.12	.09	-.02	-.09	-.14	.07
Imperialism	7.55	1.46	.37***	.45***	-.12	-.08	-.18*	-.09	-.30***	-.30***	-.03
Xenophobia	7.49	1.45	.29***	.48***	-.11	-.22**	-.29***	-.20*	-.44***	-.35***	.01
Entitled	7.49	1.35	.24**	.33***	-.16	-.13	-.21**	-.21*	-.45***	-.22**	.05
No empathy	7.40	1.45	.18*	.17*	-.21*	-.02	-.05	.00	-.24**	-.17*	.16
Sexual sadism	7.39	1.66	.09	.10	.02	.08	.08	.02	-.03	.00	.13
Strategic lying	7.16	1.40	.21*	-.06	.06	.18*	.13	.10	.04	-.03	.19*
No accountability	7.14	1.32	.18*	.04	-.01	.00	-.06	-.11	-.18*	-.13	.05
Judgmental	7.01	1.25	.25**	.13	.00	.06	-.02	-.02	-.13	-.17*	.19*
Property damage	6.98	1.26	.09	-.09	.16	.20*	.22**	.21*	.20*	.02	.31***
Infidelity fantasies	6.97	1.52	.12	-.01	.13	.11	.17*	.14	.14	.00	.18*
Spiteful	6.90	1.29	.23**	.09	.13	.21*	.17*	.21*	.11	.02	.12
Disloyal	6.90	1.32	.19*	-.06	.14	.24**	.18*	.06	.11	.04	.11
Only cares about themselves.	6.85	1.35	.27***	.18*	-.05	.09	-.10	-.05	-.20*	-.16*	.10
Hypocrisy	6.84	1.32	.14	.06	.04	.17*	.02	-.09	-.08	.11	.02
Harms environment	6.81	1.57	.30***	.27**	.00	.00	-.13	-.06	-.19*	-.22**	.10
Pathological lying	6.75	1.49	.04	-.14	.09	.12	.08	.16	.14	.06	.35***

	Badness		Moral Foundations						Conservatism		
Item	*M*	*SD*	*Care*	*Equality*	*Prop*	*Loyalty*	*Authority*	*Purity*	*Social*	*Economic*	*BPE*
Arrogance	6.66	1.41	.22**	.13	.08	.09	-.04	-.05	-.08	-.12	.10
Selfish	6.65	1.47	.12	-.05	.06	.18*	.18*	.17*	.09	-.04	.22**
Capital punishment	6.63	1.73	.11	.06	-.10	.06	.00	.05	-.04	-.08	.06
Lawbreaker	6.62	1.52	.06	-.26**	.11	.24**	.22**	.11	.25**	.07	.21*
Ideological violence	6.47	1.63	.17*	.09	-.06	.08	.03	.05	-.02	-.10	.18*
Greedy	6.41	1.46	.18*	.13	-.09	-.06	-.09	-.15	-.20*	-.12	.06
Power seeking	6.26	1.61	.24**	.17*	-.16	-.01	-.04	-.06	-.22**	-.15	.11
No gratitude	6.12	1.45	.15	-.05	.01	.23**	.13	.04	-.02	.02	.25**
Unreliable	6.11	1.34	.08	.09	.07	.08	.13	-.09	-.10	.03	.11
Pyromania	6.05	1.69	.07	-.15	.08	.23**	.21**	.27***	.12	.05	.22**
Gossips	5.95	1.47	.05	-.15	.24**	.28***	.10	.17*	.21*	.09	.23**
Kleptomania	5.95	1.67	-.04	-.32***	.21*	.28***	.32***	.38***	.31***	.19*	.30***
Impulsive	5.70	1.44	-.01	-.11	.08	.05	.09	.10	-.02	.03	.08
Conformity	5.55	1.54	.09	-.03	.00	.03	-.04	-.06	-.06	.02	.14
Reclusive	5.49	1.97	.09	-.06	-.05	.19*	.01	-.03	-.03	-.02	.16
Temper	5.44	1.53	-.03	-.11	-.02	-.04	-.02	.00	-.11	.00	.09
Bragging	5.43	1.40	.15	-.05	.02	.09	.04	-.03	-.15	-.05	.14
Nationalism	5.32	1.90	.20*	.32***	-.13	-.17*	-.33***	-.21*	-.38***	-.21*	.04
Attention seeking	5.12	1.44	.14	-.06	.01	.02	-.04	-.12	-.12	.00	.11
Loses friends	5.11	1.53	.01	-.18*	-.12	-.01	.04	-.04	-.03	-.05	.13
Sexual fetishes	5.07	1.87	-.01	-.06	-.04	.17*	.12	.21**	.03	.11	.23**
Prostitution	4.99	2.06	.04	-.27**	.22**	.33***	.36***	.35***	.30***	.28***	.16
Substance addiction	4.47	1.69	-.13	-.25**	.05	.16*	.26**	.31***	.24**	.24**	.09
Unpatriotic	3.56	1.59	-.14	-.39***	.10	.31***	.29***	.26**	.33***	.30***	.28***
Atheism	2.81	1.75	-.14	-.29***	.04	.26**	.23**	.27***	.34***	.24**	.14

III. STUDY 2: EMPRICALLY DRIVEN CATEGORY AND DIMENSIONAL MODELS OF BAD PERSON CONCEPTS

To learn whether people distinguish between different types of bad people, Study 2 utilized a similarity-sorting task to derive a taxonomy of *bad person* concepts. Previous research had empirically derived a moral person taxonomy based on behaviors alone and found three categories (Landy & Bartels, 2018). The present study expanded on this approach by including (alongside behaviors) perceived beliefs, desires, and emotions of the targets, based on the characteristics provided in Study 1. To identify candidates for the basic categories of bad people in the given population and determine which traits were perceived to be most central to each category, I conducted multiple types of clustering analysis. To reduce and interpret the dimensions people use to distinguish immoral person characteristics from each other, I conducted a dimensional reduction analysis. The clustering and dimensional reduction analyses generated categorical and dimensional models of moral characteristics that were then tested in the final study.

Method

Participants

One-hundred and thirty-one undergraduate students were recruited from the University of Oregon Psychology/Linguistics Human Subjects pool for a 1-hour in-person study and were compensated with course credit. Ten participants were removed from the sample for taking the survey in less than 17 minutes, which was determined to be too quick to read every item. The sample size was determined a priori based on the recommendation of at least 20 participants per expected cluster, though a smaller sample is needed for more separate clusters and for the fuzzy

clustering algorithm (Dalmaijer et al., 2022). Past research using this method has used a comparable sample size or smaller (Brewer et al., 1981; Pilkonis, 1988; Walker & Pitts, 1998).

The sample demographics were similar to Study 1. Students ranged from 18 to 28 (M = 19.00, SD = 1.39) and were majority white, with 60.3% reporting as White, 17.6% reporting as Hispanic, 12.4% reporting as Asian, 7.4% reporting as Black, and 2.3% reporting as biracial or "other." Additionally, 65.3% of participants reported that they identified as a woman and 33.1% reported that they identified as a man, and 1.6% identified as nonbinary or "other." The sample leaned liberal, with about 61.2% reporting as politically liberal and 5.8% reporting as politically conservative.

Procedure and Materials

Participants completed a computerized similarity-sorting task on large monitors in a computer lab[5]. When the participants arrived at the laboratory in small groups, the experimenter demonstrated the sorting task with physical cards and described the task. Participants were told that the task measured how many types of people they perceive, and that they would be sorting negative characteristics into piles, such that each pile represented a different type of person. The experimenter refrained from describing the traits using moral descriptors (e.g., bad, evil,

[5] The similarity-sorting task was computerized by using Pick, Group, & Rank questions on Qualtrics, with custom javascript to save the participants' categorization solution to embedded survey data and to hide groups until participants clicked on an 'add a pile' button so that participants could choose how many groups to make. Custom css script enhanced the user experience of the task and emulated the experience of the original task with physical cards.

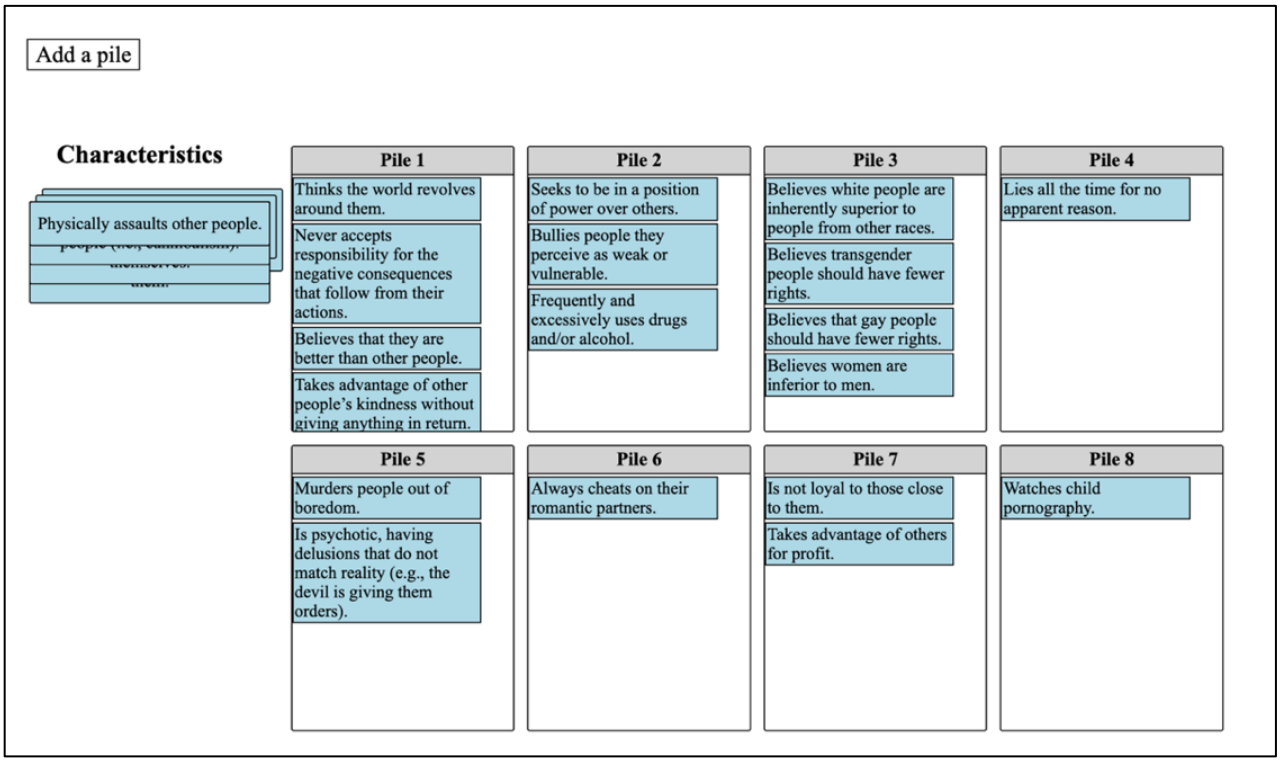
immoral) as to not bias the participants' responses. Participants were directed to make as many piles as they saw fit, with as many characteristics per pile as they saw fit.

Participants first completed a practice trial to adjust to the task where they sorted 10 animal characteristics (e.g*., excellent swimmers*, *four limbs*, *can fly*) into piles "based off whether [they] think that the characteristics would normally show up in the same type of animal." Once they sorted all the characteristics, they were given an example of a three-category solution and told that they may have had more or fewer categories, depending on how they defined their categories.

On the next page, participants read a short description of the study that repeated the researcher's verbal instructions to motivate them to complete the very cognitively demanding task and orient them to the main study. On the following page, the screen displayed a pile of 50 blue rectangles representing index cards with negative moral person attributes printed on them (generated in Study 1), one gray box labeled "Pile 1," and a button that said, "add a pile" (Figure 1). The blue cards were "shuffled" and laid on top of each other such that each participant saw the cards in random order and could drag them into piles one at a time. The computers were large enough that participants could create at least 20 piles comfortably without needing to scroll. If a pile had more items than could be viewed, participants were directed to hover over the box and scroll to review the items. When all cards were sorted, they were able to click to the next page after agreeing that they had reviewed their piles, and they felt confident that each represented a different type of person they perceive.

Figure 1

Computerized Similarity-Sorting Task



Next, participants were shown a list of the characteristics they chose for their *first pile* and were asked if they could further divide the pile into two or more piles, such that each pile represented a type of person. If they indicated yes, they were displayed a second screen that allowed them to do a second sort, such that the characteristics from the first pile were located to the left of the screen on blue cards, and participants could create as many new piles as they saw fit and put as many cards in each new pile as they saw fit. Then, on a new page, participants were displayed a screen with their Pile 1 characteristics printed and the fill-in-the blank question "What would you call someone who had all of these characteristics?" On the following page, participants were asked to generate new characteristics for their pile. This task was repeated until all of their piles (if desired) were further divided into more piles.

Then, a page was displayed to participants that printed their initial piles, including the contents and the name they gave each pile. They were asked to consider whether any of their original piles could be combined to create a bigger pile that represents a type of person. Then, they completed a third sorting task with the names of their groups on blue cards, with the ability to create as many piles as they saw fit, with as many initial groups per pile as they saw fit.

Finally, participants rated all 50 characteristics based on "the extent to which you would consider someone a bad person if they expressed the characteristic" on a scale that ranged from 1 (*definitely not a bad person*) to 9 (*definitely a bad person*) and answered standard demographics.

Results

Number of Groups

On average, participants created 6.74 (*SD* = 2.84) piles in the initial sort, with a minimum of 2 piles and a maximum of 18 piles. The distribution was skewed to the right, suggesting that the typical person initially distinguished between 5 to 7 types of people from the characteristics given (Figure 2). Further, after giving participants the opportunity to further divide their initial piles, they created on average of 9.26 (*SD* = 4.52) groupings, with a minimum of 3 piles and a maximum of 23 piles. When given the opportunity to combine their initial piles, they created on average 3.55 (*SD* = 1.42) groupings, from a minimum of 1 pile to a maximum of 8 piles.

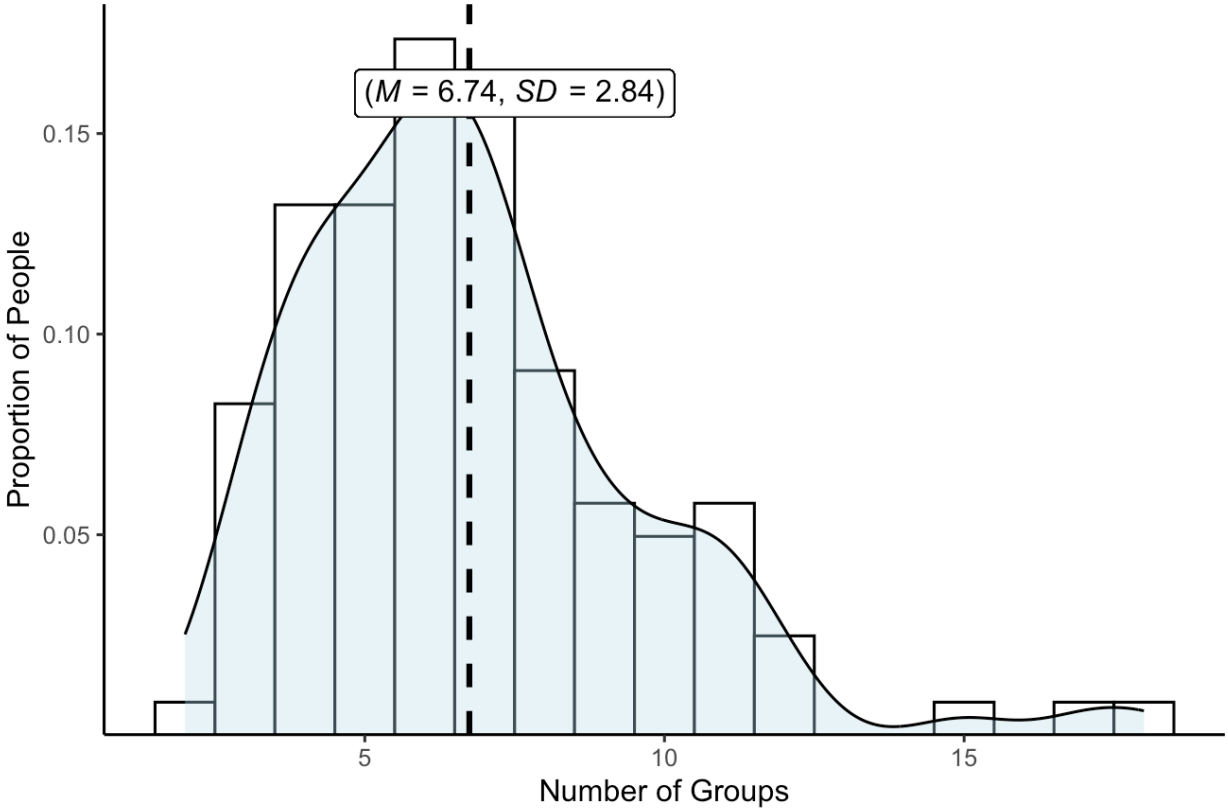
Group Names

For each initial group, participants were prompted to generate a label for the pile based on what they would call someone who had all characteristics listed. The names were cleaned by removing any modifiers or articles (e.g., "a very evil psychopath" turned into "psychopath") and correcting spelling errors. If a participant wrote several names (e.g., "a racist, homophobe, and

transphobe" or "a narcissistic psychopath") or gave a paragraph describing the person, the name was counted as a unique entry. Overall, the most common category names were "narcissist," "psychopath," and "abuser," and each were used by about a third of the participants (Table 4).

Figure 2

Distribution of Groups Created in Initial Sort



Similarity Matrix

Next, I created a 50 x 50 triangular similarity matrix where each element represents the number of participants who grouped two characteristics in the same pile, such that high numbers indicated that two characteristics were frequently in the same pile and 0 indicated that two characteristics were never put in the same pile. I also created a total similarity matrix that sums the frequencies from the initial and secondary sorts, removing any redundant entries (i.e.,

participants who chose not to split or sort a pile). Next, I created a distance matrix from the similarity matrices using the `dist` function from the `stats` package in R, where distance between each item pair was modeled by Euclidean distance.

I ran the analyses with both the distance matrix from the initial sort and distance matrix from the total sort. However, the results were extremely similar, and therefore I am only reporting the results using the similarity matrix from the initial sort.

Table 4

Most Frequent Labels Given to Describe Initial Groupings

Label	Frequency	Percent
Narcissist	41	33.9%
Psychopath	37	30.6%
Abuser	36	29.8%
Addict	20	16.5%
Conservative	20	16.5%
Psychotic	20	16.5%
Selfish	15	12.4%
Pedophile	14	11.6%
Manipulator	12	9.9%
Bigot	9	7.4%
Cannibal	9	7.4%
Bully	8	6.6%
Criminal	8	6.6%
Greedy	8	6.6%
Sociopath	8	6.6%
Dictator	7	5.8%
Insecure	7	5.8%
Mentally ill	7	5.8%
Predator	7	5.8%
Creep	6	5.0%
Entitled	6	5.0%
Self-centered	6	5.0%
Liar	5	4.1%
Racist	5	4.1%
Sadist	5	4.1%

Deriving a Category Model of Bad Person Concepts

Hierarchical Cluster Analysis. Agglomerative hierarchical cluster analysis clusters items into groups by identifying items that are close together based on the imputed distance matrix, and then merges the nearest items iteratively until all items are part of the same cluster. The analysis results in a hierarchical structure of groups that can be modeled by a dendrogram, such that the y-axis represents the Euclidean distance between clusters.

For the hierarchical cluster analysis, I used the `hclust` function from the `stats` package in R and used Ward's minimum variance method to link clusters. Ward's method prioritizes minimizing the within-cluster variance when merging two clusters together, which results in more compact clusters compared to other popular methods like single, complete, or average linkage. The `agnes` function from the `cluster` package calculated that the Ward's method had the highest agglomerative coefficient for the data, which indicates that it produced the strongest clustering structure (i.e., within-cluster similarity is high and between-cluster similarity is low) compared to the single, complete, and average linkage methods.

Next, to determine the optimal number of clusters, I used the `fviz_nbclust` function from the `factoextra` package in R to model the within-cluster variance (i.e., WSS) and between-cluster variance (i.e., silhouette width). The within-cluster sum of squares measures how compact each cluster is and can be visualized by an elbow plot. The plot demonstrated that there are three tightly knit clusters, and that additional clusters do not provide a dramatic reduction in variance (Figure 3). The between-cluster variance can be modeled by average silhouette width, which is calculated by averaging the silhouette width for each item. Silhouette width values that are close to one indicate that the item is well suited for its own cluster and is far from neighboring clusters. The plot of average silhouette width by cluster agreed that three was the

optimal number of clusters, such that additional clusters decreased the between-cluster variance (Figure 4).

Figure 3

Elbow Plot of Within-Cluster Variance for Initial Sort

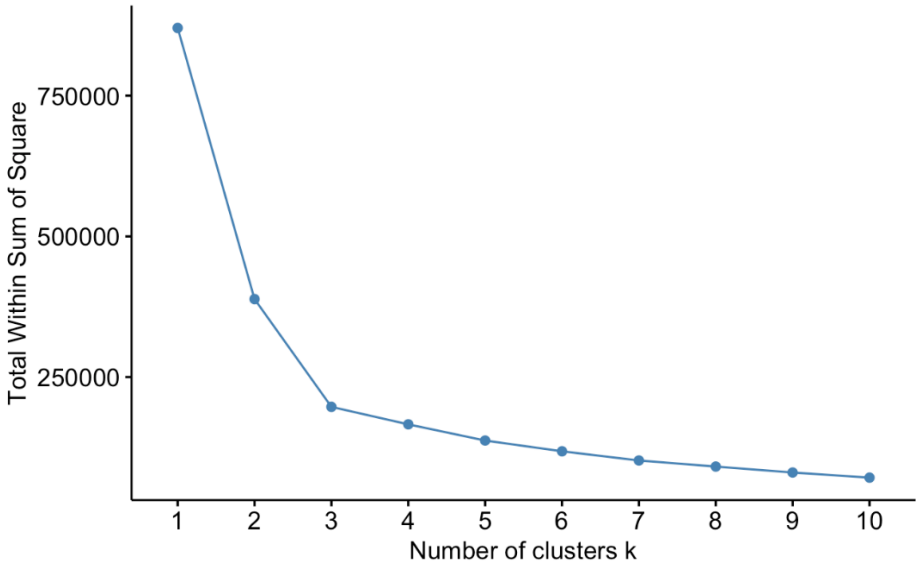
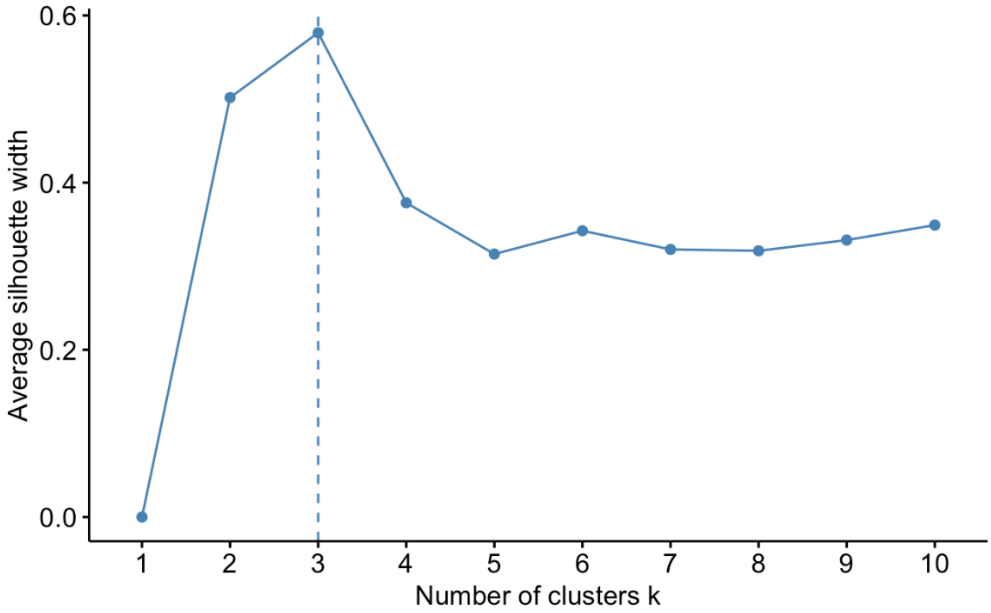


Figure 4

Plot of Average Silhouette Width for Initial Sort



Even though the three-cluster solution minimizes within-cluster variance and maximizes between-cluster variance, that does not necessarily mean that people perceive only three types of bad people and do not make further meaningful distinctions; however, it does provide strong support for the hypothesis that people perceive more than one type of bad person. Given that most participants distinguished between 5 to 7 types of people in their initial sort, I also examined larger solutions, as these solutions are expected to better represent natural, everyday distinctions. For visualizing the cluster solutions with a dendrogram, I used the `fviz_dend` function from the `factoextra` package in R.

The three-cluster solution distinguished between three types of people: *a violent person*, *a bigoted person*, and *a selfish person* (see Figure 5). The seven-cluster solution distinguished between seven types of people: *a psychopath*, *a criminal*, *an abuser*, *a bigot*, *a bully*, *a narcissist*, and *a manipulator* (see Figure 6). In the ten-cluster solution, *pedophile* broke off from *abuser*, addict broke off from *bully*, and *scammer* (i.e., manipulation for money) broke off from *manipulator* (see Figure 7). The names from the seven-cluster and ten-cluster solutions were all frequently given by participants to describe their initial groupings (except for *scammer*, which participants may have described instead as *greedy*), reflecting that these category distinctions may be naturally used, or *basic categories.* Overall, I determined that the seven-cluster solution appeared to best model the natural category distinctions, as it was the most interpretable and the best fit for the generated names. The category structure will be tested on new data in Study 3.

Figure 5

Dendrogram of Hierarchical Cluster Analysis from Initial Sort Using Ward's Method (k = 3)

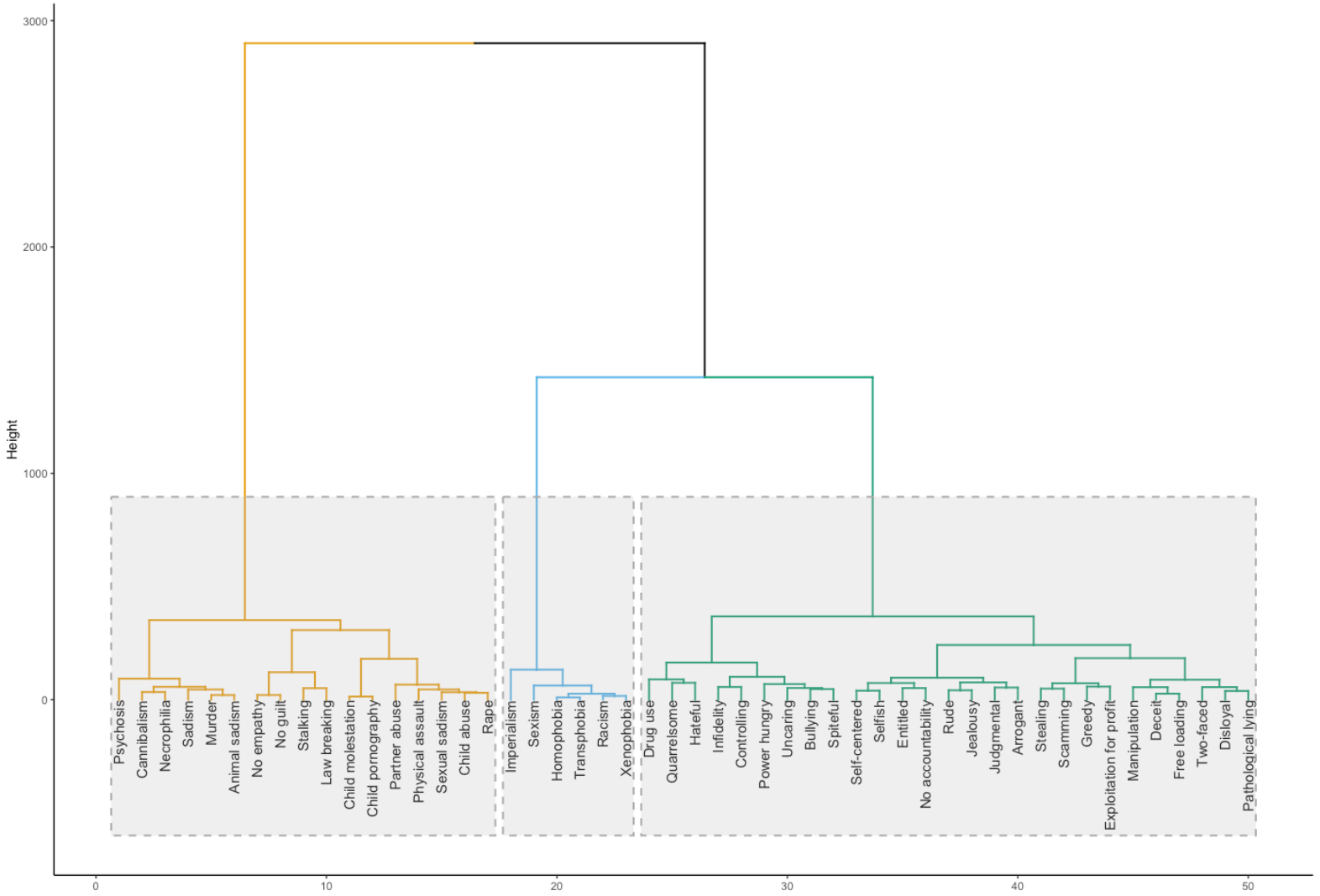


Figure 6

Dendrogram of Hierarchical Cluster Analysis from Initial Sort Using Ward's Method (k = 7)

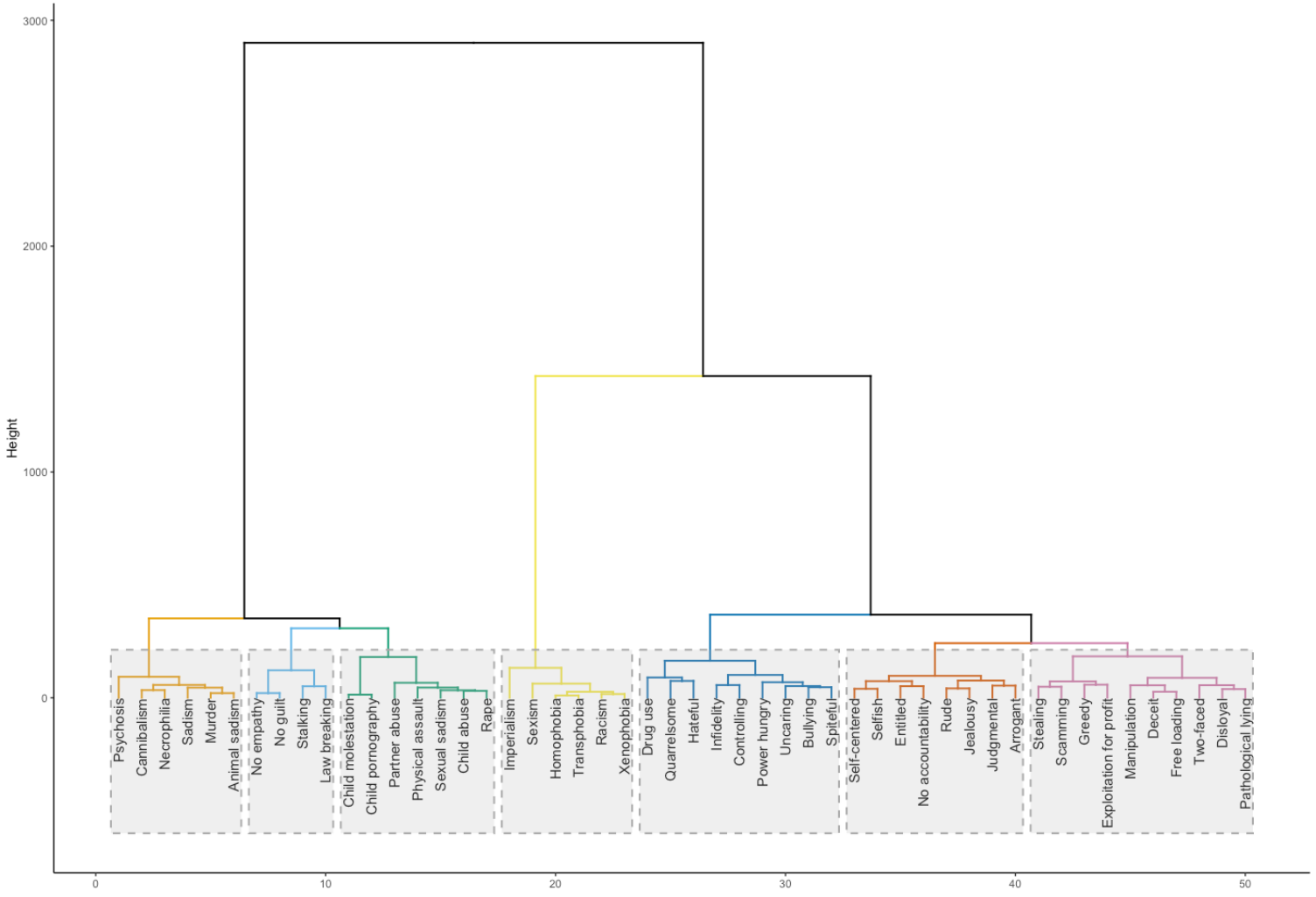
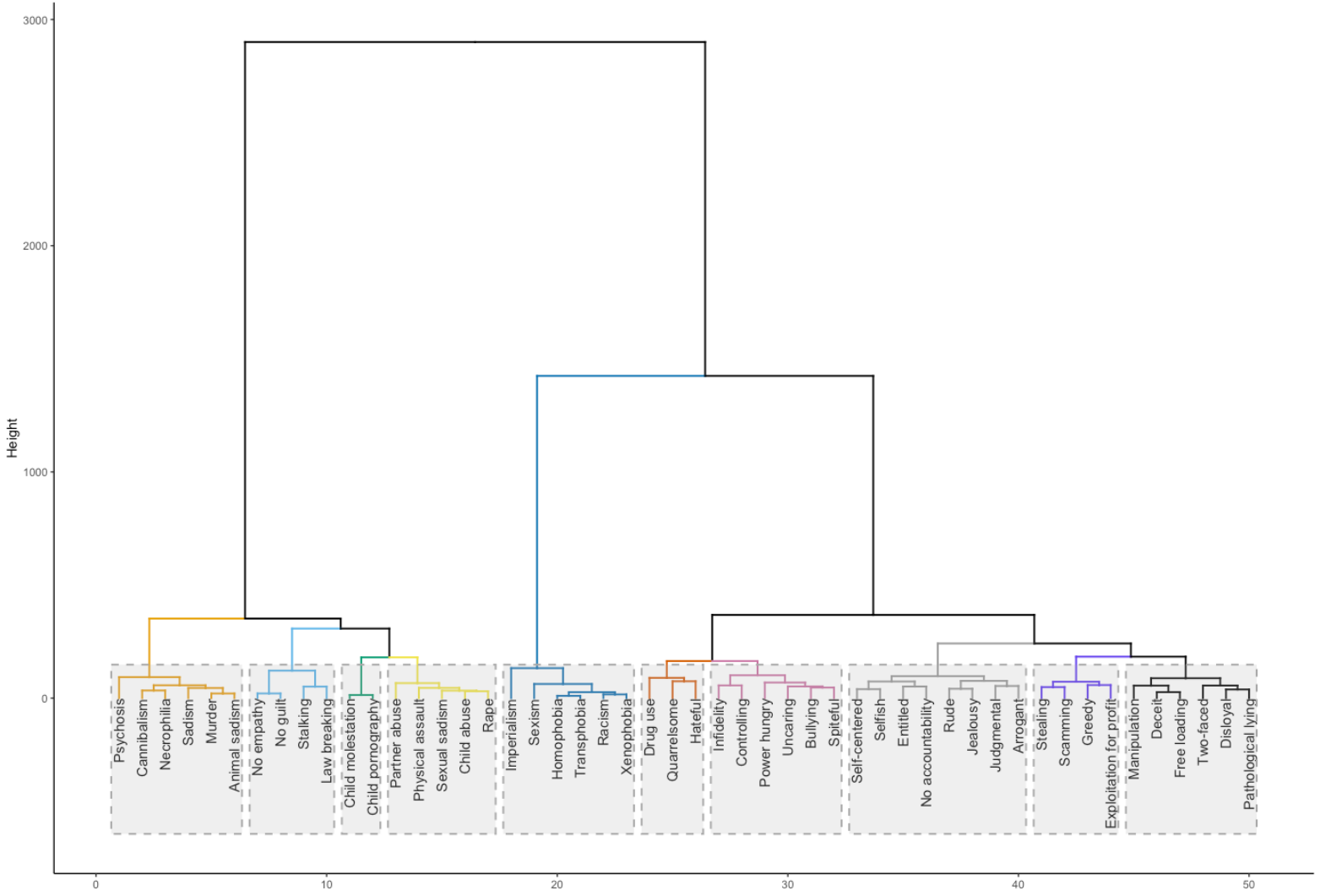


Figure 7

Dendrogram of Hierarchical Cluster Analysis from Initial Sort Using Ward's Method (k = 10)



Categorization Agreement. Next, I ran a descriptive analysis to examine the extent to which representations of different types of bad people are shared. By relying solely on aggregate-level statistics, we may misunderstand the data if a pattern emerges at the group-level but not at the individual level. In other words, the mental representation of a type of bad person could emerge at the aggregate level, even though individual participants were unlikely to create the category in the sorting task. We cannot rely on the labels that participants gave the categories to determine category agreement, as two participants who have the same name for a category may not have the same mental representation of that category, and two people could have the same mental representation of a category but name it differently.

I evaluated clustering agreement for the seven-cluster solution by describing the maximum number of cards from a given cluster that participants put in the same group (see Figure 8 for distributions). For example, if a participant put all six psychopath traits from the seven-cluster solution in the same group in their initial sort, then they received a score of "6" for psychopath agreement[6]. If all psychopath traits were put in different piles, then the participant received a score of "1" for psychopath agreement. I found that participants had the highest agreement for the bigot, psychopath, and abuser clusters. The bigot cluster achieved 78% agreement, the psychopath cluster achieved 59% agreement, and the abuser cluster achieved 37%

[6] The number of cards was used in this analysis instead of percent agreement to make it clear that some groupings were larger than others. The size of the group is important to consider when determining which groups were most reliably formed. For example, there is more evidence for a prototype if 90% of people sorted all the cards in the same pile if there were 10 cards compared to if there were 3 cards representing that prototype.

agreement, when agreement is defined by participants sorting all or all minus one of the cards from the category solution in the same pile. These agreeability ratings may reflect that traits of a prototype that are less central may be consistent with several prototypes. For instance, many different types of bad people are perceived to lack empathy, and thus, the lacks empathy card was not consistently sorted into the same pile.

Fuzzy C-Means Analysis. To determine which characteristics are more central to each cluster, I next ran a fuzzy c-means analysis, which is a variant of a k-means analysis. A k-means analysis requires an input of number of clusters, and then the algorithm iteratively places and then updates the location of the cluster centroids until the algorithm converges on a solution. A fuzzy c-means analysis is similar to a k-means analysis, but items can be assigned to more than one cluster with varying degrees. Thus, a fuzzy c-means analysis will allow the bad person attributes to be characteristic of more than one category and provide insight into which characteristics are the most central to the prototype. I chose to allow characteristics to be members of multiple clusters because that is consistent with the mental representation of natural categories (Medin & Schaffer, 1978; Medin & Smith, 1984; Rosch, 1978).

I conducted a fuzzy c-means analysis using the `fcm` function in the ` ppclust` package in R. Since the location of the initial centroids can influence the final membership values, I set the parameter `nstarts` to 20, which runs the algorithm with 20 different start locations and chooses the most optimal solution. Twenty starts was chosen because it created a stable solution while taking a reasonable amount of time to compute. Table 5 shows the membership values of each characteristic across the seven clusters. Membership values close to one are extremely close to the centroid of the cluster, indicating that they are a central aspect of the prototype, whereas values close to zero are very far away from the centroid of the cluster.

Notably, some characteristics belong to multiple categories with varying degrees. For example, someone who "cannot resist urges to steal items they don't need" is close to the *bully*, *manipulator*, and *narcissist* centroids to varying degrees, but it is relatively far away from any given centroid compared to other characteristics. Additionally, "wants to invade, occupy, and conquer other countries" is very weakly related to the *bigot* and the *bully*, suggesting that it may be a better fit for its own category that was not detected in this study. *Dictator* was a popular category label that would fit this item, and if there had been more prototypical dictator items in the pool (e.g., "leads a genocide"), we may have seen a dictator category emerge.

Figure 8

Sorting Agreement for Seven-cluster Solution

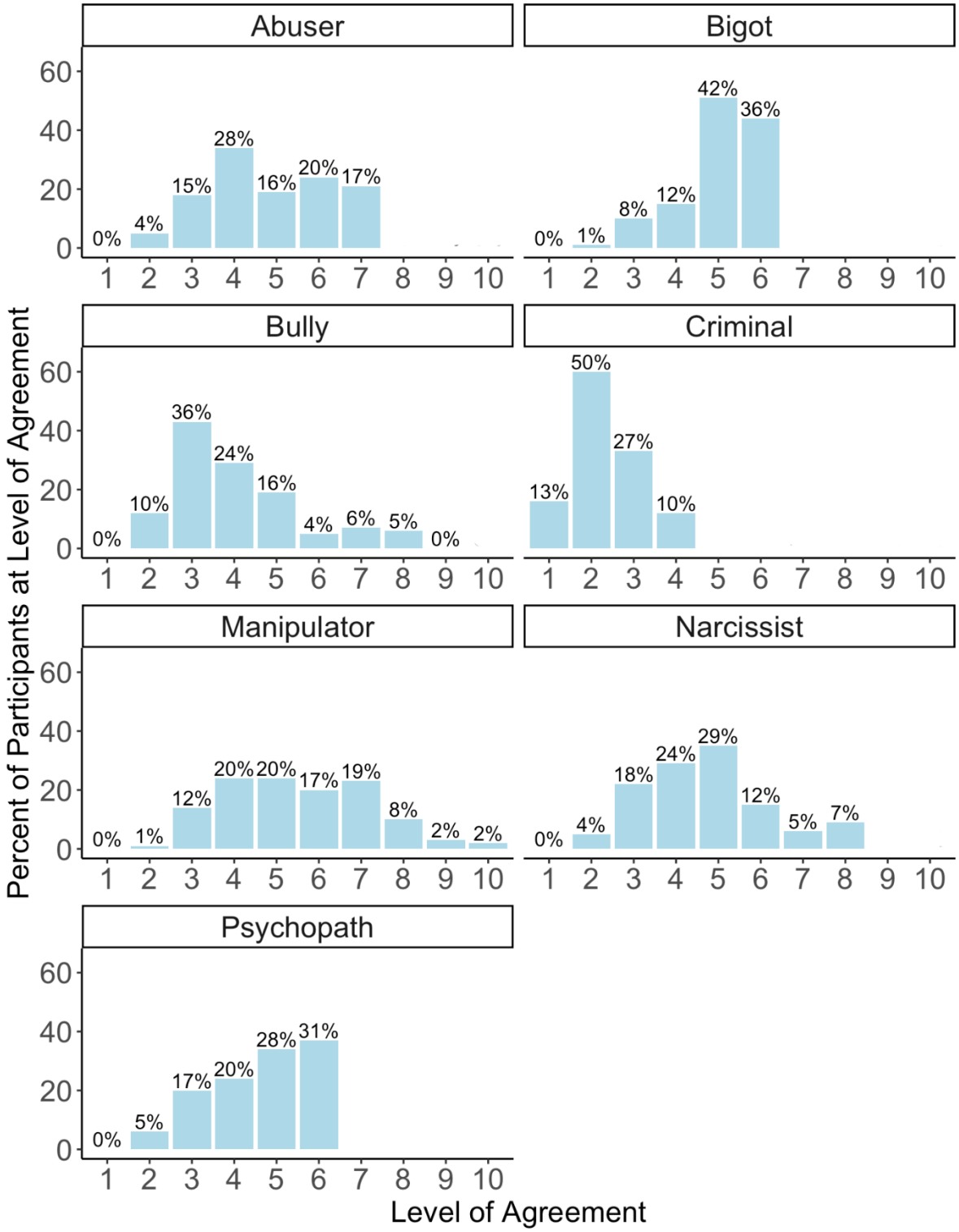


Table 5
Membership Values from C-Means Analysis

	Cluster Name						
Characteristic	Narcissist	Criminal	Bigot	Bully	Psychopath	Abuser	Manipulator
Self-centered	**0.75**	0.01	0.01	0.06	0.00	0.01	0.16
Judgmental	**0.73**	0.01	0.01	0.07	0.00	0.01	0.17
Selfish	**0.72**	0.01	0.01	0.05	0.00	0.00	0.20
Arrogant	**0.63**	0.02	0.02	0.10	0.01	0.01	0.21
Two-faced	**0.58**	0.01	0.01	0.06	0.00	0.01	0.33
Rude	**0.58**	0.01	0.01	0.10	0.00	0.01	0.29
Jealousy	**0.50**	0.01	0.01	0.08	0.00	0.00	0.40
Law breaking	0.00	**0.96**	0.00	0.01	0.01	0.01	0.01
Stalking	0.01	**0.94**	0.00	0.01	0.01	0.02	0.01
No guilt	0.03	**0.48**	0.02	0.05	0.10	0.28	0.03
No empathy	0.03	**0.46**	0.02	0.05	0.11	0.30	0.03
Drug use	0.08	**0.46**	0.04	0.22	0.04	0.07	0.10
Racism	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
Xenophobia	0.00	0.00	**0.99**	0.00	0.00	0.00	0.00
Homophobia	0.00	0.00	**0.99**	0.00	0.00	0.00	0.00
Transphobia	0.00	0.00	**0.98**	0.00	0.00	0.00	0.00
Sexism	0.04	0.04	**0.80**	0.04	0.02	0.03	0.04
Imperialism	0.17	0.11	**0.25**	0.20	0.04	0.06	0.16
Controlling	0.06	0.03	0.01	**0.79**	0.01	0.01	0.09
Spiteful	0.08	0.02	0.01	**0.75**	0.00	0.01	0.13
Uncaring	0.07	0.02	0.01	**0.77**	0.00	0.01	0.12
Power hungry	0.14	0.02	0.01	**0.61**	0.01	0.01	0.20
Quarrelsome	0.10	0.11	0.02	**0.59**	0.02	0.03	0.13
Infidelity	0.14	0.02	0.01	**0.50**	0.01	0.01	0.31
Stealing	0.14	0.03	0.01	**0.47**	0.01	0.01	0.32
Bullying	0.17	0.02	0.01	**0.46**	0.01	0.01	0.33
Hateful	0.10	0.26	0.04	**0.38**	0.03	0.05	0.13
Necrophilia	0.00	0.01	0.00	0.00	**0.97**	0.02	0.00
Murder	0.00	0.01	0.00	0.00	**0.95**	0.03	0.00
Animal sadism	0.00	0.01	0.00	0.00	**0.95**	0.03	0.00
Cannibalism	0.00	0.01	0.00	0.00	**0.94**	0.03	0.00
Sadism	0.01	0.03	0.01	0.01	**0.80**	0.15	0.01
Molest Child	0.01	0.04	0.01	0.01	**0.67**	0.24	0.01
Child Porn	0.01	0.04	0.01	0.01	**0.64**	0.27	0.01
Physical assault	0.00	0.01	0.00	0.00	0.02	**0.96**	0.00
Sexual sadism	0.00	0.01	0.00	0.00	0.03	**0.96**	0.00
Child abuse	0.00	0.01	0.00	0.00	0.03	**0.95**	0.00
Rape	0.01	0.03	0.01	0.01	0.11	**0.84**	0.01

	Cluster Name						
Characteristic	Narcissist	Criminal	Bigot	Bully	Psychopath	Abuser	Manipulator
Partner abuse	0.02	0.13	0.02	0.02	0.08	**0.72**	0.02
Psychosis	0.02	0.12	0.02	0.02	0.23	**0.57**	0.02
Pathological lying	0.15	0.01	0.01	0.08	0.00	0.00	**0.74**
Manipulation	0.14	0.01	0.01	0.15	0.00	0.01	**0.68**
No accountability	0.25	0.01	0.01	0.09	0.00	0.00	**0.64**
Disloyal	0.31	0.01	0.01	0.08	0.00	0.00	**0.59**
Greedy	0.24	0.01	0.01	0.15	0.00	0.01	**0.57**
Exploitation	0.27	0.02	0.01	0.12	0.01	0.01	**0.56**
Entitled	0.38	0.01	0.01	0.09	0.00	0.00	**0.51**
Deceit	0.38	0.01	0.01	0.08	0.01	0.01	**0.50**
Free loading	0.39	0.02	0.01	0.08	0.01	0.01	**0.48**
Scamming	0.17	0.04	0.02	0.36	0.01	0.01	**0.39**

Note. The bold text indicates that the characteristic's membership value was highest for the cluster represented by the column. For instance, the characteristic *self-centered* had the highest membership value for the *narcissist* cluster.

Deriving a Dimensional Model of Bad Person Concepts

Multidimensional scaling analysis represents distances between items by modeling those distances in a dimensional space. The algorithm reduces dimensionality down to a researcher-identified number of dimensions, while preserving the distances represented in the original distance matrix. A scree plot of eigenvalues indicated that a three-dimensional solution fit the data best, as additional dimensions only provided marginal additional explanation of variance in the pairwise distances. The first dimension explains the most variance between pairwise distances, followed by the second, then the third. The ` seqmds.stress` function from the ` seqhandbook` package yielded a stress value of .11, indicating an adequate overall fit of the three-dimensional solution (Wickelmaier, 2003). In contrast, the one-dimensional solution had a stress value of .41 and the two-dimensional solution had a stress value of .15.

The multidimensional scaling analysis provided coordinates in three-dimensional space for each immoral characteristic while maintaining the original distances generated from the first round of the sorting task. In the three-dimensional space, small distances between points indicate that two characteristics are perceived to be likely to occur in the same person, as they were often sorted into the same pile. In contrast, two characteristics that are far away are perceived to be unlikely to occur in the same type of person, as they were often sorted into different piles.

In this context, the dimensions can provide insight into *why* certain immoral characteristics were correlated or uncorrelated. In other words, interpreting the dimensions can explain why certain immoral characteristics are perceived differently from other immoral characteristics. The indication that the best fit solution was three-dimensional implies that there are more dimensions relevant to the perception of bad people than a single moral valence dimension that is used to measure moral character in the literature.

It may be difficult to interpret the dimensions of a multidimensional scaling analysis and there is a strong risk of overinterpretation (Borg et al., 2013). For example, the solution may need to be rotated from the output given to meaningfully interpret the axes. However, the three dimensions commonly found in studies of person perception (Brambilla et al., 2021; Brambilla & Leach, 2014; Goodwin, 2015; Leach, 2007): morality, competence, and sociability fit these data reasonably well. Characteristics on one end of the first dimension, e.g., murder, rape, are perceived as more immoral than characteristics on the other end, e.g., arrogant, judgmental (Figure 9). Further, characteristics on one end of the second dimension are perceived as requiring high intelligence and skill, e.g., deceit, manipulation, while characteristics on the other end are perceived as born of ignorance and stupidity, e.g., racism, homophobia (Figure 9). Finally, characteristics located on one extreme of the third dimension may be perceived as descriptive of

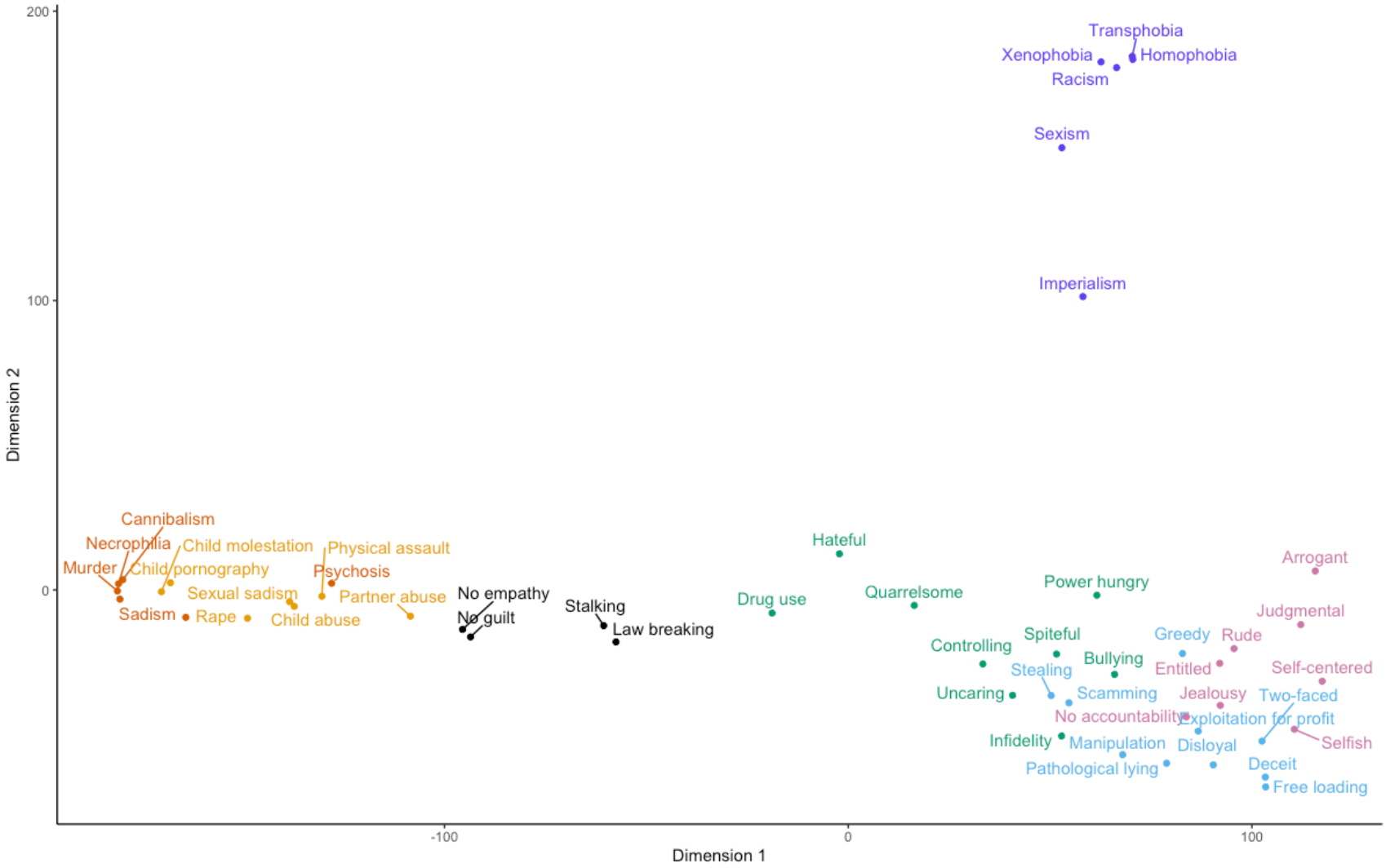
someone who is highly sociable and integrated into society, e.g., substance abuse, infidelity, partner abuse, while traits on the other end may be perceived as characteristics of loners or lunatics who are not integrated into society, e.g., cannibalism, necrophilia (Figure 10).

Discussion

In Study 2, participants completed a computerized similarity-sorting task, where they grouped 50 negative characteristics into piles representing different types of people. I examined several solutions and determined that a seven-cluster solution best modelled natural distinctions, revealing that the population of interest perceived seven types of bad people: *psychopaths*, *criminals, abusers, bigots, bullies, narcissists*, and *manipulators*. Notably, there was the most consensus amongst respondents in identifying *psychopath, bigot*, and *abuser* clusters. Thus, these results suggest that there are more perceived categories of *bad people* than have been previously found from behaviors alone (k = 3; Landy & Bartels, 2018) or that we might have expected based on theories from contemporary moral (e.g., Moral Foundations Theory, k = 5 or 6; Atari et al., 2022; Graham et al., 2013) or personality (e.g., The Dark Triad or Tetrad, k = 3 or 4; Mededović & Petrović, 2015; Paulhus & Williams, 2002) psychologists, and fewer than the 30 types of flawed individuals that Theophrastus wrote about two millennia ago.

Figure 9

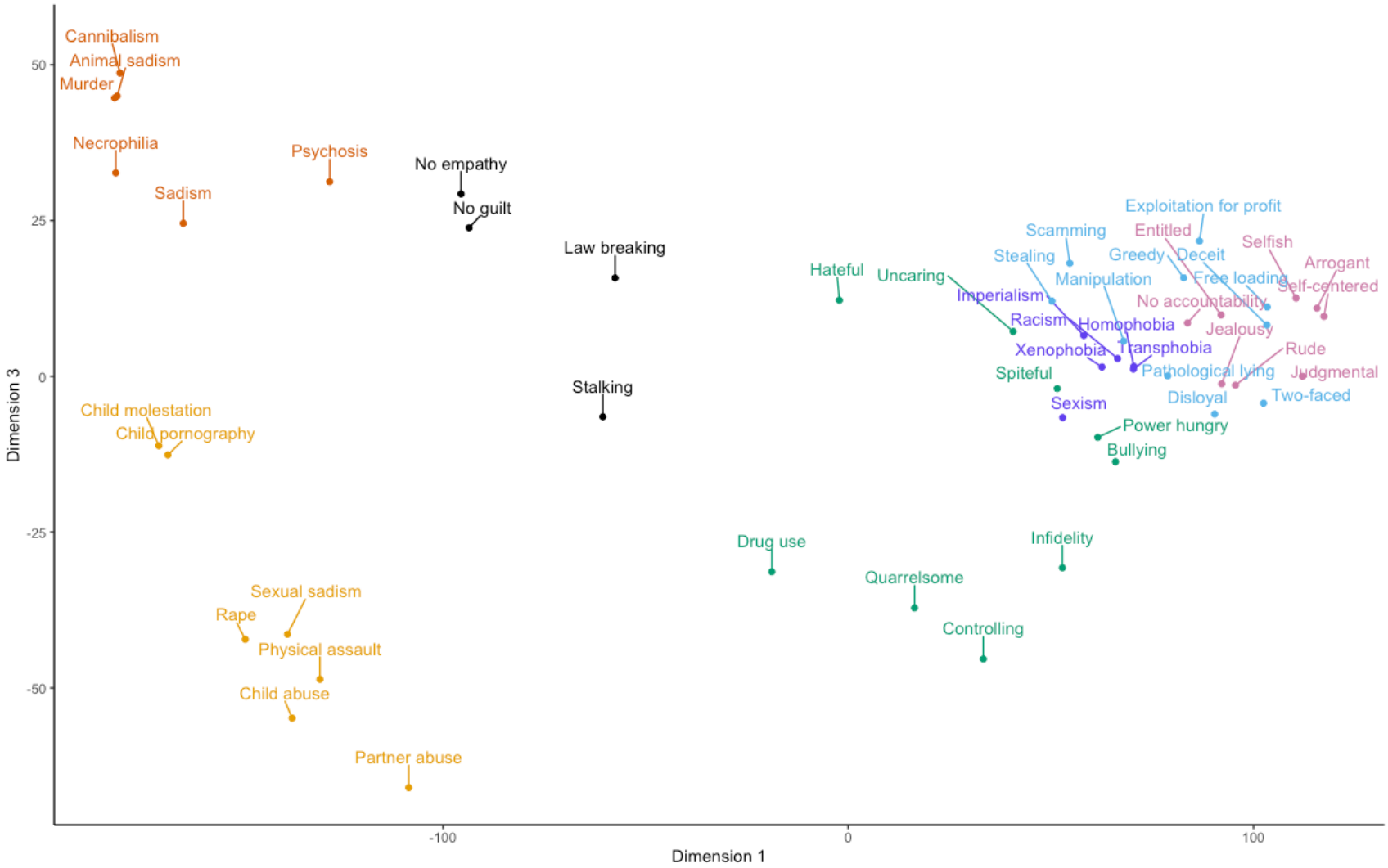
Three-Dimension Solution from Multidimensional Scaling Analysis (Dimension 1: Morality, Dimension 2: Competence)



Note. Each color represents a different cluster from the seven-cluster solution.

Figure 10

Three-Dimension Solution from Multidimensional Scaling Analysis (Dimension 1: Morality, Dimension 3: Sociability)



Note. Each color represents a different cluster from the seven-cluster solution.

The Seven Deadly "Sinners" came closest to the *number* of perceived bad person categories found in this study, though the content is a poor match. Rather, the types of categories formed appear to be based more on the contemporary media that the participants have consumed rather than on the reality of underlying personality structure. For instance, characterizations of the psychopath were weirder than how modern personality psychologists define psychopathy (Dolan & Fullam, 2006; Hare et al., 1990; Holmqvist, 2008) and sadist (the fourth trait in the dark tetrad) did not emerge as a category (Mededović & Petrović, 2015).

Additionally, multidimensional scaling analysis was conducted to generate a dimensional model of the immoral characteristics. The analysis identified that a three-dimensional model best fit the data. I interpreted the dimensions as the three dimensions of person perception -- morality, competence, and sociability – implying that these are the three most important social judgments for distinguishing bad people from each other. An alternate interpretation is that the third dimension may reflect "weirdness" or "insanity," where cannibalism and necrophilia are considered acts from people without sound mind, and partner abuse and rape are committed by those with full control over their actions. Weirdness is expected to be highly correlated with sociability, and therefore I chose to interpret the dimension in the way that is consistent with the vast literature on person perception (Brambilla et al., 2021; Brambilla & Leach, 2014; Goodwin, 2015; Leach, 2007). I tested the fit of the morality, competence, and sociability dimensions with new data in the third and final study.

IV. STUDY 3: TESTING MODELS OF IMMORAL INFERENCES

The purpose of the final study is to determine whether a prototype model, dimensional model, or dual prototype-dimensional model can best explain how people infer that someone possesses immoral characteristics that were not observed. Additionally, I tested the category structure and dimensions derived in the previous study and examined whether exemplars from four moral person categories derived in the previous study vary in perceived morality, competence, sociability, gender, race, age, physical appearance, ability to change their immoral characteristics, and how comfortable respondents are interacting with them. The aim of this final study is to examine the effect that prototypes have on moral person perception and to examine the properties of four of the most consistent immoral prototypes that emerged in the population of interest: *psychopaths, abusers, bigots*, and *narcissists*.

To test a prototype model, I examined whether exemplars are more likely to be inferred to have unobserved characteristics when the characteristics are prototypical of the exemplar's moral category compared to when the characteristics are prototypical of different moral person categories. My hypotheses follow from prototype theory, which states that category members are transferred unobserved characteristics prototypical of the category (Cantor & Mischel, 1977; Osherson et al., 1990; Rosch, 1978; Smith & Zarate, 1992).

H1: I hypothesize that the superordinate categories *good person* and *bad person* will act as prototypes, such that perceivers will be more likely to infer that targets have unobserved immoral characteristics if the targets are initially described as having immoral characteristics compared to when they are initially described as having moral characteristics.

H2: I hypothesize that the basic categories will act as prototypes, such that perceivers will be more likely to infer that targets have immoral characteristics consistent with a basic

category (e.g., abuser) if the target was described as prototypical of that category compared to prototypical of competing basic categories (e.g., narcissist, bigot, and psychopath).

Next, I tested dimensional models by examining whether characteristics are more likely to be inferred to an exemplar if the characteristic is close in proximity to the target on one (unidimensional model) or multiple (multidimensional model) trait dimensions. For example, a unidimensional model would predict that an exemplar is more likely to be inferred an immoral characteristic when the characteristic and exemplar are close in proximity on the moral dimension (e.g., both the characteristic and exemplar are considered "moderately immoral"), compared to when the characteristic and exemplar are far apart on the moral dimension.

H3: I hypothesize that the Euclidean distance between the morality of the exemplar and the morality of the unobserved characteristic predicts the likelihood that the exemplar will be inferred to possess the unobserved characteristic (a unidimensional model).

H4: I hypothesize that the Euclidean distance between the morality, competence, and sociability of the exemplar and the morality, competence, and sociability of the unobserved characteristic predicts the likelihood that the exemplar will be inferred to possess the unobserved characteristic (a multidimensional model).

Additionally, I examined whether there was evidence that both prototypes and dimensions (i.e., a mixed model) uniquely predict inferences about immoral characteristics. To test which model performed best at explaining why exemplars are inferred immoral characteristics, I compared model fit estimates that were corrected (R^2) and uncorrected for parsimony (e.g., AIC and BIC). I hypothesized that all models would perform better than the unidimensional model found in the literature and that the multidimensional-prototype dual model would perform best, as it was the most complex, where characteristics are inferred when they are

prototype consistent and when they are close to the target on morality, competence, and sociability (even when they are not consistent with the prototype).

Method

Participants

Three hundred and twenty-two participants were recruited from the University of Oregon Psychology/Linguistics Human Subjects Pool to participate in a 45-minute online study in exchange for course credit. Ten participants were removed for rating at least one bad protagonist as morally good or the good protagonist as morally bad. These participants were assumed to be either inattentive or to have moral values that significantly diverged from the population of interest. An additional 62 participants were removed for completing the study in less than half (18 minutes) or more than double (72 minutes) the 50% trimmed mean duration (36 minutes). These participants were assumed to have not read every question carefully or to have completed the study while distracted.

The final sample consisted of 250 participants, who were mostly female (68.8% female, 26.8% male, and 3.2% nonbinary), liberal (68.0% extremely or moderately liberal, 25.2% moderate, and 6.8% moderately conservative), white (71.2% White, 10.0% Asian, 8.0% Hispanic, 5.2% Black, 1.2% Native American, and 4.4% Biracial or Other), young adults between the ages of 18 and 28 (M = 19.34, SD = 1.49). An a priori power analysis on simulated data using the `simR` package in R concluded that a sample size of 200 would be large enough to detect small effects at .8 power for the fixed effects in the planned mixed effects models with characteristics and participants as random factors.

Research Design

The present experiment is a repeated measures design with six fully counterbalanced conditions. Participants learned about six different exemplars in random order: a prototypical *abuser*, a prototypical *bigot*, a prototypical *narcissist*, a prototypical *psychopath*, a prototypical *bad person* (superordinate category), and a prototypical *good person* (superordinate category). Each exemplar was defined by a set of three characteristics from a pool of four characteristics. For an example of an exemplar defined by characteristics prototypical of the *narcissist* prototype, see Figure 11.

Figure 11

Example Instructions for Introducing an Exemplar

Riley **thinks the world revolves around them**.
Riley **is very judgmental and critical of others**.
Riley **believes that they are better than other people**.

Instructions: Please take a few moments to picture Riley clearly in your mind, including what you think they would look like and how they would act in different situations. When the arrow appears and you have a clear picture of Riley, continue on with the experiment.

Four versions of the experiment were generated such that different participants would see exemplars defined by different characteristics but would see the same exemplars as ¼ of participants (Table 6). For all exemplars, participants rated the likelihood that the exemplar would have one characteristic prototypical of an abuser, one characteristic prototypical of a bigot, one characteristic prototypical of a psychopath, and one characteristic prototypical of a narcissist. The likelihood rating task was fixed for each participant such that each participant rated the same four characteristics for each exemplar, and participants rated characteristics that did not define any exemplar they were exposed to. Since I am most concerned with measuring

the fixed effects of prototype and distance rather than the random effects of the items used to define the exemplars or the participants in the study, this design was chosen to limit random error while ensuring that the effect is not a result of the specific items used in the study.

Table 6

Research Design Demonstrating How Exemplars Are Defined

Exemplar	Version 1	Version 2	Version 3	Version 4
Abuser (A)	A = {a1, a2, a3}	A = {a1, a3, a4}	A = {a1, a2, a4}	A = {a2, a3, a4}
Bigot (B)	B = {b1, b2, b3}	B = {b1, b3, b4}	B = {b1, b2, b4}	B = {b2, b3, b4}
Psychopath (P)	P = {p1, p2, p3}	P = {p1, p3, p4}	P = {p1, p2, p4}	P = {p2, p3, p4}
Narcissist (N)	N = {n1, n2, n3}	N = {n1, n3, n4}	N = {n1, n2, n4}	N = {n2, n3, n4}
Bad Person (E)	E = {e1, e2, e3}	E = {e1, e3, e4}	E = {e1, e2, e4}	E = {e2, e3, e4}
Good Person (G)	G = {g1, g2, g3}	G = {g1, g3, g4}	G = {g1, g2, g4}	G = {g2, g3, g4}
Likelihood Ratings	{a4, b4, p4, n4}	{a2, b2, p2, n2}	{a3, b3, p3, n3}	{a1, b1, p1, n1}

Note. Exemplars are denoted with capital letters and sets of characteristics are denoted with lower case letters.

Materials

Exemplars. For each participant, the exemplars were assigned random (without replacement) names from a pool of six unisex names: Alex, Blake, Casey, Dakota, Jordan, and Riley. Thus, different participants saw different names assigned to each exemplar. Naming the exemplars helped clarify that participants were rating people (not characteristics) and made it

easy to recognize when they began rating a new individual. The random assignment of names ensured that the effects observed were not due to the specific name chosen for the exemplar.

The characteristics that defined the *abuser*, *bigot*, *narcissist*, and *psychopath* exemplars were determined via the four highest membership values in the given category from the seven-solution fuzzy cluster analysis in Study 2 (Table 5). The superordinate bad person characteristics were chosen from a list of characteristics that had a < .5 membership value on their strongest category and had varying degree of membership with multiple categories. Thus, these characteristics were considered prototypical of several types of bad people and were expected to be prototypical of the superordinate bad person category (e.g., lacks empathy, hateful). The superordinate good person characteristics were taken from a research study on the prototypical traits of a good person (Lapsley & Lasky, 2001) and altered to fit the style of the other items. The pool of characteristics used can be found in Table 7.

Likelihood Ratings. To measure whether participants will attribute unobserved characteristics to exemplars, participants rated the likelihood that each exemplar would have additional, unobserved characteristics on an eleven-point scale from 1 (*extremely unlikely*) to 11 (*extremely likely*). For each of the six exemplars, participants rated the likelihood of the same four characteristics. The set of unobserved characteristics varied by the version (see Table 6). For example, people who saw version one rated the likelihood that each exemplar would have the characteristics a4, b4, p4, and n4, as these characteristics were not used to describe the *abuser* (A), *bigot* (B), *psychopath* (P), or *narcissist* (N) in this version.

Dimensions of Person Perception. To test dimensional models and to test for differences between prototypes, the three commonly used dimensions of person perception (morality, competence, and sociability) were measured with items similar to those from Goodwin

et al., (2014), who tested the orthogonality of the positive versions of the traits. The traits were chosen carefully so that they could apply to both people and characteristics (i.e., behaviors, beliefs). For example, both a person and a characteristic can be described as *just*. Bipolar, eleven-point scales were used to detect small differences between negative people and characteristics[7]. All items were rated on a scale from 1 = *extremely (negative form of trait)* to 11 = *extremely (positive form of trait)*. The moral dimension was measured with responses to *bad-good*, *unjust-just*, *unprincipled-principled, harmful-helpful*, and *immoral-moral*. The competence dimension was measured with responses to *unintelligent-intelligence*, *irrational-rational*, *ignorant-knowledgeable*, *impulsive-deliberate*, and *incompetent-competent*. The warmth dimension was measured with responses to *cold-warm*, *nonsocial-social*, *emotionless-emotional*, *introverted-extroverted*, and *disagreeable-agreeable.*

To evaluate the fit of each item on its construct, I ran a principal axis factoring analysis with a varimax rotation for orthogonal components on the morality, competence, and sociability ratings of exemplars and characteristics. Since there were three factors with eigenvalues greater than one, I determined that a three-factor solution best fit the data: morality, competence, and sociability. A three-factor solution explained 56.9% of the variance in character ratings and 61.5% of the variance in characteristic ratings. The scree plot indicated that there may be a fourth factor (with only a cold-warm loading), but since the eigenvalue for the fourth factor was less than .5, I disregarded this factor as a dimension.

[7] Having less points on the scale (e.g., a 5-point scale) would not be sensitive enough to detect small differences between the morality, competence, and sociability of exemplars.

Table 7

Pool of Characteristics Describing Exemplars

Category	Pool of Characteristics	Category Loading	*Badness*	
			M	***SD***
Abuser	Physically assaults people.	.96	10.30	1.37
	Needs their partner to suffer to achieve sexual gratification.	.96	9.31	2.26
	Abuses their children.	.95	10.80	.83
	Uses coercion, drugs, or physical violence to sexually assault people.	.84	10.42	1.35
Bigot	Believes white people are inherently superior to people from other races.	1.00	10.11	1.63
	Strongly dislikes immigrants and doesn't think people should move to this country.	.99	9.23	2.04
	Believes that gay people should have fewer rights.	.99	9.60	1.95
	Believes that transgender people should have fewer rights.	.98	9.37	2.20
Narcissist	Thinks the world revolves around them.	.75	6.91	1.95
	Is very judgmental and critical of others.	.73	6.37	1.82
	Always prioritizes themselves over others.	.72	6.56	2.24
	Believes that they are better than other people.	.63	6.78	1.92
Psychopath	Fantasizes about defiling or achieving sexual gratification from dead people.	.97	10.51	1.24
	Murders people out of boredom.	.95	10.90	.56
	Derives pleasure from hurting animals.	.95	10.44	1.49
	Craves eating the flesh of other people (i.e., cannibalism).	.94	10.22	1.72
Bad person	Hates the world and everyone in it.	.38	6.95	2.49
	Does not feel empathy or concern when people are suffering.	.46	9.50	1.98
	Bullies people they perceive as weak or vulnerable.	.46	8.85	2.06
	Steals items they don't need from stores.	.47	6.53	2.31
Good person	Always keeps their promises.			
	Is kind to others.			
	Acts in accordance with their moral values.			
	Is a genuine and authentic person.			

Since *irrational-rational* and *impulsive-deliberate* had very low loadings on the competence factor, and *cold-warm*, *emotionless-emotional*, and *disagreeable-agreeable* had very low loadings on the sociability factor for both exemplars and characteristics (Table 8), I did not include them in the composite scores. Thus, the moral dimension was calculated by the mean response on items *bad-good*, *unjust-just*, *unprincipled-principled, harmful-helpful*, and *immoral-moral* ($\alpha_{exemplars}$ = .89; $\alpha_{characteristics}$ = . 92), competence by the mean response on items *unintelligent-intelligence*, *ignorant-knowledgeable*, and *incompetent-competent* ($\alpha_{exemplars}$ = .77; $\alpha_{characteristics}$ = . 80), and sociability was calculated by the mean response on items *nonsocial-social* and *introverted-extroverted* ($\alpha_{exemplars}$ = .76; $\alpha_{characteristics}$ = . 75).

Essentialist beliefs. To test for differences in essentialists beliefs across different prototypes, I used a revised psychological essentialism scale (Bastian & Haslam, 2006). The original scale measured individual differences in the propensity to believe that social categories are natural and immutable. The scale was revised to measure essentialist beliefs regarding the immoral characteristics of specific people. For example, participants responded on a seven-point Likert scale to express their agreement with statements like "[Protagonist] has these characteristics because of their biological makeup" (see Appendix C). The internal consistency of the scale was poor, indicating that the items may not be measuring the same underlying construct (α = .50). Since the scale has theoretical importance and removing items would only marginally improve reliability, all six items were averaged to create a composite essentialism score.

Table 8

Factor Loadings of Judgments of Exemplars and Characteristics

	Exemplars			Characteristics		
	Morality	Competence	Sociability	Morality	Competence	Sociability
Immoral	**0.84**	0.16	0.18	**0.82**	0.26	0.23
Bad	**0.83**	0.17	0.16	**0.84**	0.26	0.15
Unjust	**0.76**	0.28	0.10	**0.80**	0.29	0.17
Harmful	**0.78**	0.11	0.16	**0.82**	0.22	0.16
Unprincipled	**0.60**	0.26	0.20	**0.63**	0.31	0.23
Unintelligent	0.14	**0.78**		0.27	**0.74**	0.15
Irrational	0.58	0.41	0.13	0.61	0.47	0.18
Ignorant	0.20	**0.70**		0.33	**0.68**	
Impulsive	0.35	0.26	0.21	0.25	0.28	0.31
Incompetent	0.20	**0.75**	0.10	0.25	**0.74**	0.24
Cold	0.55	0.19	0.35	0.69	0.22	0.25
Nonsocial	0.24		**0.86**	0.21	0.15	**0.75**
Emotionless	0.44		0.38	0.43	0.20	0.36
Introverted	0.17		**0.84**	0.13		**0.83**
Disagreeable	0.46	0.40	0.14	0.64	0.36	0.12

Note. Bold numbers indicate strong loadings on factor represented by the column (>= .6).

Social distance. To test for differences in desired social distance between prototypes, I used a revised Bogardus Social Distance Scale (Bogardus, 1933). The social distance measure asks participants to respond on a nine-point Likert scale whether they would be willing to accept the target in several relationships with increasing distance, such as a close friend, a coworker, or a visitor in their country (see Appendix D). Since the original scale was developed in 1933, the scale was adapted to modernize it and make it a better fit for the context of the study. The final distance score was calculated by taking a sum of the responses.

Procedure

Participants learned about six different exemplars, one at a time, in random order. Directly following each exemplar, participants rated the likelihood that the exemplar also had four other characteristics. After the likelihood task, participants rated their overall impression of

the exemplar by responding to questions about their morality, warmth, competence, psychological essentialism, and social distance. Next, participants described other characteristics not listed initially that they thought the exemplar would have, such as their "behaviors, emotions, beliefs, desires, motives, intentions, traits, relationships, abilities, or physical appearance." The participants were given three blanks and an "add another characteristic" button to add more blanks. If they did not add at least three characteristics, they were given an error message and instructed to add at least three. Then, participants indicated what gender, race, and age they pictured the exemplar to be, and the survey moved on to the next exemplar. After participants had evaluated all six protagonists, they judged each of the four characteristics that they rated in the likelihood task separately on the morality, competence, and sociability scales.

Results

First, to test whether a prototype model, a unidimensional model, a multidimensional model, or a dual prototype and dimensional model can best explain how people make inferences about whether others have immoral characteristics, I ran a series of linear mixed-effects models using maximum likelihood estimation to compare unnested models. All models included the participants and the unobserved characteristics as random intercepts to account for the repeated measures design and to generalize the findings beyond the participants and characteristics used in the study (Judd et al., 2012). I used the `lme4` package in R for all mixed-effects models, the ` lmerTest` package to estimate p-values for mixed-effects models, and the `emmeans` package to run post hoc tests.

I divided the data set into four parts based on whether the characteristic being rated was highly prototypical of an abuser, psychopath, narcissist, or bigot. Thus, all mixed-effects models

were run four times to independently test how people make inferences that people have characteristics prototypical of a *psychopath*, *abuser*, *bigot*, and *narcissist*.

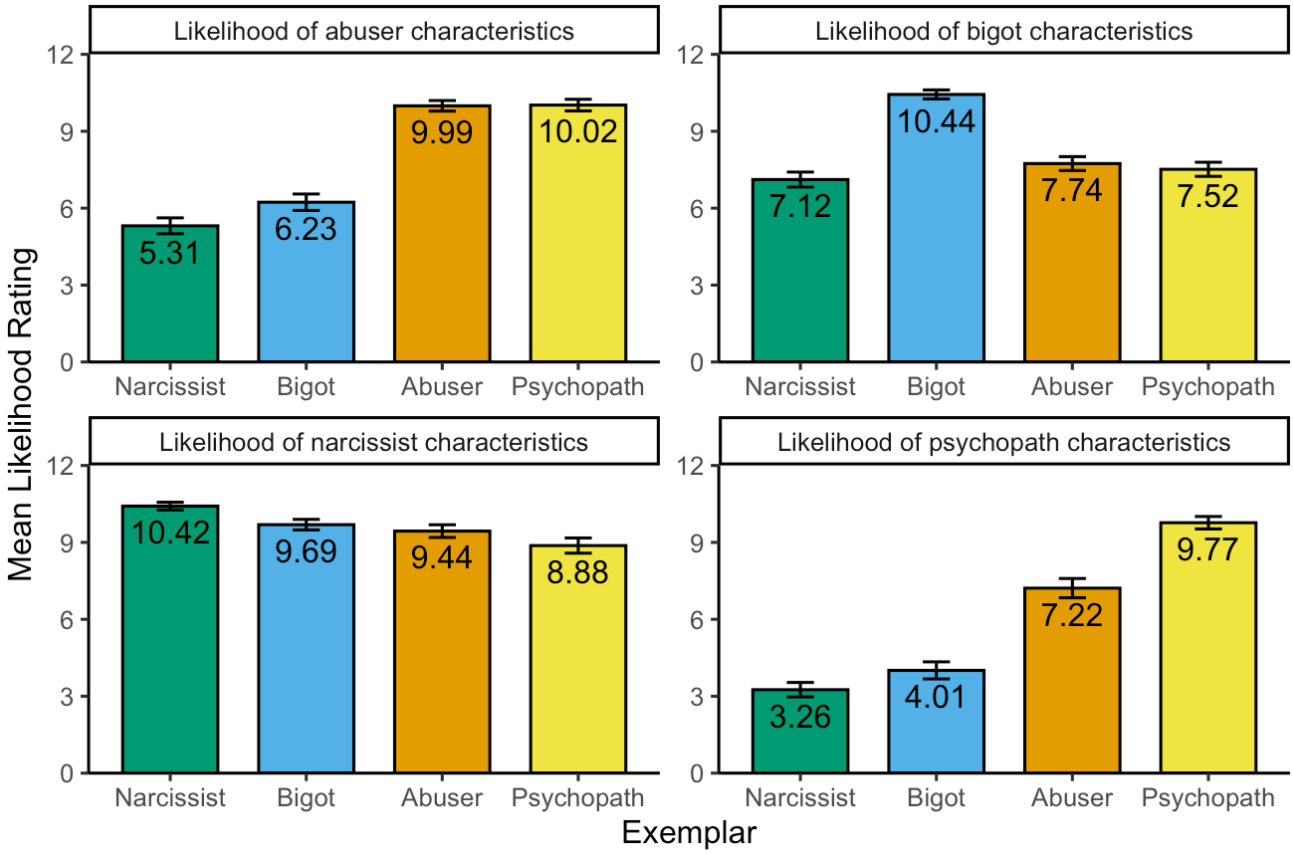
Testing A Prototype Model

Ratings of the mean likelihood that an exemplar would have an unobserved characteristic can be found in Figure 12, split by exemplar and category rated. For example, the green bar in the column chart in the top left quadrant of Figure 12 represents the mean likelihood that a narcissist exemplar would be attributed an abuser characteristic (5.31). In general, people are most likely to infer characteristics that are consistent with a prototype to exemplars of that prototype compared to exemplars of other immoral prototypes. The only exception was *abusers*, who were equally likely to be inferred *abuser* characteristics (e.g., physical violence, rape) as *psychopaths* were. For the means and standard deviations of the likelihood of every characteristic, see Appendix E.

Superordinate prototypes. First, I tested a prototype model with two superordinate prototypes *good person* and *bad person.* I hypothesized that people would rate *bad person* exemplars as more likely to have unobserved immoral characteristics compared to *good person* exemplars. In support of this hypothesis, I found exemplars with immoral characteristics (M = 7.32, SD = 2.89) were rated as significantly more likely to have unobserved immoral characteristics than exemplars with moral characteristics (M = 2.09, SD = 1.65), $t(1739) = 58.82$, $p < .001$, $d = 2.63$.

Figure 12

Likelihood Exemplars Have Unobserved Characteristics



Basic prototypes. Next, I examined whether the basic category of the exemplar (i.e., *psychopath, abuser, bigot*, or *narcissist*) predicted likelihood ratings for unobserved characteristics, using the category being rated as the reference group. In support of a prototype model, I hypothesized that exemplars would be inferred characteristics prototypical of their basic category more so than exemplars of competing bad person categories.

Psychopathic characteristics. Category membership significantly predicted inferences of psychopathic characteristics, such that characteristics prototypical of a *psychopath* (i.e., animal sadism, necrophilia, murder, and cannibalism) were more likely to be inferred to the *psychopath* exemplars ($M = 9.77$, $SD = 1.95$) than the *abuser* exemplars ($M = 7.22$, $SD = 3.02$), *bigot*

exemplars (M = 4.01, SD = 2.68), and *narcissist* exemplars (M = 3.26, SD = 2.25). See Table 9 for model coefficients. Contrary to the prediction from a simple prototype model (i.e., all competing prototypes are equally likely to be inferred), a post hoc Tukey test revealed that psychopathic characteristics were inferred significantly more to *abusers* than to *bigots*, $t(753) = 18.13$, $p < .001$, $d = 1.62$, and *narcissists*, $t(753) = 22.38$, $p < .001$, $d = 2.00$, and that psychopathic characteristics were inferred significantly more to *bigots* than *narcissists*, $t(753) = 4.25$, $p < .001$, $d = 0.38$.

Table 9

Mixed Effects Model Predicting Likelihood of Psychopath Characteristics from Category Membership

Fixed Effects	***b***	***SE***	***t***	***df***	***p***
Psychopath (Intercept)	9.71	0.69	14.07	3.15	< .001
Psychopath v. Abuser	-2.55	0.18	-14.42	747	< .001
Psychopath v. Bigot	-5.76	0.18	-32.56	747	< .001
Psychopath v. Narcissist	-6.51	0.18	-36.81	747	< .001

Abuser Characteristics. Category membership also predicted inferences of *abuser* characteristics, such that characteristics prototypical of an *abuser* (i.e., physical assault, rape, sexual sadism, and child abuse) were more likely to be inferred to the *abuser* exemplars (M = 9.99, SD = 1.66) than the *bigot* exemplars (M = 6.23, SD = 2.57) and *narcissist* exemplars (M = 5.31, SD = 2.49). See Table 10 for model coefficients. Further, *psychopaths* (M = 10.02, SD = 1.82) were just as likely to be inferred characteristics prototypical of an *abuser* as *abusers* were. A post hoc Tukey test revealed that abusive characteristics were inferred significantly more to *psychopaths* than to bigots, $t(753) = -22.05$, $p < .001$, $d = -1.98$, and *narcissists*, $t(753) = -27.41$,

$p < .001$, $d = 2.46$, and abusive characteristics were inferred significantly more to *bigots* than *narcissists*, $t(753) = 5.36$, $p < .001$, $d = 0.48$.

Table 10

Mixed Effects Model Predicting Likelihood of Abuser Characteristics from Category Membership

Fixed Effects	***b***	***SE***	***t***	***df***	***p***
Abuser (Intercept)	9.99	0.24	41.75	747	< .001
Abuser v. Bigot	-3.76	0.17	-21.89	747	< .001
Abuser v. Narcissist	-4.68	0.17	-27.25	747	< .001
Abuser v. Psychopath	0.03	0.17	0.16	747	.871

Bigoted characteristics. Category membership predicted inferences of characteristics prototypical of a *bigot* (i.e., racism, homophobia, transphobia, and xenophobia) such that people were more likely to infer bigoted characteristics to *bigot* exemplars ($M = 10.40$, $SD = 1.41$) than to *psychopath* ($M = 7.52$, $SD = 2.21$), *abuser* (M = 7.74, SD = 2.17), and *narcissist* exemplars ($M = 7.12$, $SD = 2.34$). See Table 11 for model coefficients. Post hoc Tukey tests revealed that psychopaths were not inferred significantly more bigoted characteristics compared to *abusers*, $t(753) = 1.35$, $p = .533$, $d = 0.12$, and *narcissists*, $t(753) = -2.40$, $p = .077$, $d = -0.21$, but abusers were inferred significantly more *bigot* characteristics compared to *narcissists*, $t(753) = 3.75$, $p = .001$, $d = 0.34$.

Table 11

Mixed Effects Model Predicting Likelihood of Bigot Characteristics from Category Membership

Fixed Effects	*b*	*SE*	*t*	*df*	*p*
Bigot (Intercept)	10.44	0.16	66.81	8.56	< .001
Bigot v. Abuser	-2.70	0.17	-16.20	747	< .001
Bigot v. Narcissist	-3.32	0.17	-19.95	747	< .001
Bigot v. Psychopath	-2.92	0.17	-17.55	747	< .001

Narcissist characteristics. Category membership predicted inferences of narcissist characteristics, such that characteristics prototypical of a *narcissist* (i.e., selfishness, arrogance, egocentrism, judgmental) were more likely to be inferred to the *narcissist* exemplars (M = 10.40, SD = 1.20) than the *psychopath* (M = 8.88, SD = 2.38), *abuser* (M = 9.44, SD = 1.99), and *bigot* exemplars (M = 9.69, SD = 1.66). See Table 12 for model coefficients. A post hoc test revealed that *psychopaths* were attributed significantly less narcissistic characteristics compared to *bigots*, $t(753) = 5.45$, $p < .001$, $d = 0.49$, and *abusers*, $t(753) = 3.77$, $p = .001$, $d = 0.34$, but that *bigots* and *abusers* were equally likely to be attributed narcissistic traits, $t(753) = -1.68$, $p = .334$, $d = -0.15$.

Table 12

Mixed Effects Model Predicting Likelihood of Bigot Characteristics from Category Membership

Fixed Effects	*b*	*SE*	*t*	*df*	*p*
Narcissist (Intercept)	10.42	0.18	58.68	5.49	< .001
Narcissist v. Abuser	-0.98	0.15	-6.52	747	< .001
Narcissist v. Bigot	-0.72	0.15	-4.84	747	< .001
Narcissist v. Psychopath	-1.54	0.15	-10.28	747	< .001

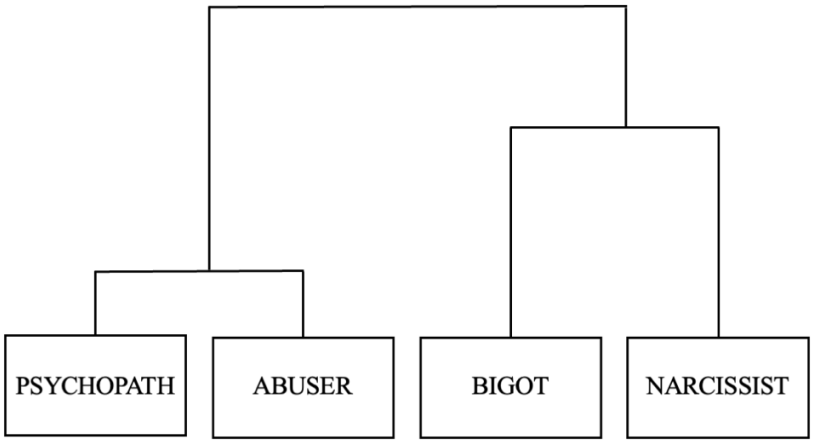
Overall, I found strong evidence that people were more likely to attribute prototype-consistent immoral characteristics to exemplars compared to prototype-inconsistent characteristics, which supports a prototype model. However, some prototypes are more similar to each other than others, suggesting that there is not a strict category structure.

Evaluating the Fit of the Empirically Derived Categorical Structure

According to the categorical structure derived from the card sorting data in Study 2: (1) *psychopaths*, *abusers*, *bigots*, and *narcissists* are subtypes of *bad people* (2) *bigots* and *narcissists* are perceived as further from *psychopaths* and *abusers* than they are to each other and vice versa, and (3) *psychopaths* and *abusers* are perceived as more similar than *bigots* and *narcissists* are (Figure 13).

Figure 13

Recreating of Category Structure Derived in Study 2



Speed of Likelihood Ratings. Prototypes of basic categories are richer in description than superordinate categories, which makes it easier to make inferences or decisions regarding members of the category (Rosch, 1978). Thus, I hypothesized that people would respond more quickly to inferences about basic prototypes (i.e., psychopath, abuser, bigot, and narcissist) than superordinate prototype (i.e., bad person). In support of this hypothesis, participants took significantly longer (in seconds) to make inferences about the bad person exemplars (M = 17.05, SD = 10.46) compared to the psychopath (M = 12.71, SD = 7.94), abuser (M = 14.06, SD = 8.44), bigot (M = 13.89, SD = 8.79), and narcissist (M = 15.45, SD = 9.28) exemplars (Table 13)[8]. Thus, I confirmed that *psychopaths*, *abusers*, *bigots*, and *narcissists* behave as basic categories that are subordinate to the superordinate *bad people* category.

Likelihood Ratings. Next, to evaluate whether (2) *bigots* and *narcissists* are perceived as further from *psychopaths* and *abusers* than they are to each other and vice versa, and (3) *psychopaths* and *abusers* are perceived as more similar than *bigots* and *narcissists* are, I examined the likelihood ratings of unobserved characteristics and found that they generally supported the category structure. In support of (2), *psychopaths* and *abusers* were more likely to be inferred each other's characteristics than to be inferred bigoted or narcissistic characteristics.

[8] To run this analysis, I removed data from 17 participants who scored five times above the trimmed group mean for at least one set of ratings. The cutoffs were 74.1 seconds for bad person ratings, 61.6 seconds for abuser ratings, 61.1 seconds for bigot, 64.2 seconds for narcissist ratings, and 56.4 seconds for psychopath ratings. Excluding participants was necessary to accurately record the time it took to complete the task, as some participants took breaks during one or more likelihood rating sessions.

However, this pattern did not apply for *bigots* and *narcissists*. *Abusers* were inferred more bigoted characteristics than *narcissists* were, and *abusers* were inferred just as many narcissistic characteristics as *bigots* were. While this pattern does support (3), the category structure is a poor fit in predicting that abusers are perceived as further away from narcissists and bigots as narcissists and bigot are from each other. While all "basic" categories do appear subordinate to "bad person," the specific structure appears overfitted to the original dataset in regards to how categories relate to each other.

Table 13

Model Fit Metrics for Models Predicting Time of Responding to Inferences from Category Membership

Fixed Effects	***b***	***SE***	***t***	***df***	***p***
Bad Person (Intercept)	17.05	0.59	28.85	904	< .001
Bad Person v. Abuser	-2.99	0.72	-4.18	928	< .001
Bad Person v. Narcissist	-1.60	0.72	-4.24	928	.025
Bad Person v. Bigot	-3.16	0.72	-2.42	928	< .001
Bad Person v. Psychopath	-4.34	0.72	-6.066	928	< .001

Testing Dimensional Models

Unidimensional Model. To test a unidimensional model of immoral inferences, I examined whether the Euclidean distance[9] between the perceived morality of the exemplar and the perceived morality of the unobserved characteristic predicted likelihood ratings of unobserved immoral characteristics. In a unidimensional model, exemplars would be more likely

[9] I also examined city block (Manhattan's) distance and maximum distance but found that these models performed equal to or worse than Euclidean distance.

to be attributed immoral characteristics that are perceived to be close in proximity on a single moral dimension (i.e., *Euclidean distance* between the morality of the target and the characteristic is near 0). For instance, a *slightly immoral person* would be more likely to be inferred *slightly immoral* characteristics than *extremely immoral* characteristics.

The Euclidean distance between the perceived morality of the person and the perceived morality of the characteristic significantly predicted the likelihood of attributing an unobserved immoral characteristic to a person for characteristics prototypical of a *psychopath*, $t(926) = -24.41$, $p < .001$, $b = -1.41$, 95%CI-1.54, -1.30], *abuser*, $t(987) = -21.09$, $p < .001$, $b = -1.15$, 95%CI[-1.26, -1.04], *bigot*, $t(995) = -7.37$, $p < .001$, $b = -0.45$, 95%CI[-0.58, -0.34], and *narcissist*, $t(996) = -3.53$, $p < .001$, $b = -0.16$, 95% CI[-0.25, -0.07]. Thus, the unidimensional model was supported across all four prototypes.

Multidimensional Model. Figures 14 displays the position of the exemplars and characteristics by category on the morality and competence dimensions, and Figure 15 displays them on the morality and sociability dimensions. The triangles represent the exemplars, and the circles represent the unobserved characteristics (open shapes represent arithmetic means). In a multidimensional model, characteristics that are closest to an exemplar should be most strongly attributed to the exemplar, whereas characteristics that are far away from the exemplar would be weakly attributed to the exemplar.

To test a multidimensional model of immoral inferences, I examined whether the Euclidean distance between the perceived morality, competence, and sociability of the exemplar and the perceived morality, competence, and sociability of the unobserved characteristic predicted inferences of the unobserved characteristic. In a multidimensional model, exemplars will be more likely to be attributed immoral characteristics that are perceived to be close in

proximity on multiple dimensions. For instance, a person who is perceived to be *slightly immoral, moderately unintelligent, and somewhat social* would be more likely to be inferred characteristics that are perceived to be *slightly immoral*, *moderately unintelligent*, and *somewhat social* and less likely to be attributed characteristics further away in the three-dimensional space.

Euclidean distance between person and characteristic in a moral, competence, and sociability three-dimensional space significantly predicted inferences of immoral characteristics for characteristics prototypical of a *psychopath*, $t(974) = -17.71$, $p < .001$, $b = -0.65$, 95%CI [-0.73, -0.57], *abuser*, $t(997) = -10.18$, $p < .001$, $b = -0.42$, 95%CI[-0.50, -0.33], *bigot*, $t(994) = -7.59$, $p < .001$, $b = -0.28$, 95%CI[-0.35, -0.21], and *narcissist* $t(996) = -2.19$, $p = .029$, $b = -0.07$, 95%CI[-0.13, -0.01]. Thus, the multidimensional model is also supported across all four prototypes.

Evaluating the Fit of the Empirically Derived Dimensional Structure

To validate the interpretation that the three dimensions from the multidimensional scaling analysis represent perceived morality, sociability, and competence of the immoral attributes, I conducted a Procrustes analysis with the `vegan` package in R. Procrustes analysis, named after the Greek mythological figure Procrustes who forced his guests to fit his iron guest bed by either stretching or cutting off their legs, enables the comparison of two sets of three dimensional points by scaling, rotating, and transforming[10] the second set to minimize the sum of square distances between the two sets (Peres-Neto & Jackson, 2001).

[10] This analysis strategy was chosen because the MDS data needs to be rotated (as MDS dimensions may need to be rotated to meaningfully interpret the axes; Borg et al., 2013) and stretched (since the MDS solution and ratings have different scales).

Figure 14

Morality and Competence Ratings for Exemplars and Characteristics

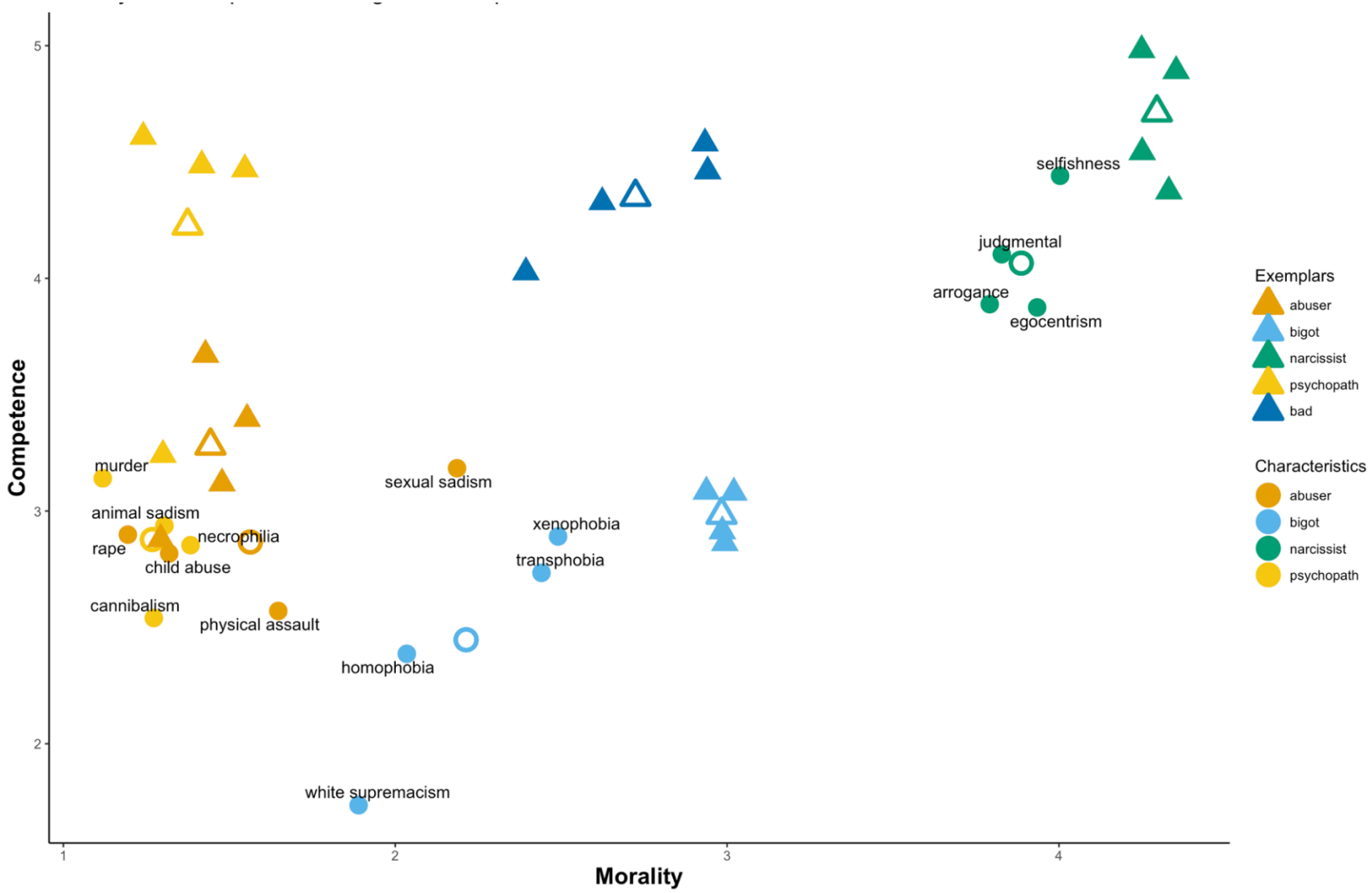
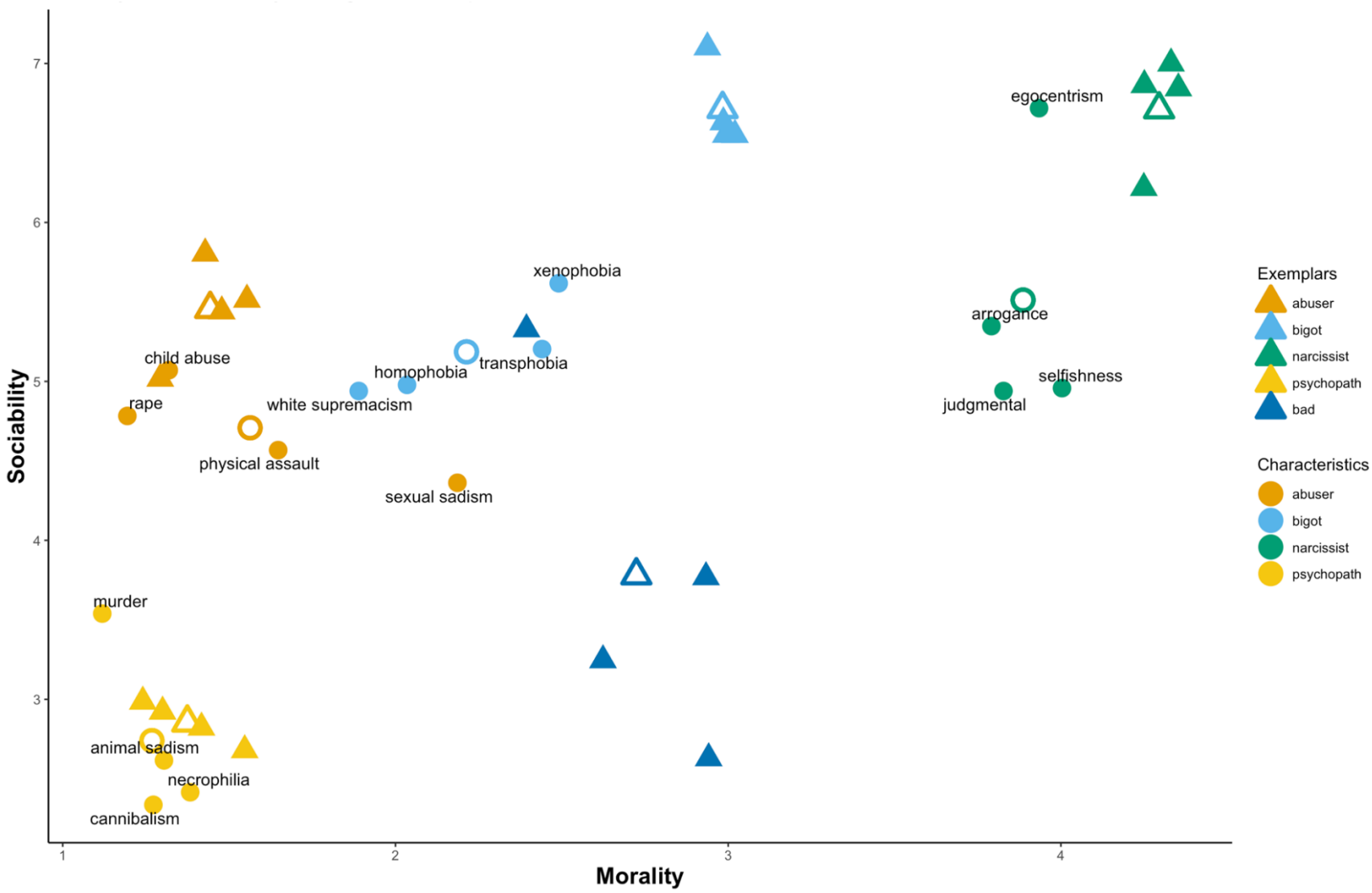


Figure 15

Morality and Sociability Ratings for Exemplars and Characteristic



I tested the fit between the participant ratings of the 16 immoral characteristics on morality, competence, and sociability from Study 3 and the spatial positions of those same 16 characteristics derived from Study 2's multidimensional scaling analysis. I found that the correlation in a symmetric Procrustes rotation was $r = .82$, indicating a strong fit between the participant ratings and the solution from the multidimensional scaling analysis. The Procrustes Sum of Squares was 0.33. A permutation test, which randomly sampled from the data 999 times to generate a Procrustes Sum of Squares distribution, indicated statistical significance compared to what would be expected by random chance ($p = .001$). See Appendix F for visualizations of the errors between the MDS solution and the expected ratings on morality, sociability, and competence. Thus, I concluded that morality, competence, and sociability are a reasonable fit for the multidimensional structure.

Comparing Prototype, Dimensional, and Dual Models of Immoral Inferences

Next, I compared the prototype model, unidimensional model, and multidimensional models explained above to two dual models: a unidimensional prototype model and a multidimensional prototype model. The marginal R^2 values quantify how much variance in likelihood ratings can be explained by the fixed effects (i.e., prototype and/ or Euclidean distance) in the model, while the conditional R^2 values quantify how much variance in likelihood ratings can be explained by both the fixed and random effects, including the effects of the participants and the items (Table 15). Overall, prototype models performed better than dimensional models in explaining how immoral characteristics are inferred to category exemplars. The model explaining variance in likelihood ratings of psychopath characteristics

performed the best (52% variance explained by category membership), followed by abuser (50%), bigot (29%), and then narcissist (8%) characteristics.

To compare the data while taking parsimony into account (i.e., preference for less predictors), I examined the AIC and BIC values for each of the five models. For all four prototypes, the unidimensional-prototype model had lower or about the same AIC and BIC values compared to other models. Thus, the model fit metrics indicate that a model that includes both prototypes and a single moral dimension fits the data best (Table 15). In further support for a dual model, model comparisons between the prototype and prototype-unidimensional model found that the addition of Euclidean distance on the moral dimension improved model fit for inferences of characteristics prototypical of a *psychopath*, $\chi^2(1) = 43.39$, $p < .001$, *abuser*, $\chi^2(1) = 44.91$, $p < .001$, and *bigot*, $\chi^2(1) = 28.44$, $p < .001$, but did not improve model fit for inferences of characteristics prototypical of a narcissist, $\chi^2(1) = 1.36$, $p = .243$. Therefore, the distance from the exemplar to the characteristic on the moral dimension explained inferences of immoral characteristics above and beyond category information. This suggests that individuals sometimes infer that exemplars of moral person categories have prototype-inconsistent characteristics when they are similar in morality, but not necessarily when they are similar in competence and sociability.

Examining Attributes of Bad Person Prototypes

Next, I examined the attributes of each prototype, including perceived gender, race, age, and physical appearance. Additionally, for each kind of exemplar, I examined beliefs regarding whether their immoral characteristics are considered fixed and innate, what kinds of relationships people would be comfortable pursuing, and what unobserved characteristics were inferred.

Table 14

Model Fit Metrics for Models Predicting Likelihood of Inference of Characteristics

Unobserved Characteristic	Model	AIC	BIC	Cond. R^2	Marg. R^2	ICC
Psychopath	Prototype	4403.2	4437.5	.70	.52	.38
	Unidimensional	4881.2	4905.7	.47	.33	.21
	Multidimensional	5068.5	5093.0	.36	.22	.18
	Prototype-Unidimensional	**4361.8**	**4401.1**	**.72**	**.54**	**.39**
	Prototype-Multidimensional	4379.5	4418.8	.71	.53	.39
Abuser	Prototype	4326.2	4360.6	.60	.50	.22
	Unidimensional	4715.6	4740.2	.36	.30	.08
	Multidimensional	4967.8	4992.3	.09	.09	.01
	Prototype-Unidimensional	**4283.3**	**4322.6**	**.62**	**.52**	**.21**
	Prototype-Multidimensional	4308.9	4348.2	.61	.50	.22
Bigot	Prototype	4253.0	4287.3	.42	.29	.19
	Unidimensional	4577.4	4601.9	.08	.05	.03
	Multidimensional	4576.2	4600.8	.13	.06	.07
	Prototype-Unidimensional	**4226.5**	**4265.8**	**.43**	**.31**	**.18**
	Prototype-Multidimensional	4237.2	4276.4	.44	.30	.20
Narcissist	**Prototype**	**4038.1**	**4072.4**	**.25**	**.08**	**.18**
	Unidimensional	4124.2	4148.7	.17	.01	.16
	Multidimensional	4121.8	4156.3	.16	< .01	.16
	Prototype-Unidimensional	**4038.7**	**4078.0**	**.26**	**.08**	**.19**
	Prototype-Multidimensional	**4036.9**	**4076.1**	**.25**	**.08**	**.18**

Note. The bold models indicate superior models for predicting unobserved characteristics that are prototypical of a moral person category.

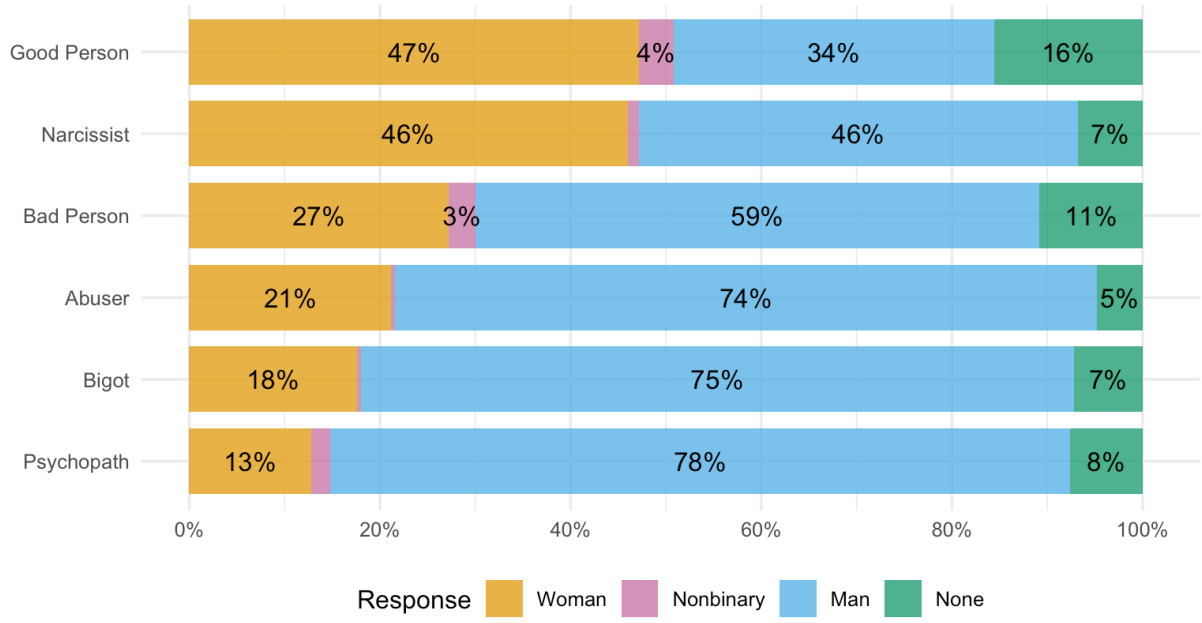
Gender. Most immoral prototypes were assumed to be men (Figure 16). There was a moderate relation between gender and type of exemplar, $\chi^2(5) = 159.92$, $p < .001$, $V = 0.35$. Psychopaths, abusers, bigots, and bad people were more likely to be pictured as men than expected if gender and exemplar were independent, and narcissists and good people were more likely to be pictured as women than expected.

Since basic categories have richer cognitive representations compared to superordinate categories, I hypothesized that people would be more likely to picture a gender for a

superordinate category versus a basic category. This hypothesis was testable because participants could opt out of choosing that they were picturing the exemplar as a man, woman, or nonbinary, by indicating that they were "not picturing a gender." There was a moderate deviation of observed values from expected values, $\chi^2(5) = 21.09$, $p < .001$, $w = 0.40$, as people were significantly more likely to say that they were not picturing a gender for good person and bad person exemplars than expected and less likely to say that they were not picturing a gender for psychopath, abuser, bigot, and narcissist exemplars than expected.

Figure 16

Imagined Gender for Each Prototype

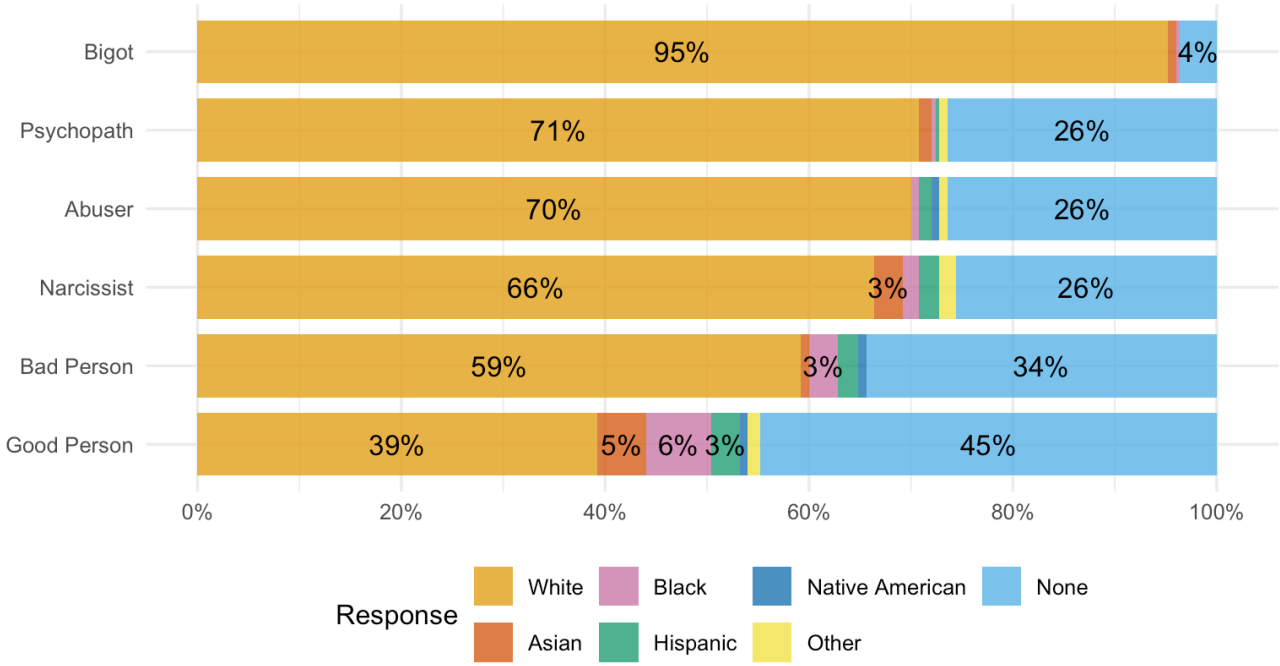


Race. Most prototypes were seen as white, or a race was not indicated (Figure 17). There was a moderate relation between race and type of exemplar, such that psychopaths, abusers, bigots, and bad people were more likely to be white than expected, and narcissists and good people were more likely to be non-white than expected, $\chi^2(5) = 98.90$, $p < .001$, $V = 0.30$.

Further, supporting the hypothesis that basic categories have richer cognitive representations, people were more likely to report that they were not picturing a race for the *good person* and *bad person* exemplars than expected and less likely to report that they were not picturing a race for the *psychopath*, *abuser*, *narcissist*, and *bigot* exemplars than expected, $\chi^2(5) = 85.77$, $p < .001$, $w = 0.46$.

Figure 17

Imagined Race for Each Prototype



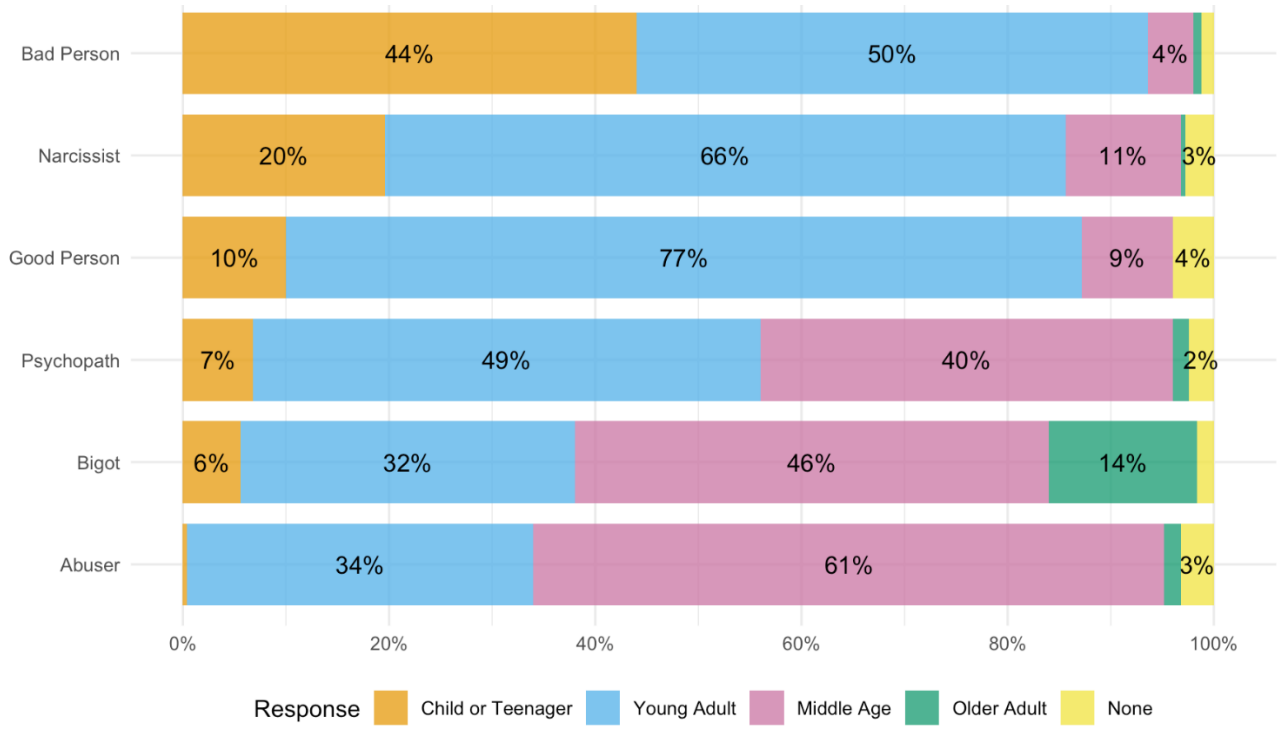
Age. Most prototypes were presumed to be young adults or middle aged (Figure 18). I conducted a final chi-squared test of independence to test for independence between age (child or teen, young adult, middle age, or older adult) and type of exemplar and found a moderate relation, $\chi^2(15) = 657.53$, $p < .001$, $V = 0.39$. Both *abusers* and *psychopaths* were more likely to be perceived as middle age than expected, *narcissists* were more likely to be perceived as children, teens, or young adults than expected, and *bigots* were more likely to be perceived as middle age or older adult than expected. Finally, *good people* were more likely to be perceived

as a young adult than expected, and the *bad people* were more likely to be perceived as a child or teenager than expected. Thus, in general, *narcissists* and superordinate prototypes were assumed to be young; *abusers* and *psychopaths* were assumed to be older (young to middle adult); and *bigots* were assumed to be the oldest prototype.

The chi-squared goodness of fit test did not find evidence that any prototypes reported significantly more or less responses of "I wasn't picturing an age" than expected under a uniform distribution, $\chi^2(15) = 5.26$, $p = .385$, $V = 0.11$. This may be because only a few respondents reported not picturing an age (0-4% per category), as many defaulted to assume that exemplars were young or middle-aged adults.

Figure 18

Imagined Age for Each Prototype



Dimensions of Social Perception. On average, *psychopaths* were seen as extremely immoral, somewhat incompetent, and moderately unsociable; *abusers* are seen as extremely immoral, moderately incompetent, and slightly unsociable; *bigots* are seen as moderately immoral, moderately incompetent, and slightly social; and *narcissists* are seen as somewhat immoral, slightly incompetent, and slightly sociable (Table 15).

As hypothesized, people rated the superordinate bad person as more negative on morality, competence, and sociability compared to the superordinate good person. This is consistent with the halo effect, such that people are inferred positive traits when they are originally described positively and negative traits when they are originally described negatively (Thorndike, 1920). In addition, I hypothesized that basic category protagonists would be perceived differently on morality, competence, sociability, essentialist beliefs, and social distance. To test for differences in morality, competence, and sociability between categories, I ran mixed effects models predicting morality, competence, sociability, essentialist beliefs, and social distance for each of the four exemplars with the `lme4` package in R, estimating the effects with maximum likelihood. To test for differences between groups, I ran post hoc Tukey tests with the `emmeans` package. The Means and Standard Deviations for each feature can be found in Table 15.

Table 15

Means (SDs) for Differences Between Cognitive Prototypes of Prototype Features

	Morality	Competence	Sociability	Essentialism	Social Distance
Psychopath	1.37[a] (0.69)	4.23[b] (2.42)	2.85[a] (1.86)	4.92[a] (0.79)	34.42[a] (3.00)
Abuser	1.44[a] (0.65)	3.28[a] (1.89)	5.46[b] (2.25)	4.37[b] (0.74)	33.20[b] (3.61)
Bigot	2.98[b] (1.24)	2.98[a] (1.42)	6.71[c] (1.94)	3.84[c] (0.74)	29.70[c] (5.88)
Narcissist	4.30[c] (1.36)	4.71[c] (1.75)	6.71[c] (2.15)	3.93[c] (0.69)	24.85[d] (6.18)
Good Person	9.69 (0.98)	8.89 (1.40)	7.98 (1.38)	4.28 (0.64)	3.38 (5.01)
Bad Person	2.72 (1.10)	4.35 (1.70)	3.78 (2.20)	3.94 (0.73)	28.36 (6.03)

Note. The subscripts for each column indicate that the prototypes were perceived to differ significantly on the characteristic in the column.

Morality. *Good people* were rated as more moral than *bad people*, $t(498) = 74.85$, $p < .001$, $d = 6.97$. *Psychopaths* were rated as significantly less moral than *bigots*, $t(747) = 20.30$, $p < .001$, $d = 1.82$, and *narcissists*, $t(747) = 36.86$, $p < .001$, $d = 3.30$, but *psychopaths* and *abusers* did not significantly differ from each other, $t(747) = 0.89$, $p = .821$, $d = .07$. Further, *bigots* were rated as more immoral than *narcissists*, $t(747) = -16.56$, $p < .001$, $d = -1.48$. Thus, consistent with the hypothesis that different types of bad people are perceived as more or less bad, exemplars of prototypes differed on their perceived morality, with *psychopaths* and *abusers* being perceived as more immoral than *bigots*, who were perceived as more immoral than *narcissists*.

Competence. As predicted, *good people* were rated as more intelligent and knowledgeable than *bad people*, $t(498) = 32.64$, $p < .001$, $d = 4.54$. Testing for differences in

basic categories revealed that *bigots* were rated as significantly less competent than *psychopaths*, $t(747) = -8.93$, $p < .001$, $d = -0.80$, and *narcissists*, $t(747) = -12.43$, $p < .001$, $d = -1.11$, but *bigots* did not significantly differ from *abusers*, $t(747) = 2.14$, $p = .143$, $d = 0.19$. Both *abusers*, $t(747) = -10.29$, $p < .001$, $d = -0.92$, and *psychopaths*, $t(747) = 3.50$, $p = .003$, $d = 0.31$, were rated as significantly less competent than *narcissists*. Thus, consistent with the hypothesis that basic prototypes vary in their perceived competence, *bigots* and *abusers* were rated as less intelligent and knowledgeable than *psychopaths*, who were rated less intelligent and knowledgeable than *narcissists*.

Sociability. As predicted *good people* were rated as more sociable than *bad people*, $t(498) = 25.60$, $p < .001$, $d = 4.20$. Post hoc tests for basic categories revealed that *psychopaths* were rated as significantly less sociable than *abusers*, $t(747) = 16.02$, $p < .001$, $d = 1.43$, *bigots*, $t(747) = 23.77$, $p < .001$, $d = 2.13$, and *narcissists*, $t(747) = 23.76$, $p < .001$, $d = 2.12$. *Abusers* were rated as significantly less sociable than *bigots*, $t(747) = -7.75$, $p < .001$, $d = -0.69$, and *narcissists*, $t(747) = -7.73$, $p < .001$, $d = -0.69$, but *bigots* and *narcissists* did not differ from each other, $t(747) = 0.01$, $p = 1.00$, $d < .01$. Thus, consistent with the hypothesis that basic prototypes differ in sociability, *psychopaths* were rated as the most introverted and reclusive prototype, whereas *abusers* were perceived as more *sociable* than *psychopaths*, and *bigots* and *narcissists* were perceived as the most sociable types of *bad person*.

Essentialism. Some immoral characteristics that others possess are perceived to be stable throughout the lifespan, hardwired in their biology, and essential to who they are, while other characteristics are perceived as transient and to have developed as a response to experiences (Heiphetz, 2020; Lynch et al., 2019; Strohminger & Nichols, 2014). This research found that people have stronger essentialist attitudes about *good people* compared to *bad people*, $t(747) =$

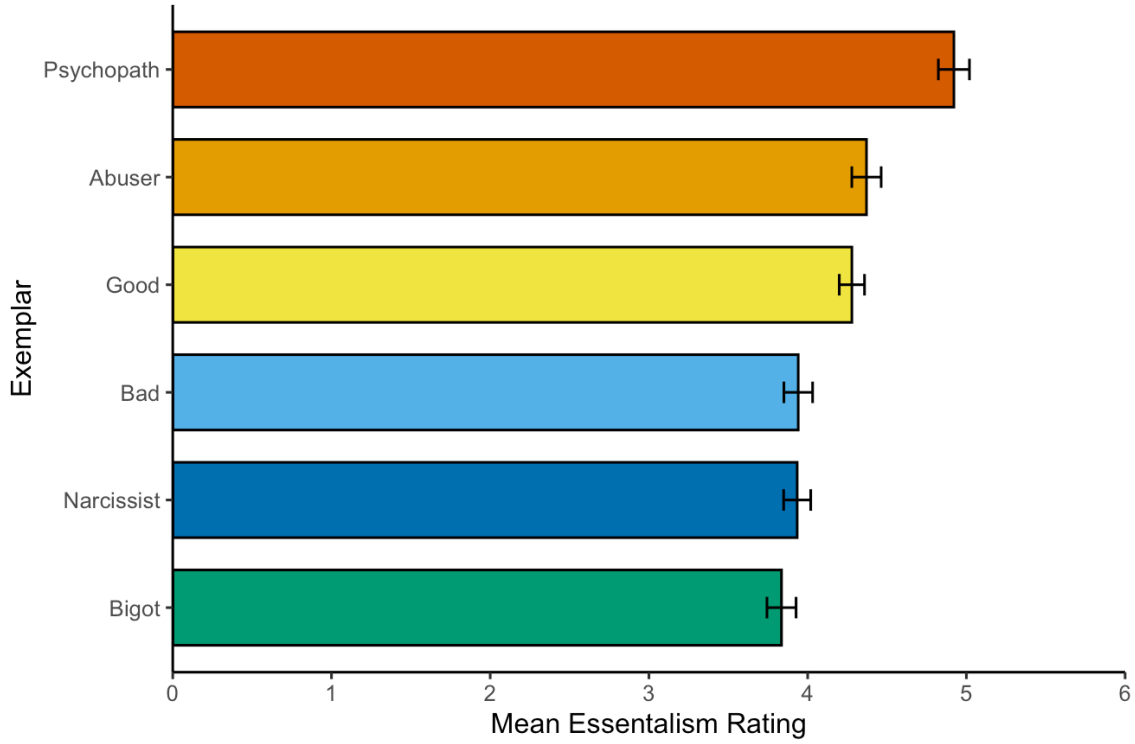
-6.39, $p < .001$, $d = -0.57$ (Figure 19). However, this pattern did not hold for all types of bad people. While people did have stronger essentialist attitudes about *good people* compared to *bigots*, $t(747) = 8.61$, $p < .001$, $d = 0.77$ and *narcissists*, $t(747) = 6.69$, $p < .001$, $d = 0.60$, they had weaker essentialist attitudes about *good people* compared to *psychopaths* ($M = 4.92$, $SD = 1.86$), $t(747) = -12.48$, $p < .001$, $d = -1.12$, and did not differ from *abusers*, $t(747) = -1.79$, $p < .001$, $d = -0.16$.

Further, participants indicated that the characteristics of *psychopaths* were more likely to be rooted in biology and immutable compared to *abusers*, $t(747) = -10.48$, $p < .001$, $d = -0.94$, *bigots*, $t(747) = -20.68$, $p < .001$, $d = -1.85$, and narcissists, $t(747) = -18.80$, $p < .001$, d = -1.68. And, respondents indicated stronger essentialists beliefs towards *abusers* compared to *bigots*, $t(747) = 10.19$, $p < .001$, $d = 0.91$, and *narcissists*, $t(747) = 8.31$, $p < .001$, $d = 0.74$, but *bigots* and *narcissists* did not differ from each other, $t(747) = -1.88$, $p = .238$, d = -0.17. Thus, people held extremely strong views that the characteristics of some types of bad people were biological driven and immutable (e.g., psychopaths, abusers) and were more likely to believe that other types were a result of learned characteristics that can be changed (e.g., narcissists, bigots).

Social Distance. An important reason that we make category distinctions is to know how to interact with category members (Rosch, 1978). Thus, I tested whether there were differences in how comfortable people were forming relationships with exemplars from different types of bad person categories (Figure 20). Consistent with the hypothesis that people generally want to approach good people and avoid bad people, respondents indicated that they would be more comfortable entering relationships of closer distance with *good person* exemplars compared to *bad person* exemplars, $t(249) = 50.97$, $p < .001$, $d = 4.56$.

Figure 19

Essentialism Ratings by Prototype

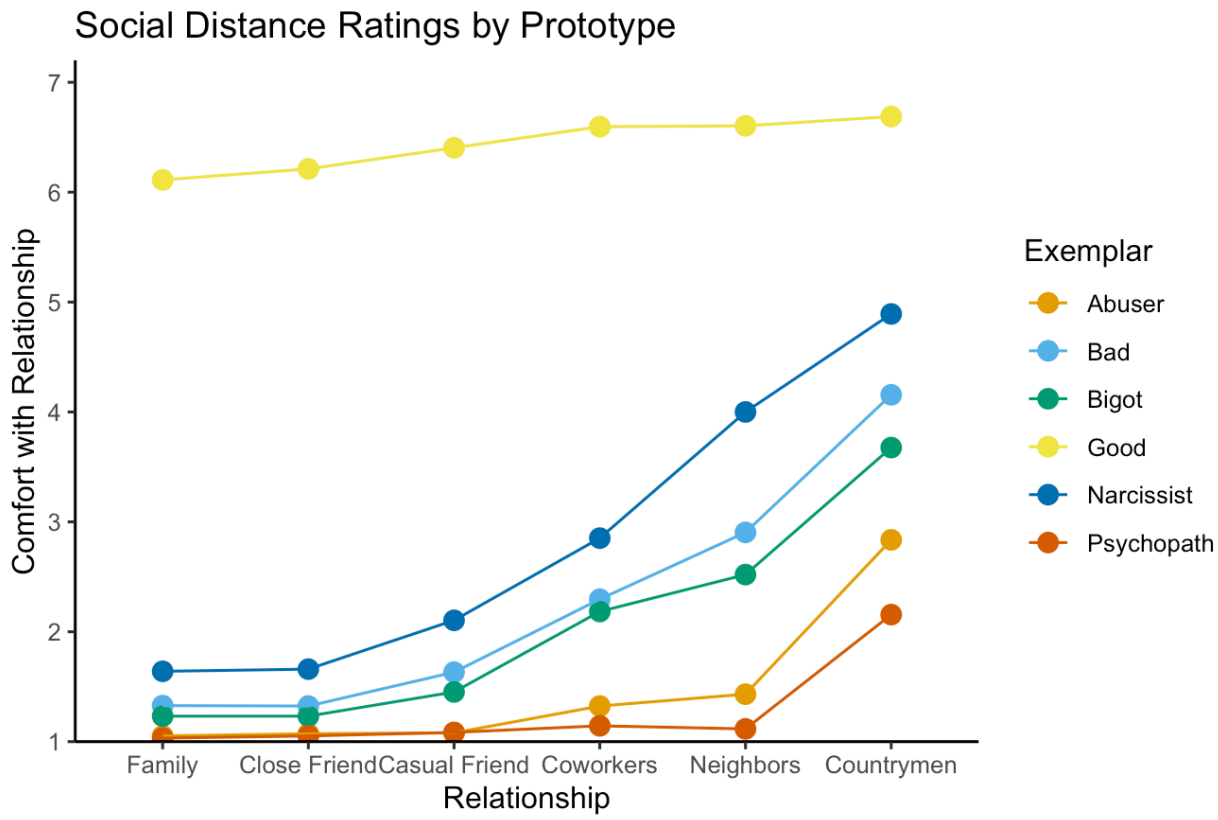


Social distance also differed by basic category, such that people preferred significantly more distance from *psychopaths* compared to *abusers*, $t(747) = -3.45$, $p = .003$, $d = -0.30$, *bigots*, $t(747) = -13.43$, $p < .001$, $d = -1.20$, and *narcissists*, $t(747) = -27.25$, $p < .001$, $d = -2.44$. They preferred significantly more distance from *abusers* compared to *bigots*, $t(747) = 9.97$, $p < .001$, $d = 0.89$, and *narcissists*, $t(747) = 23.80$, $p < .001$, $d = 2.13$, and significantly more distance from *bigots* than *narcissists*, $t(747) = 13.82$, $p < .001$, $d = 1.23$. Thus, people preferred the most distance from *psychopaths*, followed by *abusers*, then *bigots*, then *narcissists*.

Additional characteristics. After answering all other questions about the exemplar, participants were asked to list any characteristics that were not initially listed that they thought the exemplar would have, such as their "behaviors, emotions, beliefs, desires, motives, intentions, traits, relationships, abilities, or physical appearance." Participants gave on average 4.61 responses (*SD* = 1.43) per prototype.

Figure 20

Desired Social Distance by Prototype



Note. Displays the comfort participants would feel entering increasingly distant relationships with exemplars of various bad person prototypes.

To clean the responses, I used the same method from Study 1 to drop modifiers, separate compound phrases, collapse synonyms, and drop idiosyncratic responses ($n < 5$). *Bad person* exemplars were described as *depressed, insecure, selfish, mean, angry, rude, unhappy, lonely, impulsive, antisocial*, and *cold* who had endured *childhood trauma and abuse*, are *estranged from their family*, and have *few close relationships* (Table 16). In contrast, respondents described the *good person* exemplars as *caring, kind, empathetic, happy, helpful, friendly, intelligent, honest, loyal, nice, warm, trustworthy*, and *religious*, with many *close and healthy relationships* (Table 17).

Next, I examined how the basic prototypes were described. The *psychopath* exemplars were described as *mentally ill*, *psychopathic*, *antisocial*, *scary*, *emotionless*, *quiet*, *psychotic*, *weird*, *crazy*, *harmful*, *cold*, and *creepy*, who *lack empathy*, have *childhood trauma,* and *no close relationships* (Table 18). The *abuser* exemplars were described as *manipulative*, *abusive*, *aggressive*, *angry*, *easy to anger*, *mean*, *mentally ill*, *selfish*, *misogynistic*, *narcissistic*, *large*, and *impulsive* who have *substance abuse problems, lack empathy,* and were *abused as a child* (Table 19). *Bigot* exemplars were described as *politically conservative*, *white*, *racist*, *close-minded*, *uneducated*, *ignorant*, *homophobic*, *religious*, *rude*, *misogynistic*, *men*, *loud*, and *selfish* who were from the *American south* and *voted for Donald Trump* in the previous elections (Table 20). *Narcissist* exemplars were described as *selfish*, *narcissistic*, *rude*, *rich*, *self-centered*, *insecure*, *mean*, *spoiled*, *egotistical*, *arrogant*, *attractive*, *annoying*, *judgmental*, *manipulative*, who have *few close relationships* and were an *only child* (Table 21). These descriptions illuminated differences in motivations, affect, and appearance that were not apparent by the other measures in the study.

Physical appearance. I extracted the visual words from the free-listed descriptors of exemplars to investigate whether there are differences in the perceived physical appearance of each prototype. Next, I utilized DALL-E 3, an AI program that translates text input into realistic images, to generate images of each prototype. I found differences in the physical appearance of prototypes beyond their gender, race, and age, including their perceived emotions, attractiveness, weight, clothing, socioeconomic status, possessions, and cleanliness. See Table 22 for descriptions and images of basic category prototypes and Table 23 for superordinate category prototypes.

Discussion

In Study 3, participants completed a likelihood rating task to evaluate which unobserved immoral characteristics they inferred to exemplars of various bad person categories. I established that the category structure and dimensional model that I derived in the second study fit reasonably well, especially the dimensional model. And while there was strong evidence to support the hypothesis that the categories *psychopath*, *abuser*, *bigot*, and *narcissist* were subordinate to the superordinate category *bad person*, the pattern of likelihood ratings did not completely support the structure derived in Study 2, nor did it support a simple structure where all basic categories are perceived as equally dissimilar from each other.

For example, based on the category structure derived in Study 2, *narcissists* were expected to be perceived as more similar to *bigots* than *abusers* and *bigots.* However, narcissists were less likely to be inferred bigoted characteristics than *abusers*. Additionally, while *psychopaths* should have been attributed abusive characteristics less than *abusers*, *psychopaths* were just as likely to be attributed abusive characteristics as *abusers* were. These findings may suggest a poor fit of the category structure (or any category structure). Or these findings may be

a product of the items chosen and would not be replicated with different items. For instance, people may attribute the violent aspects of the abuser prototype to psychopaths, but this pattern may not hold for nonviolent aspects (e.g., substance abuse, manipulative, misogynistic). Thus, more research is required, with different items, to rule out this possibility.

Additionally, I tested a prototype model of moral character evaluation against dimensional and dual models and found that a prototype-unidimensional model performed the best overall. The superiority of the performance of the prototype-unidimensional model indicates that people infer unobserved immoral characteristics to targets when both the target and the unobserved characteristics are consistent with a cognitive prototype. In addition, even when the characteristics are not prototype-consistent, people make inferences that others have immoral characteristics when the characteristics are close to the person on the morality dimension, but not when the characteristics are close to the target on other dimensions such as intelligence or sociability. The fact that the most complex model (the dual prototype and multidimensional model) did not provide additional explanatory power ran counter to my expectations. This would imply that the stupidity or sociability of an immoral characteristic does not affect the likelihood of inference. For instance, if a person is perceived to hold ignorant, morally abhorrent beliefs, they would not necessarily be transferred other ignorant characteristics if they are not prototype-consistent.

Table 16

Participant Generated Unobserved Characteristics of Bad Person Exemplar

Characteristic	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*
Depressed	37	Childhood abuse	15	Self-centered	8	Narcissistic	6	Poor	5
Insecure	30	Estranged from family	15	Single	8	Quick to anger	6	Skinny	5
Selfish	29	Cold	13	Tall	8	Quiet	6	Stubborn	5
Mean	28	Loner	12	Atheist or not religious	8	Shy	6	Substance abuse	5
Angry	25	Mentally ill	11	Careless	8	Superiority complex	6	Uneducated	5
Rude	24	No close relationships	11	Immoral	7	Unkind	6	Adrenaline junkie	4
Unhappy	24	Dark clothing	10	Judgmental	7	Greedy	5	Bullied	4
Lonely	23	Young	10	Man	7	Homophobic	5	Goth or Emo	4
Impulsive	21	Aggressive	9	Pessimistic	7	Low self-esteem	5	Incel	4
Childhood trauma	18	Emotionless	9	Sociopath	7	Manipulative	5	Lacks empathy	4
Few close relationships	17	Hateful	13	Criminal	6	Unintelligent	5	Misogynistic	4
Antisocial	15	Racist	9	Introverted	6	Uses social media often	5	Troubled	4

Table 17

Participant Generated Unobserved Characteristics of Good Person Exemplar

Characteristic	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*
Caring	51	Healthy relationships	16	Funny	9	Respectful	7	Integrity	5
Kind	40	Trustworthy	15	Generous	9	Welcoming	7	Trusting	5
Empathetic	33	Religious	14	Patient	9	Small group of close friends	6	Understanding	5
Happy	31	Smiling	14	Gentle	8	Good student	6	Well liked	5
Helpful	29	Sweet	14	Genuine	8	Partnered	6	Ally	4
Friendly	27	Selfless	13	Good listener	8	Social	6	Clean	4
Intelligent	22	Thoughtful	13	Close with family	8	Volunteers	6	Confident	4
Many friends	21	Good childhood	12	Loving	8	Bubbly	5	Extroverted	4
Honest	20	Hardworking	12	Positive	8	Charismatic	5	Glasses	4
Loyal	20	Educated	11	Attractive	7	Considerate	5	Humble	4
Nice	19	Outgoing	11	Motivated	7	Emotional	5	Successful	4
Warm	17	Reliable	11	Open-minded	7	Good friend	5	Woman	4

Table 18

Participant Generated Unobserved Characteristics of Psychopath Exemplar

Characteristic	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*
Mentally ill	45	Harmful	16	Angry	11	Aggressive	7	Unstable	6
Psychopathic or Sociopathic	45	Cold	15	Childhood abuse	11	Gross	7	Animal abuse	5
No close relationships	25	Creepy	14	Dirty	11	Immoral	7	Awkward	6
Antisocial	23	Calculated	13	Murderer or serial killer	11	Lonely	7	Charismatic or charming	6
Scary	22	Cruel	13	Dangerous	10	Unattractive	7	Disgusting	6
Emotionless	20	Insane	13	Evil	10	Criminal	6	Horrible	6
Childhood trauma	19	Intelligent	13	Impulsive	10	Lives alone	6	Skinny	6
Quiet	19	Manipulative	13	Loner	10	Mean	6	Tall	6
Lacks empathy	17	Sadistic	13	Violent	10	Narcissistic	6	Abusive	5
Psychotic	17	Estranged from family	12	White	10	No morals	6	Apathetic	5
Weird	17	Few relationships	12	Man	9	Selfish	6	Isolated	5
Crazy	16	Single	12	Strong	7	Unkempt	6	Rapist	5

Table 19
Participant Generated Unobserved Characteristics of Abuser Exemplar

Characteristic	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*
Substance abuse	44	Large	16	Childhood trauma	12	Abusive to partner	7	Emotional	5
Manipulative	29	Lacks Empathy	15	Depressed or unhappy	11	Estranged from family	7	Evil	5
Abused as a child	26	Impulsive	14	Tall	11	Homophobic	6	Goes to bars	5
Abusive	25	Rude	14	Dangerous	10	Power hungry	6	Immoral	5
Aggressive	23	Scary	14	Poor	10	Racist	6	Stubborn	5
Angry	22	Cruel	13	Sad	10	Rapist	6	Self-centered	5
Quick or easy to anger	22	Emotionless	13	Controlling	9	Skinny	6	Tattoos	5
Mean	20	Harmful	13	Loud	9	White	6	Attractive	4
Mentally ill	20	Insecure	13	Conservative	8	Bad	5	Criminal	4
Selfish	19	Man	13	Cold	7	Dirty	5	Prison	4
Misogynistic	17	Strong	13	Sadistic	7	Disrespectful	5	Unemployed	4
Narcissistic	17	Violent	13	Unstable	7	Egotistical	5	Yells	4

Table 20

Participant Generated Unobserved Characteristics of Bigot Exemplar

Characteristic	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*
Conservative/ republican	78	Loud	20	Mean	12	Self-centered	9	Substance abuse	5
White	52	Selfish	17	Aggressive	11	Egotistical	7	Drives a truck	5
Racist	47	Southern	17	Overweight	11	Traditional	7	Unhealthy	5
Close-minded	41	Christian	16	Not exposed to minorities	10	Cold	6	Unkind	5
Uneducated	34	Friends with like-minded people	16	Opinionated	10	Insecure	6	Xenophobic	5
Ignorant	31	Angry	15	Rural	10	Lacks empathy	6	Country	4
Homophobic	27	Unintelligent	14	Transphobic	10	Rich	6	Hardheaded	4
Religious	25	Owns a gun	13	Vocal about beliefs	10	Bigoted	5	Mullet	4
Rude	25	Judgmental	13	Arrogant	9	Online	5	Quick to anger	4
Misogynistic	23	Old	13	Grew up exposed to bigoted beliefs	9	Patriotic	5	Red state	4
Voted for Donald Trump	22	Stubborn	13	Hateful	9	Privileged	5	Superiority complex	4
Man	20	Heterosexual	12	Pro life	9	Redneck	5	Unattractive	4

Table 21

Participant Generated Unobserved Characteristics of Narcissist Exemplar

Characteristic	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*	*Characteristic*	*n*
Selfish	51	Few close relationships	16	Confident	9	Quick to anger	6	Prideful	5
Narcissistic	35	Judgmental	16	Entitled	9	Short	6	Self-absorbed	5
Rude	35	Manipulative	15	Lonely	9	Stuck up	6	Poor listener	4
Rich	32	Only child	14	Attention seeking	8	Tall	6	Greek life	4
Self-centered	25	Ignorant	13	Unintelligent	8	Aggressive	5	Lazy	4
Insecure	23	Lacks empathy	13	Dumb	8	Big ego	5	Materialistic	4
Mean	23	Close-minded	12	Belittles others	7	Blonde	5	Rude to service	4
Spoiled	20	Popular	12	Critical	7	Careless	5	Snarky	4
Egotistical	19	Stubborn	12	Difficulty maintaining relationships	7	Cold	5	Uses social media often	4
Arrogant	18	Vain	12	Loud	6	Condescending	5	Talkative	4
Attractive	17	Privileged	11	Unkind	6	Conservative	5	Well dressed	4
Annoying	16	Athletic	9	Immature	6	Inconsiderate	5	Young	4

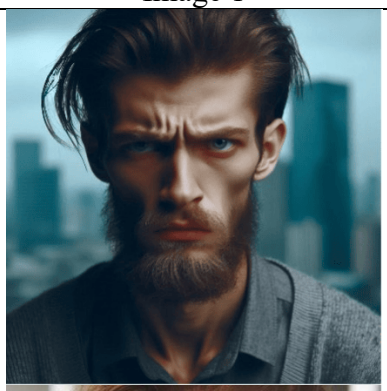
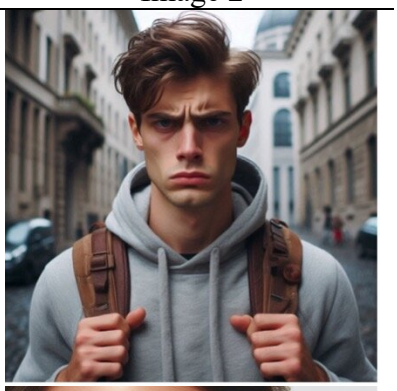
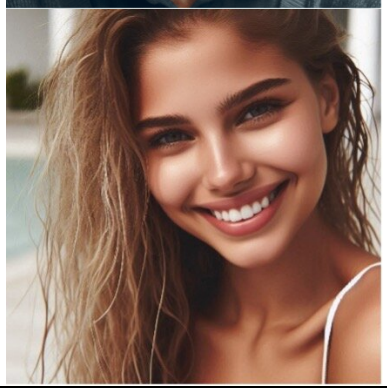
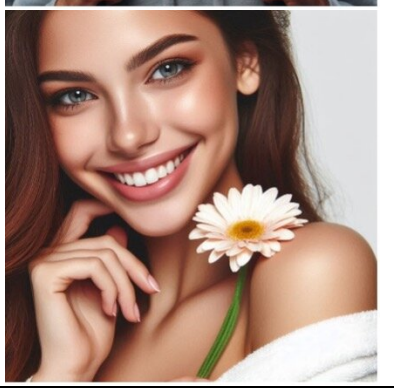
Table 22

AI Image Renderings from Free-Listed Appearance Descriptions of Basic Prototypes

Prototype	Description	Image 1	Image 2
Psychopath	*An emotionless, cold, skinny, unattractive, strong, dirty, white man with greasy, unkempt hair*		
Abuser	*An angry, large, strong, tall, poor, attractive, white man with tattoos*		
Bigot	*A Southern, old, overweight, patriotic white man who owns a gun and a pickup truck*		
Narcissist	*A rich, attractive, athletic, blonde, well-dressed man/ woman*		

Table 23

AI Image Renderings from Free-Listed Appearance Descriptions of Superordinate Prototypes

Prototype	Description	Image 1	Image 2
Bad Person	*An unhappy, angry, emotionless, tall, poor, skinny, young man*		
Good Person	*A happy, attractive, clean woman*		

However, we should be cautious not to overinterpret these findings. The prototype model performed best in this study but may not perform as well as dimensional models when targets are not exemplars of categories and/or when different characteristics are used to define the exemplars. For example, if targets are not defined by prototype-consistent characteristics, perceivers may need to rely on more effortful processes to make judgments. Additionally, these findings do not imply that competence and sociability are not important aspects of person perception. They just may not be as important when making inferences about immoral characteristics. Or, they may not be as important when making inferences about immoral information, about exemplars of cognitive prototypes.

Finally, this research uncovered differences in our perception of the morality, sociability, competence, gender, race, age, and physical appearance of different types of bad people. There are differences in how biologically driven and immutable we perceive the characteristics of each prototype to be, and the social distance we feel comfortable keeping from them. Overall, the third study found strong evidence that we have cognitive prototypes of basic categories of bad people, and that those cognitive prototypes influence the assumptions we make about the immoral characteristics of others. Further, I was able to closely examine the mental representation of four of these moral person prototypes by analyzing free responses of additional characteristics attributed to exemplars of moral person categories.

The cognitive prototype for the *psychopath* was an *emotionless, cold, skinny, unattractive, strong, dirty, white man in his 20s or 30s with greasy, unkempt hair*. The psychopath is extremely immoral, somewhat unintelligent, and moderately unsociable. He has *no close relationships,* as he is considered someone to be kept at a distance more than any other type of bad person. He is *mentally ill*, *psychopathic*, *antisocial*, *scary*, *quiet*, *unempathetic*, *psychotic*,

weird, *crazy*, *harmful*, *creepy, calculated,* and *cruel*. His characteristics are perceived as hardwired, unchangeable, and essential to his identity.

The cognitive prototype for the *abuser* is a *large, strong, tall, poor, attractive white man in his 30s or 40s with tattoos.* The abuser is extremely immoral, moderately unintelligent, and slightly unsociable. He is *manipulative*, *abusive*, *aggressive*, *angry*, *mean*, *mentally ill*, *selfish*, *misogynistic*, *narcissistic*, and *impulsive*. He has *substance abuse problems, is easy and quick to anger,* and is perpetuating *the abuse he endured as a child*. His characteristics are perceived as moderately hardwired and unchangeable, and one should strongly avoid entering romantic or platonic friendships with him.

The cognitive prototype for the *bigot* is *a Southern, old, overweight, patriotic, white man who owns a gun and a pickup truck.* The *bigot* is moderately immoral, moderately unintelligent, and slightly social. He has relationships with *like-minded people,* and is described as *politically conservative*, *racist*, *close-minded*, *uneducated*, *ignorant*, *homophobic*, *religious*, *rude*, *misogynistic*, *loud*, and *selfish*. His characteristics are perceived as the least hardwired and most changeable of the cognitive prototypes studied, as they are believed to stem from *growing up surrounded by bigoted beliefs* and from *lack of exposure to minorities*. Unlike psychopaths (and to a lesser extent, abusers), people were generally more comfortable being coworkers or neighbors with a bigot, but still uncomfortable forming close relationships like marriage or close friendship, especially relative to the good person.

Finally, the cognitive prototype for the narcissist is a *rich, attractive, athletic, blonde, well-dressed man or woman in their 20s.* The *narcissist* is somewhat immoral, slightly unintelligent, and slightly sociable. He or she is *selfish*, *narcissistic*, *rude*, *self-centered*, *insecure*, *mean*, *spoiled*, *egotistical*, *arrogant*, *annoying*, *judgmental*, and *manipulative.* His or

her immoral characteristics are generally perceived to be mutable and not hardwired, and respondents said they are more comfortable getting close to these individuals compared to the other prototypes studied. While the narcissist should certainly be avoided more than a good person, many indicated comfort with more causal acquaintanceships.

V. GENERAL DISCUSSION

In a series of three studies, I found that people distinguish between multiple types of bad people, that these different types of bad people are represented in the mind as cognitive prototypes, and that these cognitive prototypes guide the inferences we make about others' immoral characteristics. I found strong evidence for a prototype model over the unidimensional model of moral character evaluation widely used in the literature. Specifically, people attribute unobserved immoral characteristics to targets when both the targets and the characteristics are both consistent with the same cognitive prototype. I also found evidence for a unidimensional component, implying that people attribute characteristics that are similar in morality to the target, even when they are prototype inconsistent. In addition, I closely examined four cognitive prototypes recognized by the population of interest: *psychopaths*, *abusers*, *bigots*, and *narcissists* and found differences in perceived morality, intelligence, sociability, gender, age, race, physical appearance, beliefs regarding how stable or hardwired the immoral characteristics are, and desired social distance from them highlight why people form basic categories to distinguish between different types of bad people and what are the nature of those differences.

Implications

Research on moral character evaluation has been dominated by a simple, unidimensional model where targets can be placed and ranked against each other on a moral dimension from *extremely immoral* to *extremely moral* (Brandt & Reyna, 2011). The utilization of this simple model of moral character evaluation has likely been for ease of administration, but this research challenges its strength. In particular, when a moral person prototype is activated, people may assume that the target has unobserved prototype-consistent characteristics, even when those unobserved characteristics are not close in proximity to the target on the moral dimension.

Given that harmless or morally irrelevant characteristics (e.g., appearance, social identities, desires) are important features of cognitive prototypes, a target without evidence of past harm may be inferred poor character on the basis of proximity to a cognitive prototype. For example, people infer poor moral character or untrustworthiness to those who are less attractive and have more masculine and less child-like facial features (Rosa et al., 2021; Todorov et al., 2015). Additionally, behaviors that are relatively harmless (e.g., masturbating with a chicken from the supermarket) often lead to harsher moral condemnation than the action itself warrants (e.g., Tannenbaum et al., 2011; Uhlmann & Zhu, 2014). A prototype model may be able to explain why harmless behaviors or personal characteristics that are not relevant to morality can lead to a negative character evaluations. Since moral character judgements are largely reliant on a person's unobservable or unobserved immoral characteristics (e.g., future behavior, private behavior, inferred cognitions; Cohen et al., 2014), inferring that a target has immoral characteristics will lead to a negative judgment of their moral character.

Additionally, inconsistent findings in the literature may be driven by studying bad people as a monolith, rather than acknowledging that people perceive different types of bad people who are perceived and interacted with differently. For example, a previous study concluded that both adults and children view morally good characteristics as more biologically driven and stable than bad characteristics (Heiphetz, 2019). This research replicated this finding for the superordinate categories, but additionally found that people perceived the characteristics of *psychopaths* and *abusers* as more or just as biologically driven and stable as *good people*.

Additionally, people are generally biased towards perceiving unattractive people as less moral and trustworthy than attractive people (Dion et al., 1972; Klebl et al., 2021; Rosa et al., 2021). Contrary to this finding, I found that while psychopaths were described as unattractive,

other *bad person* prototypes (e.g., *narcissists, abusers*) were described as attractive. Thus, studying basic cognitive prototypes individually rather than *bad people* as a homogeneous group provides a more nuanced understanding of how people perceive and interact with different kinds of *bad people*.

Further, this work was the first rigorous investigation of bad person cognitive prototypes. Previous work had theorized distinct categories of immoral behaviors and immoral persons. For example, Moral Foundations Theory posits that there are at least six categories of moral violations: harm, fairness, proportionality, disloyalty, disobedience, and impurity (Atari et al., 2023; Graham et al., 2013). Since past research has found that liberals are almost entirely concerned with harm and fairness and unconcerned with proportionality, disloyalty, disobedience, and impurity (Graham et al., 2009), we would have expected categories to emerge in the liberal sample for a harm violator and a fairness violator (or multiple different kinds of harm and fairness violators). Another model proposed by Hartman et al. (2022) theorized that we perceive moral person categories that differ in morality and agency, and thus we would have expected categories like *opportunist* or *sinner* to emerge.

What I found did not support either of these taxonomies for moral persons. I found that liberal students distinguished between seven different types of bad people: *psychopath, criminal, abuser, bully, manipulator, bigot*, and *narcissist*. While some types of bad people are perceived to violate more fairness norms than harm norms (e.g., bigot, manipulator) and some types of bad people are perceived to violate more harm norms than fairness norms (e.g., abuser, psychopath), these categories do not align as neatly as one might expect and not all characteristics derived in Study 1 fit neatly into these two categories (e.g., cannibalism and necrophilia are purity norms).

Future research can further investigate these associations by asking respondents to rate the likelihood that each prototype would violate moral norms of from each of the six moral domains.

To the extent that our perceptions partly reflect real personality differences in others, we may have expected to observe prototypes that are caricatures of people with antagonistic personality traits. For example, research on the Dark Tetrad would predict four categories of bad people: *psychopaths*, *Machiavellians*, *narcissists*, and *sadists* (Međedović & Petrović, 2015). This prediction was partially supported, but perception differs from reality in a few important ways. First of all, psychopathy as defined by personality psychologists typically centers on impulsive behavior, emotional deficiencies, and a desire for revenge (Dolan & Fullam, 2006; Hare et al., 1990; Holmqvist, 2008), rather than unusual moral norm violations like necrophilia, murder, and cannibalism. According to the membership values from the C-Means analysis, the characteristics *lack of guilt* and *lack of empathy*, which are central to the personality trait psychopathy, were more strongly associated with the *abuser* and *criminal* prototypes than the *psychopath* prototype.

Additionally, the manipulator prototype resembles a theorized Machiavellian, who manipulates and exploits others for personal gain (Jones & Paulhus, 2014; Monaghan et al., 2016; Rauthmann & Kolar, 2012) and the narcissist prototype aligns with the narcissism trait, a person with an inflated sense of self who prioritizes themselves over others (Bushman et al., 2003; Gentile et al., 2013). A sadist category did not appear, but rather, the three sadistic items (i.e., animal sadism, sexual sadism, and sadism) were split between the psychopath and abuser prototypes. Finally, many more categories emerged that did not follow from personality theory, such as *bigots*, *bullies*, and *criminals*.

In addition, this work has important implications for how we perceive and interact with others in the real world. When prototypes are activated by harmless personal characteristics, identities, or behaviors that are characteristic of a cognitive prototype, people are wrongfully stigmatized with a *bad person* label. For instance, autistic individuals are perceived as more deceptive than neurotypical people when telling the truth because they tend to behave in unexpected or atypical ways, e.g., flat affect, gaze aversion, poor reciprocity (Lim et al., 2022). Additionally, sexual and gender minorities have a long history of being equated to *pedophiles*, *child predators*, or *groomers*, even though they do not offend at higher rates than cisgender heterosexuals (Lavietes, 2022). The attribution of immoral characteristics to these individuals may be explained by proximity to an existing cognitive prototype.

Additionally, this research has important implications for the justice system. People seek to punish others in proportion to their perceived moral character (Carlsmith et al., 2002) and perceive unfair procedures as fair when they lead to a morally desirable outcome (Skitka et al., 2008). Thus, it follows from this research that information that is harmless or irrelevant to the evidence of the case can lead to harsher punishments when they are consistent with a bad person prototype. For example, harsher sentences were recommended for defendants of financial crimes when they were perceived to have psychopathic traits, e.g., lack of remorse, lack of guilt, failure to accept responsibility, despite all other facts of the case being held constant (Cox et al., 2016), indicating that proximity to a prototype can lead to an unfair trial. In addition, since people hold extremely strong attitudes that the immoral characteristics of psychopaths and abusers are hardwired and unchangeable, people may be more likely to recommend the death penalty to defendants they think are members of these categories, even if membership is mostly or entirely based on personal characteristics or harmless actions (Martin et al., 2022).

Limitations and Future Directions

Study Good People

My dissertation project focused on the basic categories of *bad people* and should be expanded on in the future by eliciting the basic categories of *good people*. I chose to study bad people for two reasons. First, in the last two decades, moral psychologists have focused on explaining why people condemn harmless wrongs[11]: behaviors that are perceived to be morally wrong but do not harm anyone, such as consensual incest or eating your pet dog after a tragic accident (Haidt, 2001; Haidt et al., 1993). Researchers studying moral character evaluation have expanded these themes to study why people who commit harmless wrongs are condemned at a seemingly disproportionate level to the consequences of their actions (Tannenbaum et al., 2011; Uhlmann & Zhu, 2014). My work fits alongside this research as an explanation for how harmless wrongs can cause condemnation through the activation of cognitive prototypes. Second, I decided to study prototypes of *bad people* because assumptions that a person is a *bad person* based on morally irrelevant personal characteristics (e.g., sexual or gender identity, weight) or harmless behaviors can be very stigmatizing and could even impact a defendant's right to a fair trial. Thus, it made sense to start with the set of prototypes with high relevance to the literature and to society.

However, there is another significant error in moral person perception: believing that someone is moral when they are not. Trust violations can be shocking and heartbreaking, such as

[11] An alternate perspective is that there is no such thing as a harmless wrong, since people often maintain that they perceive harm in harmless wrong scenarios (Royzman et al., 2015).

.

discovering you are victim of a con, or learning a priest has molested a child. Researchers have identified factors that lead to mistaken trust (Yip & Schweitzer, 2015), and when people are given more praise than their moral behavior deserves (Schaumberg & Mullen, 2017). Identifying the structure and properties of good person categories can contribute to this developing literature by helping to identify when and why these errors are made.

By deriving *good people* basic categories, we can compare *bad person* and *good person* basic categories in number and distinctiveness. I expect the negative moral space to be more abundant (i.e., more categories), as it's especially adaptive to be able to distinguish between different types of bad people. To survive and flourish in society where some people cooperate while others harm or take advantage of others, one must make accurate predictions regarding the range of likely behaviors (e.g., aggression, manipulation, sexual violence) and targets (e.g., children, women, animals) of others. However, discerning different types of good people does not hold the same adaptive advantage.

A disparity between the number of good and bad person categories would also be consistent with a general negativity bias across different aspects of cognition (Rozin & Royzman, 2001) and specifically in moral impression formation (Coovert & Reeder, 1990; Lupfer et al., 2000; Rusconi et al., 2020). Further, since people may be perceived as *good* by default (Strohminger et al., 2017), moral character evaluations may be more frequently employed when we are presented with evidence that a target's morality is "broken" (Helzer & Critcher, 2018).

Initial research testing whether people perceive different types of good people found that people distinguish between *brave*, *caring*, and *just* people (Walker & Hennig, 2004). However, the authors limited traits to ones elicited by asking people to describe a brave person, caring

person, and just person, and the authors did not consider other important cues of moral character besides traits (e.g., cognitions, affect), so additional research on the topic is warranted to see if more than three basic categories emerge. Additionally, Hartman et al. (2022) theorize that there may be a third superordinate category relevant to morality, *victims*, that would be an interesting future topic to explore further.

Study Other Populations

A second limitation is that I used a student sample. I expect moral person concepts and their cognitive representations to vary across cultures, and so I did not attempt to derive a cross-culturally valid taxonomy of moral person concepts. Rather, I derived a taxonomy of immoral person concepts from a population of mostly liberal, white, educated psychology students, generally aged 18-24 (i.e., Gen Z). Given that the focus of my dissertation project is to demonstrate whether basic categories exist in the moral domain, examine their properties, and to test models of moral judgment, it was necessary to understand the taxonomic structure of the participants within the specific culture sampled. Due to the sampling method, I cannot do a rigorous test of differences in basic categories across generations, political affiliation, and cultures. However, Study 1 did find preliminary evidence to confirm that we should expect individual differences in conceptions of a bad person.

The moral person taxonomy derived in this research is a product of the cultural and historical period in which the data were collected, and we should not interpret it as a universal and timeless taxonomy. For instance, the increase in popularity of terms like *abuser*, *narcissist*, and *psychopath* may be shaping how young adults with access to social media form moral categories (and subsequently, make judgments of others) that is certainly ungeneralizable to a different time period or age group. Future research should derive taxonomies from other

populations that vary in moral values, experience, and domain expertise (e.g., conservatives, older adults, police officers) to learn what categories emerge in different populations and to investigate whether there are universal *bad person* categories.

An individual's cultural background and experiences (e.g., domain expertise) significantly impact their taxonomic structure in a given domain (Medin & Atran, 2004). Natural categories partially mirror the correlational structure of the environment, so people with different environments are expected to have formed different categories (Rosch, 1978). Just as someone living in the United States is expected to have a different cognitive representation of a *chair* than someone born in Japan or India based on the chairs that they have been exposed to in their environment, an undergraduate student in the United States is expected to have a different conception of a *bad person* than people in other parts of the world.

Additionally, there may be slight cultural variation in how individuals from different parts of the world categorize new stimuli into groups. In one study, participants were directed to choose whether to classify targets into one of two groups, where one group followed a unidimensional rule (i.e., group members shared a single feature with the target) and the other group shared a family resemblance (i.e., group members did not all share the same feature with the target but were holistically similar because they shared a large number of features with the target) (Norenzayan et al., 2002). The study found that East Asians were more likely to judge that a target was similar to the category if it was holistically similar (i.e., family resemblance) and Westerners were more likely to judge it as similar if individual features were similar. If this finding extends to the moral domain, East Asians may be more holistic in their classification of someone as a *bad person* or *type of bad person*, whereas Westerners may have a more rigid, rule-based inclusion criteria.

It is well established that conservatives and those outside Western cultures have different moral concerns than liberal students (Graham et al., 2009; Haidt et al., 1993; Janoff-Bulman & Carnes, 2013; Schein & Gray, 2015). For instance, we might expect to see the emergence of moral person categories dedicated to purity or divinity violators (e.g., *prostitutes*, *atheists*) in a conservative but not liberal population and the emergence of moral categories dedicated to equality violators (e.g., *bigots*) in a liberal but not conservative population.

Further, liberals tend to prioritize prescriptive regulation, focusing on what *we ought to do* to provide for ourselves, others, and our groups, while conservatives more often emphasize proscriptive regulation, concentrating on what we *ought not to do* to protect ourselves, others, and our groups (Janoff-Bulman & Carnes, 2013). For instance, conservatives are more likely to agree with items such as "It's particularly important to me to demonstrate self-control in the face of temptation." Thus, conservatives would be more likely to consider people who fail to protect themselves or others, or lack adequate self-control (e.g., *cowards, lazybones, sex offenders, welfare queens, addicts*) as immoral. These categories may have not been formed in the present research because related items were not endorsed highly nor spontaneously generated by the primarily liberal, young participants in the studies.

Additionally, I would expect conservatives to have more basic categories of bad people than liberals, which is consistent with their additional domains of moral concern (Graham et al., 2009) and heighted threat perception (Crawford, 2017; Jost et al., 2003). Thus, future research should use a large, representative sample of US adults to compare which basic categories emerge for liberals and conservatives. Additionally, future research could see to what extent basic categories are cross-culturally stable by sampling across different cultural variables.

Design an Intervention for Moral Perception Errors

Given that wrongfully believing that someone is a bad person based on personal characteristics or harmless information can be incredibly stigmatizing, an important next step is to dedicate research to preventing or reversing these errors. If we understand how someone is classified, we can prevent or change that classification. For example, by providing explicit counterevidence to the prototype, concurrently or after the prototype-consistent information, we could prevent or change a target's classification so that they are not inferred unobserved immoral characteristics and do not receive endorsements of punishment that are not in proportion to their behavior. For example, for defendants who display characteristics that would activate the *psychopath* prototype, it would be important to provide counterevidence to the prototype for the defendant to receive a fair trial. Future research should test this intervention by examining whether prototype-inconsistent information can overturn or prevent a disproportionally harsh condemnation.

Conclusion

The concept *bad person* is not a monolith, rather it is comprised of many different stones. Some bad people are seen as more malevolent, more idiotic, or more socially isolated than others. Some bad people should always be avoided, while others can be acquaintances, coworkers, or neighbors. Some bad people are believed to be born that way, while others are believed to be created through abuse or trauma. In order to know what to expect and how to act when we meet people like these, we form moral person categories (e.g., *psychopath*, *abuser*, *bigot*, *narcissist*), and represent those categories in our minds as cognitive prototypes, or average examples extracted from previous experiences with category members and augmented with social learning. When we meet someone new who is similar to our cognitive representation of a

category, we make assumptions about their unobserved or unobservable immoral attributes, and in turn, evaluate their moral character based in part on those assumptions. Because harmless or morally irrelevant information can trigger the activation of a prototype, systematic errors of assuming an innocent target is bad or assuming a guilty target is good can occur. I have found that cognitive prototypes play an important role in the evaluation of moral character and must continue to be studied to fully understand why errors occur and how to prevent or reverse them.

APPENDIX A: INITIAL 70 BAD PERSON CHARACTERISTICS

Instructions: For each of the following characteristics, please rate the extent to which you would consider someone a **bad person** (or good person) if they expressed the characteristic.

1. Does not feel guilt or remorse when they hurt others.
2. Has a temper (i.e., gets angry easily).
3. Is spiteful and vengeful.
4. Does not feel empathy or concern when people are suffering.
5. Kills people out of boredom.
6. Does not feel love or attachment towards other people.
7. Does not express gratitude when others help them.
8. Feels pleasure or delight when they hurt other people.
9. Has an obsessive desire to light things on fire (i.e., pyromania).
10. Enjoys hurting animals.
11. Has racist or white nationalist beliefs (example: white people are better than people from other races).
12. Has sexist beliefs (example: women are inferior to men).
13. Has homophobic beliefs (example: gay people should have fewer rights).
14. Has transphobic beliefs (example: transgender people should have fewer rights).
15. Acts hostile towards other races.
16. Discriminates against women.
17. Bullies or harasses gay people.
18. Endorses laws that take away rights from transgender people.
19. Believes that their country is better than other countries.
20. Endorses laws that will deport or make living in the US difficult for immigrants.
21. Endorses invading, occupying, and conquering other countries.
22. Believes that they are better than other people.
23. Only cares about themselves.
24. Never accepts responsibility for the negative consequences that follow from their actions.
25. Always wants to be the center of attention.
26. Brags constantly about their accomplishments.
27. Feels entitled to money, recognition, or sexual partners without doing anything to earn them.
28. Has very unusual sexual fetishes.
29. Fantasizes about (but doesn't act on) having sex with children.
30. Fantasizes about defiling or achieving sexual gratification from dead people.
31. Needs to fantasize about their partner suffering to achieve sexual gratification (i.e., sexual sadism).
32. Molests children.
33. Is never faithful to their sexual partners.
34. Fantasizes about sexual assaulting people.
35. Uses coercion, drugs, or physical violence to sexually assault people.
36. Has frequent fantasies about cheating on their partners.
37. Is addicted to drugs or other harmful substances.
38. Puts selfish desires over obligations to their family.

39. Tries to control and manipulate other people to get what they want.
40. Is very judgmental and mean to others.
41. Always gossips about other people.
42. Lies all the time for no apparent reason (i.e., pathological lying).
43. Cannot resist urges to steal items they don't need (i.e., kleptomania).
44. Does not believe in God or any religious or spiritual force.
45. Craves eating the flesh of other people (i.e., cannibalism).
46. Seeks to be in a position of power over others.
47. Is obsessed with having more than others have.
48. Takes advantage of other people's misfortunes for profit.
49. Values money over human lives.
50. Trades sex for money.
51. Kills people for money.
52. Scams vulnerable or naïve populations (example: the elderly) out of money.
53. Does not respect the privacy of others (examples: stalking, voyeurism).
54. Is very mean to others with less power (example: bullying).
55. Does not care about the negative impact they make to the environment or nature (example: littering).
56. Has trouble keeping friends.
57. People can never count on them to keep their promises.
58. Is not loyal to their friends and family.
59. Is not patriotic or supportive of their country.
60. Cannot control their impulses.
61. Harms their own children (examples: abuse, neglect).
62. Physically or emotionally abuses their spouse/partner.
63. Endorses violence to bring awareness to a political or moral ideology (example: terrorism).
64. Thinks people who deserve it should die.
65. Will always follow what other people do, even if means doing something wrong.
66. Believes violence can be justified if it makes the world a better place.
67. Is violent or destructive towards property.
68. Is willing to cheat and lie to get ahead.
69. Has no regard for the law.
70. Does not follow the moral standards that they endorse for others (i.e., hypocrisy).

APPENDIX B: FINAL 50 BAD PERSON CHARACTERISTICS

1. Physically and/or emotionally abuses their partner/ spouse (i.e., domestic violence). (partner abuse)
2. Abuses their children (e.g., assault, neglect, emotional manipulation). (child abuse)
3. Is not loyal to those close to them. (disloyal)
4. Bullies people they perceive as weak or vulnerable. (bullying)
5. Always cheats on their romantic partners. (infidelity)
6. Believes that gay people should have fewer rights. (homophobia)
7. Does not feel empathy or concern when people are suffering. (no empathy)
8. Murders people out of boredom. (murder)
9. Wants to invade, occupy, and conquer other countries. (imperialism)
10. Seeks to be in a position of power over others. (power hungry)
11. Believes women are inferior to men. (sexism)
12. Believes white people are inherently superior to people from other races. (racism)
13. Is always looking to instigate a fight. (aggressive)
14. Tries to control what others (e.g., their children, their partner) do. (controlling)
15. Does not care about other people. (does not care about others)
16. Thinks the world revolves around them. (self-centered)
17. Feels entitled to money, recognition, or sexual partners without doing anything to earn them. (entitlement)
18. Hoards far more resources than they could ever need for themselves. (greedy)
19. Hates the world and everyone in it. (hateful)
20. Does not respect the privacy of others (examples: stalking, voyeurism). (intrusive)
21. Is very judgmental and critical about other people. (judgmental)
22. Has no regard for the law. (lawbreaker)
23. Is willing to lie to get ahead. (deceitful)
24. Manipulates other people to get what they want. (manipulation)
25. Is rude to other people. (rude)
26. Never accepts responsibility for the negative consequences that follow from their actions. (no responsibility)
27. Does not feel any guilt or remorse when they hurt others. (no guilt or remorse)
28. Physically assaults other people. (physical violence)
29. Molests children. (child molestation)
30. Watches child pornography. (pedophilia)
31. Uses coercion, drugs, or physical violence to sexually assault people. (rape)
32. Is psychotic, having delusions that do not match reality (e.g., the devil is giving them orders). (psychotic)
33. Derives pleasure from hurting animals. (animal sadism)
34. Needs their partner to suffer to achieve sexual gratification. (sexual sadism)
35. Feels pleasure or delight when they hurt other people. (sadism)
36. Always prioritizes themselves over other people. (selfish)
37. Is spiteful and vengeful. (spitefulness)
38. Steals items they don't need from stores. (kleptomania)
39. Believes that they are better than other people. (superiority complex)
40. Believes transgender people should have fewer rights. (transphobia)

41. Is nice to you but talks badly about you behind your back. (two-faced)
42. Strongly dislikes immigrants and doesn't think people should move to their country. (xenophobia)
43. Craves eating the flesh of other people (i.e., cannibalism). (cannibalism)
44. Scams vulnerable or naive populations (example: the elderly) out of money. (scamming)
45. Fantasizes about defiling or achieving sexual gratification from dead people. (necrophilia)
46. Takes advantage of others for profit. (exploitation)
47. Extremely envious of other people's happiness and success. (jealousy)
48. Lies all the time for no apparent reason. (lying)
49. Takes advantage of other people's kindness without giving anything in return. (free loading)
50. Frequently and excessively uses drugs and/or alcohol. (drug use)

APPENDIX C: ESSENTALIST BELIEFS SCALE

Instructions: Please rate the extent to which you agree with the following statements.

1. [Exemplar] has these characteristics because of their biological makeup.
2. [Exemplar] has these characteristics because of the circumstances they grew up in. (Reverse)
3. [Exemplar] will have these characteristics their entire life.
4. From this information, I can predict how [exemplar] would behave in many other situations.
5. I think that [exemplar] could change these characteristics if they wanted to. (Reverse)
6. These characteristics are extremely important to who [exemplar] is as a person.

APPENDIX D: SOCIAL DISTANCE SCALE

Instructions: Please rate the extent to which you agree with the following statements.

1. I would want [exemplar] to marry someone in my family.
2. I would want to be close friends with [exemplar].
3. I would want to be acquaintances or casual friends with [exemplar].
4. I would want to be coworkers with [exemplar].
5. I would want to live in the same neighborhood as [exemplar].
6. I would want to live in the same country as [exemplar].

APPENDIX E: MEAN AND STANDARD DEVIATION TABLES FOR LIKELIHOOD RATINGS

Table 24

Likelihood Means (Standard Deviation) of Having Psychopath Characteristics

	Exemplar					
Characteristic	Psychopath	Abuser	Bigot	Narcissist	Bad	Good
Animal Sadism	10.58 (0.76)	8.71 (2.62)	5.99 (2.51)	4.61 (2.26)	6.68 (2.52)	1.49 (1.09)
Necrophilia	10.22 (1.13)	8.71 (1.98)	4.42 (2.61)	3.44 (2.37)	5.71 (2.82)	1.53 (1.22)
Murder	9.36 (2.39)	6.75 (2.67)	3.09 (2.01)	2.77 (1.87)	5.06 (2.87)	1.22 (0.74)
Cannibalism	8.79 (2.48)	4.43 (2.56)	2.24 (1.83)	2.00 (1.56)	3.60 (2.62)	1.40 (1.14)
Psychopath	9.77[a] (1.95)	7.22[b] (3.02)	4.01[c] (2.68)	3.26[d] (2.25)	5.32 (2.91)	1.41 (1.06)

Table 25

Likelihood Means (Standard Deviation) of Having Abuser Characteristics

	Character Exemplar					
Characteristic	Psychopath	Abuser	Bigot	Narcissist	Bad	Good
Physical Assault	10.54 (1.32)	10.63 (1.38)	7.27 (2.18)	5.31 (2.50)	8.61 (2.31)	1.98 (1.63)
Rape	10.20 (1.52)	9.81 (1.76)	6.72 (2.47)	5.91 (2.31)	7.06 (2.46)	1.93 (1.69)
Sexual Sadism	10.02 (2.31)	9.74 (1.41)	5.16 (2.57)	4.67 (2.68)	6.86 (2.61)	1.69 (1.17)
Child abuse	9.34 (1.85)	9.83 (1.86)	5.72 (2.55)	5.25 (2.38)	7.98 (2.26)	1.56 (0.96)
Abuser	10.02[a] (1.82)	9.99[a] (1.66)	6.23[b] (2.57)	5.31[c] (2.49)	7.62 (2.50)	1.79 (1.41)

Table 26

Likelihood Means (Standard Deviation) of Having Bigot Characteristics

	Character Exemplar					
Characteristic	Psychopath	Abuser	Bigot	Narcissist	Bad	Good
Transphobia	7.97 (2.20)	8.22 (1.97)	10.66 (1.36)	7.22 (2.19)	8.51 (1.82)	3.15 (2.41)
Xenophobia	7.03 (2.15)	7.31 (2.05)	10.58 (0.77)	7.56 (2.38)	7.45 (2.40)	2.16 (1.73)
Homophobia	7.52 (2.03)	7.94 (2.20)	10.39 (1.66)	7.07 (2.12)	7.22 (2.45)	2.13 (1.45)
White Supremacism	7.59 (2.42)	7.48 (2.36)	10.10 (1.64)	6.57 (2.64)	6.00 (2.54)	1.98 (1.18)
Bigot	7.52^{b} (2.21)	7.74^{b} (2.17)	10.40^{a} (1.41)	7.12^{c} (2.34)	7.30 (2.47)	2.34 (1.79)

Table 27

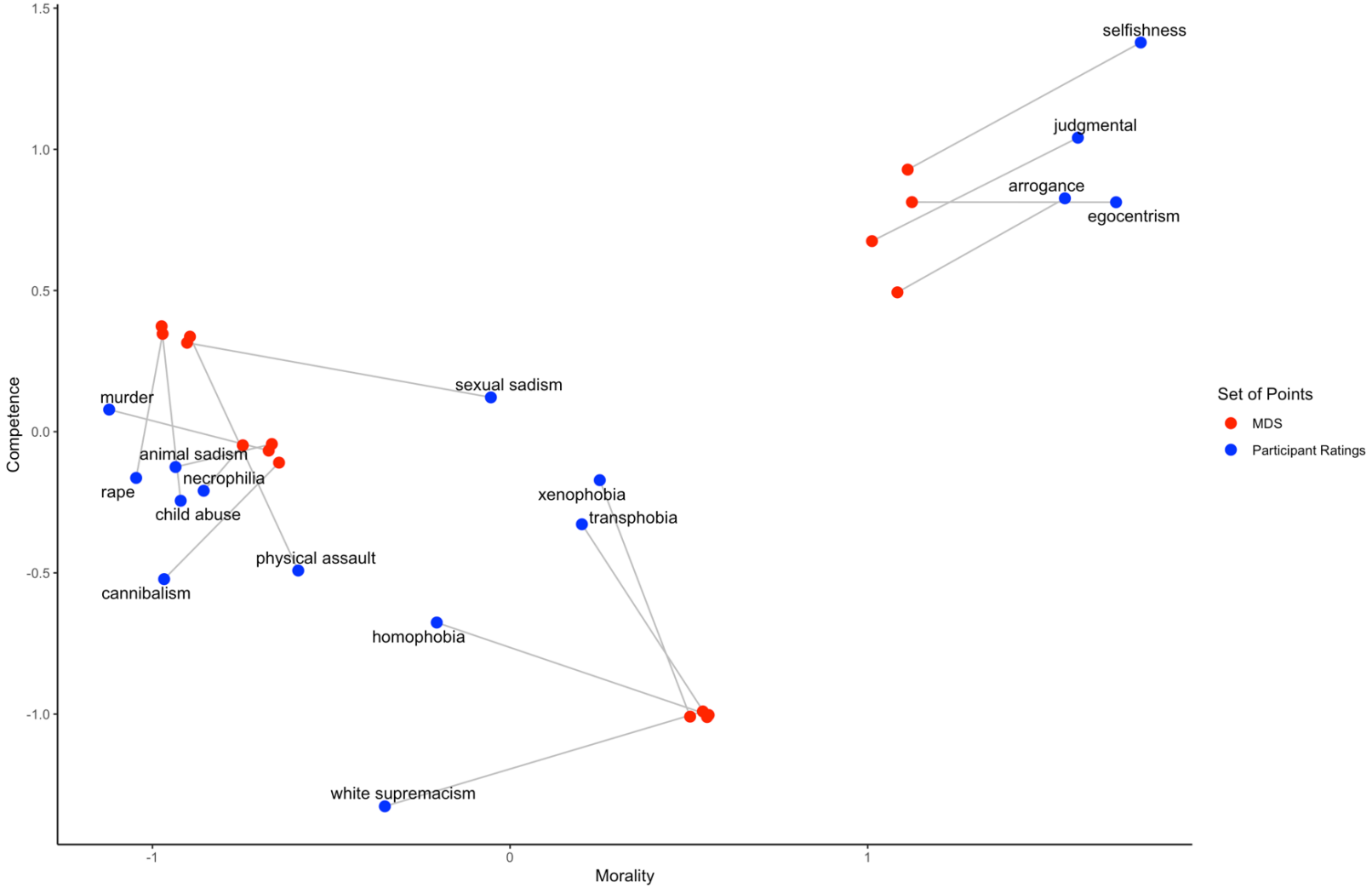
Likelihood Means (Standard Deviation) of Having Narcissist Characteristics

	Character Exemplar					
Characteristic	Psychopath	Abuser	Bigot	Narcissist	Bad	Good
Selfish	10.05 (2.10)	10.20 (1.96)	9.24 (1.72)	10.61 (0.74)	9.54 (2.14)	3.25 (1.91)
Arrogance	8.71 (2.29)	9.10 (2.61)	10.33 (1.23)	10.42 (1.19)	8.58 (2.74)	3.16 (2.08)
Egocentric	8.64 (2.03)	9.44 (1.98)	8.83 (1.74)	10.39 (1.56)	8.67 (2.25)	2.33 (1.48)
Judgmental	8.14 (2.70)	9.07 (1.61)	10.34 (1.37)	10.24 (1.13)	9.41 (2.27)	2.55 (1.68)
Grand Mean	8.88[c] (2.38)	9.44[b] (1.99)	9.69[b] (1.66)	10.40[a] (1.20)	9.02 (2.40)	2.83 (1.84)

APPENDIX F: ERRORS FROM PROCRUSTES ANALYSES

Figure 21

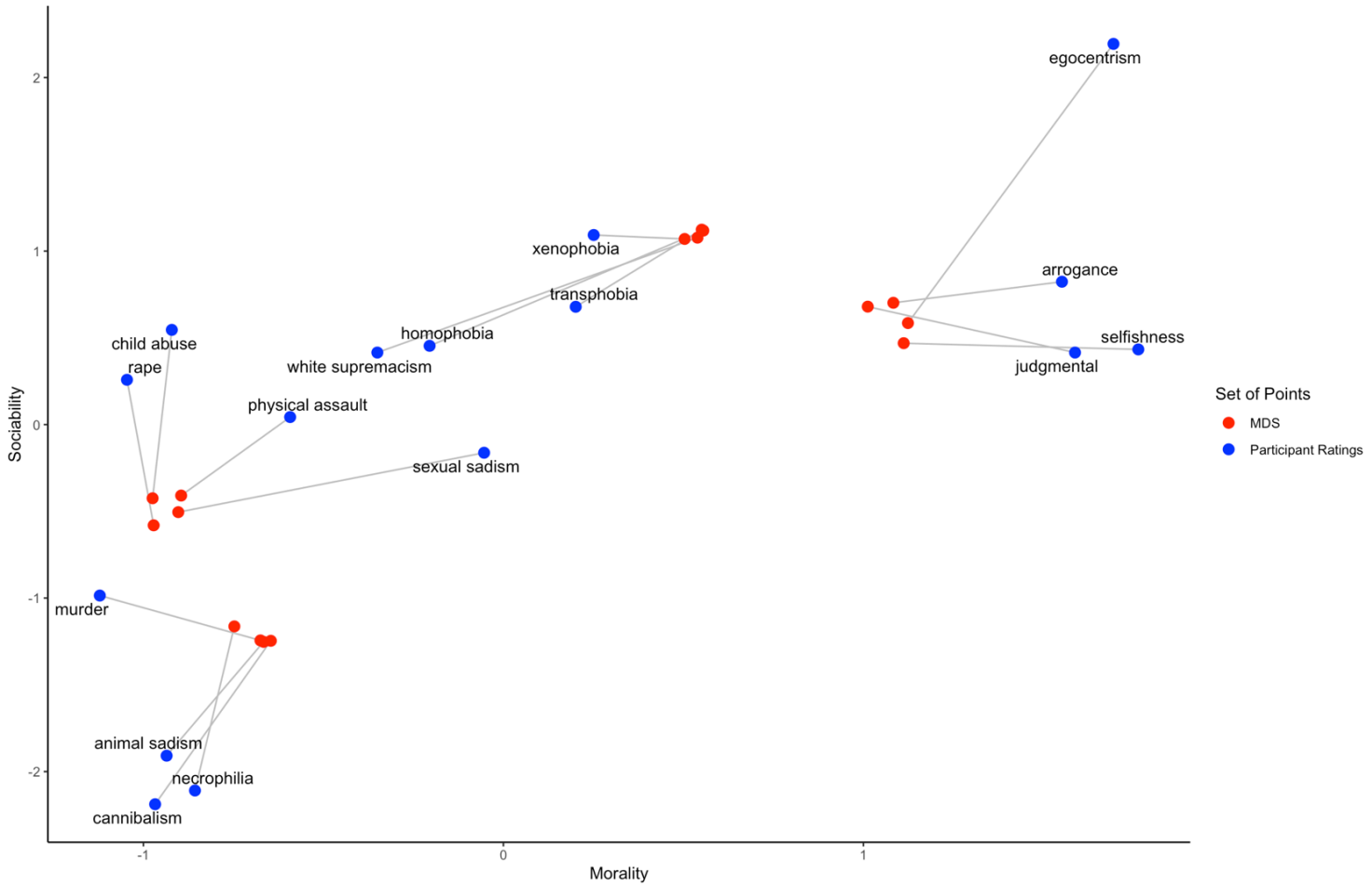
Errors from Procrustes Analysis (Morality and Competence Dimensions)



Note. This figure displays differences between the multidimensional scaling solution (Study 2) and participant ratings (Study 3).

Figure 22

Errors from Procrustes Analysis (Morality and Sociability Dimensions)



Note. This figure displays differences between the multidimensional scaling solution (Study 2) and participant ratings (Study 3).

REFERENCES CITED

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A. & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

Abele, A. E. & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in experimental social psychology*, 50, 195–255. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Anderson. (1970). *Theophrastus: The Character Sketches*. Kent State University Press.

Anderson, R. A., Kamtekar, R., Nichols, S. & Pizarro, D. A. (2021). "False positive" emotions, responsibility, and moral character. *Cognition*, *214*, 104770. https://doi.org/10.1016/j.cognition.2021.104770

Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T. & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. Journal of Social and Personality Psychology, 125, 1157-1188. https://doi.org/10.1037/pspp0000470

Bastian, B. & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228–235. https://doi.org/10.1016/j.jesp.2005.03.003

Baumeister, R. F. & Campbell, W. K. (1999). The intrinsic appeal of evil: Sadism, sensational thrills, and threatened egotism. *Personality and Social Psychology Review*, *3*, 210–221. https://doi.org/10.1207/s15327957pspr0303_4

Bloomfield, M. (1952). *The seven deadly sins: An introduction to the history of a religious concept with special reference to medieval english literature*. Michigan State College Press.

Bodenhausen, G. V., Kang, S. K. & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & N. Macrae (Eds.) *The Sage handbook of social cognition* (pp. 318–336). SAGE Publications Ltd.

Bogardus, E. S. (1933). A social distance scale. *Sociology & Social Research*, *17*, 265–271.

Borg, I., Groenen, P. J. F. & Mair, P. (2013). Applied Multidimensional Scaling. *SpringerBriefs in Statistics*. https://doi.org/10.1007/978-3-642-31848-1

Brambilla, M. & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*, 397–408. https://doi.org/10.1521/soco.2014.32.4.397

Brambilla, M., Sacchi, S. & Rusconi, P. (2021). The primacy of morality in impression development: Theory, research, and future directions. *Advances in experimental social psychology*, 64, 187–262. https://doi.org/10.1016/bs.aesp.2021.03.001

Brandt, M. J. & Reyna, C. (2011). The chain of being: A hierarchy of morality. *Perspectives on Psychological Science*, *6*, 428–446. https://doi.org/10.1177/1745691611414587

Brewer, M. B., Dull, V. & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology*, *41*(4), 656–670. https://doi.org/10.1037/0022-3514.41.4.656

Brunell, A. B., Gentry, W. A., Campbell, W. K., Hoffman, B. J., Kuhnert, K. W. & DeMarree, K. G. (2008). Leader emergence: The case of the narcissistic leader. *Personality and Social Psychology Bulletin*, *34*(12), 1663–1676. https://doi.org/10.1177/0146167208324101

Buckels, E. E., Jones, D. N. & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, *24*, 2201–2209. https://doi.org/10.1177/0956797613490749

Bushman, B. J., Bonacci, A. M., Dijk, M. V. & Baumeister, R. F. (2003). Narcissism, Sexual Refusal, and Aggression: Testing a Narcissistic Reactance Model of Sexual Coercion. *Journal of Personality and Social Psychology*, *84*, 1027–1040. https://doi.org/10.1037/0022-3514.84.5.1027

Cantor, N. & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, *35*(1), 38–48. https://doi.org/10.1037/0022-3514.35.1.38

Carlsmith, K. M., Darley, J. M. & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*(2), 284–299. https://doi.org/10.1037/0022-3514.83.2.284

Cohen, T. R. & Morse, L. (2014). Moral character: What it is and what it does. *Research in Organizational Behavior*, *34*, 43–61. https://doi.org/10.1016/j.riob.2014.08.003

Coovert, M. D. & Reeder, G. D. (1990). Negativity effects in impression formation: The role of unit formation and schematic expectations. *Journal of Experimental Social Psychology*, *26*, 49–62. https://doi.org/10.1016/0022-1031(90)90061-P

Cox, J., Edens, J. F., Rulseh, A. & Clark, J. W. (2016). Juror perceptions of the interpersonal-affective traits of psychopathy predict sentence severity in a white-collar criminal case. *Psychology, Crime & Law*, *22*, 721–740. https://doi.org/10.1080/1068316X.2016.1174864

Crawford, J. T. (2017). Are conservatives more sensitive to threat than liberals? It depends on how we define threat and conservatism. *Social Cognition*, *35*(4), 354–373. https://doi.org/10.1521/soco.2017.35.4.354

Critcher, C. R., Helzer, E. G. & Tannenbaum, D. (2020). Moral character evaluation: Testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology*, *87*, 103906. https://doi.org/10.1016/j.jesp.2019.103906

Cuddy, A. J. C., Fiske, S. T. & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology*, *40, 61-149*. https://doi.org/10.1016/s0065-2601(07)00002-0

Dalmaijer, E. S., Nord, C. L. & Astle, D. E. (2022). Statistical power for cluster analysis. *BMC Bioinformatics*, *23*. https://doi.org/10.1186/s12859-022-04675-1

Dion, K., Berscheid, E. & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285–290. https://doi.org/10.1037/h0033731

Dolan, M. & Fullam, R. (2006). Face affect recognition deficits in personality-disordered offenders: Association with psychopathy. *Psychological Medicine*, *36*, 1563–1569. https://doi.org/10.1017/s0033291706008634

Dopkins, S. & Gleason, T. (1997). Comparing exemplar and prototype models of categorization. *Canadian Journal of Experimental Psychology*, *51*, 212–230. https://doi.org/10.1037/1196-1961.51.3.212

Everett, J. A. C. (2013). The 12-item social and economic conservatism scale (SECS). *PLoS ONE*, *8*. https://doi.org/10.1371/journal.pone.0082131

Fiske, S. T., Cuddy, A. J. C. & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*, 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*, 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gentile, B., Miller, J. D., Hoffman, B. J., Reidy, D. E., Zeichner, A. & Campbell, W. K. (2013). A test of two brief measures of grandiose narcissism: The narcissistic personality inventory-13 and the narcissistic personality inventory-16. *Psychological Assessment*, *25*, 1120–1136. https://doi.org/10.1037/a0033192

Goodwin, G. P. (2015). Moral Character in Person Perception. *Current Directions in Psychological Science*, *24*, 38–44. https://doi.org/10.1177/0963721414550709

Goodwin, G. P., Piazza, J. & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168. https://doi.org/10.1037/a0034726

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P. & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Psychology, 47,* 55-130. https://doi.org/10.1016/b978-0-12-407236-7.00002-4

Graham, J., Haidt, J. & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046. https://doi.org/10.1037/a0015141

Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B. H. & Fraley, R. C. (2015). Narcissism and leadership: A meta-analytic review of linear and nonlinear relationships. *Personnel Psychology*, *68*, 1–47. https://doi.org/10.1111/peps.12072

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834. https://doi.org/10.1037//0033-295x.108.4.814

Haidt, J., Roller, S. H. & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613–628. https://doi.org/10.1037/0022-3514.65.4.613

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D. & Newman, J. P. (1990). The revised psychopathy checklist: Reliability and factor structure. *Psychological Assessment*, *2*, 338–341. https://doi.org/10.1037/1040-3590.2.3.338

Hartman, R., Blakey, W. & Gray, K. (2022). Deconstructing moral character judgments. *Current Opinion in Psychology*, *43*, 205–212. https://doi.org/10.1016/j.copsyc.2021.07.008

Heiphetz, L. (2019). Moral essentialism and generosity among children and adults. *Journal of Experimental Psychology: General*, *148*(12), 2077–2090. https://doi.org/10.1037/xge0000587

Heiphetz, L. (2020). The Development and Consequences of Moral Essentialism. *Advances in Child Development and Behavior*, *59*, 165–194. https://doi.org/10.1016/bs.acdb.2020.05.006

Helzer, E. G. & Critcher, C. R. (2018). What do we evaluate when we evaluate moral character. In Gray K. & Graham, J. (Eds.), *Atlas of Moral Psychology* (pp. 99–107). New York, NY: Guilford Press.

Henss, R. (1998). Type nouns and the five factor model of personality description. *European Journal of Personality*, *12*(1), 57–71. https://doi.org/10.1002/(SICI)1099-0984(199801/02)12:1<57::AID-PER293>3.0.CO;2-%23

Holmqvist, R. (2008). Psychopathy and affect consciousness in young criminal offenders. *Journal of Interpersonal Violence*, *23*, 209–224. https://doi.org/10.1177/0886260507309341

Janoff-Bulman, R. & Carnes, N. C. (2013). Surveying the moral landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review*, *17*, 219–236. https://doi.org/10.1177/1088868313480274

John, O. P., Hampson, S. E. & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology*, *60*(3), 348–361. https://doi.org/10.1037/0022-3514.60.3.348

Jones, D. & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, *21*, 28–41. https://doi.org/10.1177/1073191113514105

Jost, J. T., Glaser, J., Kruglanski, A. W. & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Judd, C. M., Westfall, J. & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Klebl, C., Rhee, J. J., Greenaway, K. H., Luo, Y. & Bastian, B. (2021). Beauty Goes Down to the Core: Attractiveness Biases Moral Character Attributions. *Journal of Nonverbal Behavior*, 1–15. https://doi.org/10.1007/s10919-021-00388-w

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C. & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Landy, J. F. & Bartels, D. M. (2018). An empirically-derived taxonomy of moral concepts. *Journal of Experimental Psychology: General*, *147*, 1748–1761. https://doi.org/10.1037/xge0000404

Landy, J. F., Piazza, J. & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and Social Psychology Bulletin, 42*(9), 1272-1290. https://doi.org/10.1177/0146167216655984

Lapsley, D. K. & Lasky, B. (2001). Prototypic Moral Character. *Identity*, *1*(4), 345–363. https://doi.org/10.1207/s1532706xid0104_03

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*(2), 234-249. https://doi.org/10.1037/0022-3514.93.2.234

Lim, A., Young, R. L. & Brewer, N. (2022). Autistic adults may be erroneously perceived as deceptive and lacking credibility. *Journal of Autism and Developmental Disorders*, *52*(2), 490–507. https://doi.org/10.1007/s10803-021-04963-4

Lupfer, M. B., Weeks, M. & Dupuis, S. (2000). How pervasive is the negativity bias in judgments based on character appraisal? *Personality and Social Psychology Bulletin*, *26*, 1353–1366. https://doi.org/10.1177/0146167200263004

Lynch, J. M., Lane, J. D., Berryessa, C. M. & Rottman, J. (2019). How information about perpetrators' nature and nurture influences assessments of their character, mental states, and deserved punishment. *PLoS ONE*, *14*, 1–20. https://doi.org/10.1371/journal.pone.0224093

Macrae, C. N. & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Marcus, D. K. & Zeigler-Hill, V. (2015). A big tent of dark personality traits. *Social and Personality Psychology Compass*, *9*(8), 434–446. https://doi.org/10.1111/spc3.12185

Marcus, D. K., Zeigler-Hill, V., Mercer, S. H. & Norris, A. L. (2014). The psychology of spite and the measurement of spitefulness. *Psychological Assessment*, *26*(2), 563–574. https://doi.org/10.1037/a0036039

Martin, J. W., Charles, S. & Heiphetz, L. (2022). Essentialist views of criminal behavior predict increased punitiveness. In J. Musolino, J. Sommer, & P. Hemmer (Eds.), *The Cognitive Science of Belief* (1st ed., pp. 254–276). Cambridge University Press. https://doi.org/10.1017/9781009001021.019

Martin, J. W. & Heiphetz, L. (2021). "Internally wicked": Investigating how and why essentialism influences punitiveness and moral condemnation. *Cognitive Science*, *45*. https://doi.org/10.1111/cogs.12991

McHugh, C., McGann, M., Igou, E. R. & Kinsella, E. L. (2022). Moral judgment as categorization (MJAC). *Perspectives on Psychological Science*, *17*, 131–152. https://doi.org/10.1177/1745691621990636

Mededović, J. & Petrović, B. (2015). The dark tetrad: Structural properties and location in the personality space. *Journal of Individual Differences*, *36*, 228–236. https://doi.org/10.1027/1614-0001/a000179

Medin, D. L. & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review*, *111*(4), 960–983. https://doi.org/10.1037/0033-295x.111.4.960

Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238. https://doi.org/10.1037/0033-295x.85.3.207

Medin, D. L. & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 241–253. https://doi.org/10.1037/0278-7393.7.4.241

Medin, D. L. & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*, 113–138. https://doi.org/10.1017/cbo9780511801938.011

Melnikoff, D. E. & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences*, *115*, E592–E600. https://doi.org/10.1073/pnas.1714945115

Minson, J. A. & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, *3*, 200–207. https://doi.org/10.1177/1948550611415695

Monaghan, C., Bizumic, B. & Sellbom, M. (2016). The role of Machiavellian views and tactics in psychopathology. *Personality and Individual Differences*, *94*, 72–81. https://doi.org/10.1016/j.paid.2016.01.002

Norenzayan, A., Smith, E. E., Kim, B. J. & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, *26*(5), 653–684. https://doi.org/10.1207/s15516709cog2605_4

Osherson, D. N., Smith, E. E., Wilkie, O., López, A. & Shafir, E. (1990). Category-Based Induction. *Psychological Review*, *97*(2), 185–200. https://doi.org/10.1037/0033-295x.97.2.185

Parfitt, C. H. & Alleyne, E. (2020). Not the sum of its parts: A critical review of the MacDonald triad. *Trauma, Violence, and Abuse*, *21*, 300–310. https://doi.org/10.1177/1524838018764164

Paulhus, D. & Williams, K. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Peres-Neto, P. R. & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, *129*(2), 169–178. https://doi.org/10.1007/s004420100720

Pilkonis, P. A. (1988). Personality prototypes among depressives: Themes of dependency and autonomy. *Journal of Personality Disorders*, *2*(2), 144–152. https://doi.org/10.1521/pedi.1988.2.2.144

Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363. https://doi.org/10.1037/h0025953

Rauthmann, J. F. & Kolar, G. P. (2012). How "dark" are the Dark Triad traits? Examining the perceived darkness of narcissism, Machiavellianism, and psychopathy. *Personality and Individual Differences*, *53*, 884–889. https://doi.org/10.1016/j.paid.2012.06.020

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Rhodes, M. & Moty, K. (2020). What is social essentialism and how does it develop? In M. Rhodes (Ed.), *Advances in child development and behavior* (Vol. 59, pp. 1–30). Academic Press. https://doi.org/10.1016/bs.acdb.2020.05.001

Rosa, A. O.-L., Villacampa, J., Corradi, G. & Ingram, G. P. D. (2021). The creepy, the bad and the ugly: Exploring perceptions of moral character and social desirability in uncanny faces. *Current Psychology*. https://doi.org/10.1007/s12144-021-01452-w

Rosch, E. (1978). Principles of categorization. In *Cognition and categorization* (pp. 27–48).

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Royzman, E. B., Kim, K. & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, *10*(4), 296–313. https://doi.org/10.1017/s193029750000512x

Rozin, P. & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Rusconi, P., Sacchi, S., Brambilla, M., Capellini, R. & Cherubini, P. (2020). Being honest and acting consistently: Boundary conditions of the negativity effect in the attribution of morality. *Social Cognition*, *38*, 146–178. https://doi.org/10.1521/soco.2020.38.2.146

Saucier, G. (2003). Factor Structure of English-Language Personality Type-Nouns. *Journal of Personality and Social Psychology*, *85*(4), 695–708. https://doi.org/10.1037/0022-3514.85.4.695

Schaumberg, R. L. & Mullen, E. (2017). From incidental harms to moral elevation: The positive effect of experiencing unintentional, uncontrollable, and unavoidable harms on perceived moral character. *Journal of Experimental Social Psychology*, *73*, 86–96. https://doi.org/10.1016/j.jesp.2017.06.016

Schein, C. & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*, 1147–1163. https://doi.org/10.1177/0146167215591501

Schein, C. & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*, 32–70. https://doi.org/10.1177/1088868317698288

Skitka, L. J., Bauman, C. W. & Mullen, E. (2008). Morality and justice: An expanded theoretical perspective and empirical review. In Hegtvedt, K.A. and Clay-Warner, J. (Eds.) *Justice* (Advances in Group Processes, Vol. 25, pp. 1-27). Emerald Group Publishing Limited. https://doi.org/10.1016/S0882-6145(08)25001-0

Smith, E. R. & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*(1), 3–21. https://doi.org/10.1037/0033-295x.99.1.3

Stearns, D. C. & Parrott, W. G. (2012). When feeling bad makes you look good: Guilt, shame, and person perception. *Cognition and Emotion*, *26*, 407–430. https://doi.org/10.1080/02699931.2012.675879

Stellar, J. E. & Willer, R. (2018). Unethical and inept? The influence of moral information on perceptions of competence. *Journal of Personality and Social Psychology*, *114*(2), 195–210. https://doi.org/10.1037/pspa0000097

Strohminger, N. & Nichols, S. (2014). The essential moral self. *Cognition*, *131*, 159–171. https://doi.org/10.1016/j.cognition.2013.12.005

Tannenbaum, D., Uhlmann, E. L. & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, *47*, 1249–1254. https://doi.org/10.1016/j.jesp.2011.05.010

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 25–29.

Tilburg, W. A. P. van, Igou, E. R. & Panjwani, M. (2022). Boring people: Stereotype characteristics, interpersonal attributions, and social reactions. *Personality and Social Psychology Bulletin*, *49*(9), 1329–1343. https://doi.org/10.1177/01461672221079104

Todorov, A., Olivola, C. Y., Dotsch, R. & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Uhlmann, E. L. & Zhu, L. L. (2014). Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science*, *5*, 279–285. https://doi.org/10.1177/1948550613497238

Uhlmann, E. L., Zhu, L. L. & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, *126*, 326–334. https://doi.org/10.1016/j.cognition.2012.10.005

Vel-Palumbo, M. de, Howarth, L. & Brewer, M. (2019). Once a sex offender always a sex offender? Essentialism and attitudes towards criminal justice policy. *Psychology, Crime & Law*, *25*, 421–439. https://doi.org/10.1080/1068316X.2018.1529234

Walker, L. J. & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology*, *86*, 629–647. https://doi.org/10.1037/0022-3514.86.4.629

Walker, L. J. & Pitts, R. (1998). Naturalistic conceptions of moral maturity. *Developmental Psychology*, *34*, 403–419. https://doi.org/10.1037/0012-1649.34.3.403

Webster, R. J. & Saucier, D. A. (2013). Angels and demons are among us: Assessing individual differences in belief in pure evil and belief in pure good. *Personality and Social Psychology Bulletin*, *39*, 1455–1470. https://doi.org/10.1177/0146167213496282

Wickelmaier, F. (2003). *An Introduction to MDS*. Sound Quality Research Unit.

Yip, J. A. & Schweitzer, M. E. (2015). Trust promotes unethical behavior: excessive trust, opportunistic exploitation, and strategic exploitation. *Current Opinion in Psychology*, *6*, 216–220. https://doi.org/10.1016/j.copsyc.2015.09.017